2003

Rodent sperm analysis in field-based ecological risk assessment: pilot study at Ravenna army ammunition plant, Ravenna, Ohio

Lawrence V. Tannenbaum
US Army Center for Health Promotion and Preventive Medicine

Matthew Bazar
US Army Center for Health Promotion and Preventive Medicine

Melanie S. Hawkins
US Army Center for Health Promotion and Preventive Medicine

Barney W. Cornaby
Science Applications International Corporation

Elizabeth A. Ferguson
US Army Corps of Engineers, Louisville District

See next page for additional authors

ELSEVIER

Environmental Pollution 123 (2003) 21–29

ENVIRONMENTAL POLLUTION

www.elsevier.com/locate/envpol

Rodent sperm analysis in field-based ecological risk assessment: pilot study at Ravenna army ammunition plant, Ravenna, Ohio☆

Lawrence V. Tannenbaum[a,*], Matthew Bazar[a], Melanie S. Hawkins[a], Barney W. Cornaby[b], Elizabeth A. Ferguson[c], L. Chantelle Carroll[d], Patrick F. Ryan[b]

[a]US Army Center for Health Promotion and Preventive Medicine, 5158 Blackhawk Road, Aberdeen Proving Ground, Aberdeen, MD 21010-5403, USA
[b]Science Applications International Corporation, 151 Lafayette Drive, Oak Ridge, TN 37830, USA
[c]US Army Corps of Engineers, Louisville District, 600 Martin Luther King, Jr. Place, PO Box 59, Louisville, KY 40201-0059, USA
[d]C eco, 7007 Mahoning Avenue, Warren, OH 44481, USA

Received 17 April 2002; accepted 6 September 2002

"Capsule": *Rodent sperm analysis is a visable method for use in field studies of risk at contaminated sites.*

Abstract

Ecological risk assessment (ERA) guidance recommends that field-truthing efforts proceed when modeled hazard quotients (HQs) suggest that toxicological effects are occurring to site receptors. To date, no field methods have been proposed by the regulatory community that can lead to definitive determinations of acceptable or unacceptable risk for birds and mammals, the two terrestrial classes of receptors that are commonly assessed using the HQ method. This paper describes rodent sperm analysis (RSA) as a viable method to be applied in the field at sites with historical contamination. RSA is capable of detecting biological differences that bear on reproduction, a highly regarded toxicological endpoint of concern in USEPA Superfund-type ERAs. The results of RSA's first application at a study site are reported and discussed. The paper also provides the rationale for RSA's efficacy in the context of Superfund and other environmental cleanup programs, where limited time and money are available to determine and evaluate the field condition.
Published by Elsevier Science Ltd.

Keywords: Ecological risk assessment; Field-truthing; Sperm; Rodents; Reproduction

1. Introduction

The current state of practice in screening-level and baseline ecological risk assessments (ERA) is the computation of hazard quotients (HQs) for a series of receptors (e.g., plants, invertebrates, birds, and mammals) that are representative of the site of interest. HQs (ratios of an animal's estimated daily dietary dose of a chemical to a reputedly safe dose of the same chemical) have notable limitations, and can only serve as mere risk screening tools (USEPA, 1989; Bartell, 1996). By themselves, HQs do not demonstrate that receptors are actually at risk, and consequently HQs alone cannot justify a remedial action (e.g. excavation of soils) to ensure receptor protection, although some regulators may elect to use the HQ ratio this way. Recent EPA guidance (USEPA, 1997, 1998) acknowledges the imprecision in ERA modeling, and recommends that modeled results (that anticipate toxicological effects occurring in the wild on the basis of threshold HQ exceedance) be compared to field data as a check on whether the understanding of site conditions was correct. To date, efforts to do so have been limited to the comparison of modeled and measured tissue concentrations in plants and mice (Alsop et al., 1996), in order to address model over and underprediction. No formal field-truthing methodologies for the verification of modeled toxicological *effects* in terrestrial systems have been proposed or endorsed by the regulatory community. It would be prudent though, to develop such methods given the uncertainty in current modeling efforts, and given the frequency with which HQs are found to exceed the effects threshold of 1.0 (Duke and Taggart, 2000). This paper examines a field-truthing method, rodent sperm analysis (RSA), that documents the consequences of chemical exposure on reproductive

☆ The opinions or assertions contained herein are the views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

* Corresponding author. Tel.: + 1-410-436-5210; fax: 1-410-436-8170.
E-mail address: lawrence.tannenbaum@amedd.army.mil (L.V. Tannenbaum).

0269-7491/03/$ - see front matter Published by Elsevier Science Ltd.
PII: S0269-7491(02)00360-3

endpoints. The results of the method's test case at contaminated sites at Ravenna Army Ammunition Plant (RVAAP) in rural northeastern Ohio, suggest that RSA has the potential to address HQ concerns at other sites.

The RSA method is predicated on three principal assumptions. First, if after several decades of exposure to contaminated site media, animals do not display physiological evidence of reproductive impact, it is reasonable to anticipate that future reproductive impacts will not arise. In most cases, contamination at Superfund and similar sites is historical in nature (i.e., decades old). The short life spans of birds and mammals have allowed for many or multiple generations (possibly numbering in the hundreds) of exposure to contaminated media. Second, where an observed reduction in the quality of conventional sperm parameters (e.g. sperm count, sperm motility, sperm morphology) is noted at the site of interest, and the site and the matched reference location differ only in their soil chemistries, the observed sperm effects are due to the chemical-in-soil influences. These conventional sperm parameters when impaired were shown by Chapin et al. (1997) to correlate with reduced reproductive success. Third, when small rodents at a contaminated site are deemed to have impaired reproductive capability on the basis of lesser quality sperm parameters, by implication, other site terrestrial receptors have the potential to be experiencing similar reduced reproductive success. This is a conservative assumption, because the degree of direct soil contact in wider-ranging mammals and birds (e.g., deer and hawks) is substantially less than that of small-ranging small rodents, and because biomagnifying compounds are not fully addressed by the method. Such assumptions must be made, however, if an ERA is to be brought to closure in a reasonable time frame. The method acknowledges that some chemicals may or may not interfere with reproduction, rather, they may affect other endpoints. Additionally, it is not known if decades-old and weathered chemicals in soil and that (especially organic chemicals) may have an ongoing capacity to interrupt reproduction.

The RSA method as applied in the RVAAP pilot study was intended to detect if any reproductive effects might be occurring in small mammals and other organisms with similar exposures. These receptors were the ones evaluated in the Phase II ERA (SAIC, 1999) for RVAAP's Winklepeck Burning Grounds (WBG). In conjunction with a weight of evidence approach (USEPA, 1999; e.g. small mammal species composition measures were also considered), the RSA method fostered comparisons of reproductive measures between contaminated study sites and clean reference sites. The screening-level assessment resulted in HQs that were frequently in the high multiple 100s and occasionally exceeding 1000 (SAIC, 1999). The HQs reflect chemical concentrations in soil resulting from historical site operations dating back to 1941. Site use at the 200-acre WBG included open burning (i.e. on slag-covered bare ground) for melting explosives out of heavy artillery projectiles, and waste disposal of explosives (i.e. RDX, TNT, Composition B, black powder, propellant), antimony sulfide, lead oxide, lead thiocyanate, sludge, sawdust from the installation's load lines, and domestic wastes. The open burning and detonation activities resulted in residual chemicals and metallic munitions fragments remaining on as many as 70 burning "pads" ranging in size from 50 feet×70 feet to 75 feet×110 feet (Fig. 1).

2. Materials and methods

2.1. Study sites

Three WBG study sites, posing the greatest potential chemical stress to site receptors on the basis of HQ magnitude alone were selected (SAIC, 1999, 2000). The sites were geographically distinct from each other, such that small rodent home ranges at each did not overlap. Each study site consisted of two adjacent burning pads (pad pair numbers 37 and 38, 58 and 59, 66 and 67; Fig. 1) that for multiple receptors, shared high HQs for either metals, explosives, or a mixture of these chemical groups (Table 1). Corresponding reference sites (on RVAAP but more than one mile beyond the WBG boundary) for each two-pad grouping were selected based on similar soil, site history and other characteristics, during a preliminary field reconnaissance effort in the spring of 2000 (SAIC, 2000). Specific criteria for selection included hydrology, soil type, topography, site-use history, degree of maintenance (i.e. mowing), and plant community type. Care was taken to ensure that the reference sites offered the same level of resources as the study sites with regard to their ability to attract and support White-footed mice (*Peromyscus leucopus*) and Meadow voles (*Microtus pennsylvanicus*) and other animal life. An earlier small mammal survey (Carroll, 1999) indicated these were the two most numerous small rodent species at RVAAP, although the survey had not been extended to WBG.

2.2. Animal trapping

Four consecutive trap nights at each of three sites (at a time) were envisioned, with the reasonable expectation that this would result in the minimum number of target animals (27 each of White-footed mice and Meadow voles for each site) being trapped to support a rigorous statistical comparison. The 27 adult males corresponds to an alpha of 5%, a statistical power of 95%, and a 1:1 ratio of significant difference (Sign. Diff.) to coefficient of variation (CV). Employing this ratio, whereby the

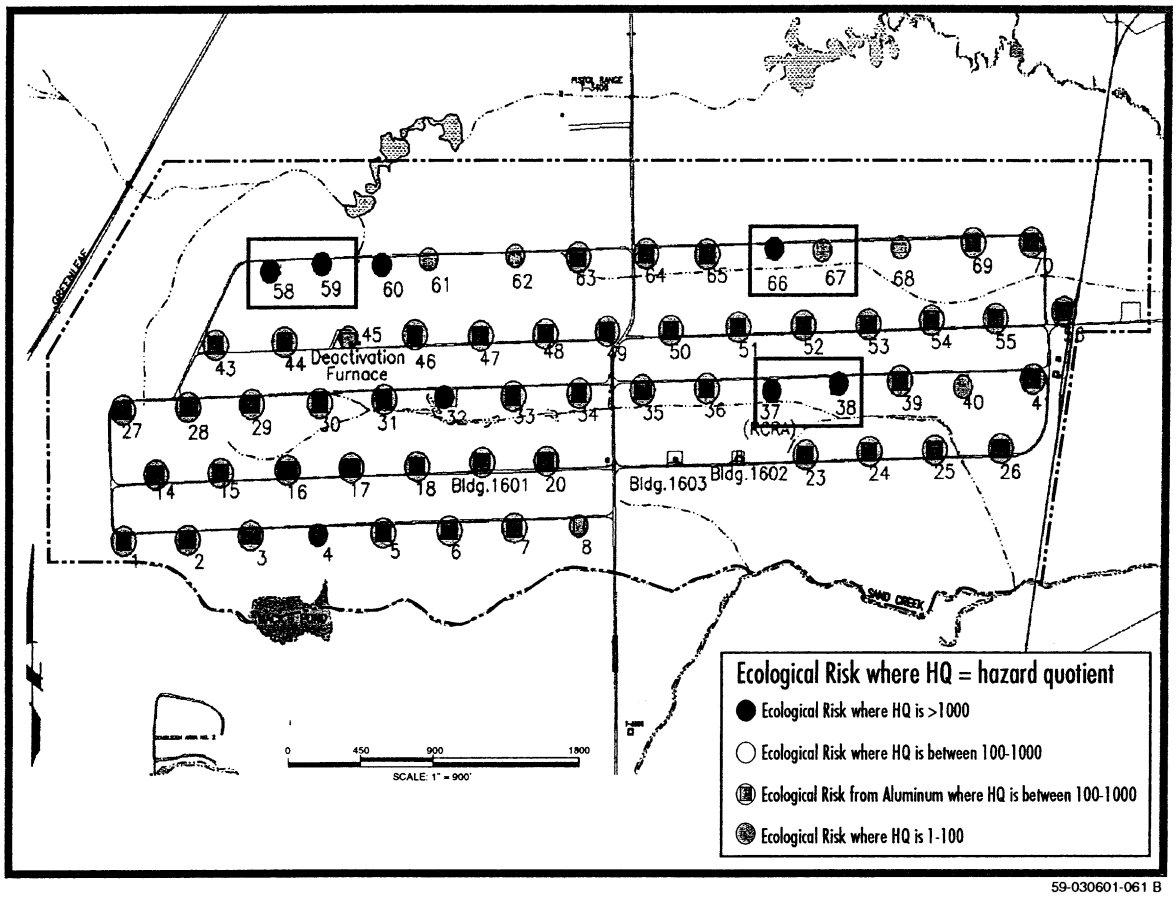

Fig. 1. Schematic of the Winklepeck Burning Grounds site at Ravenna Army Ammunition Plant (RVAAP). Some 70 burning pads, each measuring roughly 50 feet×70 feet to 75 feet×110 feet are aligned in rows. The magnitude of the higher hazard quotients for the pads is depicted with a pattern (see legend). The three two-pad complexes that constituted the field study sites are demarcated (rectangles). Reference sites lie beyond the WBG boundary and are not shown here.

Table 1
Highest receptor-specific hazard quotients[a] by receptor at the burning pad sites[b,d]

Pad pairs	Robin	Hawk	Owl	Shrew	Fox	Cottontail rabbit
37 & 38	2650 – cadmium	13 – zinc	12 – zinc	71 – cadmium	31 – thallium	4 – thallium
	1890 – lead	7 – lead	6 – lead	64 – thallium	2 – zinc	4 – arsenic
	45 – zinc	4 – DBP[c]	3 – DBP[c]	92 – lead	–	–
58 & 59	1300 – lead	60 – zinc	55 – zinc	212 – cadmium	7 – thallium	12 – antimony
	242 – cadmium	3 – mercury	4 – lead	19 – antimony	7 – zinc	4 – arsenic
	54 – zinc	5 – lead	2 – mercury	12 - thallium	3 – cadmium	4 – cadmium
66 & 67	1280 – lead	60 – zinc	7 – zinc	706 – TNT	361 – TNT	709 – TNT
	54 – zinc	5 – lead	4 – lead	207 – HMX	229 – RDX	208 – HMX
	16 – barium	–	–	446 – RDX	106 – HMX	448 – RDX

[a] The three highest chemical-specific hazard quotients are shown.
[b] Corresponding toxicological effects for all hazard quotients, except those for RDX, are reproductive. The study endpoint for RDX is increased organ weight.
[c] Di-n-butylphthalate.
[d] Aluminum at numerous pads and for almost all receptors had hazard quotients in the multiple hundreds. Aluminum was determined to not be a contaminant of concern.

Sign. Diff. is selected relative to the CV rather than independent of it, allowed the sample size to be determined without knowing the measured CV for the sperm parameters from the field. Statistically for a normally distributed sperm parameter, the selection of 20% (of the parameter mean) as the Sign. Diff., means that more than 99% of the results would fall within 2.5 standard deviations of the mean, and that one standard deviation would represent about 20% of the parameter's range.

Animal trapping began in mid-spring 2000 after the season's first litters had matured to adulthood (verifiable by pelage). Sherman live traps, baited with a rolled oats and peanut butter mix and a sweet feed mix for horses, were randomly placed in preferable habitat within a 50-m radius from the center of each of the pads (the three pad pairs). Seventy-five traps were set out on each pad (150 per site), effectively saturating the trapping area. During the field effort, heavy rains flooded the soils of both the burning pad and reference sites, causing rodents to migrate to higher ground and necessitating a delay of a second trapping event. All non-target animals trapped were identified to species, weighed using a Pesola scale, and released. All female juvenile and sub-adult White-footed mice and Meadow voles and juvenile males were weighed in the field and also marked on the top of the head with a small dab of colored nail polish prior to release, so that animals trapped on subsequent nights would not be double-counted. Females were additionally assessed for reproductive state (lactating or pregnant).

2.3. Sperm measurement

Target animals (i.e. adult male White-footed mice and Meadow voles) were transported in their traps on the day following capture to the field laboratory. Target animals were euthanized with carbon dioxide, and liver weights were first recorded. For the assessment of sperm motility, the right vas deferens was surgically removed with care to minimize blood contamination. The excised tissue was immediately placed into a pre-warmed suspension medium containing 3 ml of phosphate buffered saline with 1% bovine serum albumin, and given a 3-min "swim-out" period to allow sperm to enter the medium. A 100 μm cannula was then inserted into the medium to obtain a sample, and the cannula inserted into the retractable stage of a Hamilton-Thorne integrated visual optics system (IVOS) sperm analyzer, for a general examination of sperm on the analyzer's main unit color monitor. The analyzer was preset to automatically move the stage to five different fields along the length of the cannula and to store each motion image (uniquely identified by study number, animal number, and cannula field number) on a Hewlett-packard write-once optical disk, creating a permanent record for precise image reproduction and retrieval. Several weeks later, each image was recalled from the optical disk and analyzed for motile and non-motile cells. A percentmotility for all five recorded fields was determined for each animal, and other motility parameters (including straight-line, curvilinear, and path velocities; and progressive motility and cross-beat frequency) were also calculated.

The left epididymis was also removed following animal euthanization, and frozen on dry ice. It was used to determine total sperm count and sperm abnormality (the percentage of misshapen sperm). The epididymis was thawed and the caudal section removed and weighed in order to report the total count as millions of sperm/gram of caudal epididymal tissue. It was then homogenized and a 100 μl sample added to a vial containing a fluorescent dye (Hoechst dye H33342) to stain the DNA in the sperm head, in order to prevent surrounding debris from being counted as sperm. A 9 μl sample was added to a slide which was cover-slipped, secured to the retractable stage, and then loaded into the IVOS. The analyzer automatically counted the stained sperm heads for 20 fields per slide, minimizing the sperm cell distribution variance within single samples.

For sperm morphology, two slides were prepared from the epididymal sample prior to homogenization, and later stained with 5% eosin and cover-slipped for microsopic evaluation. Two-hundred sperm cells were evaluated with reverse phase/dark field microscopy (×40 objective) for head and tail abnormalities (size, shape, and double heads/tails), with the results reported as the percentage of the 200 sperm that were abnormal. The procedures followed for the evaluation of sperm count, sperm motility, and sperm morphology are those of Pathology Associates, A Charles River Company 2002.

3. Results

3.1. Primary assessment metrics—sperm parameters

Due to the weather and reduced trapping success, there were not enough adult male White-footed mice to perform pair-wise statistical comparisons between burning pad pairs and corresponding reference sites as planned. The burning pads and the reference sites were each pooled to preserve statistical integrity (described later). It was only possible to statistically compare the sperm parameters of mice because of the low vole captures. Consistent with previous reports in the literature (Zenick et al., 1994), the sperm parameter with the greatest variability, as illustrated by the CV, was sperm count (Tables 2 and 3). Table 3 reports the results of the Wilcoxon rank sum test comparing medians of sperm parameters, and body and normalized liver weights of

Table 2
Sperm parameters comparison in field collected White-footed mice

Sperm parameter	Pooled reference areas					Pooled burning pads				
	n	Range	Mean	S.D.	CV	*n*	Range	Mean	S.D.	CV
Sperm count (10^6 sperm/g tissue)	8	1178.8–2241.9	1670	353.8	21.2	6	1229.5–1901.7	1409	309	21.9
Sperm motility (percent)	8	94–99	98.4	1.77	1.8	5[1]	98–100	99.2	0.84	0.8
Sperm abnormality (percent)	8	NA	0.0	0	NA	6	0–1	0.0	0	NA

1 = one sperm sample was lost, and did not allow for the motility measure; NA not applicable; *n* = sample size; S.D. = standard deviation; and CV = coefficient of variation.

Table 3
Statistical analysis of sperm parameter comparison

Sperm parameter	Mean of reference areas	Mean of burning pads	Percent difference[a]	Percent difference/CV[b]	Agreement with expected direction of difference[c]	Probability that observed difference resulted from chance[d]
Sperm count	1670	1409	−16.71	−0.78	Yes	0.114
Sperm motility	98.4	99.2	0.84	0.55	No	0.093
Sperm abnormality	0.0	0.0	NA	NA	No	NA

[a] Calculated as mean of pooled burning pads minus mean of pooled reference areas/overall mean, multiplied by 100.
[b] Calculated as the pooled standard deviation/overall mean.
[c] The expectation is that contaminants at the burning pads will reduce sperm count, reduce sperm motility, and increase the percentage of mis-shapen sperm.
[d] Probability calculated using one-sided Wilcoxon exact test assuming a power of 95%.

the mice. There was no statistically significant difference between burning pads and reference sites for either sperm parameter (sperm count and sperm motility; sperm abnormality was 0% for both groups), although the lesser sperm count of the burning pad mice (a reduction of 16.7%) was in the direction expected (i.e. exposures to contaminants are known to reduce sperm production and will be explained in Section 4.3). However, variability of burning pad mice was less than the reference site mice, as demonstrated by a smaller range of values for two of the parameters (count and motility). The only attribute with a statistically significant difference was liver weight, with livers of mice from the burning pads 17.9% heavier than those of reference site mice. However, this difference disappears when the liver weight is normalized by body weight.

The data was further evaluated to determine the minimum detectable difference between groups using the Wilcoxon Rank Sum Test. Table 4 lists the biological attributes in order of decreasing power to detect a 20% difference between group means at an alpha of 5%. Power ranged from 100 to 91% for sperm morphology, motility, body weight, and liver weight, with the power decreasing as the CV increased, respectively. It was, therefore, possible to detect a minimum difference of 20% between groups for these four measures. A detectable difference of 20% was not possible for sperm count, which had a 48% power. However, the power to detect a minimum difference of 30% at an alpha of 5% for sperm count was 77%.

3.2. Small mammal species composition and reproductive status

A total of 152 small mammals representing 10 species were captured in the study during the two trapping sessions of 8 and 6 days, respectively (Table 5). The majority of these animals (58%) were target species. White-footed mice were present in nearly equal numbers at the burning pads and the reference sites, while Meadow voles were five times more numerous at the burning pad sites. The reference sites had four non-target species present that were absent from the burning pads, and the burning pad sites had two non-target species present that were absent from the reference sites. For four of these non-target species, only one or two individuals constituted the captures. The greatest disparity in species composition between the burning pads and the reference sites, were captures only at the reference sites of Short-tailed shrews (*Blarina brevicauda*) and Eastern chipmunks (*Tamias striatus*).

Table 6 provides the age structure and sex ratios for the target species. No apparent differences in age structure and sex ratios were evident for White-footed mice between the burning pads and reference sites. Meadow vole age and sex ratios were similar to those of White-footed mice at the burning pad sites, however, the low number of captures at the reference sites did not allow a similar comparison for Meadow voles. Of the 12 female adult and sub-adult White-footed mice trapped at the reference sites, two (17%) were pregnant, and four

Table 4
Estimated probability to detect a difference between biological attributes

Biological attribute	Average attribute value over all areas	CV over all areas	Power at 5% α level			Power at 5% α level		
			For Sign. Diff. equal to CV	For 20% Sign. Diff.	For 30% Sign. Diff.	For Sign. Diff. equal to CV	For 20% Sign. Diff.	For 30% Sign. Diff.
Sperm								
Count (10^6 sperm/g tissue)	1558	21.55	53	48	77	68	64	88
Motility (%)	98.7	1.52	48	100	100	65	100	100
Body weight (grams)	21.73	10.34	48	93	100	65	97	100
Normalized Liver Weight[a]	0.053	13.16	48	79	98	65	89	99

[a] Normalized liver weight as liver weight (g)/body weight (g).

Table 5
Numbers of small mammal species trapped

Species trapped	Burning pads	Reference sites	Total number trapped
White-footed mouse[a]	29	33	62
Meadow vole[a]	22	4	26
Eastern cottontail rabbit	2	2	4
Deer mouse	1	–	1
Masked shrew	1	–	1
Short-tailed shrew	–	17	17
Eastern chipmunk	–	36	36
Meadow jumping mouse	1	1	2
Southern flying squirrel	–	2	2
Woodland vole	–	1	1
Total number animals trapped	55	96	152

[a] Excludes recaptures.

Table 6
Age structure and sex ratios of key species[a]

Species	Adults		Sub-adults		Juveniles	
	Males	Females	Males	Females	Males	Females
White-footed mice						
Reference sites	8	9	5	3	3	5
Burning pads	7	5	3	3	7	4
Meadow voles						
Reference sites	1	1	0	0	2	0
Burning pads	4	11	0	3	3	1

[a] Figures are numbers of animals trapped.

(33%) were lactating. Of the eight female adult and sub-adults trapped at the burning pads, one (13%) was pregnant, and three (38%) were lactating. Thus, female reproductive status was similar.

4. Discussion

In theory when HQs above 1.0 are computed and cannot be rationalized through uncertainties or the conservative nature of the method used, a more focused, field-based effort to establish whether signs of stress or impact are evident in site receptors should proceed.

4.1. Rationale for using small mammals

Given the impracticability (e.g. logistics, cost, potential to decimate a population) of working with large, wide-ranging and higher trophic level terrestrial species, it is only feasible to collect and utilize small mammals. Small rodents are advantageous for study because they are generally plentiful in most habitats, relatively easy to capture, and trapping, handling, and euthanizing methods are commonly approved by institutional animal care and use committees. Within an ERA context, they are additionally advantageous for use because of their small or limited home ranges, which can virtually guarantee that trapped specimens are coming in contact with sites of interest. Having the most direct contact with soil compared to other mammals and birds, and life spans that rarely exceed one year, small rodents are the species of choice. They are maximally exposed to site contaminants and have had multiple opportunities (i.e. successive generations) to display critical defects.

Certain small mammals, however, are not as appropriate for this type of study, especially where a minimum number of specimens is needed to ensure adequate statistical power for data analysis. For instance, shrews are not as suitable because of their exceedingly high metabolism that necessitates continuous feeding; these species cannot be expected to survive more than a few hours after trapping unless special precautions are taken (Butterfield et al., 1981; Shore et al., 1995; Little and Gurnell, 1989; Kutzageorgis and Mason, 1997). In summary, most small mammal species are ideal target organisms.

4.2. Justification of the use of RSA in Superfund-type ERAs

In the context of sites managed under the USEPA Superfund program and similarly structured environ-

mental programs where time and money are limiting resources, the ultimate concern for any species within an ERA context is that it be able to survive and reproduce. This is reflected in the wording of assessment endpoints and the selection of toxicity reference values (TRVs). When there are several TRVs available for a given chemical of concern, each with a different corresponding toxicological effect (e.g. excess liver weight, reproductive effect, excess salivation), the risk assessor will typically use the TRV based on reproductive endpoints. Evaluation of reproductive endpoints is also common to most routine field-based toxicity tests, regardless of the medium (e.g. soil, surface water) they address. However, the well being of terrestrial birds and mammals cannot be reasonably addressed by standardized soil toxicity tests (e.g. lettuce seed germination or earthworm survival). Other types of field tests are needed for them.

Direct evaluation of reproductive parameters using RSA is an alternative to population studies that require prolonged periods of time in order to supply reliable data (Krebs, 1989; Seber, 1982), which Superfund and other related programs cannot support. Population monitoring (censusing) often requires many measurements over a multi-year period, and even then, one cannot usually distinguish between population impacts and natural differences of population cycling. Spatially, species presence–absence data are difficult to relate to chemical causation in any statistically rigorous way (Strayer, 1999), although such information may contribute to weight of evidence arguments. By contrast, RSA provides a direct measure of a population's ability to reproduce through time, despite the fact that the information is collected in a single sampling event.

4.3. Evidence of causal relationship of chemicals to reproductive metrics

Although the RSA method does not absolutely require that any of a site's chemicals of concern be known reproductive toxins, there is a substantial body of evidence demonstrating that sperm parameter impairment and other related reproductive effects are chemically based. Direct evidence of sperm parameter effects in rodents exists for the metals, aluminum (Llobet et al. 1994), arsenic (Pant et al., 2001), and lead (Wadi and Ahmad, 1999), and for the explosives 1,3-DNB (Linder et al., 1986; Cody et al., 1981) and 1,3,5-TNB (Reddy et al., 1996). Aside from many of the above compounds being commonly encountered at contaminated sites (including army sites), all of these were identified in soils of the burning pads of the WBG study, thereby substantiating that sperm parameters, as surrogate measures of reproductive success, were appropriate for study. Additional substantiation of the appropriateness of tracking sperm quality in rodents, derives from other laboratory studies, where chemical exposure may produce a host of reproductive toxicological endpoints. These include fewer successful matings, fewer litters, smaller litters, and also seminiferous tubule degeneration or reduced seminiferous tubules. Although these studies do not evaluate sperm directly, it is very likely that the endpoints measured (and particularly those concerning seminiferous tubules) are sperm-mediated. The explosives 2,4,6-TNT and RDX, detected at the WBG burning pads, have both been shown to cause seminiferous tubule effects (Dilley et al., 1982; Levine et al., 1984; Lish et al., 1984, respectively) in addition to testicular atrophy and testicular degeneration.

4.4. Rationale for assessing male reproduction

Female reproductive success measures would be highly desirable in a field-truthing effort. However, they involve additional complications because nearly all necessitate mating studies. Wild rodents have a history of not breeding well under indoor conditions, and adequate numbers of animals may not be able to be procured from the field to provide for a sufficiently large mating study (the case at RVAAP's WBG). Hence, it is common practice to pair two females with one male when such studies are conducted. Sufficient time would also have to be allowed to elapse in order for usable data to be gleaned from females, as their reproductive biology (e.g., estrous cycles) is easily offset by the stresses of having been trapped, conveyed to indoor housing, and handled. Evaluating *any* female reproductive measure, including those that do not require mating studies (e.g. estrous cycle length), assumes that animals will acclimatize well; necessitates additional time, cost, and animal housing; and shows the potential to allow for animal recovery with the obscuring of effects. In contrast, conventional sperm parameters (count, motility, morphology) which have been shown to be affected by chemical exposures and to correlate with reduced reproductive success (Chapin et al., 1997), are unaffected in trapped and handled rodents. In addition to collection of the RSA data being time- and cost-effective, population impacts are limited because only males are sacrificed.

4.5. The utility of RSA measures as assessment endpoints

A critical issue that biologists face with interpreting field-based data is how much of a difference between groups is biologically or ecologically significant. In the case of conventional sperm parameters, the literature provides thresholds for parameter impairment at which reduced reproductive success can be expected to occur in the field (Chapin et al., 1997). Numerous studies with mice and rats indicate that these species are robustly

fertile, and that sperm count needs to be reduced approximately 80% or more before reproductive success is compromised (Meistrich et al. 1994; Bucci and Meistrich, 1987; Gray et al., 1992). In a comprehensive multi-generational study with Swiss mice (Chapin et al., 1997), normal fertility was maintained in treated animals until motility declined to approximately 40–50%, and furthermore, all treated groups with less than 37% motile sperm had fewer pups than normal. In this same study, sperm morphology was found to be the most sensitive sperm parameter, with clearly associated reduced fertility occurring when the control range (2–12% abnormally shaped) was only slightly exceeded (16%). With further study in wild rodents, these various degrees of difference could be narrowed and bona fide field-based thresholds could be developed with RSA methods.

4.6. Interpretation of results of the RVAAP pilot study

The animal capture information can be used to support the information provided by the primary assessment (i.e. sperm parameter) metrics. The capture information indicates that the key species are not being excluded from, and are not avoiding the WBG burning pads, reputedly one of the most contaminated portions of RVAAP. Although the captured animal numbers are small, the field measurements show that females are not any less reproductively active at the burning pads than at the reference sites, based on the percentages of lactating and pregnant individuals.

Results of the sperm analysis show that male White-footed mice are not reproductively impaired, because there were no statistically significant differences ($P > 0.05$) between groups (i.e. count and motility). The CVs for all three sperm parameters (see Table 2) of mice trapped at both the burning pads and the reference sites are consistent with those reported in the literature for rodents (Zenick et al., 1994), with count being the most variable, motility having only slight variability, and an abnormality rate as low as 1%. Because rodents produce 10-20 times more sperm than needed to ensure full reproductive success (Meistrich et al., 1994; Bucci and Meistrich, 1987; Gray et al., 1992), it is a safe assumption that the approximate 17% reduction in the sperm count of WBG mice is inconsequential, even had the difference been statistically significant. Collectively, the trapping results and sperm parameters for mice mean reproductive success for these terrestrial receptors at WBG.

5. Conclusions

Results from the RSA method allowed us to arrive at a determination of acceptable risk for mammals at the WBG burning pads. Although desirable numbers of target animals were not captured, the small CVs for the sperm parameters allowed good statistical confidence in the study results. RSA indicated that reproductive effects, as an assessment endpoint, were not evident in the exposed population despite the fact that the HQ calculations of the initial desktop assessment had indicated otherwise.

The finding of no unacceptable risk lends support to those contentions that ecological HQs are misleading numbers because they overestimate the prevalence of toxicological effects in the field (USEPA, 1997, 1998; Alexander, 2000; Tannenbaum, 2001; Tannenbaum et al., in press). The results also suggest that there is more to be gained by advancing to the field for a verification endeavor, rather than conducting mathematically focused second and third tier ERAs.

Although small mammals at the contaminated sites did not display any adverse reproductive effects, a significant difference ($P <= 0.05$) in liver weights was evident. When liver weights were normalized to body weight, however, differences were not evident. Regardless, fresh liver weights in captured small rodents is a useful as a measure of exposure and liver measurement should remain as a fixed component of the RSA method.

We believe that at sites where the in-place contamination is one or more decades old, and where birds and mammals consequently have had multigenerational exposures, that detrimental effects should be either present or not present. In the latter case, such effects either never occurred, or did first occur but were followed by a period of ecological recovery. The RVAAP pilot study was designed to detect critical detrimental (i.e. reproductive) effects if such were present. Based on the study's outcome, we conclude that RSA represents an reasonable, practical and cost-attractive field-oriented technique. The field method allows for much better closure on animal aspects of ERA risk than depending on HQs and other mathematical predictions.

Acknowledgements

The work for this project was performed under the direction of the US Army's Operations Support Command. The authors wish to thank the other members of the Ravenna Ecological Risk Assessment Work Group, listed here in alphabetical order W. Glen Beckham, Jr., Dr. David J. Brancato, James P. Groton, W. Kevin-Jago, John P. Jent, Stephen V. Mitz, Timothy M. Morgan, Mark C. Patterson, Dr. Judith C. Pennington, Stephen B. Selecman, LTC Thomas A. Tadsen, Robert W. Whelove, Jr., Paul L. Zorko. The authors also thank Mr. Dennis E. Druck for his careful review of the manuscript.

References

Alexander, M., 2000. Aging, bioavailability, and overestimation of risk from environmental pollutants. Environmental Science and Technology 34 (20), 4259–4265.

Alsop, W.R., Hawkins, E.T., Stelljes, M.E., Collins, W., 1996. Comparison of modeled and measured tissue concentrations for ecological receptors. Human and Ecological Risk Assessment 2 (3), 539–557.

Bartell, S.M., 1996. Ecological/environmental risk assessment principles and practices. In: Kolluru, R.V., Bartell, S.M., Pitblado, R.M., Stricoff, R.S. (Eds.), Risk Assessment and Management Handbook for Environmental, Health, and Safety Professionals. McGraw-Hill, New York , NY, USA, pp. 10.24–10.31.

Bucci, L.R., Meistrich, M.L., 1987. Effects of busulfan on murine spermatogenesis: cytotoxicity, sterility, sperm abnormalities, and dominant lethal mutations. Mutation Research 176, 259–268.

Butterfield, J., Coulson, J.C., Wanless, S., 1981. Studies on the distribution , food, breeding biology, and relative abundance of the pigmy and common shrews (*Sorex minatus* and *S. araneus*) in upland areas of northern England. Journal of Zoology (London) 195, 169–180.

Carroll, C.A., 1999. A survey of small mammals of the Ravenna Arsenal (for the) Ohio Department of Natural Resources. .

Chapin, R.E., Sloane, R.A., Haseman, J.K., 1997. The relationships among reproductive endpoints in Swiss mice, using the Reproductive Assessment by Continuous Breeding database. Fundamental and Applied Toxicology 38, 129–142.

Cody, T.E., Witherup, S., Hastings, L., Stemmer, K., Christian, R.T., 1981. 1,3-Dinitrobenzene: toxic effects in vivo and in vitro. Journal of Toxicology Environmental Health 7, 829–848.

Dilley, J.V., Tyson, C.A., Spanggord, R.J., Sasmore, D.P., Newell, G.W., Dacre, J.C., 1982. Short term oral toxicity of 2,4,6-trinitrotoluene in mice, rats and dogs. Journal of Toxicology and Environmental Health 9, 565–585.

Duke, L.D., Taggart, M., 2000. Uncertainty factors in screening ecological risk assessments. Environmental Toxicology and Chemistry 19 (6), 1668–1680.

Gray, L.E., Marshall, J.O., Setzer, R., 1992. Correlation of ejaculated sperm numbers with fertility in the rat. Toxicologist 12, 433.

Krebs, C.J., 1989. Ecological Methodology. Harper and Row, New York, NY. (Chapters 2–4).

Kutzageorgis, G.C., Mason, C.F., 1997. Small mammal populations in relation to hedgerow structures in an arable landscape. Journal of Zoology (London) 242, 425–434.

Levine, B.S., Furedi, E.M., Gordon, D.E., Lish, P.M., Barkley, J.J., 1984. Subchronic toxicity of trinitrotoluene in Fischer-344 rats. Toxicology 32, 253–265.

Linder, R.E., Hess, R.A., Strader, L.F., 1986. Testicular toxicity and infertility in male rats treated with 1,3-dinitrobenzene. Journal of Toxicology and Environmental Health 19, 477–489.

Lish, P.M., Levine, B.S., Furedi, E.M., Sagartz, E.M., Rac, V.S., 1984. Determination of the chronic mammalian toxicological effects of RDX: twenty-four month chronic toxicity/carcinogenicity study of hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) in the B6C3F1 hybrid mouse. Phase VI, vol. 1. AD A160774. IIT Research Institute Chicago, IL. US Army Medical Research and Development Command, Frederick, MD.

Little, J., Gurnell, J., 1989. Shrew captures and rodent field studies. Journal of Zoology (London) 218, 329–331.

Llobet, J.M., Colomina, M.T., Sirvent, J.J., Domingo, J.L., Corbella, J., 1994. Reproductive toxicology of aluminum in male mice. Fundamental and Applied Toxicology 25, 45–51.

Meistrich, M.L., Kasai, K., Olds-Clarke, P., MacGregor, G.R., Berkowitz, A.D., Tung, K.S.K., 1994. Deficiency in fertilization by morphologically abnormal sperm produced by azh mutant mice. Molecular Reproduction and Development 37, 69–77.

Pant, N., Srivastava, S.P., Murthy, R.C., Kumar, R., 2001. Male reproductive effect of arsenic in mice. Biometals 14 (2), 113–117.

Pathology Associates A Charles River Company, 2002. Maryland Operations, Standard Operating Procedure No. 909.3, 909.4, 909,5, 909.8.

Reddy, T.V., Daniel, F.B., Olson, G.R., Wiechman, B.H., Reddy, G., 1996. Chronic toxicity studies on 1,3,5-trinitrotoluene in Fischer-344 rats. USEPA, Cincinnati, OH. ADA315216, US Army Medical Research and Materiel Command, Frederick, MD.

Science Applications International Corporation, 1999. Draft Final Phase II Remedial Investigation Report for the Winklepeck Burning Grounds at the Ravenna Army Ammunition Plant, Ravenna, Ohio. various pagination.

Science Applications International Corporation, 2000. Sampling and analysis plan and site safety and health plan addendum no. 2 for the biological measurements at Winklepeck Burning Grounds at Ravenna Army Ammunition Plant at Ravenna, Ohio. various pagination.

Seber, G.A.F., 1982. The Estimation of Animal Abundance and Related Parameters. Griffin, London.

Shore, R.F., et al., 1995. Capture success for pigmy and common shrews (Sorex minatus and S. araneus) in Longworth and pitfall traps. Journal of Zoology (London) 237, 657–662.

Strayer, D.L., 1999. Statistical power of presence–absence data to detect population declines. Conservation Biology 5, 1034–1038.

Tannenbaum, L.V., 2001. Background ecological risk: to weigh in or not, that is the key question. Human and Ecological Risk Assessment 7 (1), 217–219.

Tannenbaum, L. V., Johnson, M. S. and Bazar, M., Application of the hazard quotient method in remedial decisions: a comparison of human and ecological risk assessments. Human and Ecological Risk Assessment (in press).

US Environmental Protection Agency, 1989. Risk assessment guidance for Superfund, vol 1. Human Health Evaluation Manual (Part A), Interim Final. (EPA/540-1-89/002) USEPA, Washington DC.

US Environmental Protection Agency, 1997. Ecological risk assessment guidance for Superfund: Process for designing and conducting ecological risk assessments. Interim Final. (EPA/540-R-97-006) USEPA, Washington, DC.

US Environmental Protection Agency, 1998. Guidelines for ecological risk assessment. (EPA/630/R-95/002F) USEPA, Washington, DC.

US Environmental Protection Agency, 1999. Issuance of Final Guidance: Ecological Risk Assessment And Risk Management Principles for Superfund Sites. OSWER Directive 9285.7-28P.

Wadi, S.A., Ahmad, G., 1999. Effects of lead on the male reproductive system in mice. Journal of Toxicology and Environmental Health Part A 56 (7), 513–521.

Zenick, H., Clegg, E.D., Perrault, S.D., Klinefelter, G.R., Gray, L.E., 1994. Assessment of male reproductive toxicity, a risk assessment approach. In: Hayes, A.W. (Ed.), Principles and Methods of Toxicology, third ed. Raven Press, New York, NY, USA. (Chapter 27).