Regulatory Toxicology and Pharmacology 56 (2010) 51–53

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Advancing the science of EPA guidelines for sponsor-financed topical insect repellent efficacy studies

Linda J. Young *, Celia B. Fisher

University of Florida, 404 McCarty C, P.O. Box 110339, 32611-0339 Gainesville, FL, USA
Fordham University, Center for Ethics Education, Dealy Hall, 441 East Fordham Road, Bronx NY 10458, USA

ARTICLE INFO

Article history:
Received 21 September 2009
Available online 13 November 2009

Keywords:
Censored data
Sample size
Mean–variance relationship
Statistical models

ABSTRACT

In 1999, the US Environmental Protection Agency (EPA) published guidelines for product performance testing of skin-applied insect repellents, which provide guidance for topical insect repellent efficacy studies. EPA subsequently uses these sponsor-financed studies in their evaluation of proposed label claims. This paper reviews some of the statistical flaws in the proposed revisions to these guidelines and suggests possible improvements. This review is important because EPA's revisions to the 1999 guidelines do not address these issues.

1. Introduction

Mosquitoes, ticks, and biting flies can pose serious health risk by transmitting diseases like malaria, Lyme disease, West Nile virus, St. Louis encephalitis, and infections at the site of bites. Topical insect repellent products are designed to reduce consumer exposure to these diseases by repelling insects or deterring insect feeding on the surfaces to which the product is applied. The Environmental Protection Agency (EPA) under the Federal Insecticide, Rodenticide, and Fungicide Act (FIFRA, 1996) is responsible for insect repellent product registration and regulation. The EPA-approved label indicates that the product is safe and effective if used under the directions specified on the label. The label also indicates the protection time (PT) claimed for the product, which is how long the consumer can expect the product to continue being effective once it has been applied to the skin.

EPA, like other government agencies, depends upon, and issues guidelines for, industry-sponsored research that will inform regulatory decisions affecting the health and safety of the public, including those related to insect repellants and their label claims. Over the past four years the EPA Office of Pesticide Programs (OPP) has been working to revise its 1999 Guidelines for Product Performance Testing of Skin-Applied Insect Repellents (EPA, 1999). The revised guidelines are intended to provide sponsors and other third-parties with recommendations for improved methods of repellent efficacy testing. In October 2008, OPP presented draft guidelines (EPA, 2008a) to the Human Studies Review Board (HSRB), an independent board established through regulation 40 C.F.R. 26.1603 (EPA, 2006) to review the science and ethics of human research regulated by the Agency, including third-party human research on the safety and efficacy of insect repellants. Although the HSRB noted that the draft had achieved many of the Agency's goals in providing updated, informative, and clear guidance to third-party investigators, the Board also recognized deficient areas requiring improvement to ensure well-designed protocols that would generate data robust enough to support accurate product labels. Drawing upon transcripts, minutes, and reports from HSRB meetings, this critique draws attention to two such critical areas: sample size selection and statistical analysis of product efficacy. Highlighting the need for further revision of the guidelines is timely since OPP has repeatedly raised the possibility that these deficiencies would remain in the final version and addressed only at a later date (EPA, 2008b, 2009).

2. Selection of sample size

The 2008 draft guidelines alert investigators to the need to and methods for determining the sample size required to meet study objectives. However, the document does not address problems in sample size estimation identified in protocols reviewed by the HSRB to date. For insect repellent studies, if the objective is to estimate the mean PT of the product, then the sample size is determined from the precision with which the mean is to be estimated (e.g., ±2 h with 95% confidence) and an estimate of the variance of PT. For sponsor-funded insect repellent studies presented to the HSRB from 2006 to 2009, sample size determination has largely depended on a single paper by Rutledge and Gupta

* Corresponding author. Fax: +1 352 392 8555.
E-mail address: LJYoung@ufl.edu (L.J. Young).

0273-2300/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.yrtph.2009.11.002

(1999, hereafter referred to as RG) that calculated samples sizes based on mean PTs of 0.48–8.50 h, though only one study had a mean PT of more than 7 h. Although the RG paper was important and innovative for its time, as early as 2000 the EPA FIFRA Scientific Advisory Panel (EPA, 2000, SAP, Report No. 2000-02) noted that "more test subjects may be necessary to test repellents with longer durations of repellency" (p. 48) than those in RG. In this section we detail the pitfalls of applying the RG data for sample size estimations, and we argue that EPA should revise guideline language to guard against the continued use of assumptions by third-parties based on these data.

2.1. Calculation of means and standard deviations

To obtain an estimate of the standard deviation (SD) of PT for sample size determination, RG identified 22 previously published studies on topical insect repellents. In 16 studies, RG ostensibly combined information from different repellents, from different concentrations of the same repellent, or from different tests, often with significantly different PTs and consequently different SDs, to produce a single mean and a single SD for each study. As an example, the data used by RG from one study (Reifenrath & Akers, 1981) are presented in Table 1. Based on those data, RG reported a mean PT of 8.5 hours, namely the average of the PTs from the *two different* repellents (Sulfonamide and SRI-6). Although RG stated that the estimated SD of PT was taken to be the square root of the among-subjects mean square, we the authors of this present paper have been unable to replicate their analysis. Further, RG reported an estimated SD of 4.40, a value higher than the SD for either repellent, as again from Table 1. Instead, the mean and variance from each repellent and from each concentration of the same repellent should have been reported. The incorrectly combined estimates were later used to develop the model upon which sample size justifications presented to the HSRB have subsequently been based, thus further compounding the prior errors.

2.2. Modeling

RG deviated from accepted statistical practice in two ways in calculating sample sizes for different precision requirements: (1) modeling the relationship in mean and variance and (2) assessing the appropriateness of using a single model for field and laboratory data. They first used a linear model to describe the relationship between SD and mean. Then, based on this model, they estimated the SD for a given mean PT to insert in the sample size calculations. The problem with this is that the relationship between the mean and the variance (not the SD) is usually modeled, and the SD determined from the estimated variance. More importantly, SDs of PTs were assumed to be normally distributed. If the response were normally distributed, the estimated variance would have a gamma distribution, which would, in consequence, lead to a different model and possibly better fit. Modeling the relationship between mean and variance has a rich history in several branches of sciences, including entomology (see for example Taylor, 1984; Taylor et al., 1979). Quadratic and power models, not linear ones, have been consistently found to fit entomological data well.

Table 1
Means and Standard Deviations Representing Hours of CPT for Two Different Insecticide Repellents Calculated in Combination by RG*

	Repellents	
Volunteer	Sulfonamide	SRI-6
1	10	12
2	8	12
3	6	12
4	4	4
Mean	7	10
Standard Deviation	2.58	4

* Data from Table 2 of Reifenrath & Akers, 1981.

2.3. Combining field and laboratory data

RG compared the means of mean PT from different locations (state/country) and from different settings (field/laboratory), found no significant differences, and concluded that a single model is adequate for all locations and settings. However, comparing the means is not the same as comparing the models because means are often comparable when models are dramatically different. A more methodologically sound approach would have been to fit a regression line to the field data, fit another regression line to the laboratory data, and test whether the intercepts and/or slopes of the two regression lines differed significantly, which would indicate a need for separate regression lines for field and laboratory studies.

2.4. Implications

Current EPA guidelines state that at least five subjects are needed when the product label is to claim 1–4 h PT, and at least ten subjects are required when the label is to claim 5 or more hours. Although these ensure that the sample size is at least at some minimal level, how do we know whether these sample sizes are sufficient? This recommendation is not based on the flawed RG work and the statistical basis for this claim, if any, is unclear. More importantly, the revision does not provide any new or more accurate guidance for investigators that will lead to more adequate sample size determinations. Without legitimate sample size criteria to assure that statistical statements are correct, the guidelines will continue to reflect EPA's willingness to accept data from sponsor-financed studies using sample sizes that risk producing imprecise estimates of mean PTs for use in product labeling.

3. Statistical analysis of product efficacy

For sponsor-funded studies, the 1999 guidelines recommended using the first confirmed bite (FCB), defined as the time to the first bite followed by a subsequent bite within 30 min. The most recent draft guidelines recommend use of first confirmed landing with *intent to bite* instead of actual biting. Here we refer to both measures as FCB. The time from which the product was applied until the FCB is called the complete protection time (CPT). In this section, we examine the statistical analysis of CPT, whether the endpoint is based on an actual confirmed bite or on the first confirmed landing with intent to bite.

3.1. Censored data

In most sponsor-financed completed protocols brought before the HSRB to date, the study actually ended before a FCB for most, if not all, study participants. In these instances, the CPT is assumed to be at least as long as the study, although how much longer is unknown. Under these circumstances, the data are said to be *censored*. When data are heavily censored, the use of standard statistical methods for uncensored data is problematic. First, whenever investigators substitute the length of the study, say 10 h, as the CPT for subjects who did not have a FCB by then, the estimate of the mean is too small because actual CPTs could have been substantially longer had it been possible to continue observing each subject until an FCB. Not only is the mean underestimated, so too is the SD. This is important to consumer protection because underestimation of the SD results in a confidence interval that is

too short, thereby leading to the interpretation that the mean PT is estimated more precisely than it actually is. Although the draft guidelines note that censored data "may compromise the validity of test results", investigators are only required to describe how the sample size was determined and how possible premature withdrawal of subjects from the test will be treated (p. 46). As a result, the current draft guidelines permit continued use of the same statistically imprecise methods of analyzing censored data that have been used in protocols critiqued by the HSRB over the past 3 years. For example, in one sponsor-financed study of 12 subjects (Spero, 2008, Study A-117a) an FCB was observed at 3 h for 1 subject, but for the 11 subjects who did not have a FCB, the study cut-off time (10 h) was used in place of their FCBs. The researchers then calculated means and SDs, and concluded that the mean PT was 8 h ± 2 h, with 95% confidence, a conclusion statistically unjustified for the reasons indicated above.

3.2. Implications

Good statistical alternatives exist for heavily censored data. For example, if mean PT is used, more sophisticated statistical methods for censored data can be used. As an example, assuming that the probability distribution of time until failure is known, maximum likelihood methods and the EM-algorithm can be used to obtain estimates of the mean and its standard error, as long as at least a moderate number of observations are not censored. As the number of censored values increases, the standard error also increases, reflecting more uncertainty in the estimated mean. But perhaps a more important consideration for these kinds of data sets is whether the mean PT is really the label information that is most meaningful to consumers. In the A-117a study cited above 11 of 12 subjects (92%) had $\geqslant$ 8 h of CPT. Setting a lower, one-sided confidence interval on the population proportion with a CPT of $\geqslant$ 8 h, the investigators could have accurately stated with 95% confidence that at least 78% of people using the product could expect to have $\geqslant$ 8 h of CPT from mosquitoes, a statement this would be more meaningful to the individual consumer. Estimating the proportion protected for a specified PT would eliminate the problem of censored data, and the study's length would then be the hypothesized protection time. However, it is also important to note that larger sample sizes are generally required to estimate proportions than to estimate means.

4. Conclusion

Lisa Jackson, the new EPA Administrator has pledged to strengthen science at EPA through decisions that "reflect the expert judgment of the agency's career scientists and independent advisors" (Jackson, 2009). The scientific weaknesses in sponsor-financed insect repellent efficacy studies identified by the HSRB to date underscore the need to infuse rigor into all aspects of the proposed revised insect repellent guidelines, including those areas highlighted in this paper. At recent HSRB meetings, OPP has claimed that time pressures and lack of statistical staff expertise underlies their willingness to issue a version of the current draft of the insect repellant guidelines prior to addressing deficiencies in sample size selection and analysis of product efficacy. We believe, based upon our analysis in this paper, that the better approach would be to provide OPP with the human capital required to develop complete state-of-the-science guidelines for insect repellent product performance testing now. Until this is achieved, the commitment of career scientists and independent advisory boards to ensure that EPA-approved labels provide the public with directions for product use based on empirically sound science will not be fully realized.

5. Note

Linda J. Young is a member and Celia B. Fisher is past Chair of the EPA Human Studies Review Board (HSRB). The views expressed in this article are solely those of the authors and do not necessarily represent those of the HSRB.

References

Federal Insecticide, Fungicide and Rodenticide Act (FIFRA), amended 1996. 7 U.S.C. §136 et seq.

Jackson, L.P., January 23, 2009. Memo to EPA employees. Available from: <http://www.epa.gov/administrator/memotoemployees.html>.

Reifenrath, W.G., Akers, W.A., 1981. Field testing of repellents against *Anopheline* mosquitoes. Mosquito News 41, 276–280.

Rutledge, L.C., Gupta, R.K., 1999. Variation in the protection periods of repellents on individual human subjects: an analytical review. Journal of the American Mosquito Control Association 15, 348–355.

Spero, N., 2008. Evaluation of the efficacy of personal repellents against mosquitoes in the laboratory. ICR, Inc. Available from: http://www.epa.gov/osa/hsrb/files/meeting-materials/jun-24-25-2008-public-meeting/mrid47397701primaryreport_of_icr_a117(4-7-08).pdf. Accessed June 24, 2009.

Taylor, L.R., 1984. Assessing and interpreting the spatial distributions of insect populations. Annual Review of Entomology 29, 321–357.

Taylor, L.R., Woiwood, I.P., Perry, J.N., 1979. The negative binomial as a dynamic ecological model for aggregation, and the density dependence of k. Journal of Animal Ecology 48, 289–304.

US Environmental Protection Agency, 1999. Product Performance Test Guidelines OPPTS 810.3700 Insect Repellents for Human Skin and Outdoor Premises. Available from: <http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/810_Product_Performance_Test_Guidelines/Drafts/810-3700.pdf>. Accessed June 24, 2009.

US Environmental Protection Agency, 2000. SAP Report No. 2000-02: Sets of Scientific Issues Being Considered by theEnvironmental Protection Agency Regarding:Session I—Implementation Plan for Probabilistic Ecological Assessment: A Consultation Session II - Insect Repellent Product Performance Testing Guideline Evaluation. FIFRA Scientific Advisory Panel Meeting, April 5–7. Available from: <http://www.epa.gov/scipoly/SAP/meetings/2000/april/freportapril572000.pdf>. Accessed June 24,2009.

US Environmental Protection Agency, 2006. Protections for Subjects in Human Research. Fed. Reg. 71.24. pp. 6137–6176.

US Environmental Protection Agency, 2008a. EPA Product Performance Test Guidelines: Insect Repellents to be applied to Human Skin. Draft, September, 2008. Available from: <http://www.epa.gov/OSA/hsrb/files/meeting-materials/oct-21-22-2008-public-meeting/guidelines-for-performance-testing-of-skin-applied-insect-repellents-9.23.08.pdf>. Accessed June 24, 2009.

US Environmental Protection Agency, 2008b. Minutes of the United States Environmental Protection Agency (EPA) Human Studies Review Board (HSRB). October 21–22, 2008. Available from: <http://www.epa.gov/osa/hsrb/files/hsrb-oct2008-meeting-minutesfinal12208.pdf>.

US Environmental Protection Agency, 2009. Minutes of the United States Environmental Protection Agency (EPA) Human Studies Review Board (HSRB). June 24–25, 2009. Available from: <http://www.epa.gov/osainter/hsrb/files/hsrb_june09_mtg_minutesfinal.pdf>.