Published in final edited form as:

Radiat Res. 2013 December ; 180(6): 567–574. doi:10.1667/RR13429.1.

Practical Advice on Calculating Confidence Intervals for Radioprotection Effects and Reducing Animal Numbers in Radiation Countermeasure Experiments

Reid D. Landes^{a,1,2}, Shelly Y. Lensing^a, Ralph L. Kodell^a, and Martin Hauer-Jensen^b

^aDepartment of Biostatistics, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205

^bDivision of Radiation Health, University of Arkansas for Medical Sciences, and Central Arkansas Veterans Healthcare System, Little Rock, Arkansas 72205

Abstract

The dose of a substance that causes death in $P\%$ of a population is called an LD_P , where LD stands for lethal dose. In radiation research, a common LD_P of interest is the radiation dose that kills 50% of the population by a specified time, i.e., lethal dose 50 or LD_{50} . When comparing LD_{50} between two populations, relative potency is the parameter of interest. In radiation research, this is commonly known as the dose reduction factor (DRF). Unfortunately, statistical inference on dose reduction factor is seldom reported. We illustrate how to calculate confidence intervals for dose reduction factor, which may then be used for statistical inference. Further, most dose reduction factor experiments use hundreds, rather than tens of animals. Through better dosing strategies and the use of a recently available sample size formula, we also show how animal numbers may be reduced while maintaining high statistical power. The illustrations center on realistic examples comparing LD_{50} values between a radiation countermeasure group and a radiation-only control. We also provide easy-to-use spreadsheets for sample size calculations and confidence interval calculations, as well as SAS® and R code for the latter.

INTRODUCTION

Experiments are routinely conducted to learn the radiation dose that is lethal to 50% of a population – LD_{50} . Additionally, researchers often compare LD_{50} between two treatment groups with a dose reduction factor (DRF) – a ratio of two LD_{50} . Typically, the LD_{50}

©2013 by Radiation Research Society. All rights of reproduction in any form reserved.

¹Address for correspondence: Department of Statistics, Radiation Effects Research Foundation, Hijiyama-koen 5-2, Minami-ku, Hiroshima-shi 732-0815, Japan; rldandes@uams.edu and rldandes@rerf.or.jp. .

²Written while at the Department of Biostatistics, University of Arkansas for Medical Sciences, Little Rock, AR 72205-7199.

SUPPLEMENTARY INFORMATION Supplementary material mentioned through out this article; <http://dx.doi.org/10.1667/RR13429.1.S1>.

Supplementary spreadsheet based on logistic regression. Excel file with several spreadsheets, some containing data and others “tools” that one can enter routine statistical results to obtain the results described in this article; <http://dx.doi.org/10.1667/RR13429.1.S2>.

Supplementary spreadsheet based on probit regression. Excel file with several spreadsheets, some containing data, and others “tools” into which one can enter routine statistical results to obtain the results described in this article; <http://dx.doi.org/10.1667/RR13429.1.S3>.

Supplementary material – “R” code accompanying text. Computer code for generating the data, performing the analyses and displaying much of the results in this article. This computer code is for a statistical software entitled “R”; <http://dx.doi.org/10.1667/RR13429.1.S4>.

Supplementary material – “SAS” code accompanying text. Computer code for generating the data, performing the analyses and displaying much of the results in this article. This computer code is for a statistical software entitled “SAS”; <http://dx.doi.org/10.1667/RR13429.1.S45>.

associated with irradiated animals receiving a radiation countermeasure is divided by the LD_{50} associated with a radiation-only control group. Hence, an effective countermeasure will have $DRF > 1$, that is, a higher (better) LD_{50} than control.

Studies evaluating radiation countermeasure efficacy occur with some regularity, but inference on LD_{50} and DRF does not. Using Google Scholar®, we searched journals with “radiation research” in the title over the years 2003–2013 for articles having “dose reduction factor,” “dose modification factor” or “radiation protection factor” anywhere in the article. We found 22 articles that used animal lethality experiments to evaluate efficacy of a radiation countermeasure (1–22; see Supplementary Information section for links to supplementary materials with summaries of LD_{50} and DRF results). Sixteen studies (73%) did not provide statistical inference for their reported dose reduction factor. Additionally, 10 of the 22 studies used statistically inefficient designs for detecting $DRF > 1$. Any concern over the lack of formal statistical inference or efficient designs can perhaps be somewhat alleviated when considering the median number of animals used in these studies was 185, with a minimum of 60 and maximum of 620. Still, using relevant statistical methods would improve reported inferences while substantially reducing the numbers of animals required.

Confidence intervals (CIs) provide a range of plausible values for a parameter of interest, like LD_{50} or DRF . Planners of future studies and importantly, regulatory agencies often want to know worst- and best-case scenarios, CIs can better inform these scenarios than point estimates (i.e., single “best” values) alone, even when point estimates are accompanied by a P value from significance testing. Several methods exist for calculating CIs for LD_{50} and by extension, $DRFs$ (23–25). Some have slightly better statistical properties than others; some are easier to calculate than others. Radiation researchers not specializing in statistics would need and would benefit from an easy-touse CI method that has good statistical properties.

When a dose reduction factor is truly greater than 1, applying the same radiation doses to both treatment groups is statistically inferior for estimating dose reduction factor compared to staggering the doses between the groups. All 22 studies mentioned above clearly expected a $DRF > 1$, but nearly half used a “same-dose” design. An explanation of how staggered-dose designs improve power for detecting $DRF > 1$ may benefit those who conduct these experiments.

Recently, Kodell *et al.* (26) produced a sample-size formula for detecting $DRF > 1$. The current article expands on that work to help researchers plan studies with appropriate power, using sample sizes that may be notably smaller than those historically used. Having a sample-size formula is helpful, but there are several inputs into the formula that may not be well understood by non-statisticians, therefore, some examples should encourage its use among radiation researchers.

Using realistic examples, we illustrate (i) how to obtain confidence intervals for LD_{50} and DRF , (ii) how staggered-dose designs improve statistical power over same-dose designs and (iii) how to use previous data to design a prospective radiation countermeasure experiment with appropriate power (usually between 0.80–0.90). Underlying these objectives, we hope to convince the reader that sound statistical methods can substantially reduce the numbers of animals used in radiation countermeasure studies – an important goal in terms of both animal welfare and cost. Implementing these methods is easy with the Excel spreadsheets provided (see Supplementary Information section for the links to supplementary material Excel spreadsheets), the examples herein are provided in the spreadsheets and will assist users in understanding how the tools may be used for their specific situation.

CONFIDENCE INTERVALS FOR LD₅₀ and DRF

Our hypothetical experiment seeks to compare the total-body irradiation dose that, by day 30, kills 50% (henceforth LD₅₀) of animals given or not given a radioprotectant. We let $R = 0$ or $R0$ represent unprotected animals, and $R = 1$ or $R1$ represent animals receiving radioprotectant. The parameter of primary interest is the dose reduction factor: $DRF = LD_{50(1)}/LD_{50(0)}$. Table 1 contains hypothetical data from such an experiment. The data elements needed for estimating LD₅₀ and DRF are treatment group assignment ($R0$ and $R1$), log-radiation dose ($\log X$), group size (N) and the number of animals dead by the designated time point (Y). (Although technically redundant, two variables, $R0$ and $R1$ are used to identify the two treatments when we specify two equivalent models for estimating, first, LD₅₀ and then, DRF.) Table 1 also includes the radiation doses (X) in Gy and “TargetP” – a variable to be explained below. We present Table 1 in a format suitable for most statistical software.

Statistical Analysis

We model Y with probit regression³ that accounts for radioprotectant, $R0$ and/or $R1$ and log-radiation dose, $\log X$. The regression assumes the slope on $\log X$ is the same for both treatment groups, but allows their LD₅₀ to differ. We consider two equivalent statistical models of Y . They both return the same predictions, but one is more useful for estimating LD₅₀ and the other for DRF.⁴ Fitting these models with the maximum likelihood method⁵ returns two things: regression estimates and a covariance matrix for the regression estimators. Point estimates for LD₅₀ and DRF are functions of the regression estimates. For confidence intervals, we recommend Wald’s method because it is formula-based, and has good statistical properties when compared to other CI methods.⁶ Wald’s CIs for LD₅₀ and DRF are functions of both the regression estimates and certain variances/covariances of those estimates (i.e., select elements of their covariance matrix).

LD₅₀ Model

Enter $R0$, $R1$ and $\log X$ into the model to obtain $\alpha_0 R0 + \alpha_1 R1 + \beta \log X$. By default, most statistical software routines will insert an additional “intercept” parameter into this model, unless the user specifies it be excluded. This model should exclude the default intercept parameter. Statistical software should then provide estimates a_0 , a_1 and b , for α_0 , α_1 and β , respectively, and likely upon request (i.e., typically not by default), the estimated covariance matrix,

$$V = \begin{bmatrix} \nu_{00} & \nu_{01} & \nu_{0b} \\ \nu_{01} & \nu_{11} & \nu_{1b} \\ \nu_{0b} & \nu_{1b} & \nu_{bb} \end{bmatrix}.$$

³Others may use logistic regression. Methods recommended in this article are valid with logistic regression, too.

⁴Technically, the models are reparameterizations of each other. Parameterizing the models as we will show below allows us to use essentially the same confidence interval formula. Using the same model to estimate both LD₅₀ and DRF would lead to more involved confidence interval formulae.

⁵Many statistical software packages use the maximum likelihood method as default when fitting the regression; most have maximum likelihood as an option. This article assumes maximum likelihood estimation; results/formulae in this article differ for other estimation methods (e.g., exact, Bayes, etc.).

⁶Landes RD, Kodel RL, Lensing SY, Hauer-Jensen M. Recommendations for estimators of LD₅₀ and dose reduction factors in animal lethality studies. Presented at the 58th annual meeting of the Radiation Research Society, San Juan, Puerto Rico (2012, October).

The diagonal elements of V are variances for a_0 , a_1 and b . Off-diagonal elements are covariances of the two implicated estimators. For example, v_{bb} is the variance of b and v_{1b} is the covariance between a_1 and b .

DRF Model

Enter $R1$ and $\log X$ into the model to obtain $\alpha_0 + \alpha^*_1 R1 + \beta \log X$. Unlike the above LD_{50} model, this model should *include* the (often) default intercept parameter. Here, α_0 is the default intercept; its estimate should be a_{20} – the same as from the LD_{50} model. The α^*_1 is the difference between radioprotection and none; its estimate, a^*_1 , should equal $a_1 - a_0$ from the LD_{50} model. The estimated covariance matrix is

$$V^* = \begin{bmatrix} \nu_{00} & \nu_{01}^* & \nu_{0b} \\ \nu_{01}^* & \nu_{11}^* & \nu_{1b}^* \\ \nu_{0b} & \nu_{1b}^* & \nu_{bb} \end{bmatrix}.$$

We provide formulas for point and interval estimators for log- LD_{50} and log-DRF in Table 2. Table 3 contains results from statistical analyses of our example data. Taking the antilog of resulting estimates gives the point and interval estimates of LD_{50} and DRF. For the $R0$ group, LD_{50} was $10^{0.878} = 7.55$ Gy, and the 90% CI was $(10^{0.862}, 10^{0.894})$ or (7.27, 7.83). And for the $R1$ group, LD_{50} (90% CI) was 8.99 (8.66, 9.33) Gy. The DRF when adding radioprotectant was $10^{0.076} = 1.19$ with a 90% CI of (1.13, 1.25). Besides quantifying the precision with which we know a DRF or LD_{50} , CIs may be used for hypothesis testing. From the example above, because the lower bound of the 90% CI is greater than 1.1, then a 0.05 significance level test of $DRF = 1.1$ (a one-sided null hypothesis) is rejected. Of course, the usual null hypothesis that $DRF = 1$ is also rejected. Our supplementary materials include a spreadsheet (see Supplementary Information section) that the user may enter needed elements from Table 2 formulas to arrive at results like those in Table 3, as well as SAS and R code, which also produce Table 3 results.

STAGGERED-DOSE VS. SAME-DOSE DESIGNS

In experiments evaluating the dose reduction factor of a countermeasure, staggering radiation doses between the control and treatment groups to achieve (approximately) equal target lethality results in “staggered-dose” designs. “Same-dose” designs occur when both groups receive the same radiation doses and inadvertently, different target lethality. Compared to same-dose designs, staggered-dose designs tend to better ensure variability in lethality across the radiation doses. It is this lethality variability across radiation doses that gives understanding into how lethality is related to radiation dose (see Supplementary Information section for an example found in the supplementary materials).

Consider a hypothetical researcher studying a promising radiation countermeasure,⁷ based on previous experiments, he believes the LD_{50} for untreated animals is about 7.3 Gy. He also knows that a dose reduction factor of at least 1.2 is clinically meaningful and hopes to find the countermeasure to have such a dose reduction factor. Due to budget constraints, he has only 60 mice to work with and most studies he has read about in the past use at least 5 radiation doses, so he chooses to expose both untreated and treated animals to total-body

⁷We base this example on 20 studies identified with “dose reduction factor” in our literature review. The minimum number of animals used was 60. The median number of radiation doses examined was 5. The median LD_{50} for untreated animals was 7.28 with an interquartile range (IQR) of 6.62 – 8.43. The median DRF was 1.22 with IQR of 1.12 – 1.30. The median slope on probit regression lines was 25.3 with IQR of 17.5 – 42.3. After the initial review, we added “dose modification factor” and “radiation protection factor” to the search strategy, picking up an additional two studies (5, 12).

irradiation doses of 6, 7, 8, 9 and 10 Gy. Suppose the *true* control LD₅₀ is 7.28 and the *true* DRF is 1.22, both unbeknownst to the researcher. Table 4 contains the most probable result for each treatment × dose group (i.e., each radiation-dose group within a treatment). If these were the observed data, the 6 dead control mice receiving 10 Gy provide little additional information to that already provided by the 6 dead control mice receiving 9 Gy. Similarly, the 0 dead treated mice receiving 6 and 7 Gy tell us little more than the 0 dead treated mice receiving 8 Gy. Repeated 0% or 100% responses at adjacent extreme doses provide very little useful information to inform the dose-response relationship (see Supplementary Information section for an example found in the supplementary materials).

The researcher's assignment of the same doses to both groups is statistically inefficient for detecting a clinically meaningful dose reduction factor (and actually estimating DRF). Using information from other experiments (i.e., not the experiment being designed), the researcher had sensible estimates for two of three factors needed for designing a statistically efficient dosing strategy: (i) a control LD₅₀ estimated from previous experiments and (ii) a clinically meaningful dose reduction factor. The remaining factor is (iii) an estimated log-dose slope – how slowly or quickly lethality increases with increasing radiation dose; i.e., an estimated β from the LD₅₀ and DRF models above. Assuming $\beta = 20$ for our hypothetical researcher, along with his estimated LD₅₀₍₀₎ = 7.3 Gy and DRF = 1.2, Table 1 contains a more efficient dose design where the *X* column denotes doses in Gy. Importantly, the control and treated groups have the *same targeted* % lethality (see TargetP in Table 1), which result in *staggered doses*. The supplementary materials contain a spreadsheet and R code in which the user can input an estimated control LD₅₀ and slope, an anticipated DRF, select lethality targets and treatment × dose sample size to produce radiation doses (Gy) and the expected number of deaths for each dose.

When the true DRF is greater than 1, same-dose designs have less power for detecting DRF > 1 than staggered-dose designs. Figure 1 compares power between a staggered-dose design and a same-dose design. Two important points should be taken from this figure: first, using 30 animals ($N = 3$ per treatment × dose group) in a staggered-dose design for our hypothetical experiment is estimated to have power close to 0.90 for detecting a DRF = 1.2 compared to 50 animals in a same-dose design (i.e., a 40% reduction in animal numbers when going from a same-dose to staggered-dose design). Second, with 60 animals in this realistic example, the same-dose design has over 0.95 power for detecting a clinically meaningful difference. We reiterate that 60 animals was the *minimum* number used in the 22 studies we reviewed. Therefore, with proper planning, experiments may require substantially fewer animals than what has commonly been used.

CALCULATING SAMPLE SIZE

We now show how to calculate the sample size required to achieve adequate statistical power when testing whether the dose reduction factor of a potential countermeasure is clinically meaningful with the following example.

As in the examples above, one set of animals will receive a radiation countermeasure and a control set will not. The countermeasure's dose reduction factor is the statistical parameter of interest. We want to design an experiment that will have about 0.90 power to detect a DRF of 1.2 with a 0.05 significance level, one-sided test. We will consider experiments with 3, 5 and 7 doses per treatment and use the experiment that requires fewest animals.

Kodell *et al.* (26) developed the following sample size formula (Eq. 14, pp. 242) for these types of experiments. The formula assumes (i) two treatment groups (e.g., control and radioprotective countermeasure) and that each group has (ii) the same slope, b , (iii) the same

number of animals, N , at each radiation dose and (iv) the same target lethalties (i.e., uses a staggered-dose design).

$$N \cong \frac{(t_{f,\alpha} + t_{f,\beta^*})^2 \left(2 / \sum_{i=1}^g \omega_i \right)}{(\text{blog} \rho)^2} \quad (1)$$

Here,

- N is the sample size for a treatment $\times$ dose group;
- g is the number of dose groups, assumed equal for each treatment;
- $t_{f,\alpha}$ and t_{f,β^*} are the $100(1 - \alpha)^{\text{th}}$ and $100(1 - \beta^*)^{\text{th}}$ percentiles from a t distribution having f degrees of freedom, where
- $f = 2g - 3$,
- α is the significance level, and
- $1 - \beta^*$ is the power, thus β^* is the probability of a Type II error;
- w_i is a weight, related to the target lethality of dose i for $i = 1$ to g ;
- b is an estimated slope; and
- ρ is the DRF to be detected.

From our initial information, we have $\alpha = 0.05$, $\beta^* = 1 - 0.90$, $\rho = 1.2$, and $g = 3, 5$ or 7 . The values of $t_{f,\alpha}$ and t_{f,β^*} are readily available from tables or t statistics' calculators on the internet. For w , a function of the lethality targets,⁸ we choose the lethality targets for the 3-, 5- and 7-dose designs that are presented in Table 5; see supplementary materials

(Supplementary Information section) for calculation of $\sum_{i=1}^g$ for the designs in Table 5. The last piece is b . To estimate b , we use previous data from control mice, exposed to total body irradiation, all treated under protocols equivalent to the planned experiment. We observed the following lethality per total mice: 1/16 at 7 Gy, 3/16 at 8 Gy, 3/8 at 8.5 Gy, 15/24 at 9 Gy, 6/8 at 9.5 Gy and 15/16 at 10 Gy. Using probit regression of the proportion dead on $\log_{10}$ -dose, our b (95% CI) was 20.26 (12.61, 27.91). Table 5 contains sample size calculations for 3-, 5- and 7-dose designs assuming a slight, moderate and steep slope (b) represented by our point estimate and confidence limits.

We learn two important concepts from Table 5. First, within each column, we see that increasing the number of doses decreases N – the number of required animals per treatment $\times$ dose. The number of doses is something that researchers can control. Hereafter, given a fixed number of animals, it is usually better to have fewer animals over more doses than more animals over fewer doses. Second, within each row, as the probit log-dose slope, b , moves away from 0, required animal numbers tend to decrease. However, the actual slope is unknown. Using sample sizes based on the slope furthest from 0 (e.g., $b = 27.91$ above) represents the “best-case scenario” in DRF experiments, and is seldom realistic; save if it is in line with past estimates from multiple similar experiments. Usually, the point estimate (e.g., $b = 20.26$ above) is most sensible in terms of statistically justifying the number of animals. But using a b closer to 0 (e.g., $b = 12.61$ above) may also be justified. For example, the point estimate may be notably higher than expected or higher than other past estimates;

⁸For probit regressions, the weight for target lethality, P_i is $w_i = [\phi\{\Phi^{-1}(P_i)\}]^2 / P_i(1 - P_i)$, where $\phi(*)$ and $\Phi(*)$ are, respectively, the probability density function and cumulative distribution function for the standard normal distribution. For logistic regressions, the weight is $w_i = P_i(1 - P_i)$.

or the implications of *not* detecting a true DRF > 1 (a Type II statistical error) are severe. In cases like these two, choosing the more shallow slope may be prudent.

By its very structure, Eq. (1) offers a way to help relieve the dilemma of which slope to choose. When using probit regression, the denominator – the only place where the slope (b) and DRF (ρ) appear – is the square of the effect size.⁹ The effect size is $(b \log \rho)$. If uncertain about the anticipated slope and/or DRF, entering a reasonable effect size will return an appropriate sample size given the other inputs in Eq. (1). Since “reasonable” depends on context, careful consideration of the future experiment’s context is important. Seventeen of the 22 DRF studies we reviewed had enough information to obtain effect size estimates, ranging from 0.75 to 3.77 with a median of 1.67.

Returning to our example in Table 5, we see that given a dose reduction factor of 1.2, the estimated slope from the fitted line, $b = 20.26$, corresponds to an effect size of 1.6 SD units. This is a reasonable target to detect. An effect size of 1.0, corresponding with the most shallow slope, $b = 12.61$, requires substantially more animals. Conversely, the largest effect size of 2.2 (and steepest slope of $b = 27.91$) needs noticeably fewer animals to maintain desired power. If a 2.2 SD effect size is considered relevant, then for the same total of about 30 animals for each of the three designs (Table 5), we recommend a 3-dose or 5-dose design over a 7-dose design. A design with only 1 or 2 animals per group is generally not recommended, even though in theory it will provide nominal power (also see ref. 26). Correctly reporting how we arrived at our sample size (including references) is important for regulatory bodies (e.g., institutional animal care and use committees) and funding agencies and is appreciated in journal articles. Ideally, reports should provide all inputs needed to arrive at our calculated sample size; thus making calculations reproducible for those who need or want to verify. From Table 5, we choose the 5-dose design assuming a $b = 20.26$. Emphasizing all inputs with italics, we may report our sample size calculation as follows.

“Assuming the *probit* slope of $\log_{10}$ -radiation dose is 20.26, and the true DRF is 1.2, and with control and treated animals exposed to LD_5 , $LD_{27.5}$, LD_{50} , $LD_{72.5}$ and LD_{95} , then 4 animals for each treatment $\times$ dose combination (40 in total) will provide 0.90 power on a one-sided 0.05-significance level test.”

DISCUSSION

LD_{50} and DRF estimation are primary objectives in radiation countermeasure research as mandated by the U.S. FDA (28, 29). Sound statistical methods aimed at these goals are not easily implemented with conventional statistical software. Through examples, we described and demonstrated here how to compute confidence intervals (CIs) for LD_{50} and DRFs. We illustrated how staggered-dose designs provide more efficiency than same-dose designs. And to researchers outside of statistics, we introduced a sample size formula specially suited for designing DRF experiments with appropriate power. Finally, we provided supplementary easy-to-use Excel spreadsheets, as well as SAS and R code files (see Supplementary Information section), which will facilitate implementation of the methods described here. Although we have emphasized radiation research, these methods can be useful in other settings that seek to estimate the relative potency of drugs or vaccines.

The vast majority of DRF studies seek to detect whether the proposed countermeasure provides a meaningful increase in LD_{50} , with “meaningful” preferably defined in each case (for example DRF > 1.1). Typically, investigators have used large numbers of animals to

⁹For probit regression (as described in this article), $1/b$ is the standard deviation of the log-dose distribution; hence, the effect size is expressed in SD units. The effect size for logistic regression differs and is beyond the scope of this article.

answer this question. Since formal statistical inference for dose reduction factor has not been the norm, perhaps a reason researchers have used these large numbers of animals to feel relatively sure their point estimates were precise. We have addressed this important question through analysis of CIs for DRF. With CIs also providing a range of plausible values for the true DRF. Although precise (i.e., narrow) CIs are preferred, they come at a high price – many expended animals. The diminishing return of excessive animals is clear when we focus on the motivating question rather than on high precision. The sample size formula illustrated in this paper shows we can address the primary question and dramatically reduced the numbers of animals traditionally used, oftentimes by greater than half.

The sample size obtained from the sample size formula (Eq. 1) implemented in the spreadsheet is only as good as the estimates inputted by the researcher. Researchers can draw from their own and others' experience in informing the estimates of lethality (e.g. LD₅₀) and log-dose slope. Although data may not be available for the countermeasure agent of interest, data from similar experiments may be informative (see supplementary materials for LD₅₀, slopes and DRFs from the review of studies described in the Introduction section). As a starting point the researcher should ask the following for each treatment group (e.g., countermeasure and control): "What is the highest radiation dose that will likely leave all of the animals alive?" and "What is the lowest radiation dose that will likely kill all of the animals?" Any doses more extreme will likely be noninformative and a waste of resources. Because these questions can be asked for each treatment group individually, staggered-dose designs are a natural (and efficient) result. However, despite researchers' best efforts, their low and high-dose estimates are still guesses, albeit educated ones. For this reason, it is important to evaluate a range of scenarios, not only expected ones, but particularly realistic worst-case scenarios. Designing efficient studies requires careful thought and time, but in the long term will save even more time and resources.

Inherent in the sample size formula are the following design recommendations;

- Use staggered-dose designs rather than same-dose designs. The former are more efficient, with efficiency increasing as the true DRF moves further from 1 and as the slope steepness increases.
- For a set number of animals, use more rather than fewer radiation doses (see also 30, 31). This comes with the caveat that having very small numbers of animals at each dose (e.g., 1 or 2) may not be as powerful in practice as it is in theory.
- Use at least 3 radiation doses per treatment, with the middle dose at the estimated or projected LD₅₀ for that treatment. Note: though placing all of the animals for each treatment at its true LD₅₀ will yield the most precise estimate of DRF, the true LD₅₀ is not known; hence our recommendation.

It is important to note that these design features are all under the researcher's control, and should be capitalized upon.

Beyond the scope of this article are the following theoretical considerations. Sometimes with probit and logistic regression modeling there is no unique maximum likelihood solution (usually indicated with a warning message by the software). In those cases, Wald intervals fail to exist or are bounded by 0 or infinity. For these instances, other confidence interval methods may provide valid intervals, such as the inverted likelihood ratio test, Bayesian, or bootstrap intervals. In addition, it is important to remember that Eq. (1) is based on what is called the delta method, but we are recommending using Wald confidence intervals. The two methods differ slightly, so actual power may not be the claimed power, but Kodell *et al.* (26) have demonstrated good performance for Wald tests using this formula. Finally, we suggest that researchers keep in mind that their experiments contain rich survival information that is

reduced when examining LD₅₀ and DRF endpoints; other survival analyses may be more informative (e.g. 32).

We hope we have convinced the reader that too many animals are often being unnecessarily sacrificed, given that study objectives can be achieved with fewer animals. The 4 “R”s in animal research are:

- Replacement (of animals with alternative techniques);
- Refinement (of design and analysis);
- Reduction (of animal numbers); and
- Responsibility (for conducting the best studies).

Using our methods and tools, our hope is that researchers can easily accomplish 3 of these 4 “R”s.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was partially supported by the Translational Research Institute (TRI), grant UL1TR000039 through the NIH National Center for Research Resources and National Center for Advancing Translational Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. MJH received support from U19 AI67798 (NIAID) and the Veterans Administration.

REFERENCES

1. Monobe MS, Koike A, Uzawa. Effects of beer administration in mice on acute toxicities induced by X rays and carbon ions. *J Radiat Res.* 2003; 44:75–80. [PubMed: 12841603]
2. Samarth RM, Kumar A. Radioprotection of Swiss albino mice by plant extract *Mentha piperita* (Linn.). *J Radiat Res.* 2003; 44:101–109. [PubMed: 13678338]
3. Song JY, Han SK, Bae KG, Lim DS, Son SJ, Jung IS, et al. Radioprotective effects of ginsan, an immunomodulator. *Radiat Res.* 2003; 159:768–774. [PubMed: 12751959]
4. Jagetia GC, Baliga MS, Venkatesh P, Ulloor JN. Influence of ginger rhizome (*Zingiber officinale* Rosc) on survival glutathione and lipid peroxidation in mice after whole-body exposure to gamma radiation. *Radiat Res.* 2003; 160:584–592. [PubMed: 14565823]
5. Satyamitra M, Um Devi P, Murase H, Kagiya VT. *In vivo* postirradiation protection by a vitamin E analog, α -TMG. *Radiat Res.* 2003; 160:655–661. [PubMed: 14640781]
6. Anzai K, Furuse M, Yoshida A, Matsuyama A, Moritake T, Tsuboi K, et al. *In vivo* radioprotection of mice by 3-methyl-1-phenyl-2-pyrazolin-5-one (Edaravone; Radicut®), a clinical drug. *J Radiat Res.* 2004; 45:319–323. [PubMed: 15304976]
7. Jagetia GC, Baliga MS, Venkatesh P. Influence of seed extract of *Syzygium Cumini* (Jamun) on mice exposed to different doses of γ -radiation. *J Radiat Res.* 2005; 46:59–65. [PubMed: 15802860]
8. Krishna, Kumar A. Evaluation of radioprotective effects of *Rajgira* (*Amaranthus paniculatus*) extract in Swiss albino mice. *J Radiat Res.* 2005; 46:233–239. [PubMed: 15988142]
9. Sharma M, Kumar M. Radioprotection of Swiss albino mice by *Myristica fragrans* houtt. *J Radiat Res.* 2007; 48:135–141. [PubMed: 17314471]
10. Anzai K, Ikota N, Ueno M, Nyui M, Kagiya TV. Heat-treated mineral-yeast as a potent post-irradiation radioprotector. *J Radiat Res.* 2008; 49:425–430. [PubMed: 18480592]
11. Gaur A. Ameliorating effects of genestein: Study on mice liver glutathione and lipid peroxidation after irradiation. *Iran J Radiat Res.* 2010; 7:187–199.

12. Brown SL, Kolozsvary A, Liu J, Jenrow JA, Ryu S, Kim JH. Antioxidant diet supplementation starting 24 hours after exposure reduces radiation lethality. *Radiat Res.* 2010; 173:462–468. [PubMed: 20334518]
13. Zhang C, Lin J, Cui J, Li B, Liu C, Wang J, Gao F, Cai J. Radioprotection of bone marrow hematopoiesis by CpG-oligo-deoxynucleotides administered to mice after total-body irradiation. *J Radiat Res.* 2011; 52:828–833. [PubMed: 22104272]
14. Satyamitra MM, Kulkarni S, Ghosh SP, Mullaney CP, Condliffe D, Srinivasan V. Hematopoietic recovery and amelioration of radiation-induced lethality by the vitamin E isoform δ -tocotrienol. *Radiat Res.* 2011; 175:736–745. [PubMed: 21434782]
15. Satyamitra MM, Lombardini E, Graves J III, Mullaney C, Ney P, Hunter J, et al. A TPO receptor agonist, ALXN4100TPO, mitigates radiation-induced lethality and stimulates hematopoiesis in CD2F1 mice. *Radiat Res.* 2011; 175:746–758. [PubMed: 21476857]
16. Singh VK, Ducey EJ, Fatanmi OO, Singh PK, Brown DS, Purnal A, et al. CBLB613: A TLR 2/6 agonist, natural lipopeptide of *Mycoplasma arginini* as a novel radiation countermeasure. *Radiat Res.* 2012; 177:628–642. [PubMed: 22175300]
17. Singh VK, Christensen J, Fatanmi OO, Gille D, Ducey EJ, Wise SY, et al. Myeloid progenitors: A radiation countermeasure that is effective when initiated days after irradiation. *Radiat Res.* 2012; 177:781–791. [PubMed: 22559205]
18. Peebles DD, Soref CM, Copp RR, Thunberg AL, Fahl WE. ROS-scavenger and radioprotective efficacy of the new PrC-2010 aminothiols. *Radiat Res.* 2012; 178:57–68. [PubMed: 22702647]
19. Asadullina NR, Usacheva AM, Gudkov SV. Protection of mice against X-ray injuries by the post-irradiation administration of inosine-5'-monophosphate. *J Radiat Res.* 2012; 53:211–216. [PubMed: 22510593]
20. Ghosh SP, Kulkarni S, Perkins MW, Hieber K, Pessu RL, Gambles K, et al. Amelioration of radiation-induced hematopoietic and gastrointestinal damage by Ex-RAD® in mice. *J Radiat Res.* 2012; 53:526–536. [PubMed: 22843617]
21. Medhora M, Gao F, Fish BL, Jacobs ER, Moulder JE, Szabo A. dose-modifying factor for captopril for mitigation of radiation injury to normal lung. *J Radiat Res.* 2012; 53:633–640. [PubMed: 22843631]
22. Wang B, Tanaka K, Morita A, Ninomiya Y, Maruyama K, Fujita K, et al. Sodium orthovanadate (vanadate), a potent mitigator of radiation-induced damage to the hematopoietic system in mice. *J Radiat Res.* 2013 doi: 10.1093/jrr/rrs140.
23. Racine A, Grieve AP, Flühler H. Bayesian methods in practice: Experiences in the pharmaceutical industry. *Journal of the Royal Statistical Society-Series C.* 1986; 35:93–150.
24. Williams DA. Interval estimation of the median lethal dose. *Biometrics.* 1986; 42:641–645. [PubMed: 3567296]
25. Kelly GE. The median lethal dose-design and estimation. *Journal of the Royal Statistical Society-Series D.* 2000; 50:41–50.
26. Kodell RL, Lensing SY, Landes RD, Kumar KS, Hauer-Jensen M. Determination of sample sizes for demonstrating efficacy in radiation countermeasures. *Biometrics.* 2010; 66:239–248. [PubMed: 19432769]
27. Abdelbasit KM, Plackett RL. Experimental design for binary data. *J Am Stat Assoc.* 1983; 78:90–98.
28. FDA (U. S. Food and Drug Administration). New drug and biological drug products: evidence needed to demonstrate effectiveness of new drugs when human efficacy studies are not ethical or feasible. *Federal Register.* 2002; 67:37988–37998. [PubMed: 12049094]
29. FDA (U.S. Food and Drug Administration). Guidance for Industry: Animal models – essential elements to address efficacy under the Animal Rule (draft guidance). Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. 2009. www.fda.gov/cber/guidelines.htm
30. Brown BW. Planning a quantal assay of potency. *Biometrics.* 1966; 22:322–329. [PubMed: 5961448]

31. Karger CP, Hartmann GH. Determination of tolerance dose uncertainties and optimal design of dose response experiments with small animal numbers. *Strahlentherapie und Onkologie*. 2001; 177:37–42. [PubMed: 11200111]
32. Landes RD, Lensing SY, Kodell RK, Hauer-Jensen M. Statistical analysis of survival data from radiation countermeasure experiments. *Radiat. Res.* 2012; 177:546–554. [PubMed: 22401302]

FIG. 1.

Estimated power as a function of sample size (N), compared between a staggered-dose design (solid line) and a same-dose design (dashed line) for a 5-dose experiment. The staggered-dose design has radiation doses for R_0 and R_1 given in Table 1; the same-dose design has the R_0 doses in Table 1 assigned to both groups. Power is based on a significance level of 0.05 for a one-sided Wald test with true $LD_{50} = 7.3$ Gy for R_0 , true probit log-dose slope $\beta = 20$ and true $DRF = 1.2$. N is for a radiation dose group within a treatment, and is assumed constant across all radiation doses and for both treatments, so total sample size is $N/\text{dose} \times 5 \text{ doses/treatment group} \times 2 \text{ treatment groups} = 10N$.

TABLE 1Hypothetical Data Set for Comparing LD₅₀ Between Radioprotectant and Control Groups

<i>R0</i>	<i>R1</i>	<i>TargetP</i>	<i>X</i>	<i>logX</i>	<i>Y</i>	<i>N</i>
1	0	10	6.30	0.7992	0	6
1	0	30	6.87	0.8371	1	6
1	0	50	7.30	0.8633	0	6
1	0	70	7.75	0.8894	5	6
1	0	90	8.46	0.9274	6	6
0	1	10	7.56	0.8784	0	6
0	1	30	8.25	0.9163	1	6
0	1	50	8.76	0.9425	3	6
0	1	70	9.31	0.9687	4	6
0	1	90	10.15	1.0066	5	6

Notes. *R0* and *R1* indicate absence and presence of radioprotectant, respectively; *X* and *logX* are, respectively, the radiation dose (Gy) and log₁₀-dose corresponding to the indicated % target lethality (*TargetP*); *Y* is how many of *N* animals were dead by day 30. Radiation doses (*X*) are for a staggered-dose design when the estimated LD₅₀ for control is 7.3 Gy, the anticipated DRF is 1.2, and an estimated probit log-dose slope, *b*, is 20. To adjust the doses for a different control LD₅₀, multiply *all* doses by (new control LD₅₀/7.3). To adjust radioprotectant doses for a different DRF, multiply *only* radioprotectant doses by (new DRF/1.2). To adjust the doses for a different estimated *b*, see the supplementary materials. We present Table 1 in a format suitable for most statistical software.

TABLE 2

Point and Interval Estimators for log-LD₅₀ and log-DRF

	Model	
	LD₅₀ (remove default intercept)	DRF (keep default intercept)
Software output	a_0, a_1, b, V	a_0, a^*_1, b, V^*
Parameter	log-LD _{50(k)} for $k = 0, 1$	log-DRF
Point estimator	$-\frac{a_k}{b}$	$-\frac{a^*_1}{b}$
Margin of error (MoE)	$z_{\alpha} \sqrt{\frac{a_k^2}{b^2} \left(\frac{v_{kk}}{a_k^2} + \frac{v_{bb}}{b^2} - \frac{2v_{kb}}{a_k b} \right)}$	$z_{\alpha} \sqrt{\frac{a^{*2}_1}{b^2} \left(\frac{v^*_{11}}{a^{*2}_1} + \frac{v_{bb}}{b^2} - \frac{2v^*_{1b}}{a^*_1 b} \right)}$
Interval estimator	$-\frac{a_k}{b} \pm MoE_k$	$-\frac{a^*_1}{b} \pm MoE$

Notes. For significance level, α , z_{α} is the $100(1 - \alpha)$ th percentile of a standard normal distribution, and is appropriate for a confidence interval (CI) having a confidence coefficient of $100(1 - 2\alpha)\%$. For 90%, 95% and 99% CIs, $z_{\alpha} = 1.645, 1.960$ and 2.576 , respectively.

TABLE 3Statistical Analyses Applied to Example Data to Produce 90% CIs for log-LD₅₀₍₀₎, log-LD₅₀₍₁₎ and log-DRF

	Model	
	LD ₅₀	DRF
Software output	$a_0 = -26.59, a_1 = -28.89, b = 30.29$	$a_0 = -26.59, a^*_1 = -2.30, b = 30.29$
	$V = \begin{bmatrix} 34.69 & 37.65 & -39.61 \\ & 41.05 & -43.09 \\ & & 45.34 \end{bmatrix}$	$V^* = \begin{bmatrix} 34.69 & 2.96 & -39.61 \\ & 0.44 & -3.48 \\ & & 45.34 \end{bmatrix}$
Point estimates	log-LD ₅₀₍₀₎ = 0.878 log-LD ₅₀₍₁₎ = 0.954	log-DRF = 0.076
Margin of error for 90% CI	MoE ₀ = 0.016 MoE ₁ = 0.016	MoE = 0.023
90% CI	for R0 (0.862, 0.894) for R1 (0.938, 0.970)	(0.053, 0.099)

Notes. 90% CIs are appropriate for conducting $\alpha=5\%$ significance level, one-sided tests. Parameter estimates and covariances can vary slightly from one software package to another.

TABLE 4

Most Probable Percentage Lethality (Number of Dead Animals Out of 6 per Treatment $\times$ Dose Group), Assuming Typical Experimental Conditions (see footnote 7)

Percentage lethality (number dead)	Radiation Dose (Gy)				
	6	7	8	9	10
Control	0 (0)	33 (2)	83 (5)	100 (6)	100 (6)
Treated	0 (0)	0 (0)	0 (0)	50 (3)	100 (6)

TABLE 5

By The Sample Size Formula of Kodell *et al.* (26), Calculated Samples Sizes for Detecting a DRF=1.2 with About 0.90 Power of a 0.05 Significance Level, One-Sided Test for The Indicated Slopes (b), When Using Staggered-Dose Designs With 3, 5 and 7 Doses Per Treatment

Number of doses {target % lethality}	Some inputs for Eq. (1)	Estimated probit log-dose slope (b) (effect size in SD units: $b \log \text{DRF}$)		
		12.61 (1.0)	20.26 (1.6)	27.91 (2.2)
3 {13.6, 50, 86.4} ^a	$t_{3,.05} = 2.353$	$N = 22.08 \rightarrow 23$ Total $N = 138$	$N = 8.55 \rightarrow 9$ Total $N = 54$	$N = 4.51 \rightarrow 5$ Total $N = 30$
	$t_{3,.10} = 1.638$			
	$\sum_{i=1}^3 w_i = 1.447$			
5 {5, 27.5, 50, 72.5, 95}	$t_{7,.05} = 1.895$	$N = 9.98 \rightarrow 10$ Total $N = 100$	$N = 3.87 \rightarrow 4$ Total $N = 40$	$N = 2.04 \rightarrow 3$ Total $N = 30$
	$t_{7,.10} = 1.415$			
	$\sum_{i=1}^5 w_i = 2.201$			
7 {5, 20, 35, 50, 65, 80, 95}	$t_{11,.05} = 1.796$	$N = 6.12 \rightarrow 7$ Total $N = 98$	$N = 2.37 \rightarrow 3$ Total $N = 42$	$N = 1.25 \rightarrow 2$ Total $N = 28$
	$t_{11,.10} = 1.363$			
	$\sum_{i=1}^7 w_i = 3.270$			

^aTarget lethalties are D-optimal, a 3-dose, symmetric design [see ref. (27)].