# Principles Of Animal Use For Gerontological Research

**Richard A. Miller1 and Nancy L. Nadon2**
1Department of Pathology and Geriatrics Center, University of Michigan School of Medicine, Institute of Gerontology, and Ann Arbor VA Medical Center, Ann Arbor MI 48109 and 2National Institute on Aging, Bethesda MD 20892

## Abstract

This essay presents some practical advice and suggestions for those who wish to use mice and rats in experiments on the biology of aging. Ten principles set forth guidance on choice of ages, choice of stocks, the importance of specific-pathogen-free status, the uses of necropsy data, the dangers of pooling samples from different individuals, planning ahead for loss of aged mice to death and disease, the use of cost-adjusted power calculations, and the dangers of inferring causal associations from correlated age effects. A concluding executive summary distills these tips into six bullets and 111 words.

List of abbreviations:
IL, interleukin
IL-2, Interleukin 2
NIA, National Institute on Aging
SPF, specific-pathogen free

This article is intended to convey some rules-of-thumb for investigators just starting to think about the design of experiments on aging using mice and rats. The principles stated below reflect the opinions of the authors, based on years of experience in rodent-based work in gerontology and molecular biology. However, to avoid excessive circumlocution of the "in my opinion it may be helpful to" variety, the style is deliberately imperative, modeled on examples set by writers of style manuals (1) and columns of advice for the lovelorn. In addition the word "mice" will be used throughout to mean "mice and rats," except in those cases where mice are different from rats.

### First Principle: Don't use mice that are too old.
Many beginners ask for the oldest available mice for their initial age effect study under the presumption that the old ones will show bigger differences from young controls, and thus produce significant results quicker. This is rarely a good idea for several reasons:

(a) Old mice are usually sick (even if not quite dead yet). If a trait differs from young controls only in the last 10% of the cohort to drop off, then it's hard to be confident that the change is a result of aging rather than of the advanced disease or diseases most typical in the stock under study. After all, the aging process, which creates decrepit old mice from healthy, fit, young ones, takes many months to do this, and age-dependent changes in many cells, tissue, and organ systems can usually be demonstrated well before the median survival time for the species or stock. If your

assay shows no change at 18, 22, 26, or 30 months of age (in a stock with a median survival of 24 months), then demonstrating a change in 34 month old animals may well be due to sickness per se. Judicious selection of ages for initial exploratory work may depend on the specific characteristics of the stock to be used, and stocks with median survivals of 22 months or of 30 months may call for adjustments of ages selected for initial examination.

(b) Old mice are very expensive, particularly if you want them to be disease-free. The problem is that the real production cost of mice rises not linearly with chronologic age, but instead in proportion to the mortality rate, i.e. as an exponential function of age. If half your mice live to age 24 months, then producing a single 24 month old mouse requires you to pay someone to house two mice for 24 months, one of which has just died. If only 10% of the mice survive to age 32 months, then the real cost of each 32 month old mouse is the cost of raising 10 mice for anywhere from 18 to 32 months to get the one alive at 32 months. And then that one mouse, when you do the necropsy, may well turn out to have advanced neoplasia.

To illustrate the projected costs, at one well-known Midwestern Medical Center, animal users are charged $0.58/cage/day for cages of 4 mice. At this price it costs $106 to grow a mouse for two years; but because half the mice die, the cost of a live 2 year old mouse is twice as high, or about $212. Because half of the two-year old mice are found to have advanced neoplasia even at a cursory necropsy, the cost of a more-or-less tumor-free 2 year old mouse is another two-fold higher, or $414. The nominal cost of a 32 month old mouse (at $0.58/cage/day) is $140, but adjusting for attrition and disease gives a real cost closer to $1,400 each.

The NIA Office of Biological Resources provides highly subsidized animals, even with the recent price increase; for a two-year old C57BL/6 mouse, for example, the cost charged to investigators is a mere $72, well under the local production cost. The real subsidy (production cost minus cost to you) rises exponentially with age, however, and therefore it costs much more to raise a very old mouse in one's own facility than it does to buy them from NIA. While NIA continues to raise these very old mice, despite their exceptionally high cost, in order to provide investigators maximal flexibility in designing their experimental protocols, investigators who do use this scarce resource are still confronted with the problems imposed by their very high rates of concurrent illness.

**Second Principle: don't use mice that are too young.** The beginner also begs for the youngest possible controls, often just weaned, again on the grounds that these are more likely to show big differences from old mice. The catch is that these mice are no more typical of "young adults" than, say, your typical nine-year old human person. Even 2-3 month old mice, the biological equivalent of teen-agers and college freshmen, i.e. technically postpubertal but hardly adult, are still in the throes of complex maturational changes we may wish to distinguish from the aging process. Although differences between 2 month old and 5 month old mice may be well worth studying, and in some cases may be highly relevant to aging (thymic involution, for example, is well advanced by 5 months of age), a conservative strategy might be to

use mice aged 4-6 months as the "young" control group in experimental comparisons.

**Third Principle: don't use too few age groups.** It is certainly cheaper to do your initial survey experiment with only two age groups, e.g. 6 and 24 months, but there are major risks of missing interesting effects, and equal risks of reporting positive results whose significance would be routinely mis-interpreted without data on intermediate ages. Figure 1 shows some hypothetical situations of this kind, in which the inclusion of one or more intermediate age groups radically alters the picture that would have been obtained by study of extreme ages only.

The take-home message is: if you start out knowing nothing about the effects of age on your measurement of interest, try an initial survey from young adult to the median survival age, say 6, 12, 18, and 24 months of age for most strains. If there really is no change between 6 and 24 months of age, then you're welcomed to look at older animals, but remember that the older they get, the more likely it is that any effect you're seeing reflects the diseases and debilities that are likely to kill the mice in a few more months.

**Fourth Principle: the mice must be specific pathogen free, and you have to be able to prove it.** Specific pathogen free (SPF) colonies are not free of all disease, and are not free of all infectious agents, but they are free of a well-defined group of known murine (or ratoid) pathogens that routinely plague the old-fashioned "conventional" mouse colony. The most common mouse pathogens endemic in conventional colonies are Sendai virus, coronavirus (mouse hepatitis virus), and pinworm, while rat colonies are most often infected with sialodacryoadenitis virus, Sendai, and Mycoplasma pulmonis (Robert Dysko, DVM, personal communication). The methods required for maintenance of an SPF rodent vivarium are beyond the scope of this handbook, and are well reviewed elsewhere (2). Key steps include: never allowing in animals, from any source, that have not been proven to be SPF; never taking animals out of the colony and allowing them to return; never allowing visitors, even site visitors and deans, to enter the colony; never letting anyone into the colony who does not need to be there; and the routine but obsessive use of precautions like gloves, gowns, and shoe covers. Some facilities use an expensive barrier system, in which each room has a clean door (through which supplies and clean personnel enter) and a dirty door (through which used cages leave). Others have almost equal luck with commercially available single-cage filter bonnets that greatly reduce the risk of cage-to-cage transmission of airborne microbes.

These operating procedures, while necessary, are not sufficient to earn the right to refer to your colony as SPF; the colony must be proven to be free of key pathogens by routine testing of surveillance mice, for example on a quarterly basis. A minimal surveillance program involves introducing several cages of new mice, without filter bonnets, into each room every three months, using a stock (CD-1, for example) known to be susceptible to many common pathogens. These mice are then tested after 90 days, a period of time sufficient to allow them to become infected by anything in circulation, and to have developed antibodies to the infectious agent. The mice are then euthanized, examined carefully for evidence of intestinal and external parasites,

and their sera tested (often by a commercial laboratory) for evidence of anti-viral antibodies specific for the agents of interest. Periodic histopathological analysis is also recommended. If all tests come back negative, then you can call the colony SPF. Positive responses should induce a panicky feeling and a vigorous re-testing effort, and repeat positives usually require that the entire affected colony be discarded and re-derived.

If this surveillance system is not practicable in a given institutional vivarium, a useful alternative is to keep a small number of weanlings from breeding cage until they are about 3 months old and then to send these to a commercial facility for testing. Since the breeding pairs are usually long-term residents of the facility, and newborn pups are particularly sensitive to infection, this procedure is more sensitive than buying mice and allowing them to simply to reside in the colony for a few months without contact with the local residents. The optimal situation for maintaining an SPF colony combines the use of filter bonnets with the use of sentinel mice. Since filters work very well in preventing spread of airborne pathogens, exposure of the sentinel animals to potentially infectious agents requires a procedure in which the used bedding from a pool of cages be thoroughly mixed, and then added to the cages containing the sentinels. (A policemouse's life is not a happy one.) Sentinels who put up with this treatment for several months are then volunteered for necropsy and serological analysis. In these circumstances a stray positive result may not require sacrifice of the entire colony, since it is more likely that the infection has been confined to one or two cages; detailed follow-up studies may document good health for the majority of mice in the room. To be effective, this system requires good record-keeping in order to trace all cages that a sentinel has had contact with.

Why go through this hassle? "After all," the scientist stuck with a conventional colony might rationalize, "people are not free of all infectious agents; I'm just trying to more closely mimic the real world situation." The basic problem is that the intensity, variety and prevalence of infection in any given conventional colony may well change from month to month and year to year, and is likely to differ greatly from one colony to another. Because many infections can alter a mouse's immune, hepatic, endocrine, digestive, pulmonary, and neurological responses, studies carried out on conventional colonies can prove very difficult to reproduce in another, or even in the same, laboratory. In some cases allegations of age effects on variables of interest have proven to occur only in conventional colonies (3), and are thus likely to reflect unsuspected influences of one or more uncharacterized infectious agents than of aging itself. Most effects of this kind doubtless go undetected, since few workers routinely use mice from two distinct colonies, one conventional and the other SPF, but it seems likely that many of the unnerving conflicts among reports in the gerontological literature may reflect variations in colony pathogen status.

Successful maintenance of an SPF colony also requires sufficient discipline to prevent the importation of new mouse stocks from uncertified suppliers. A well-run colony will usually permit unfettered importation of mice from only a very small number of commercial vendors, vendors that routinely submit clean bills of health with all shipments. A request for permission to bring in animals from an uncertified vendor or another research institution should trigger a process in which the sender is

required to document the health status of the animals, and in which even allegedly clean animals are kept in a separate quarantine facility (or building) until tested locally for pathogens before they or their offspring are introduced into the general population. Importation of a stock that cannot be proven specific pathogen-free ordinarily requires long-term quarantine or rederivation of the stock by caesarian delivery and foster nursing.

Two common mistakes:

*Scenario #1* - the vivarium manager tells you it's an SPF colony, because the facility only buys from SPF suppliers. So why spend the money to test this? Four years ago, however, your technician visited a pet store on the way to work and every cage has had Sendai for four years - your laboratory mice are about as SPF as the ones in your basement. Test quarterly, and you can proudly report the clean bill of health in every paper and every grant proposal.

*Scenario #2* - you have a conventional colony, and you're not proud of it, but you buy the SPF mice from NIA and let them sit in the colony for just a week or two before use. This approach, which is remarkably common, just about guarantees that the mice used in your tests are infected with something; blessed with an SPF upbringing, they have no protective antibody titers, and are thus sitting ducks for whatever virus happens to be in your colony during their initiation into the tough realities of real world infection. Even if the mice will be housed in the animal facility short-term, it is well worthwhile to use an SPF facility or at least a quarantine room and filter bonnets.

*Scenario #2b*- it is not a good idea to try to sneak past this problem by using the mice the day they arrive off the delivery van. Shipping is very tough on mice, and the stress has an impact on adrenal size, steroid hormone levels, immunity, and other organ. No one would write a protocol that began "Prior to their use in experiments, the mice were placed in a shipping container without access to their usual sources of food and water, and then flown 1000 miles in a dark, cold, noisy plane, followed by interstate truck shipmentŒ". It's a good idea to let the mice sit after arrival, for at least a week or preferably two, before use. But do it in an SPF colony.

**Fifth Principle: Don't bet the farm on C57BL/6 mice.** Don't bet it on F344 rats, either. Inbred strains, developed over the last 80 years for their usefulness in transplantation and cancer research, have become the de facto standard strains for research in most other areas, including aging. Sixty percent of the rats ordered from the NIA colony are of the F344 stock, despite the well-known problems with this strain (4), and 40% of the mice used are C57BL/6. The F1 hybrid mice, which in some respects have much to recommend them when compared to inbred animals, account for only 15% of the mouse orders. This is a shame, whose implications for aging research have been reviewed in detail elsewhere (4-6). In brief, the case against inbreds includes the following counts:

All mice in an inbred stock are genetically identical. It's therefore impossible to be certain that conclusions based on an inbred stock will apply equally to any other inbred stock without doing the study all over again.

Inbred stocks are not only homogeneous, they are also weird, debilitated, and short lived. Creation of an inbred stock involves forced homozygosity at all loci. This is a highly selective process, because the inbreeding process frequently creates genotypes that impair viability and fertility; in fact, most brother-sister mated families eventually die out, with the few surviving families becoming the "standard" inbred lines we all know and love. F1 hybrids created by a cross between two different inbred lines are almost invariably longer-lived than either of the two parents (7), consistent with the notion that the homozygous condition produces an animal of lower quality. Many inbred lines are famed for properties that clearly count as strain-specific oddities: the chronic renal disease and high lymphoma incidence of the F344 rat, the 100% incidence of thymic lymphoma in the AKR/J mouse, the near 100% incidence of reticulum cell sarcoma in SJL/J mice, and many other similar peculiarities. Table 1 shows a series of anecdotes: the take-home message is that individual inbred lines may have a very high incidence of lesions that are rarely seen in other inbred lines. These obvious illnesses, and other idiosyncrasies less obvious to the naked eye, could in principle wreak havoc on the process by which general conclusions are inferred from a limited data set. In some cases, of course, genetic identity is critical to the experimental plan; these situations include protocols that involve transfer of tissues from one mouse to another, and those where the goal of the study involves analyses of inter-strain variation. But in many other instances the use of an inbred strain reflects mere custom rather than a careful decision among alternatives. To read more along these lines, check out (5).

So what is a gerontologist to do? There are a number of possible pathways through the current difficulties:

(a) If you have to use genetically homogeneous animals, prefer F1 hybrids to inbred mice. Although each individual F1 stock is genetically uniform, at least you've ducked the homozygosity problem, and F1 mice are in general longer-lived, hardier beasts.

(b) Replicate key findings in multiple stocks. Once you've invested three years in proving something in CB6F1 mice, it may seem a waste of time to spend another few months checking the main points in two other F1 lines, but in the long run this may be more productive than spending the rest of your career chasing a finding that turns out only to apply to CB6F1.

(c) Consider the use of an animal stock with controlled heterogeneity, such as mice bred by a four way cross (Example: CB6F1 mothers crossed to C3D2F1 fathers). Such a cross yields an arbitrarily large group of full sibs; no two mice are genetically

identical, but each mouse shares half of its genetic code with any other randomly chosen animal in the pool. There is a slowly growing literature demonstrating the usefulness of such heterogeneous lines in aging research, and the NIA Office of Biological Resources plans to add such mice to their contract colonies soon.

**Sixth Principle: Do at least a quickie autopsy on each old mouse; if you can afford it, pay a pro to do a gross necropsy.** Even a pathological novice can take a quick look: is the spleen three times normal size? Are there little white bumps in the lung or liver? Battle-scars, probably infected, on the backs and legs? Skipping this precaution may make it easier to call the mice "apparently healthy" in the materials and methods section, but at the cost of making the results much harder to replicate and interpret. Tumors are common in aged rodents, and small tumors may not pose a problem for many experiments. However, as a rule, the simplest, and in some ways the best, plan is to toss out the data from mice that have large lesions - though of course this can be a very expensive rule to follow, particularly if you've ignored the First Principle.

Having a look yourself is better than not looking, but better still is to give the animal, after you've taken the tissue you need, to a veterinary pathologist or technician and ask them to have a look. This is fairly inexpensive, you get the written report back in a few weeks, and then you can go back through your notes and discard the data that came from the ones later found to have a serious illness.

Best of all is a histopathological autopsy. A thorough job costs $50 - $100/case; a quick microscopic look at the obvious lesions can cut the cost to $25 or so. This is very important if you are characterizing a new model (does the drug you administer to prevent neurodegeneration increase the incidence of liver abnormalities?). But if your main goal is to eliminate data from sick mice, the gross inspection is often adequate and much cheaper.

An alternate, more informative, approach would be to consider whether the presence of a specific form of illness modifies the age effect or treatment effect under study. Limited statistical power - are the differences between tumor-ridden and tumor-free mice big enough to be worth separate analysis? - and the difficulties of deciding the extent to which similar disease states can safely be lumped together can greatly complicate this variety of analysis. In any case it is important for the investigator to state explicitly the criteria used for elimination (or stratification) of individual animals, and the proportion of animals the met the inclusion criteria.

At what ages is the yield worth the cost? This will vary from strain to strain, and if you find that 95% of your 16 month old mice are free of lesions that might compromise your interpretation, then you may want to skip paying for the gross inspections and do it yourself. As a rule of thumb, we try to get a gross necropsy on mice over 18 months and rats 24 months or older. Your mileage may vary.

**Seventh principle: Don't pool unless you absolutely have to.** There are two potential problems with pooling cells and tissues. The first is that pooling rapidly increases the chance that the cells or tissues under study contain abnormal cells from a diseased subject. In a study of immunity, for example, the inclusion of even a fairly small proportion of lymphoma cells in a pool of otherwise normal cells may

strongly influence the result of the analysis. If sick mice make up 1/3rd of the population, each pool of 5 mice is very likely (83%) to include at least one diseased animal.

The second issue is a statistical one: the assessment of a statistical hypothesis ("mice aged 18 months express, on average, more of this gene than mice aged 6 months") depends on the number of individual mice, or pools of mice, tested independently. Thus an experiment in which a pool of 20 young mice is compared to a pool of 20 old mice has no greater statistical power than an experiment comparing one young to one old animal, i.e. none at all. The fewer mice used in each pool (ideally one mouse per pool), the more statistical power is achieved for the available mouse budget. Putting in the extra effort to miniaturize your test system to the point where it can be performed with material from a single mouse pays off handsomely in the long run.

**Eighth Principle: buy extra old mice to compensate for death and disease.** If you need 20 mice aged 24 months, don't buy 20 mice aged 24 months, because when you get around to using them 2 months later you'll have 18 live ones, of which only 10 will be free of visible lesions at necropsy. If you want 20 mice at age 24 months, buy 30 mice at age 22 months; use them 2 - 6 weeks later and discard the ones with lesions.

**Ninth Principle: Do a cost-adjusted power analysis and save a bundle.** OK, you know how to do a power analysis. You call the local statistician, and indicate that you're trying to figure out if male mice have more muscles than female mice, and that the three males you've tested so far have 1200 ug worth of the muscle in question, with a standard deviation of 200 mg, and that you'd consider it worth knowing if the sex difference were as great as 200 mg, so how many mice do you need to use? Would 10 of each group do the trick? The statistician plugs the values into a secret program, and tells you that if you use a $p < 0.05$ criterion for significance (and who wouldn't?), then with $N = 10$ per group you've got only a 58% chance of getting a significant result if the real difference is 200 mg. To get an 80% chance of detecting a difference of 200 mg, you're going to need to include 17 mice in each group. You then go out and buy some extra mice.

OK, next you want to know if the size of this muscle varies with age rather than with gender. So you do the same calculation, and get the same result: to get 80% power for detecting a difference between young and old of 200 mg, given the same assumptions above, you're going to need 17 young mice and 17 old mice. You go out and buy these 34 mice; actually you buy 17 young and more than 17 old, since some of the old animals will have to be discarded when you find out they have tumors.

Mistake: you've just wasted some money. The power analysis was done to calculate the minimum number of animals, but what you really want to do is get the maximal amount of statistical power per dollar spent. Because the old mice cost a lot more than the young ones, the cheapest way to get this statistical power is to buy slightly fewer old mice, and a good deal more young ones. If, for example, the cost of studying each young mouse (purchase cost plus cost of doing the assay) is $18, and the cost of studying each old mouse is $106, then the optimal solution is to buy 29

young mice and 12 old ones. If you buy 17 of each, you'll spend $2108; if you buy 29 plus 12, you spend $1797. You can allocate the $311 to the next experiment, or give your tech a well-earned raise.

Actually, the real savings can be very high indeed, particularly if you have to grow the mice yourself rather than obtain them from the NIA's highly subsidized colonies, and particularly if you count in the cost of the mice you couldn't use because they had serious disease. The real cost of a two year old mouse is not $106, but the $212 you spent to grow two of them, the one that died last week and one you've still got, or often the $424 you need to spend to get a tumor-free mouse. At $424 per old mouse, the minimal cost is to study 50 young and 10 old mice, at a cost of $5,242 and a savings of $2,272. At the NIA subsidized cost, you'll pay a mere $72 a head for the old mice, or $144 per tumor-free old mouse, and $22 per young one; so you'll save a mere $454 - less cost, and less savings than the home-grown variety, but still worth the cost of writing down the following formulas, which were derived by Andrzej Galecki of the University of Michigan's Geriatrics Center and Institute of Gerontology:

$$N_Y = 0.5 * N_e * (1 + SQRT(C_O/C_Y)$$

$$N_O = 0.5 * N_e * (1 + SQRT(C_Y/C_O)$$

Here's what to do: first do the regular old power analysis, that tells you have many mice you'll need if you use equal numbers of mice in each group. This was $N_e = 17$ mice per group in the example shown above. Then figure out how much it costs to do the assay for each young mouse, including purchase costs, supply costs, tech time, overhead; this number is CY, the cost per young mouse. Calculate CO, the cost per old mouse, in the same way, and be sure to throw in the adjustment for the number of mice you'll need to discard for disease. Then plug in the values and calculate NY, the number of young mice to buy, and NO, the number of old mice to buy.

**Tenth Principle: don't misinterpret artifactual correlations due to age effects.**This isn't really a principle of animal use, because it applies with equal force to studies of human aging, but the error is so common that it's worth noting in this context. Consider, for example, a hypothetical situation in which an investigator is interested in testing the hypothesis that the relatively low ability of T cells from old mice to secrete Interleukin-2 contributes to their low ability to make antibodies when injected with tumor cells. The investigator tests a group of mice, containing equal numbers of old and young animals, for both IL-2 and antibody production, and obtains the results shown in the left panel of Figure 2. He interprets this as good support for the idea that low IL-2 levels are indeed associated with poor antibody production, and since he knows from prior work that T cells from old mice do indeed make less IL-2 than cells from young mice, he concludes that the low IL-2 levels may underlie the poor antibody production.

The error in this inference is that the correlation between IL-2 and antibody production could well reflect the common influence on aging on both outcomes, rather than any direct connection between IL-2 production and antibody responses. The righthand panel of Figure 2 shows the same data set, but with triangles used to indicate the data from the old mice and circles to show data from young

donors. *Within* each group, there is no correlation between IL-2 and antibody, and the impression of a correlation conveyed by the first figure reflects the age influence on both traits. The older mice may also, compared to young animals, have more cataracts, weaker muscles, and a preference for classical music over grunge rock; plots of these measures against IL-2 production in a mixed-age group will also show excellent correlations that do not have tell us much about causal relationships.

## Executive Summary:

Use youngish mice for your "old" groups, and use older young adults instead of adolescents as young controls. Use some mice in the middle range, too.

Make sure they're specific pathogen-free, and free of the most obvious tumors and other significant diseases.

Try not to pool unless absolutely necessary.

Don't put clean SPF rodents into a conventional colony; the power of prayer, though redoubtable in some earthly domains, does not always fully prevent infections in rodents.

Shake the inbred habit – this is not your father's rodent. Pick F1s if you have to use genetically homogeneous stocks, and make sure to confirm your key findings in multiple strains.

*Table 1: Disease Incidence in Inbred Rodents: Some Sobering Anecdotes*

| Strain | Disease | Age | Incidence | Citation |
|---|---|---|---|---|
| C3H/He | Hepatoma | 14 | 85% | (8) |
| A/He | Pulmonary Adenoma | 18 | 90% | (8) |
| BALB/c | Lymphoma | 13 | 44% | (8) |
| SJL/J | Reticulum cell sarcoma | 13 | 91% | (8) |
| C57BR | Pituitary tumor | "old" | 33% | (8) |
| BALB/c female | Ovarian granulosa tumor | ?? | 76% | (8) |
| F344 male | Leydig cell tumor | 24 | 99% | (8) |
| F344 | glomerulonephropathy | 27 | >90% | (9) |
| > | | | | |
| (AxB6)F1 | At least one neoplasm | 18-24 | 25% | (10) |
| | | 30-35 | 70% | (10) |
| | | 36-41 | 94% | 10 |

*Moral 1: if you need 25 "disease-free" old inbred or hybrid mice, you had better obtain >> 25 mice.*
*Moral 2: Aged animals of Strain A may differ in known and unknown ways from those in Strain B, due in part to disease effects.*

**Legends to figures:** Some reasons to include middle-aged animals in a survey experiment. Panel A shows the hypothetical results of a study looking at three interleukins in the blood of young and old rats. The authors conclude that all three interleukins decline in parallel with aging - perhaps they share some common control mechanism on which one could base a Program Project application? Panel B shows a more comprehensive survey of the same topic conducted at another laboratory: IL-X drops early in adult life (a maturational effect), IL-Y shows a progressive decline through the lifespan, and IL-Z drops only when the animals get ill in very old age. Panel C shows three hypothetical reports of age effects on Ig-X levels in serum. Lab #1 says there's no change; Lab #2 says the levels go down; Lab #3 says the levels go up. Panel D shows the real hypothetical data set from which each lab drew its misleading conclusion: use of a wider and higher resolution set of age groups would have avoided much embarrassment and needless confusion.

Misleading inferences from age-confounded correlations. This hypothetical data set represents an attempt to test the idea that the low levels of IL-2 production seen in old mice contribute to poor antibody production in these mice. Unsophisticated investigators might conclude, from the left panel, that there is indeed a strong correlation between IL-2 production (already known to decline with age) and antibody production, and thus claim that their hypothesis is supported. The right panel, showing old and young individuals separately, reveals the fallacy: the correlation results from the high age-sensitivity of both measured traits. Any two traits strongly influenced by aging would generate a similar correlation, even if neither trait had any direct mechanistic relation to the other.