FUNDAMENTAL AND APPLIED TOXICOLOGY 15, 651–665 (1990)

Semen Analysis and Fertility Assessment in Rabbits: Statistical Power and Design Considerations for Toxicology Studies

JACQUELINE WILLIAMS,*,[1] BETH C. GLADEN,† STEVEN M. SCHRADER,‡ TERRY W. TURNER,‡ JERRY L. PHELPS,* AND ROBERT E. CHAPIN*

**Developmental and Reproductive Toxicology Group, National Toxicology Program, and †Statistics and Biomathematics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709; and ‡Experimental Toxicology Branch, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, Ohio 45226.*

Received December 18, 1989; accepted July 6, 1990

Semen Analysis and Fertility Assessment in Rabbits: Statistical Power and Design Considerations for Toxicology Studies. WILLIAMS, J., GLADEN, B. C., SCHRADER, S. M., TURNER, T. W., PHELPS, J. L., AND CHAPIN R. E. (1990). *Fundam. Appl. Toxicol.* **15,** 651–665. Semen analysis is commonly used in evaluating human response to reproductive toxicants. Serial semen samples can be collected from rabbits and fertility assessed by artificial insemination, hence this species is potentially well suited for male reproductive toxicity studies that might be extrapolated to humans. However, the size and cost of rabbits often restricts the number of animals used, reducing the sensitivity of such studies. Therefore, it was of interest to optimize study design for semen analysis and fertility assessment in rabbits. Semen samples were collected weekly from sexually mature New Zealand white rabbits and a range of parameters was analyzed (Semen—pH, volume, osmolality; Sperm—number and concentration, morphology, viability, percentage motility, motion characteristics; Seminal plasma—fructose, citric acid, carnitine and protein concentrations, acid phosphatase activity). Male fertility was assessed by inseminating female rabbits with the minimum number of motile sperm required for normal fertility, determined to be one million. The within- and between-buck variabilities were determined for all parameters and used to calculate the statistical power of different study designs. The variability of sperm number and concentration was decreased when measured in four ejaculates collected within a short period of time rather than in a single ejaculate; this was not true of other endpoints measured. In addition, use of preexposure observations further increased the statistical power for all of the parameters. These data can be used to determine the optimum design for studies of male reproductive toxicity using rabbits, with particular regard to cost and the number of animals used.

Rodents are the species most frequently used in routine reproductive toxicology studies. However, of the reproductive parameters readily measured in such animals (e.g., tissue weights, histology, epididymal sperm characteristics, blood hormone profiles, and fertility)(Clegg *et al.*, 1986) only a few can be assessed in humans (e.g., testis size *in situ,* hormone profiles)(see Working, 1988). Semen analysis, frequently used to monitor human reproductive function, provides valuable information about the reproductive tract (Grunfeld, 1989). Since ejaculated semen can not be easily collected from rodent species, direct comparison of reproductive function or toxicity between humans and rodents is not always possible.

The rabbit, the smallest common laboratory species from which serial semen samples can be readily collected, is potentially well

[1] Present address: ICI PLC, Central Toxicology Laboratory, Alderley Park, Nr Macclesfield, Cheshire, SK10 4TJ, United Kingdom.

0272-0590/90 $3.00

suited for male reproductive toxicology studies (reviewed by Amann, 1982). Additionally, male fertility can be assessed by artificial insemination of the female with known numbers of sperm (Tesh and Tesh, 1971). Hence, the fertilizing capacity of sperm from treated males can be evaluated, a question not readily addressed in humans or rodents. However, the expense of purchasing and maintaining rabbits, and test material availability, very often limits the number of animals used in toxicology studies. One consequence of using small numbers of animals per group is a reduction in statistical power (the probability of detecting a change of defined magnitude). For example, Desjardins *et al.* (1968) demonstrated that by using five rabbits per group and collecting semen samples for 20 weeks there was only a 25% chance of detecting a 50% change in sperm production. Therefore, it is desirable to optimize study design so as to maximize its power while minimizing the number of animals required. Unfortunately, there is limited information available on study designs for rabbit male reproductive toxicology studies, notable exceptions being the studies by Berndtson *et al.* (1989) and Desjardins *et al.* (1968).

The aim of the present study was to assess the within- and between-rabbit variability for various parameters analyzed in semen. The endpoints chosen were based on their origin within the male reproductive tract and, hence, can be considered as potential functional markers for such sites. In addition, since fertility assessment in animal species is biologically insensitive due to the vast excess of sperm produced (Working, 1988), we attempted to determine the minimum number of sperm required for normal fertility in the New Zealand white (NZW) rabbit. The coefficients of variation for all of these evaluations were then used to calculate the power of different study designs.

The data reported herein should allow for a more appropriate choice of study design for using rabbits in male reproductive toxicity studies.

MATERIALS AND METHODS

Animals

Sexually mature noninbred NZW rabbits, supplied by Hazelton Research Products, Inc. (Denver, PA), were acclimated for 2 weeks. The male rabbits (mean age = 9 months, range = 8–10 months, and mean body weight = 4.3 kg) were trained, for 1–2 months, to give semen. Virgin female rabbits were 6–7 months old (mean body weight = 3.9 kg). Animals were housed individually in stainless steel cages (dimensions, 24 × 24 × 14.5 in) under standard conditions (temperature, 65 ± 2°F, 40 ± 20% relative humidity and 12-h light:12-h dark cycle). Food (5322 Certified Rabbit Chow, Purina Mills, Inc., St Louis, MO) and deionized water were available *ad libitum.*

Chemicals

All biochemicals and chemicals, purchased from J. T. Baker Chemical, Co. (Phillipsburg, NJ), Mallinckrodt, Inc. (Paris, KY), or Sigma Chemical, Co. (St. Louis, MO), were of the highest purity available.

Semen Collection

Semen was collected using an artificial vagina and teaser doe (Tesh and Tesh, 1971). Since sexual preparation enhances the quantity of semen produced (MacMillan and Hafs, 1967), each male was allowed three false mounts of the female before collecting semen. For multiple ejaculates (20–30 min between successive ejaculates for any individual male), the males were sexually prepared for only the first two samples, since the libido of some males was decreased if false mounting was allowed for further ejaculates.

Semen Analysis

Any gel plug present in the semen was immediately removed, since it contains high concentrations of citric acid (Holtz and Foote, 1978) which would increase the variability of this endpoint. The pH (using a micro-electrode; Microelectrodes, Inc., Londonderry, NH) and osmolality (using a vapor pressure osmometer; Wescor, Inc., Logan, UT) of whole semen (indicative of accessory sex gland activity) were measured within 5–10 min of sample collection. Ejaculate volume was determined to the nearest 0.05 ml using a plastic syringe of 1 or 3 ml capacity.

TABLE 1

PARAMETERS USED FOR COMPUTERIZED ANALYSIS OF RABBIT SPERM (CELLSOFT SYSTEM)

System	Parameter	Value
Main	Number of frames to analyze	15
	Number of frames/sec	30
	Video standard	A
	Minimum number of points for motility	2
	Minimum number of points for velocity	4
	Maximum velocity, μm/sec	250
	Threshold velocity, μm/sec	20
	Cell color	white
	Pixel scale	0.662
	Cell size range, pixels	15–40
Lateral head amplitude	Minimum number of points	7
	Minimum velocity, μm/sec	20
	Minimum linearity	1.5
Circular motion	Minimum number of points	4
	Minimum velocity, μm/sec	20
	Maximum radius, (μm)	80

(a) Motility Assessment

All equipment in contact with ejaculated semen was maintained at 37°C. Within 5–10 min of collection, semen was diluted (typically 1:20) in Ham's F-10 tissue culture medium without protein. Five microliters of diluted semen, placed into a 10-μm Makler chamber (Zygotek Systems, Inc., Springfield, MA), was examined by microscopy using a 10X negative phase objective and 10X ocular attached to a video camera and recorder. Eight fields, at predetermined sites, were video recorded for 10 sec per field onto RXPRO videotapes, and samples identified by a concurrent audio recording. Curvilinear and straight line velocities (Vcl, Vsl), amplitude of lateral head displacement (ALH), linearity (LIN), beat cross frequency (BCF), and the percentage circular were determined with the CellSoft system (Cryo Resources, New York, NY), using the settings listed in Table 1. The number of sperm analyzed per sample from a videotape ranged from 2 to 185 (median = 59). Hinting *et al.* (1988) reported that analysis of 20 motile sperm is sufficient to achieve reliable estimation of motility characteristics.

(b) Sperm Concentration and Total Number

Semen was diluted (typically 1:200) in phosphate-buffered saline (PBS) and 10% neutral buffered formalin. The concentration of sperm was determined by the hemocytometer method and counting two chambers per sample (Freund and Carol, 1964). The total number of sperm ejaculated was calculated from the sperm concentration and ejaculate volume data.

(c) Sperm Viability

The physiological and structural integrity of the sperm membrane was assessed using a modification of the hypoosmotic swelling test (Jeyendran *et al.*, 1984). Semen was diluted 1:20 in a hypoosmotic solution (11.8 mM sodium citrate, 19.3 mM D-fructose; osmolality 75 mOSM) and incubated at 37°C for 1 h. The cells were fixed by adding 8 μl of 50% glutaraldehyde. The percentage of sperm with swollen (coiled) tails was assessed using phase contrast microscopy, and counting a minimum of 300 sperm per sample, whenever possible, i.e., in 93% of the samples.

(d) Percent Motility

The percentage of motile sperm in a sample was determined from the videotapes. The total number of sperm in a visual field was determined, the tape advanced several frames, and sperm in their original positions were counted as nonmotile. Whenever possible, 200 sperm were counted per sample, i.e., in 61% of the samples.

(e) Morphology

Semen was diluted (typically 1:150) in PBS, and 20 μl smeared onto a clean glass microscope slide. The smears were than airdried at 37°C. The sperm were stained in a single step with a trypan blue, naphthol yellow, and eosin-Y solution (0.8, 0.4, and 0.2% w/v, respectively, in 1% acetic acid) as described by Unnithan (1976). Two hundred sperm for each rabbit were examined under phase contrast at 400× and classified as normal or abnormal.

(f) Biochemical Measurements

Semen was kept at room temperature for less than 45 min before centrifugation. Previous data (not shown) indicated that this procedure did not affect the biochemical parameters evaluated (see below). Semen was centrifuged at 2000*g* for 15 min at 4°C, and the supernatant (seminal plasma) used for biochemical analysis. Total and tartrate-resistant acid phosphatase (AP) activities were analyzed in fresh samples, and the remaining seminal plasma was stored frozen until assayed for fructose, citric acid, carnitine, and protein. Seminal plasma was

deproteinized (except for protein analysis) prior to assay as described by Mollering (1985), with one exception: for neutralization, 0.6 M perchloric acid was added to 0.75 M potassium carbonate (ratio 3.8:1). For the protein, fructose, citric acid, and carnitine assays, concentrations were determined, in duplicate, by reference to standard curves (protein: 0–0.6 mg/ml, fructose: 0–20 mM, citric acid: 0–20 mM, and *L*-carnitine: 0–15 mM).

(i) Acid phosphatase activity. (origin = epididymis; Jones, 1974): The method used was adapted from Moss (1985). Fifty microliters of diluted (1:10 in 45.9 mM citrate buffer, pH 4.9) seminal plasma was added to 250 μl of substrate solution (10.7 mM 4-nitrophenyl phosphate; total AP activity) or substrate plus inhibitor solution (75 mM L-tartaric acid; tartrate-resistant AP activity). Following incubation for 75 min at 37°C, the reaction was stopped by the addition of 2 ml 0.1 M NaOH solution. The increase in absorbance at 405 nm was determined by reference to a blank, prepared with water instead of seminal plasma.

(ii) Protein concentration. The protein concentration (indicative of accessory sex gland function in general) was analyzed according to Bradford (1976) using bovine serum albumin as a standard.

(iii) Fructose concentration. The concentration of fructose (origin = prostate; Holtz and Foote, 1978) was determined by an endpoint assay with sorbitol dehydrogenase (SDH), modified from the method of Anderson *et al.* (1979). Five hundred microliters of deproteinized sample was added to a reaction mixture containing 0.56 μmol β-NADH in 0.1 M sodium phosphate buffer, pH 6.8, in a final volume of 2 ml. Samples were incubated for 90 min at 25°C in the presence of 3.8 units of sheep liver SDH, and the decrease in absorbance was measured at 340 nm.

(iv) Citric acid concentration. Citric acid (origin = glandular vesicularis; Holtz and Foote, 1978) concentrations were determined as described by Mollering (1985) except that the final incubation time was 30 min.

(v) Carnitine concentration. Carnitine (origin = epididymis; Hinton *et al.*, 1979) concentrations were determined as described by Marquis and Fritz (1964). Two hundred to seven hundred microliters of deproteinized sample was assayed in 0.1 M Tris–HCl buffer, pH 7.5, containing 0.2 μmol 5,5,-dithiobis(2-nitrobenzoic acid), 0.5 μmol acetyl coenzyme A (lithium salt), and 0.2 units of pigeon breast muscle carnitine acetyltransferase in a final volume of 2 ml. The increase in absorbance at 412 nm was measured following incubation for 30 min at 25°C.

Artificial Insemination

The day before insemination serum progesterone concentrations were determined by radioimmunoassay (Coat-a-Count, Diagnostic Products Co., Los Angeles, CA) as a test for pseudopregnancy (Browning *et al.*, 1980). Pseudopregnant rabbits (progesterone concentration > 1 ng/ml) were eliminated from the study.

Ovulation was stimulated by an ear vein injection of 30 units hCG (Pregnyl, Organon Inc., West Orange, NJ) per female, 1–2 hr before insemination. Semen was collected (no more than an hour before insemination) and the sperm concentration determined. The percentage of motile sperm was determined by counting nonmotile sperm in a hemocytometer at 37°C, cooling the sample on ice (to immobilize the sperm), and then counting the total number of sperm in the same fields of view. The semen was then diluted in oviduct simulating medium (Alvarez and Storey, 1982) to the desired concentration of motile sperm (see below). Each female rabbit received 0.25 ml of this diluted semen intravaginally using glass insemination pipettes.

A titration curve to determine the minimum number of motile sperm required for normal fertility was constructed by inseminating 4–20 females with a total of 0.25, 0.5, 0.75, 1, 2, 5, 10 or >200 million motile sperm from 1 male. The variation in the number of implantation sites resulting from the insemination of one million motile sperm from different males was also determined, by inseminating 2–8 females for each of an additional 28 males.

On Day 13 of gestation (insemination day = Day 0), females were killed with an overdose of anesthetic (T61, Taylor Pharmacal Co., Decatur, IL) and the total number of embryos and resorptions determined.

Statistical Analysis

Coefficients of variation were calculated using standard analysis of variance techniques. The VARCOMP procedure of SAS was used to calculate the within-buck and between-buck components of variance (SAS, 1985). Power and sample size calculations were performed using routine assumptions and methods, and the formulae used are shown in the Appendix. All measurements were assumed to be normally distributed with constant within-buck and between-buck variances. Some measurements deviated from normality and were transformed. However, results with the transformed versions were not different from those obtained with the untransformed ones, so, for simplicity, only the latter are presented. Means were assumed to be constant within a specified treatment group and time period. Power calculations were based on the usual assumption that variances could be estimated well enough to be considered known. We considered first a design with a control group and a treated group; the corresponding test was assumed to be based on the difference between the control and treated means (essentially a *t* test). We then considered a design with a preexposure period, one group (control) would be untreated throughout the study and the second group

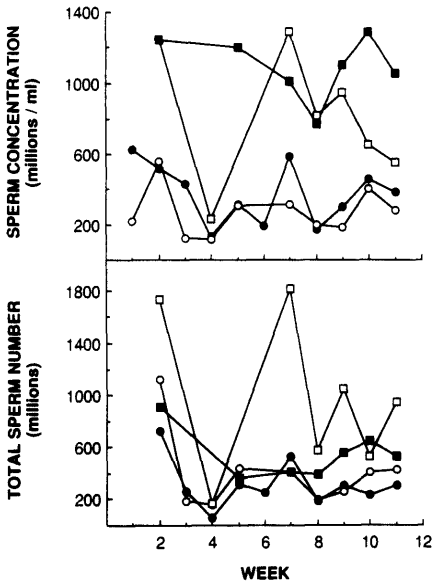

FIG. 1. The variability of sperm concentration and number in single ejaculates. One ejaculate per male rabbit was collected once a week. Each curve represents data from an individual animal.

would be untreated in the first period and treated in the second part. The corresponding test was assumed to be based on a comparison of the change between periods rather than on the means of the groups. All powers were calculated for a significance level of 0.05.

RESULTS

Semen Analysis

One ejaculate per week was collected from four rabbits for 11 weeks. Sperm concentration and total sperm number exhibited a large variation both within and between rabbits over time (Fig. 1). The CVs for within a rabbit (W) were 53% (sperm concentration) and 71% (total sperm number), and for between rabbits (B) were 54 and 47%, respectively.

In an attempt to reduce this variability in sperm number, multiple ejaculates were collected from each rabbit within a short period of time. At least 90% of the sperm obtained in six ejaculates (collected 15–20 min apart) was present in the first four samples (data not shown). Figure 2 shows that the total number of sperm and the sperm concentration in four ejaculates are considerably less variable [CVs sperm number: 23% (W) and 20% (B); sperm concentration: 17% (W) and 47% (B)] than those determined in a single ejaculate.

The effect of collecting multiple ejaculates (four samples per male, $n = 10$) twice weekly for 5 weeks on the overall variability was also determined. For both sperm concentration and total number in the first and total ejaculates, there was no advantage in collecting samples twice a week. Indeed, in some cases, the variability was actually increased (data not shown).

Using the data illustrated in Figs. 1 and 2, the power of two study designs was calculated using the formulae derived in the Appendix. Design 1 is a conventional design involving a control group and various treatment groups. Design 2 includes a preexposure period in

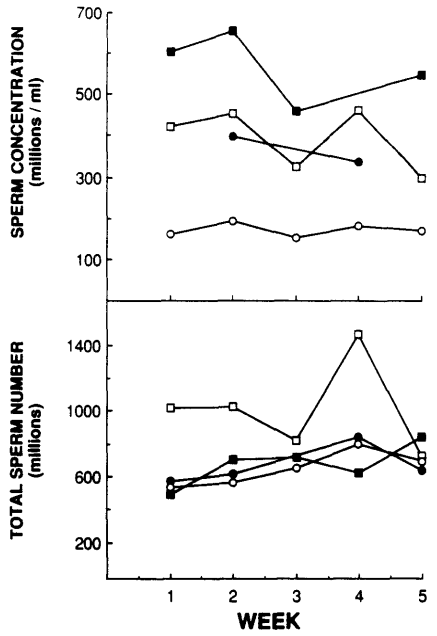

FIG. 2. The variability of sperm concentration and number in four ejaculates. Semen was collected from each rabbit (four ejaculates per male) once a week. The time interval between successive ejaculates for any male was 15–20 min. Each curve represents data from an individual rabbit, the same animals as in Fig. 1.

TABLE 2

STATISTICAL POWER OF VARIOUS STUDY DESIGNS

Percentage change to be detected	Total sperm number in one ejaculate		Total sperm number in four ejaculates	
	Design 1	Design 2	Design 1	Design 2
10	0.06	0.07	0.12	0.21
20	0.10	0.13	0.32	0.71
30	0.16	0.23	0.61	0.96
40	0.24	0.38	0.85	1.00
50	0.35	0.54	0.96	1.00

Note. The values in the table are the probabilities of detecting each respective percentage change in total sperm number, which is shown for two study designs comparing one ejaculate with four ejaculates. Both designs use five animals per group, with semen samples collected once a week for 15 weeks. Design 2 includes a 5-week preexposure period followed by a 10-week exposure (postexposure) period. The calculations used the formulae derived in the Appendix and the data shown in Figs. 1 and 2.

which baseline data are obtained for all animals, followed by a treatment or treatment/posttreatment phase. Thus, in addition to a concurrent control group each animal effectively acts as its own control. For design 1, using five animals per group and measuring sperm number in single ejaculates collected once a week for 15 weeks, there is only a small chance (35%) of detecting even a substantial change (50%) produced by a treatment (Table 2). This probability is considerably increased (to 96%) by evaluating the total sperm number in four ejaculates. A further increase in power is achieved by using design 2 (Table 2).

The smaller the change to be detected and the higher the power required, the more animals per group are needed, regardless of endpoint or design (Table 3). However, for the same change and power, these results again emphasize the advantage of design 2 over design 1, and measuring the total number of sperm in four ejaculates compared with one ejaculate.

Another major influence on statistical power is the number of observations made per animal (equivalent to time in a longitudinal study). In design 1, using a 10-week exposure period, there is only a 31% chance of detecting a 20% change in the total number of sperm in four ejaculates (Table 4). However, increasing the length of the preexposure period (N1) in design 2 has a large influence on power (Table 4).

Additional endpoints, evaluating accessory sex gland function or sperm characteristics (see Table 5), were measured in the first and total (four) ejaculates collected once or twice a week from 10 rabbits. Overall, there was only a small reduction, if any, in the variability of any of the biochemical endpoints measured either in four ejaculates per week or in the first of the ejaculates collected twice weekly, compared with the first of four ejaculates collected once a week (data not shown). Hence, these endpoints were subsequently measured only in the latter.

The mean values and the within- (W) and between-buck (B) variabilities of the various parameters measured in semen are presented in Table 5. The B/W ratio (Table 5) can be used to determine the best study design to use for a defined set of conditions. In general, when the between-buck variability is large, design 2 is preferable, and when the within-buck variability is large, design 1 is better. For example, for a total study length of 15 weeks with five animals per group, design 2 (where N1 = 5 weeks and N2 = 10 weeks) is preferable to design 1 (see Appendix) when the $B/$

TABLE 3

THE NUMBER OF ANIMALS REQUIRED PER GROUP TO ACHIEVE POWER 0.8 FOR DIFFERENT STUDY DESIGNS

Percentage change to be detected	Total sperm number in one ejaculate		Total sperm number in four ejaculates	
	Design 1	Design 2	Design 1	Design 2
10	394	234	71	26
20	99	59	18	7
30	44	26	8	3
40	25	15	5	2
50	16	10	3	2

Note. The values shown are the number of animals required per group for two study designs and power of 0.8. The designs use groups of equal size and semen samples were collected once a week for 15 weeks. Design 2 uses a preexposure period of 5 weeks followed by a 10-week exposure (postexposure) period. The calculations used the formulae derived in the Appendix and the data shown in Figs. 1 and 2.

W ratio is >0.483 for all of the parameters except for the sperm characteristics derived from the CellSoft System. In the latter case, design 2 is preferable when the B/W ratio is >0.068 (assuming that 50 sperm [i.e., the number of observations] per rabbit are analyzed each time). Thus, under these conditions, design 2 is more suitable for all of the endpoints shown in Table 5.

TABLE 4

THE INFLUENCE OF LENGTH OF SAMPLING TIME ON STATISTICAL POWER FOR DETECTING A 20% CHANGE IN THE TOTAL NUMBER OF SPERM IN FOUR EJACULATES FOR TWO STUDY DESIGNS

Length of preexposure period in weeks (N1)	Length of exposure (or postexposure) period in weeks (N2)
N1	N2 = 10
0	0.31
2	0.42
4	0.64
6	0.76
8	0.82
10	0.86

Note. Statistical power is shown for two study designs using five animals per group. For design 1, in which there is no preexposure period, N1 = 0. Design 2 includes a preexposure period (N1) followed by an exposure (postexposure) period (N2). The calculations used the data presented in Fig. 2.

The power and minimum detectable effect (the smallest percentage change from control which can be detected) of both study designs for all of these parameters are shown in Table 6. It is noteworthy that for all of the endpoints shown, design 2 is more powerful and sensitive than design 1. A major assumption made for design 2 is that there are no age or seasonal effects, otherwise the power would be reduced. For all semen endpoints analyzed over a 6-month period, there were no differences due to age (11–16 months) or season (August–January; data not shown). Similar findings were reported by Macari and Machado (1978).

Fertility Assessment

It is well recognized that many laboratory animals produce far more sperm than are actually required to be maximally fertile (Working, 1988). Therefore, the chances of detecting decreased male fertility are improved by reducing the number of sperm inseminated. In our studies, the mean number of fetuses and resorptions (produced by in-

TABLE 5

THE WITHIN- AND BETWEEN-RABBIT VARIABILITY FOR VARIOUS PARAMETERS IN SEMEN FROM UNTREATED RABBITS

Parameter	Mean	% CV		
		Between (B)	Within (W)	B/W
Accessory sex gland function				
[Fructose] mM	11.6	43	25	1.7
Volume (four ejaculates) ml	1.62	36	26	1.4
Tart-res AP units/liter	58.8	32	23	1.4
AP units/liter	86.9	27	20	1.4
[Protein] mg/ml	15.1	21	17	1.3
[Carnitine] mM	2.01	43	33	1.3
% Inhibition (AP)	33.2	22	21	1.0
Tart-res AP units/g	3.94	21	27	0.8
Volume (1st ejaculate) ml	0.61	29	39	0.7
AP units/g	5.91	18	24	0.7
[Citric acid] mM	11.0	35	58	0.6
pH	7.14	2	3	0.6
Osmolality mOSM	257	3	5	0.6
Testicular function				
[Sperm] four ejaculates millions/ml	445	34	24	1.4
Total sperm four ejaculates millions	698	35	32	1.1
[Sperm] 1st ejaculate millions/ml	441	43	43	1.0
Total sperm first ejaculate millions	262	47	52	0.9
Sperm function				
% Viable	59.0	26	22	1.2
% Circular	23.9	30	39	0.8
% Motile	87.2	7	10	0.7
% Abnormal forms	13.0	19	34	0.6
Vcl, μ/sec	117	11	33	0.3
Vsl, μ/sec	108	12	36	0.3
ALH	2.3	16	56	0.3
LIN	9.1	3	16	0.2
BCF	15.0	3	33	0.1

Note. Semen was collected from 42 rabbits (four ejaculates/male) once a week for 6 weeks. The time interval between successive ejaculates for any one male was 20–30 min. The results shown are for the first ejaculates unless otherwise stated. The values within each category are ranked according to the B/W ratio.

semination with greater than 200 million motile sperm) was 8.88 ± 1.02 (mean ± SE, n = 16 females). When female rabbits were inseminated with 10, 5, 2, or 1 million motile sperm (from the same male), the number of implantation sites was not significantly less than this value. However, when the inseminations used less than one million motile sperm there was a significant reduction in the total number of fetuses and resorptions (Fig. 3).

The variability in the number of implantation sites produced by insemination with one million motile sperm from an additional 28 males was also assessed. The overall mean number of fetuses and resorptions was 7.16, and the within-buck and between-buck CVs were 16 and 44%, respectively (n = 29 males and 121 females). The statistical power of fertility using artificial insemination with one million motile sperm was calculated using these data and the formula for study design 1

TABLE 6

THE STATISTICAL POWER AND MINIMUM DETECTABLE EFFECT OF TWO STUDY DESIGNS FOR VARIOUS PARAMETERS IN RABBIT SEMEN

Parameter	Power[a]		Minimum detectable effect[b]	
	Design 1	Design 2	Design 1	Design 2
Accessory sex gland function				
pH	1.00	1.00	4	3
Osmolality	1.00	1.00	5	4
[Protein]	0.41	0.98	33	14
AP units/liter	0.28	0.93	41	17
% Inhibition (AP)	0.39	0.90	34	18
Tart-res AP units/liter	0.21	0.85	49	19
AP units/g	0.51	0.81	29	20
[Fructose]	0.14	0.78	66	21
Vol four ejaculates	0.17	0.76	56	22
Tart-res AP units/g	0.40	0.72	33	23
[Carnitine]	0.14	0.55	66	27
Vol first ejaculate	0.23	0.42	46	32
[Citric acid]	0.16	0.22	58	48
Testicular function				
[Sperm] four ejaculates	0.19	0.81	53	20
Total sperm four ejaculates	0.18	0.56	54	27
[Sperm] 1st ejaculate	0.13	0.36	67	36
Total sperm 1st ejaculate	0.12	0.26	74	43
Sperm function				
LIN	1.00	1.00	5	2
Vcl	0.92	1.00	17	4
BCF	1.00	1.00	5	4
Vsl	0.90	1.00	18	5
% Motile	1.00	1.00	12	9
ALH	0.63	1.00	25	7
% Viable	0.30	0.89	40	18
% Abnormal forms	0.42	0.51	32	29
% Circular	0.22	0.43	47	32

[a] The values shown under "power" are the probabilities of detecting a 20% change from the control mean.

[b] The values shown under "minimum detectable effect" are the smallest percentage change from the control mean which can be detected with 80% power. Both designs assume seven rabbits per group and semen samples collected once a week for 15 weeks. Design 1 has no preexposure period. Design 2 has a 5-week preexposure period and a 10-week exposure (postexposure) period. The results are calculated from the data in Table 5. The values within each category are ranked according to the power of Design 2.

(see Appendix). The number of females inseminated per male is the number of samples per male. Due to the large variability, the power of artificial insemination with one million motile sperm is very low (Table 7). For example, using 3 females per male and 6 males per treatment group (i.e., a total of 24 animals per group), the smallest change in litter size that can be detected with an 80% probability is 49%. Increasing the number of females inseminated per male has only a small effect; even with 24 females for each of 6 males, the smallest change detectable is 31%. If design 2 is considered, with the fertility of the males being tested prior to exposure, there is no increase in statistical power (data not shown).

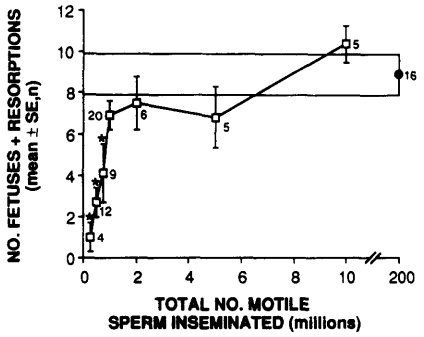

FIG. 3. Artificial insemination titration curve. Female rabbits were artificially inseminated with different numbers of motile sperm from one male rabbit as shown. Females were killed on Gestation Day 13 and the uterine contents examined. All values are means ± SEM, *n*, number of females inseminated. Student's *t* test, $*p < 0.05$, compared with the value for >200 million motile sperm.

DISCUSSION

This report is the first one to detail the results of a full semen analysis in rabbits, evaluating many aspects of male reproductive function, such as the testis, accessory sex gland function (general and site-specific), and sperm characteristics. Computer assisted semen analysis has not previously been applied extensively to rabbit sperm. The advantage of a sophisticated characterization of sperm movement is that factors important in fertility could be identified. Overall, the importance of such a detailed semen analysis is that direct comparisons can be made with humans as an aid in cross-species extrapolation.

The data and/or formulae presented can be used to optimize study designs with rabbits in male reproductive toxicity studies for efficient use of available resources. Ideally, pilot data on variability should be collected before a study, as it is influenced by many different factors (e.g., environmental conditions, genetic factors, assay variation, etc.). Such data, perhaps in conjunction with the reported values, would enable an approximation of the required study conditions. Clearly, there is no single optimum design for all parameters described. However, each study can be tailored specifically to address a particular aspect of male reproductive function via semen analysis. Such an approach is useful when information about a toxicant is available from studies with other species and/or with structurally related chemicals.

Sperm concentration and total number were among the most variable parameters analyzed in rabbit semen. The present results demonstrated that collecting four ejaculates/male once a week was the least variable of the collection frequencies evaluated. Desjardins *et al.* (1968) also noted that intensive ejaculation frequencies decreased variability in sperm number compared to that of single ejaculates. For both collection frequencies (single vs multiple ejaculates) sperm concentration was less variable than total sperm number. However, the latter is more biologically relevant, as sperm concentration depends upon accessory sex gland fluid secretion (Amann, 1982).

In addition, the present results demonstrated that the inclusion of a preexposure period into the study design increased the statistical power for all semen parameters. In assessing a toxic insult on spermatogenesis in the rabbit via semen analysis, a treatment/

TABLE 7

ARTIFICIAL INSEMINATION OF FEMALE RABBITS WITH ONE MILLION MOTILE SPERM: STATISTICAL POWER

% Change in litter size	Number of females per male		
	3	6	24
10	0.09	0.11	0.15
20	0.21	0.30	0.46
30	0.41	0.57	0.80
40	0.63	0.82	0.96
50	0.82	0.95	0.99

Note. The values shown are the probabilities (statistical power) for detecting a change in litter size of defined magnitude. The number of male rabbits per treatment group = 6. The number of females inseminated per male is equivalent to the number of observations per male (see Appendix for formula).

posttreatment period of 9–10 weeks is the minimum time that should be considered. This period is the estimated production time of an ejaculated sperm from a spermatogonium in this species (Amann, 1982).

The overall variability of a study is comprised of biological and analytical variation, and the latter should be reduced as much as possible. The inter- and intra-assay variations of the biochemical assays used in the present evaluations were small, i.e., <10%. The analytical variation for computer-assisted semen analysis is also low (e.g., Ginsburg *et al.*, 1988). It is well known that the hemocytometer method for counting sperm is variable (Freund and Carol, 1964). This variability can be reduced by one person counting many chambers per sample, e.g., eight, although this is very labor intensive. Unfortunately, automated methods (Coulter Counter, and CellSoft) were unsuitable in our hands for counting rabbit sperm (data not shown) due to the presence of nonsperm particulates in the semen. Hence, until improved methods become available, the present analytical variation must be accommodated into the study design.

Fertility assessment is an integral part of a male reproductive toxicity study, since the ability to fertilize ova is the functional test of sperm and is not, as yet, predictable from semen analysis (Amann, 1989). The present study demonstrated that one million motile sperm can be considered to be the minimum number of sperm for normal reproductive capacity of the NZW rabbit, in close agreement with previous reports (e.g., Austin, 1948; Wales *et al.*, 1965). However, this number will vary with species, strain, and possibly with individual animals. Ideally, although not practical, a titration curve should be constructed for each male. A more feasible method would be to evaluate each male's fertility at, e.g., 1 and 0.5 million motile sperm, to more accurately define the minimum number required. This procedure could be used to evaluate preexposure fertility, i.e., design 2, which, although not statistically advantageous, is biologically advantageous in identifying infertile males.

Since most animal species produce a large excess of sperm, it has been suggested that a >90% decrease in sperm fertilizing capacity would have to occur before any change in fertility could be detected (Aafjes *et al.*, 1980; Working, 1988). Indeed, our data suggest that a reduction of >99.6% in the number of motile sperm would have to occur before any decrease in fertility could be detected in the NZW rabbit.

The large variation in fertility assessment by artificial insemination with one million motile sperm is comprised of male and female components. If the within-female variability is less than that for between-females, then it might be beneficial to reuse females for assessing male fertility. However, the disadvantages of reusing females include extra housing costs (larger cages for bearing young, longer housing periods) and multiparous does tend to have decreased fertility (Wales *et al.*, 1965).

Despite the low power of fertility assessment by artificial insemination with reduced numbers of sperm, the biological sensitivity is increased (see also Amann, 1982; Working, 1988). Thus, it is surprising that when male fertility was assessed in reproductive toxicity studies using rabbits, few, if any, attempts were made to adequately reduce the number of sperm inseminated. Some studies used natural mating (e.g., John *et al.*, 1983; Rao *et al.*, 1982; Foote *et al.*, 1986), despite its poor success rate (Foote *et al.*, 1986). Even when artificial insemination was used, excess sperm were still inseminated (e.g., Foote *et al.*, 1986; Chang *et al.*, 1980). Hence, it is possible, when negative effects on fertility were reported (Chang *et al.*, 1980; Foote *et al.*, 1986; John *et al.*, 1983), that the fertility test used was biologically insensitive to detect subtle effects.

Statistical power provides information on the probability of detecting a change of defined magnitude if it occurs. The actual change produced by a toxicant will depend

upon the biological susceptibility of the parameter in question. The likelihood of a change occurring can not be gained from variability data, but only from experiments with male reproductive toxicants. Such studies are currently underway with rabbits. Alterations in seminal characteristics that have previously been reported for various reproductive toxicants in the rabbit have been substantial for certain endpoints. For example, Rao *et al.* (1982) reported a 57% decrease in sperm count following inhalation exposure to dibromochloropropane (10 ppm). Fox *et al.* (1963) noted that tretamine produced a 500% increase in morphologically abnormal sperm. As data variability was not reported, it is not possible to determine the power of the study designs used and, therefore, to comment on the biological versus statistical sensitivity for these endpoints. It is feasible that smaller changes in other endpoints in these studies were not detected due to variability, thus, emphasizing the need for optimizing study design so that male reproductive toxicants do not go undetected in the rabbit.

In summary, the rabbit is awaiting full utilization as a model for human response to male reproductive toxicants. Every effort should be made to optimize study design, i.e., to maximize the chance of detecting adverse effects. We have presented variability data for semen analysis and fertility assessment and its effects on the power of various study designs, which should prove useful in achieving such an aim.

APPENDIX

1. Notation for a Single Group

To develop the notation, first consider a single group of R rabbits with N samples each. The jth measurement from the ith rabbit is denoted x_{ij}. Rabbits are assumed to differ in their mean; for the ith rabbit, the mean of the measurements is denoted μ_i. The variance of x_{ij} about this mean is denoted W^2; this within-rabbit variability is assumed to be the same for all rabbits. The means μ_i themselves have a distribution; their mean is denoted μ, and their (between-rabbit) variance is denoted B^2. Then overall, the mean of x_{ij} is μ and the variance of x_{ij} is $W^2 + B^2$. The within-rabbit coefficient of variation (CV) is defined to be W/μ; the between-rabbit CV is similarly defined as B/μ.

Now construct the grand mean $\bar{x}$, defined simply as the sum of all the observations divided by the number of observations:

$$\bar{x} = \frac{\sum_{i=1}^{R} \sum_{j=1}^{N} x_{ij}}{RN}.$$

Simple calculations show that the mean, variance, and CV of $\bar{x}$ are:

$$E(\bar{x}) = \mu$$

$$\mathrm{Var}(\bar{x}) = \frac{W^2}{RN} + \frac{B^2}{R}$$

$$\mathrm{CV}(\bar{x}) = \sqrt{\frac{(W/\mu)^2}{RN} + \frac{(B/\mu)^2}{R}}.$$

2. Testing for Differences Using Design 1

Now suppose that we have two groups of R rabbits and we have N samples from each rabbit (for a grand total of $2RN$ samples). Suppose that the means of the two groups may differ but the variances do not. Using notation analogous to that used above, we will have:

$$E(\bar{x}_1) = \mu_1$$

$$E(\bar{x}_2) = \mu_2 = \theta\mu_1$$

$$\mathrm{Var}(\bar{x}_1) = \mathrm{Var}(\bar{x}_2) = \frac{W^2}{RN} + \frac{B^2}{R}.$$

We wish to test the hypothesis that the two means are equal. We will test by looking at the differences in the two means and rejecting when the difference is too large; if $N = 1$, this is a t test. We have:

$$E(\bar{x}_2 - \bar{x}_1) = \mu_2 - \mu_1$$

$$\mathrm{Var}(\bar{x}_2 - \bar{x}_1) = 2\left(\frac{W^2}{RN} + \frac{B^2}{R}\right).$$

Define the quantity Z_1 as the ratio of the mean of the difference to the standard deviation of the difference:

$$Z_1 = \frac{E(\bar{x}_2 - \bar{x}_1)}{\mathrm{SD}(\bar{x}_2 - \bar{x}_1)} = \frac{\theta - 1}{\sqrt{2}\mathrm{CV}(\bar{x}_1)} = \frac{(\theta - 1)\sqrt{R}}{\sqrt{2}\sqrt{\frac{B^2}{\mu_1^2} + \frac{W^2}{N\mu_1^2}}}.$$

Let α be the probability of a type 1 error (the probability of wrongly rejecting the null hypothesis of equality). Define C as $\Phi^{-1}(1 - \alpha/2)$, where Φ is the standard normal (Gaussian) distribution function. Conventionally, α is chosen to be 0.05, which gives $C = 1.96$; other values can be looked up in tables of Φ or obtained from computer programs. The power of the test (the probability of correctly rejecting the null hypothesis) will then be $\Phi(Z_1 - C) + \Phi(-Z_1 - C)$.

3. *Testing for Differences Using Design 2*

Now suppose that the two groups of rabbits consist of one group which is untreated throughout the study and a second group which is untreated in the first part of the study and treated in the second part. Suppose that each rabbit has N_1 samples in the first period and N_2 in the second, giving a grand total of $2R(N_1 + N_2)$ samples. Note that the sum $N_1 + N_2$ here is analogous to N in the previous design. We allow for the possibility that means vary with both treatment and period; for example, if the vehicle has an effect itself, even the control group will be changed in the second period. Variances are again assumed independent of treatment or period. Letting the first subscript denote group and the second denote period, we have:

$$E(\bar{x}_{11}) = E(\bar{x}_{21}) = \mu_1$$

$$E(\bar{x}_{12}) = A\mu_1$$

$$E(\bar{x}_{22}) = \theta A\mu_1.$$

We test for differences between groups by comparing $\bar{x}_{22} - \bar{x}_{21}$ to $\bar{x}_{12} - \bar{x}_{11}$. In the previous section, we compared the measurements in group 1 to those in group 2; now we compare the changes in group 1 to those in group 2. By looking only at the changes each rabbit undergoes, the fact that each rabbit may have a different baseline becomes unimportant. Define Z_2 in a manner similar to Z_1 above:

$$Z_2 = \frac{E((\bar{x}_{22} - \bar{x}_{21}) - (\bar{x}_{12} - \bar{x}_{11}))}{\mathrm{SD}((\bar{x}_{22} - \bar{x}_{21}) - (\bar{x}_{12} - \bar{x}_{11}))}.$$

Straightforward calculation shows that:

$$Z_2 = \frac{A(\theta - 1)\sqrt{R}}{\sqrt{2}\sqrt{\frac{B^2}{\mu_1^2}(1 - A)^2 + \frac{W^2}{\mu_1^2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}.$$

Again, the power of the test is given by $\Phi(Z_2 - C) + \Phi(-Z_2 - C)$. Note that if there are no differences between the two periods, so that $A = 1$, then we have:

$$Z_2 = \frac{(\theta - 1)\sqrt{R}}{\sqrt{2}\sqrt{\frac{W^2}{\mu_1^2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}.$$

and the between-rabbit standard deviation B plays no role.

4. *Comparison of the Two Designs*

Suppose that the number of rabbits and the total number of samples per rabbit is fixed and you wish to determine whether design 1 or design 2 is better. The power of the test associated with design j is $\Phi(Z_j - C) + \Phi(-Z_j - C)$. Normally, one of these terms will be quite small. If treatment decreases the measurements, so that $\theta < 1$, then $Z_j < 0$ and power is approximately $\Phi(-Z_j - C)$. Conversely, if $\theta > 1$, power is approximately $\Phi(Z_j - C)$. In either case, the approximate power

for design 1 will be less than that for design 2 if $|Z_1| < |Z_2|$. Recall that:

$$Z_1 = \frac{(\theta - 1)\sqrt{R}}{\sqrt{2}\sqrt{\frac{B^2}{\mu_1^2} + \frac{W^2}{(N_1 + N_2)\mu_1^2}}}$$

$$Z_2 = \frac{A(\theta - 1)\sqrt{R}}{\sqrt{2}\sqrt{\frac{B^2}{\mu_1^2}(1 - A)^2 + \frac{W^2}{\mu_1^2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}.$$

Thus design 1 will be less powerful if:

$$\left[\frac{1}{N_1} + \frac{1}{N_2} - \frac{A^2}{N_1 + N_2}\right]W^2 < [A^2 - (1 - A)^2]B^2.$$

If $A = 1$, this condition simplifies to:

$$\left[\frac{1}{N_1} + \frac{1}{N_2} - \frac{1}{N_1 + N_2}\right]W^2 < B^2.$$

We prefer the first design when B/W is small and the second when B/W is large.

ACKNOWLEDGMENTS

The authors greatly appreciate the technical assistance expertly provided by J. D. Allen, L. Harris, M. Harris, E. Haskins, L. Judd, and W. Purdie, and also thank Sharyn Wilkins for the excellent preparation of the manuscript.

REFERENCES

AAFJES, J. H., VELS, J. M., AND SCHENCK, E. (1980). Fertility of rats with artificial oligozoospermia. *J. Reprod. Fertil.* **58,** 345–351.

ALVAREZ, J. G., AND STOREY, B. T. (1982). Spontaneous lipid peroxidation in rabbit epididymal spermatozoa: Its effect on sperm motility. *Biol. Reprod.* **27,** 1102–1108.

AMANN, R. P. (1982). Use of animal models for detecting specific alterations in reproduction. *Fundam. Appl. Toxicol.* **2,** 13–26.

AMANN, R. P. (1989). Can the fertility potential of a seminal sample be predicted accurately? *J. Androl.* **10,** 89–98.

ANDERSON, R. A., REDDY, J. M., OSWALD, C., AND ZANEVELD, L. J. D. (1979). Enzymic determination of fructose in seminal plasma by initial rate analysis. *Clin. Chem.* **25,** 1780–1782.

AUSTIN, C. R. (1948). Number of sperms required for fertilization. *Nature (London)* **162,** 534–535.

BERNDTSON, W. E., NEEFUS, C., FOOTE, R. H., AND AMANN, R. P. (1989). Optimal replication for histometric analysis of testicular function in rats or rabbits. *Fundam. Appl. Toxicol.* **12,** 291–302.

BRADFORD, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal. Biochem.* **72,** 248–254.

BROWNING, J. Y., KEYES, P. L., AND WOLF, R. C. (1980). Comparison of serum progesterone, 20α-dihydroprogesterone, and estradiol-17β in pregnant and pseudopregnant rabbits: Evidence for postimplantation recognition of pregnancy. *Biol. Reprod.* **23,** 1014–1019.

CHANG, M. C., GU, Z., AND SAKSENA, S. K. (1980). Effect of gossypol on the fertility of male rats, hamsters and rabbits. *Contraception* **21,** 461–469.

CLEGG, E. D., SAKAR, C. S., AND VAYTEK, P. E. (1986). Assessment of reproductive risks. *Biol. Reprod.* **34,** 5–16.

DESJARDINS, C., KIRTON, K. T., AND HAFS, H. D. (1968). Sperm output of rabbits at various ejaculation frequencies and their use in the design of experiments. *J. Reprod. Fertil.* **15,** 27–32.

FOOTE, R. H., SCHERMERHORN, E. C., AND SIMKIN, M. E. (1986). Measurements of semen quality, fertility and reproductive hormones to assess dibromochloropropane (DBCP) effects in live rabbits. *Fundam. Appl. Toxicol.* **6,** 628–637.

FOX, B. W., JACKSON, H., CRAIG, A. W., AND GLOVER, T. D. (1963). Effects of alkylating agents on spermatogenesis in the rabbit. *J. Reprod. Fertil.* **5,** 13–22.

FREUND, M., AND CAROL, B. (1964). Factors affecting haemocytometer counts of sperm concentration in human semen. *J. Reprod. Fertil.* **8,** 149–155.

GINSBURG, K. A., MOGHISSI, K. S., AND ABEL, E. L. (1988). Computer-assisted human semen analysis: Sampling errors and reproducibility. *J. Androl.* **9,** 82–90.

GRUNFELD, L. (1989). Workup for male infertility. *J. Reprod. Med.* **34,** 143–149.

HINTING, A., SCHOONJONS, F., AND COMHAIRE, F. (1988). Validation of a single-step procedure for the objective assessment of sperm motility characteristics. *Int. J. Androl.* **11,** 277–287.

HINTON, B. T., SNOSWELL, A. M., AND SETCHELL, B. P. (1979). The concentration of carnitine in the luminal fluid of the testis and epididymis of the rat and some other mammals. *J. Reprod. Fertil.* **56,** 105–111.

HOLTZ, W., AND FOOTE, R. H. (1978). Composition of rabbit semen and the origin of several constituents. *Biol. Reprod.* **18,** 286–292.

JEYENDRAN, R. S., VAN DER VEN, H. H., PEREZ-PELAEZ, M., CRABO, B. G., AND ZANEVELD, L. J. D. (1984). Development of an assay to assess the func-

tional integrity of the human sperm membrane and its relationship to other semen characteristics. *J. Reprod. Fertil.* **70,** 219–228.

JOHN, J. A., QUAST, J. F., MURRAY, F. J., CALHOUN, L. G., AND STAPLES, R. E. (1983). Inhalation toxicity of epichlorohydrin, effects on fertility in rats and rabbits. *Toxicol. Appl. Pharmacol.* **68,** 415–423.

JONES, R. (1974). Absorption and secretion in the cauda epididymis of the rabbit and the effects of degenerating spermatozoa on epididymal plasma after castration. *J. Endocrinol.* **63,** 157–165.

MACARI, M., AND MACHADO, C. R. (1978). Sexual maturity in rabbits defined by the physical and chemical characteristics of the semen. *Lab. Anim.* **12,** 37–39.

MACMILLAN, K. L., AND HAFS, H. D. (1967). Semen output of rabbits ejaculated after varying sexual preparation. *PSEBM* **125,** 1278–1281.

MARQUIS, N. R., AND FRITZ, I. B. (1964). Enzymological determination of free carnitine concentrations in rat tissues. *J. Lipid Res.* **5,** 184–187.

MOLLERING, H. (1985). Citrate. In *Methods of Enzymatic Analysis.* (H. U. Bergmeyer, J. Bergmeyer, M. Grassi, Eds.), Vol. VII, 3rd ed. pp. 2–12. VCH Publishers, FL.

MOSS, D. W. (1985). Acid phosphatases. In *Methods of Enzymatic Analysis.* (H. U. Bergmeyer, J. Bergmeyer, M. Grassi, Eds.), Vol. IV, 3rd ed., pp. 92–106. VCH Publishers, FL.

RAO, K. S., BUREK, J. D., MURRAY, F. J., JOHN, J. A., SCHWETZ, B. A., BEYER, J. E., AND PARKER, C. M. (1982). Toxicologic and reproductive effects of inhaled 1,2,dibromochloropropane in male rabbits. *Fundam. Appl. Toxicol.* **2,** 241–251.

SAS Institute Inc. (SAS)(1985). *SAS Users Guide: Statistics Version,* fifth ed. SAS Institute Inc., Cary NC.

TESH, S. A., AND TESH, J. M. (1971). Artificial insemination in the rabbit and its use in routine teratogenic studies. *Excerpta Medica (Amsterdam) Int. Congress Series* **220,** 332–336.

UNNITHAN, R. R. (1976). A rapid single-step stain for mammalian spermatozoa. *Indian J. Exp. Biol.* **14,** 611–613.

WALES, R. G., MARTIN, L., AND O'SHEA, T. (1965). Effect of dilution rate and of the number of spermatozoa inseminated on the fertility of rabbits ovulated with chorionic gonadotrophin. *J. Reprod. Fertil.* **10,** 69–78.

WORKING, P. K. (1988). Male reproductive toxicology: Comparison of the human to animal models. *Environ. Health Perspect.* **77,** 37–44.