Materials and Methods for the Perinatal and Three-month Toxicity Study of Triphenyl Phosphate (CASRN 115-86-6) Administered in Feed to Sprague Dawley (Hsd:Sprague Dawley® SD®) Rats

June 2026

Division of Translational Toxicology
National Institute of Environmental Health Sciences
National Institutes of Health

Research Triangle Park, North Carolina, USA

Table of Contents

Tables

Figures

About This Document

Triphenyl phosphate (TPHP) is an organophosphate flame retardant widely used in a variety of industries, with concern growing about potential health risks associated with exposure via inhalation or ingestion. A dose range-finding study in Sprague Dawley rats previously conducted by researchers in the Division of Translational Toxicology at the National Institute of Environmental Health Sciences revealed that the reproductive performance of dams was perturbed at ≥10,000 ppm TPHP following exposure via feed. In the offspring, TPHP-related toxicity (reductions in survival and body weights) was noted in pups at ≥10,000 ppm, and delays in puberty and reduced cholinesterase enzyme activity were observed starting at 3,000 ppm and 1,000 ppm TPHP, respectively.[1] These findings set the foundation for a subsequent study to evaluate the subchronic toxicity and developmental neurotoxicity of TPHP following exposure via dosed feed in male and female Sprague Dawley (Hsd:Sprague Dawley® SD®) rats. Exposure began in utero and continued through lactation and weaning, followed by exposure for a minimum of 3 months.

This document details the Materials and Methods used to conduct the 3-month study in rats and obtain the data posted on the Chemical Effects in Biological Systems (CEBS) data page.[2] A peer-reviewed journal publication on this in vivo, repeat-dose biological potency study of TPHP is in preparation, with anticipated publication in fall 2026.

Materials and Methods

Procurement and Characterization of Triphenyl Phosphate

Triphenyl phosphate (TPHP) was obtained from Acros Organics (Fair Lawn, NJ) in a single lot (lot A0343614). Identity, purity, and stability analyses were conducted at the analytical chemistry laboratory (MRIGlobal, Kansas City, MO). Reports on analyses performed in support of the TPHP studies are on file at the National Institute of Environmental Health Sciences (NIEHS).

Lot A0343614 appeared as white, crystalline, solid flakes at room temperature. The identity of lot A0343614 was evaluated using Fourier transform infrared (IR) spectroscopy, ultraviolet spectroscopy (UV/Vis), ^{1}H and ^{13}C nuclear magnetic resonance (NMR) spectroscopy, direct infusion mass spectrometry (DIMS), melting point determination, and elemental analysis. The IR, UV/Vis, ^{1}H NMR, and ^{13}C NMR spectra were consistent with the reference spectra[3-5] and with the structure of TPHP. The monoisotopic mass determined by DIMS (327.3 Da) was consistent with the molecular weight of TPHP. The melting point was determined to be consistent with the literature value.[6] Elemental analysis was performed by Galbraith Laboratories, Inc. (Knoxville, TN) to aid in identification. The relative amounts of carbon (66.20%), hydrogen (4.63%), and phosphorus (9.48%) were within <0.1% of the theoretical values.

Purity evaluation was conducted using gas chromatography (GC) with mass spectrometry (MS), GC with flame ionization detection (FID), differential scanning calorimetry (DSC), GC headspace analysis for residual solvent, and Karl Fischer titration for water content. GC/MS determined that all containers of the lot were TPHP with chromatographic purity >99%. GC/FID determined a purity of 99.85%; one impurity had a peak area >0.1% of the total integrated peak area. No residual solvents were identified using GC headspace analysis. DSC determined that the test article was 99.78% pure. Karl Fischer titration performed at Galbraith Laboratories (Knoxville, TN) yielded a water content of 0.02%. The overall purity of lot A0343614 was determined to be >99%.

The test article was stored at room temperature under inert gas.

Preparation and Analysis of Dose Formulations

Dose formulations were prepared with lot A0343614 at concentrations of 0, 30, 100, 300 1,000, and 3,000 ppm in both NIH-07 and NTP-2000 feed at the testing facility throughout the 3-month study in rats. Formulations in NIH-07 and NTP-2000 feed were stored frozen at approximately −30°C to −15°C in sealed containers protected from light and were used within 42 days and 63 days of preparation, respectively.

Prior to study start, the homogeneity and stability of dose formulations using lot A0343614 were determined by the analytical chemistry laboratory using liquid chromatography tandem mass spectrometry (LC-MS/MS). Homogeneity of TPHP in NIH-07 and NTP-2000 feed was tested at 10 ppm. All formulations analyzed were determined to be homogeneous and of appropriate concentrations. The stability of the 10 ppm dose formulation was evaluated in NIH-07 and NTP-2000 feed when stored under ambient, refrigerated, or frozen conditions. The TPHP formulation

was stable in NTP-2000 feed under all conditions; however, in NIH-07 feed, it was stable only under frozen conditions and protected from light. Stability of TPHP was confirmed for up to 42 days under frozen conditions and protected from light for NIH-07 and in ambient, refrigerated, or frozen conditions for 63 days in NTP-2000 feed.

Analysis of preadministration dose formulations was conducted on five batches of the NIH-07 feed and on three batches of the NTP-2000 feed. Analysis of the postadministration dose formulations was conducted on the same batches of NIH-07 and NTP-2000 feed used for preadministration analyses. Postadministration samples were collected from each formulation barrel and from the animal rooms at the end of the exposure period. All preadministration and postadministration samples were within 10% of the target concentration for each dose formulation, except for one batch of preadministration 3,000 ppm formulation in NIH-07 feed, which exceeded the percent of target criteria at 110.9%.

Animal Source

Time-mated (F_0) female Sprague Dawley (Hsd:Sprague Dawley® SD®) rats were obtained from Inotiv (Envigo at time of procurement; Indianapolis, IN).

Animal Welfare

Animal care and use were in accordance with the Public Health Service Policy on Humane Care and Use of Animals. The animal study was conducted in an animal facility accredited by AAALAC International. The study was approved by the Battelle (West Jefferson, OH) Animal Care and Use Committee (ACUC number T06093) and conducted in accordance with all relevant NIEHS/Division of Translational Toxicology (DTT) animal care and use policies and applicable federal, state, and local regulations and guidelines.

Exposure Concentration Selection Rationale

In a dose range-finding (DRF) study, exposure concentrations of 1,000 to 30,000 ppm TPHP were administered in feed to groups of time-mated F_0 female rats from gestation day (GD) 6 through delivery of F_1 pups and through the lactation period until weaning of the F_1 pups on postnatal day (PND) 28.[1] After weaning, F_1 pups were administered TPHP in feed at the same exposure concentration as their respective F_0 dams until study termination on PND 56. In the DRF study, TPHP increased mortality in F_0 females at ≥15,000 ppm and in F_1 males and females at ≥10,000 ppm. However, no overt toxicity was observed in F_0 females and F_1 animals exposed to 1,000 and 3,000 ppm TPHP. Exposure-related effects included lower body weights in F_1 offspring during lactation and after weaning at 10,000 ppm, as well as a delay in pubertal development (i.e., vaginal opening), and decreased brain cholinesterase in F_1 females at 3,000 ppm. TPHP-related findings at 1,000 ppm were limited to lower brain acetylcholinesterase (AChE) activity in F_1 males. Given these findings, TPHP exposure concentrations of 0, 30, 100, 300, 1,000, and 3,000 ppm were selected for the evaluation of subchronic and developmental neurotoxicity in the current 3-month study.

Three-month Study

Study Design for Rats

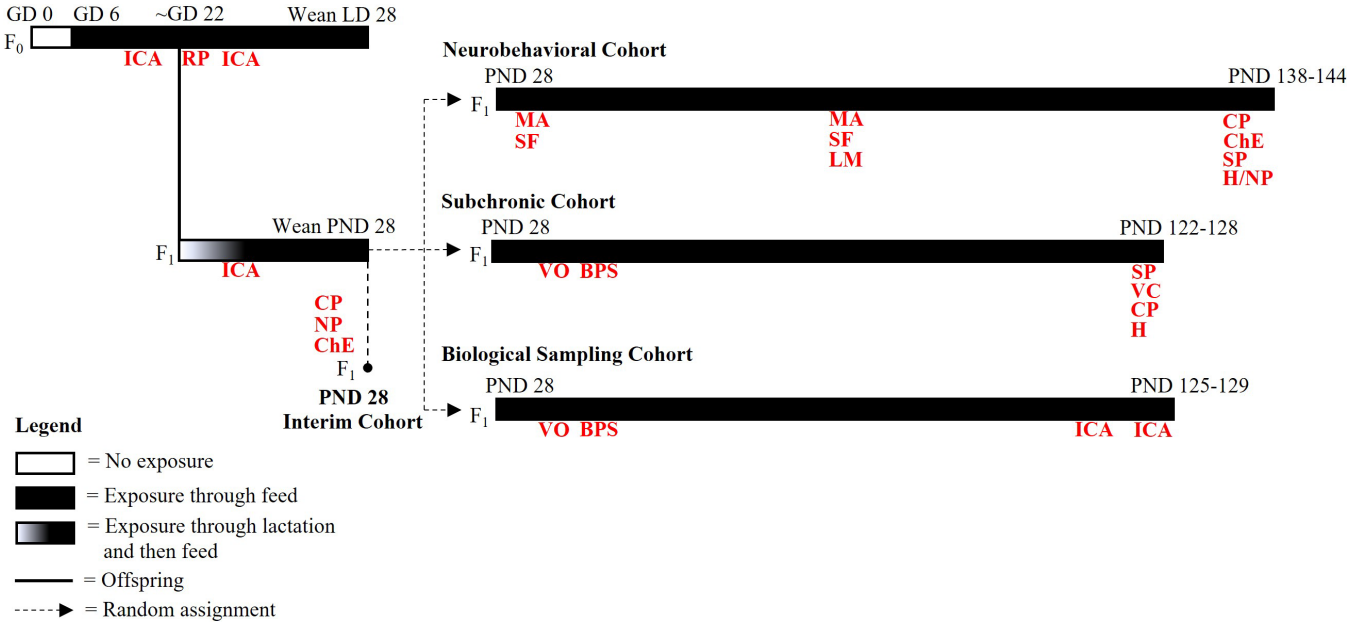

Figure 1. Design of the Perinatal and Three-month Rat Study

GD = gestation day; LD = lactation day; ICA = internal concentration assessment; RP = reproductive performance; PND = postnatal day; MA = motor activity; SF = sensorimotor function; LM = learning and memory; CP = clinical pathology; ChE = cholinesterase activity; SP = sperm parameters; H = histopathology; NP = neuropathology; VO = vaginal opening; BPS = balanopreputial separation; VC = vaginal cytology.
F_0 dams were exposed to the test article from GD 6 through weaning on LD 28 and evaluated for maternal toxicity. F_1 offspring were exposed in utero through PND 28 and evaluated for signs of toxicity. After weaning, F_1 offspring were allocated into the PND 28 interim, neurobehavioral, subchronic, or biological sampling cohorts, exposed to the test article until necropsy, and evaluated for signs of toxicity.

F_0 female Sprague Dawley rats were 12 to 14 weeks old upon receipt. Time-mated animals were received on GD 1 or 2 and held for 4 or 5 days. F_0 females were randomly assigned to one of six exposure groups on GD 5. Randomization was stratified by body weight to produce similar group mean weights using National Toxicology Program (NTP) Provantis software (Instem, Stone, UK).

F_0 female rats were quarantined for 12 to 14 days after receipt. Ten nonmated female rats received in the same shipment as the time-mated dams were used for disease monitoring. Samples were collected for serology and parasite evaluation 2 days after arrival and then rats were euthanized, necropsied, and examined for the presence of disease or parasites at 4 weeks. Blood samples were collected from each sentinel animal and allowed to clot, and the serum was separated. Additionally, fecal samples were collected and tested for endoparasites and *Helicobacter* species. All samples were processed appropriately, with serology and *Helicobacter* testing performed by IDEXX BioAnalytics (formerly IDEXX BioResearch, Columbia, MO) for determination of the presence of pathogens. Evaluation for endo- and ectoparasites was performed in-house by the testing laboratory. The health of the animals was monitored during

the study according to the protocols of the NIEHS/DTT Sentinel Animal Program. All test results were negative.

Beginning on GD 6, groups of 36 F_0 female rats were fed diets containing 0, 30, 100, 300, 1,000, or 3,000 ppm TPHP throughout gestation and lactation. Water and dosed feed were available ad libitum. F_0 females were housed individually during gestation and with their respective litters during lactation. Cages were changed weekly for pregnant dams before delivery and twice weekly for dams and their litters after lactation day (LD) 4. Natural crinkled kraft paper (Crink-l'Nest; The Andersons, Inc., Maumee, OH) was provided to F_0 females during the gestation period for environmental enrichment. The crinkled kraft paper was removed from each F_0 female's cage on GD 19 and added back to the cages of dams with litters on LD 4. F_0 females were observed twice daily for signs of mortality or moribundity. Body weights were recorded on GD 5 (for randomization), and on GDs 6, 9, 12, 15, 18, and 21. Clinical observations were recorded daily from GD 6 through delivery. Feed consumption was measured on GDs 6, 9, 12, 15, 18, and 21. Details of the study design and animal maintenance are summarized in Table 1.

The day of parturition was considered LD 0 for dams and PND 0 for pups. F_0 females that did not deliver were euthanized between GD 24 and GD 26, and the uteri were examined for evidence of implantation and resorption. Body weights and feed consumption of F_0 females that littered were recorded on LDs 1, 4, 7, 10, 13, 16, 19, 21, 25, and 28. Clinical observations were recorded daily from LD 1 through LD 28. From PND 0 through PND 28, the number and sex of pups for each litter were recorded daily. On PND 1, litter weights by sex were recorded. Clinical observations were recorded daily from PND 1 through PND 28, and pups were observed for external developmental abnormalities and the absence of a milk band from PND 1–4. F_1 pups were individually weighed on PNDs 1, 4, 7, 10, 13, 16, 19, 21, 25, and 28.

On PND 4, the litters were standardized to eight pups per litter (4/sex/litter when possible). At weaning on PND 28, offspring were randomly assigned to one of four cohorts: the PND 28 interim (2/sex/litter when possible from 19 [control group] or 20 [exposed groups] litters/group), neurobehavioral (1/sex/litter from 19 control litters or 20 exposed litters), subchronic (1/sex/litter from 19 control litters or 20 exposed litters, with males and females from different litters), or biological sampling (1/sex/litter from 10 litters/exposure group, with males and females from different litters) cohort. Five pups per sex from control litters not assigned to cohorts and five pups per sex from control litters of the subchronic cohort were used for parasite evaluation, serology, and gross observation of disease; evaluations were performed at study termination. All test results were negative. Following weaning, all F_0 females and unselected pups were humanely euthanized with carbon dioxide. F_0 females were examined for evidence of implantation, and the number of implantation sites was recorded. Weaning marked the beginning of the 3-month study. Starting on PND 28, F_1 animals (pups) were administered the same exposure concentrations that their respective dams received during gestation and lactation: 0, 30, 100, 300, 1,000, or 3,000 ppm TPHP in feed for approximately 3 months.

After weaning on PND 28, F_1 male rats were housed up to two per cage and F_1 female rats were housed up to four per cage. Water and dosed feed were available ad libitum. Cages were changed and sanitized at least twice weekly. Red rectangular polycarbonate shelters (Rat Retreats™, Bio-Serv, Flemington, NJ) were provided for environmental enrichment. F_1 rats were observed twice daily for signs of mortality or moribundity. Body weights were recorded on PND 28, once

weekly thereafter, and at study termination. Feed consumption and clinical observations were recorded weekly. Details of the study design and animal maintenance are summarized in Table 1.

Two diets were used in the study: (1) NIH-07 during the perinatal phase and (2) NTP-2000 during the postweaning phase, except for the neurobehavioral cohort, which received NIH-07 for the duration of the study. The NIH-07 diet is a higher protein diet that supports reproduction and lactation in rodents, whereas the NTP-2000 diet is a lower protein diet that decreases the incidence of chronic progressive nephropathy in adult rats.

Endocrine-sensitive and Pubertal Endpoints

Anogenital distance (AGD) and corresponding body weight (for covariate analyses) were recorded for all live F_1 pups on PND 1. The distance between the midpoint of the anal opening to the caudal edge of the genital papilla was recorded and converted from micrometers (μm) to millimeters (mm). All F_1 male pups were evaluated for retention of areolae/nipples on PND 13. Attainment of balanopreputial separation (BPS), defined as complete retraction of the prepuce from the glans penis, was evaluated in F_1 male rats in the subchronic and biological sampling cohorts beginning on PND 35 until the day of attainment, and body weights were recorded on the day of attainment. The attainment of vaginal opening (VO) was evaluated in all F_1 female pups beginning on PND 25 until the day of attainment or PND 28. Female pups assigned to the subchronic and biological sampling cohorts continued VO examinations after PND 28 until the day of attainment or at least PND 42, and the corresponding body weights were recorded upon VO attainment.

PND 28 Interim Cohort

Blood was collected from all F_1 rats in the interim evaluation cohort for clinical chemistry on PND 28 (see below for more details). Animals were anesthetized with a 70% CO_2/30% O_2 mixture and bled in a random order. Following blood collection, animals were euthanized via CO_2 inhalation and brains were collected and processed for AChE and butyrylcholinesterase (BChE) activity analysis (5/sex/exposure group from up to 5 litters/exposure group).

Neurobehavioral Cohort

Neurobehavioral assessments were performed for all rats in the neurobehavioral cohort. Testing was conducted at two ages for spontaneous locomotor activity and acoustic startle response (juvenile and adult) and at one age (adult) for the Morris water maze (MWM). Juvenile neurobehavioral assessments were conducted for all exposure groups, and the adult neurobehavioral assessments were conducted only for the 0, 30, 1,000, and 3,000 ppm groups. The individual apparatus, testing order, and experimenter history were maintained across animals, and each assessment was conducted on different days and staggered according to individual postnatal weaning dates. For the start of testing, if animals were staggered across days, the same number of animals/sex/exposure group was tested on each day. All testing was conducted between approximately 8:30 a.m. and 1:00 p.m.

Spontaneous locomotor activity (also known as motor activity) was assessed between PND 29 and PND 35 (juvenile) and again between PND 57 and PND 63 (adult) using an automated photocell apparatus (MotorMonitor II system; Kinder Scientific, Poway, CA). Rats were placed in the center of the open field arena (40 cm × 40 cm × 20 cm) containing a photocell device with a 16 × 16 sensor array configuration. Infrared photobeam breaks measured basic movements and

rearing in 5-minute epochs over a 60-minute testing period. Data collected over the first 30 minutes of testing were analyzed for juveniles and the first 45 minutes for adults.

Acoustic startle assessed sensorimotor function in rats between PND 30 and PND 36 (juvenile) and between PND 58 and PND 64 (adult) using the StartleMonitor II (load cell transducer; Kinder Scientific, Poway, CA). All exposures to the testing chamber were conducted under a constant 65 dB background noise level. At least 1 day prior to testing and on the day of testing, rats were acclimated to the restrainer and the testing chamber for 5 minutes. On the day of testing, rats were placed in the restrainer, which was connected to a sensing transducer, to measure aspects of the startle response. To evaluate habituation, a 120 dB startle stimulus was delivered for 50 trials with a 20 second intertrial interval (ITI). Startle magnitudes were sampled over a 150-millisecond sampling window within the total 500-millisecond recording window of the Kinder Scientific system. Maximum force in newtons (MaxN) was captured for each trial.

MWM assessments of spatial learning and memory started between PND 68 and PND 73 (adult). Animals were tested over four successive phases: (1) cued learning (2 consecutive days), (2) spatial acquisition learning (7 consecutive days), (3) reference memory using a probe trial on the 10th day (followed by approximately 48 hours before the final phase), and (4) reversal learning trials (3 consecutive days). The circular tank (180 cm diameter) containing a 10 cm plexiglass escape platform was filled with water (24°C–27°C) made opaque by the addition of milk powder to a depth of 28–35 cm. The pool was designated into four quadrants. During phases 1, 2, and 4, animals received four sequential trials per day with a 60 second variable ITI. For each trial, animals were placed into a distal start quadrant at the periphery of the tank with the nose facing the wall. Animals were allotted 90 seconds to find the platform. After finding the platform, they were allowed to remain on the platform for approximately 20 seconds. If an animal failed to find the platform, it was gently guided to and helped onto the platform for approximately 20 seconds.

Phase 1 (visible platform) of cued learning was conducted to confirm each animal's ability to see the platform—which was raised 1.5 cm above the surface of the water and identified by a visible flag projection—to swim and to climb onto the platform for escape. During this phase, curtains surrounded the tank to obstruct the view of distal cues. For spatial acquisition (hidden platform; phase 2), the platform was submerged 1 cm below the surface of the water within the middle of an identified escape quadrant, which was set across all training days. Approximately 24 hours following the final day of the spatial acquisition trials, a single probe trial was conducted to assess reference memory (phase 3). During the 90-second probe trial, the platform was removed from the tank, and animals were allowed to freely swim the entire tank. Approximately 48 hours following the probe trial, reversal learning (phase 4) was conducted during which the hidden platform was moved to the opposite quadrant relative to its location during phase 2. This phase assessed the ability of the animals to shift from a previously learned location and learn a new platform location. Video tracking software (ANY-Maze, Stoelting Co., Wood Dale, IL) was used to the track the movement of each animal. Mean daily latency to the platform and distance traveled were analyzed for cued, acquisition, and reversal learning trials. For the probe test, distance traveled, time spent in each quadrant, and swim speed were assessed for all quadrants in 30-second epochs and across the total session.

At the end of the 3-month study, on PND 138–144, blood was collected from selected F_1 rats in the neurobehavioral cohort for AChE and BChE activities (5/sex/exposure group); clinical

chemistry (10/sex/exposure group), including a subset of special chemistries (see below for more details); and the peripheral blood micronucleus assay (10 males/exposure group). Animals were anesthetized with a 70% CO_2/30% O_2 mixture and bled in a random order. Following blood collection, animals were euthanized via CO_2 inhalation, and tissues were collected for sperm parameters (up to 10/exposure group in the 0, 300, 1,000, and 3,000 ppm groups) and neuropathology and histopathology (up to 15/sex/exposure group).

Subchronic Cohort

General toxicity was assessed in all F_1 rats in the subchronic cohort. For 16 consecutive days before scheduled study termination, samples were collected for vaginal cytology evaluations. At the end of the 3-month study, on PND 122–128, animals were anesthetized with a 70% CO_2/30% O_2 mixture and bled in a random order. Following the collection of blood samples for clinical chemistry, hematology, and the peripheral blood micronucleus assay (see below for more details), animals were euthanized via CO_2 inhalation. The animals were necropsied in a random order and tissues were collected for sperm parameters and histopathology.

Biological Sampling Cohort

On GD 18, three F_0 females in each exposure group were euthanized via decapitation approximately 2 hours into the light cycle, and trunk blood samples for internal concentration assessment (ICA) were collected into tubes containing tripotassium ethylenediaminetetraacetic acid/sodium fluoride (K_3 EDTA/NaF). The samples were centrifuged and the plasma was harvested. The uterus of each dam was removed and opened, and amniotic fluid was collected and pooled by dam but excluded from analysis. After collection of amniotic fluid, fetuses were collected, and the entire litter (pooled two to three fetuses per vial) was flash frozen in liquid nitrogen.

On LD 4, animals were euthanized via decapitation, and trunk blood was collected approximately 2 hours into the light cycle from up to three F_0 females and their respective pups from the 30 and 100 ppm groups for ICA. Because there was an insufficient number of litters in the 0, 300, 1,000, and 3,000 ppm groups, three dams from these groups had a blood collection from the jugular vein, and they continued on the study. Pups were euthanized via decapitation and blood samples were collected from the culled pups of these dams and other selected litters from these groups for ICA; whole brains were collected for potential analysis of AChE and BChE activities but not analyzed. Blood samples were collected in tubes containing K_3 EDTA/NaF, centrifuged, and processed to plasma. The samples were flash frozen in liquid nitrogen. All samples were stored in a freezer set to maintain approximately −85°C to −60°C until shipped to an NIEHS/DTT-designated laboratory.

Approximately 1 week prior to the end of the 3-month study, on PND 117–119, urine samples were collected for TPHP analysis from F_1 rats (5/sex/exposure group). To collect urine, animals were placed in metabolism cages and given control feed (0 ppm) and water ad libitum, and samples were collected over a 24-hour period. Urine samples collected for chemical analysis were flash frozen in liquid nitrogen and stored in a freezer set to maintain approximately −85°C to −60°C until shipped to an NIEHS/DTT-designated laboratory.

At study termination on PND 125–129, blood and tissue collection began as soon as feasible into the light cycle. Blood samples were collected from the vena cava or abdominal aorta from anesthetized animals into tubes containing K_3 EDTA/NaF, centrifuged, and processed to plasma.

The samples were flash frozen in liquid nitrogen. The brain was also collected for ICA, weighed, and stored frozen. All samples were stored in a freezer set to maintain approximately −85°C to −60°C until shipped to an NIEHS/DTT-designated laboratory.

Clinical Examinations and Pathology

Cholinesterase Activity Analyses

Acetylcholinesterase (AChE) and butyrylcholinesterase (BChE) activity analyses were conducted on brain and plasma samples. Whole brains were collected on PND 4 from pups in the biological sampling cohort or from designated culled pups but not analyzed. Whole brains were also collected from selected F_1 rats in the PND 28 interim cohort. In addition, at the end of the 3-month study, whole blood was collected (1.0 mL into tubes containing K_3 EDTA) from the retroorbital plexus of select rats in the neurobehavioral cohort. Brain and blood samples were stored in a freezer set to maintain approximately −85°C to −60°C until used for AChE and BChE activity analysis. Brain samples were homogenized with phosphate-buffered saline and protease inhibitor tablets. Homogenates were rocked for an hour at 2°C to 8°C to ensure solubilization of the membranes. Homogenized samples were then centrifuged at 10,000 relative centrifugal force (RCF) for 2 minutes at 4°C. The supernatant was removed and stored at −85°C to −60°C until analysis the following day. Whole blood samples were separated into cellular and plasma fractions by centrifugation at 1,300 RCF for 10 minutes at 4°C. The cellular fraction samples underwent additional processing to prepare erythrocyte membranes and remove hemoglobin. The resulting erythrocyte membrane preparations and the plasma were stored at −85°C to −60°C until analysis. AChE activity was assessed in erythrocyte membrane preparations, whereas BChE activity was assessed in plasma using a spectrophotometric assay. Absorbance was measured on a BioTek Synergy HTX Multimode microplate reader utilizing Gen5™ Secure Data Collection and Analysis Software (Agilent, Santa Clara, CA). Cholinesterase data were subsequently reported as units of enzyme activity per gram of total protein (U/g) (brain tissue) or as units of enzyme activity per liter of whole blood or plasma (U/L).

Clinical Pathology

Blood was collected from the retroorbital plexus of all animals in the PND 28 interim cohort (on PND 28) and subchronic cohort (at the end of the 3-month study) or from the jugular vein of animals in the neurobehavioral cohort (at the end of the 3-month study) for hematology (subchronic cohort only), clinical chemistry, and special chemistries (neurobehavioral cohort only) analyses. The parameters measured are listed in Table 1. Blood for hematology was collected into tubes containing K_3 EDTA. Blood for clinical chemistry, including the special chemistries (C-peptide, insulin, leptin, nonesterified fatty acids) was collected into tubes void of anticoagulant, allowed to clot, centrifuged, and the serum harvested. Hematology parameters were analyzed using an Advia 2120i hematology analyzer (Siemens Medical Solutions USA, Inc., Malvern, PA). Clinical chemistry parameters were analyzed using a Roche cobas® c501 Chemistry Analyzer (Roche, Indianapolis, IN). C-peptide concentrations were analyzed using the singleplex Meso Scale Diagnostics (MSD) U-PLEX platform (Rockville, MD). Analyses of insulin, and leptin concentrations were performed using enzyme-linked immunosorbent assay (ELISA) kits: Millipore Sigma Rat/Mouse Insulin ELISA (Burlington, MA) and Invitrogen Rat Leptin (Waltham, MA), respectively. Nonesterified fatty acids concentrations were assessed using a three-step enzymatic assay from FujiFilm Wako Chemicals, the Fujifilm HR Series NEFA-HR (2) Assay (Lexington, MA).

Reproductive Parameters

At the end of the 3-month study, samples were collected for sperm parameters (subchronic and neurobehavioral cohorts) and vaginal cytology evaluations (subchronic cohort only) on rats in the 0, 300, 1,000, and 3,000 ppm groups. The parameters evaluated are listed in Table 1. For 16 consecutive days before scheduled study termination, the vaginal vaults of the females were moistened with saline, if necessary, and samples of vaginal fluid and cells were collected and subsequently stained. Relative numbers of leukocytes, nucleated epithelial cells, and large squamous epithelial cells were determined and used to ascertain estrous cycle stage (i.e., diestrus, proestrus, estrus, and metestrus). Males were evaluated for sperm count (up to 10/exposure group in the subchronic cohort and 6/exposure group in the neurobehavioral cohort) and motility (5/exposure group in the subchronic cohort and 10/exposure group in the neurobehavioral cohort). The left testis and left epididymis were isolated and weighed. The tail of the epididymis (cauda epididymis) was then removed from the epididymal body (corpus epididymis) and weighed. The cauda was placed in a Petri dish containing M199 solution (maintained at approximately 34°C–38°C) with 1.0% bovine serum albumin (BSA) and one to three punctures were made in the distal region of the left cauda epididymis. A small sample of the diluted sperm was loaded into a 100 µm-chambered slide for Computer Assisted Sperm Analysis using TOX IVOS. After completion of sperm motility estimates, the remainder of the diluted sperm and cauda epididymis in M199/BSA solution and the left testis were stored frozen (approximately −70°C) until enumeration of sperm concentration was performed. The left cauda epididymis in M199/BSA solution was thawed and homogenized, and the left testis was thawed, homogenized, and sonicated in deionized water with 0.05% Triton X-100. Prepared samples were mixed with a DNA-specific fluorescent dye (IDENT) to allow for sperm identification under fluorescent illumination. A small amount of the IDENT sample was loaded into a 20-µm-chambered slide and analyzed using TOX IVOS.

Histopathology and Organ Weights

At the end of the 3-month study, F_1 rats in the neurobehavioral cohort selected for neuropathological evaluation (up to 15/sex/exposure group) were anesthetized with a 70% CO_2/30% O_2 mixture, and whole-body perfusion was performed via the left cardiac ventricle or aorta. Following a saline flush, animals were perfused using fresh 4% phosphate-buffered paraformaldehyde, followed by sequential perfusion with 4% phosphate-buffered medical grade glutaraldehyde (pH 7.4). Following perfusion and refrigeration at 2°C to 8°C for at least 24 hours, brains (including olfactory bulbs) were removed and the weight and length (without olfactory bulbs) and width were recorded. Brains were placed in 10% neutral buffered formalin (NBF) before trimming, and then were embedded in paraffin, put to slide, and stained with hematoxylin and eosin (H&E). Additional slides were produced and stained with Kluver-Barrera (brain and spinal cord only). Sciatic, vagus, tibial, and lateral sural nerves were collected bilaterally. Following perfusion fixation, left-side nerves were collected, placed in 10% NBF until trimmed for paraffin embedding, and archived without histopathological examination. Right-side nerves underwent postperfusion fixation via immersion first in 4% glutaraldehyde followed by osmium fixation. Right-side peripheral nerves were then embedded in epoxy resin (epon), sectioned at ≤1 µm, and stained with Toluidine Blue.[7] Histopathological examinations were performed by the study pathologist (a board-certified veterinary pathologist) on all animals in the 0 and 3,000 ppm groups. Table 1 lists the tissues and organs examined.

Necropsies were performed on all F_1 rats in the subchronic cohort. Organ weights were determined for the brain, liver, thymus, left and right kidney, left and right testis, left and right epididymis, left and right ovary, heart, and lungs. Tissues for microscopic examination were fixed and preserved in 10% NBF (except eyes, which were first fixed in Davidson's solution, and the right testis, including the vaginal tunic and epididymis, which were first fixed in modified Davidson's solution). Tissues were then processed and trimmed, embedded in paraffin, sectioned to a thickness of 4 to 6 μm, and stained with H&E, with the exception of the testis, which was stained with Periodic Acid-Schiff/hematoxylin. Complete histopathological examinations were performed by the study pathologist on rats in the subchronic cohort in the 0 and 3,000 ppm groups. Table 1 lists the tissues and organs examined.

Microscopic evaluations were completed by the study pathologist, and the pathology data were entered into the NTP Provantis software (Instem, Stone, UK). The report, slides, paraffin blocks, residual wet tissues, and pathology data were sent to the NTP Archives for inventory and storage.

The review of the histopathology slides by an independent quality assessment (QA) pathologist was limited to the liver, which was considered a potential target organ and was examined in all animals in all exposure groups from the subchronic cohort. The histotechnique was also evaluated. The lead DTT pathologist met with the QA pathologist and reviewed the selected liver tissues and addressed any inconsistencies in the diagnoses made by the study pathologist and the QA pathologist. Representative histopathological examples of liver lesions related to chemical administration, examples of disagreements in diagnoses between the study pathologist and the QA pathologist, or lesions of general interest were presented to two additional DTT pathologists for peer review. Final diagnoses for reviewed liver lesions represent a consensus among the QA pathologist, lead DTT pathologist, and the two additional DTT pathologists. The study pathologist and the QA pathologist evaluated the slides in an informed manner (with knowledge of the animal numbers and exposure groups the animals were from). The rationale for this is presented in Sills et al.[8]

Internal Concentration Assessment

Study samples analyzed included maternal plasma and fetuses (GD 18); maternal and pup plasma (LD 4); and F_1 rat urine (PND 117–119), plasma (PND 125–129), and brain (PND 125–129). NaF was added as a preservative to stabilize TPHP during sample collection for biological fluids (i.e., urine and plasma) or during fetus and brain tissue homogenization.

TPHP and its major metabolite, diphenyl phosphate (DPHP), were quantified in samples using validated LC-MS/MS methods.

TPHP and DPHP results in urine were reported as creatinine-normalized values (nanograms of TPHP or DPHP per milligram of urinary creatinine). Urinary creatinine was determined using a qualified colorimetric assay that employed the Jaffé reaction.

Table 1. Experimental Design and Materials and Methods in the Perinatal and Three-month Feed Study of Triphenyl Phosphate

Three-month Study
Testing Facility
Battelle/AmplifyBio[1] (West Jefferson, Ohio)
Strain and Species
Sprague Dawley (Hsd:Sprague Dawley® SD®)
Animal Source
Inotiv (Envigo at time of procurement, Indianapolis, IN)
Time Held Before Study
4 to 5 days
Average Age When Study Began
F_0 females: 13 to 14 weeks
Date of First Exposure
F_0 females: August 1–3, 2021 F_1 rats: September 13–17, 2021
Duration of Exposure
F_0 females: GD 6 through LD 28 F_1 rats (PND 28 interim cohort): Perinatal through PND 28 F_1 rats (neurobehavioral cohort): Perinatal through PND 138–144 F_1 rats (subchronic cohort): Perinatal through PND 122–128 F_1 rats (biological sampling cohort): Perinatal through PND 125–129
Date of Last Exposure and Necropsy
F_0 females: September 13–17, 2021 F_1 rats (PND 28 interim cohort): September 13–17, 2021 F_1 rats (neurobehavioral cohort): January 4–7, 2022 F_1 rats (subchronic cohort): December 20–23, 2021 F_1 rats (biological sampling cohort): December 22–23, 2021
Average Age at Necropsy
F_0 females: 19 to 21 weeks F_1 rats (PND 28 interim cohort): 4 weeks F_1 rats (neurobehavioral cohort): 20 to 21 weeks F_1 rats (subchronic cohort): 18 weeks F_1 rats (biological sampling cohort): 18 weeks

[1]This study was conducted under the NIEHS contract with Battelle Memorial Institute (HHSN273201400015C). The study was initiated at the Battelle testing facility in West Jefferson, Ohio. On May 1, 2021, the testing facility was transferred to a new company, AmplifyBio. The study was transferred to AmplifyBio as a subcontract under HHS273201400015C.

Three-month Study

Size of Study Groups

F_0 females (core): 36/exposure group
F_0 females (biological sampling): 3/exposure group (GD 18) and 3/exposure group (LD 4)
F_1 rats (PND 28 interim cohort): 34–40/sex/exposure group
F_1 rats (neurobehavioral cohort): 19/sex (0 ppm) or 20/sex (30, 100, 300, 1,000, and 3,000 ppm)
F_1 rats (subchronic cohort): 10/sex/exposure group
F_1 rats (biological sampling cohort): 5/sex/exposure group

Method of Distribution

Animals were distributed randomly into groups of approximately equal initial mean body weights.

Animals per Cage

F_0 females: 1 (with litter)
F_1 rats: ≤2 (males) or ≤4 (females)

Method of Animal Identification

F_0 females: Cage card and indelible ink tail marking
F_1 rats: Appendage tattoo until PND 28, then cage card and tail tattoo

Diet

F_0 females: Irradiated NIH-07 meal feed (Zeigler Brothers, Gardners, PA), available ad libitum, changed at least weekly

F_1 rats (neurobehavioral cohort): Irradiated NIH-07 meal feed (Zeigler Brothers, Gardners, PA), available ad libitum, changed at least once weekly (individually housed animals) or twice weekly (group-housed animals)

F_1 rats (subchronic and biological sampling cohorts): Irradiated NTP-2000 meal feed (Zeigler Brothers, Gardners, PA), starting on PND 28, available ad libitum, changed at least once weekly (individually housed animals) or twice weekly (group-housed animals)

Water

Tap water (West Jefferson, OH municipal supply) via automatic watering system (Edstrom Industries, Waterford, WI), available ad libitum

Cages

Solid polycarbonate (Lab Products, Inc., Seaford, DE) changed and sanitized at least weekly for dams through LD 4, and twice weekly for post-LD 4 dams with litters and F_1 rats

Bedding

Irradiated Sani-Chips® (P.J. Murphy Forest Products Corporation, Montville, NJ), changed with cage changes

Environmental Enrichment

F_0 females and F_1 pups: Natural crinkled kraft paper (Crink-l'Nest™, The Andersons, Inc., Maumee, OH), provided during gestation until GD 19 and during lactation starting on LD 4

F_1 rats: Red rectangular polycarbonate shelters (Rat Retreats™, Bio-Serv, Flemington, NJ)

Racks

Stainless steel (Lab Products, Inc., Seaford, DE), changed and sanitized every 2 weeks

Rack Filters

Spun-bonded polyester (National Filter Media Corporation, Olive Branch, MS), changed every 2 weeks

Three-month Study

Animal Room Environment

Animals were housed across three rooms with environmental parameters within the following ranges:
Temperature: 69°F–75°F
Relative humidity: 20%–75%
Room fluorescent light: 12 hours/day
Room air changes: at least 10/hour

Exposure Concentrations

0, 30, 100, 300, 1,000, or 3,000 ppm in feed

Type and Frequency of Observation

F_0 females: Observed twice daily. Weighed upon receipt on GDs 1 or 2, GDs 5, 6, 9, 12, 15, 18, and 21, and LDs 1, 4, 7, 10, 13, 16, 19, 21, 25, and 28. Clinical observations and survival were recorded daily from GD 6 through LD 28. Feed consumption was measured and recorded on GDs 6, 9, 12, 15, 18, and 21, and LDs 1, 4, 7, 10, 13, 16, 19, 21, 25, and 28.

F_1 rats: Observed twice daily. The number and sex of pups for each litter were recorded daily from PND 0 through PND 28. Whole litter weights by sex were recorded on PND 1. Pups were individually weighed on PNDs 1, 4, 7, 10, 13, 16, 19, 21, 25, and 28, then weekly thereafter and at study termination. Clinical observations were recorded daily from PND 1 through PND 28, and weekly thereafter. Feed consumption was measured and recorded weekly.

Endocrine and pubertal endpoints: AGD on PND 1; areolae/nipple retention on PND 13; day of BPS and body weight at attainment, evaluated from PND 35 through attainment (subchronic and biological sampling cohorts); day of VO and body weight at attainment, evaluated from PND 25 through attainment (subchronic and biological sampling cohorts).

Method of Euthanasia

F_0 females (core): Carbon dioxide
F_0 female rats (biological sampling): Decapitation
F_1 fetuses (GD 18) and pups (PND 0–10): Decapitation
F_1 rats (PND 28 interim evaluation): Carbon dioxide
F_1 rats (animals selected for neuropathology from the neurobehavioral and biological sampling cohorts): Anesthetization with a carbon dioxide/oxygen mixture followed by perfusion until exsanguination
F_1 rats (all other adults): Carbon dioxide

Necropsy

F_0 females: Gross necropsies were performed on all animals. All dams were examined for evidence of implantation and the number of implantations sites recorded.

F_1 rats (subchronic cohort): Necropsies were performed on all animals. Organs weighed at study termination were: brain, liver, thymus, left and right kidney, left and right testis, left and right epididymis, left and right ovary, heart, and lungs.

Clinical Pathology

On PND 28 or at study termination, blood was collected from the retroorbital plexus (PND 28 interim and subchronic cohorts) or jugular vein (neurobehavioral cohort) for clinical chemistry (PND 28 interim, neurobehavioral, and subchronic cohorts), hematology (subchronic cohort), and special chemistries (neurobehavioral cohort) analyses.

Hematology (subchronic cohort): erythrocyte count, hemoglobin, hematocrit, manual hematocrit, mean cell volume, mean cell hemoglobin, mean cell hemoglobin concentration, white blood cell count and differential, reticulocyte count, platelet count.

Three-month Study

Clinical chemistry (PND 28 interim rats and subchronic cohort): total protein, albumin, globulin, A/G ratio, urea nitrogen, creatinine, alanine aminotransferase, aspartate aminotransferase, sorbitol dehydrogenase, alkaline phosphatase, total bile acids, total bilirubin, direct bilirubin, glucose, creatine kinase, cholesterol, triglycerides, lactate dehydrogenase, sodium, potassium, and chloride. Normal baseline total and direct bilirubin values were below the linear range of the assay and thus were not reported.

Clinical chemistry (neurobehavioral cohort): glucose, cholesterol, triglycerides, HDL cholesterol, and LDL cholesterol.

Special Chemistries (neurobehavioral cohort): insulin, C-peptide, leptin, and nonesterified fatty acids.

Neurobehavioral Assessment

F_1 rats (neurobehavioral cohort): Neurobehavioral assessments were performed for 19 to 20 rats/sex/exposure group as juveniles between PND 29 and PND 36 (all exposure groups) and as adults starting on approximately PND 60 (0, 30, 1,000, and 3,000 ppm groups). Animals completed all behavioral tests. Animals were given a ~15- to 30-minute room acclimation period for motor activity, sensorimotor function, and Morris water maze assessments.

Motor activity: Spontaneous locomotor activity was assessed between PND 29 and PND 35 and again between PND 57 and PND 63 in an open field test.

Sensorimotor function: Spontaneous startle assessed sensorimotor function between PND 30 and PND 36 and again between PND 58 and PND 64.

Learning and memory: Spatial learning and memory was assessed with the Morris water maze over 14 days starting on ~PND 68.

Histopathology

F_1 rats (neurobehavioral cohort): Neuropathology and histopathology was performed in selected animals ($n \leq 15$/sex/exposure group) in the 0 and 3,000 ppm groups. In addition to gross lesions, the following tissues were examined: brain (seven sections including: olfactory bulbs, frontal-parietal cortex including basal ganglia, midparietal cortex and thalamus, midbrain with substantia nigra and red nucleus, posterior colliculi, midcerebellum including cranial nerve VIII, and posterior medulla), colon, muscle (thigh and gastrocnemius, medial head), nerves (dorsal root ganglion, lateral sural, sciatic, tibial, trigeminal with ganglion, vagus), spinal cord (three sections including: cervical, thoracic, and lumbar), and urinary bladder. Evaluation of the autonomic ganglia was included during standard histopathological evaluation of preexisting sections of the colon and urinary bladder to assess the autonomic nervous system.[9]

F_1 rats (subchronic cohort): Complete histopathology was performed on all rats in the 0 and 3,000 ppm groups. In addition to gross lesions and tissue masses, the following tissues were examined: adrenal glands, brain (seven sections including: olfactory bulbs, frontal-parietal cortex including basal ganglia, midparietal cortex and thalamus, midbrain with substantia nigra and red nucleus, posterior colliculi, midcerebellum including cranial nerve VIII, and posterior medulla), cervix, clitoral glands, esophagus, eyes, femur, Harderian glands, heart and aorta, kidneys, large intestine (cecum, colon, rectum), small intestine (duodenum, jejunum, ileum), larynx, liver (two sections including left lateral lobe and median lobe), lung, lymph nodes (mandibular and mesenteric), mammary gland, muscle (gastrocnemius and thigh), nose (nasal cavity and nasal turbinates), nerves (dorsal root ganglion, lateral sural, sciatic, tibial, trigeminal with ganglion, vagus), ovaries, pancreas, parathyroid glands, pituitary gland, preputial glands, prostate, salivary glands, seminal vesicles, skin (dorsal), spinal cord (three sections including: cervical, thoracic, and lumbar), spleen, stomach (forestomach and glandular), testis with epididymis (right), thymus, thyroid gland, trachea, urinary bladder, uterus, and vagina.

Three-month Study

Sperm Parameters and Vaginal Cytology

At study termination, sperm samples were collected from male rats in the in the subchronic and neurobehavioral cohorts in the 0, 300, 1,000, and 3,000 ppm groups for sperm count and motility evaluations. The following parameters were evaluated: spermatid heads per testis and per gram testis, and epididymal spermatozoal motility and concentration. The left cauda, left epididymis, and left testis were weighed. Vaginal samples were collected for up to 16 consecutive days before study termination from female rats in the subchronic cohort in the 0, 300, 1,000, and 3,000 ppm groups for vaginal cytology evaluations. The following parameters were evaluated: cycle length, number of cycles, and time spent in each specific stage of the estrous cycle.

Internal Concentration Assessment

F_0 females and F_1 pups: Dam plasma, amniotic fluid (pooled by litter), and fetuses (pooled by litter) (n = 3 dams/exposure group) were collected at GD 18, but amniotic fluid was excluded from the analysis. On LD 4, plasma was collected from dams and their respective pups (n ≤ 3 dams/exposure group for the 30 and 100 ppm groups). On the morning of LD 4, plasma was collected from dams in the 0, 300, 1,000, and 3,000 ppm groups via a survival blood collection from the jugular vein, and PND 4 pup plasma samples were collected from the culled pups of these dams and other selected litters from these groups.

F_1 rats (biological sampling cohort): Urine samples were collected from select animals (n = 5 rats/sex/exposure group) 1 week prior to the end of the 3-month study over a 24-hour period. At termination of the 3-month study, all animals were anesthetized, and blood was collected from the vena cava or abdominal aorta.
All samples were analyzed to determine TPHP concentration using a validated analytical method published previously (under peer review).

Acetylcholinesterase and Butyrylcholinesterase Activity

F_1 rats (PND 28 interim cohort): Whole brain samples were collected from selected animals (n = 5/sex/exposure group) at PND 28 for cholinesterase activity analyses.

F_1 rats (neurobehavioral cohort): At study termination, blood was collected from the retroorbital plexus of selected animals (n = 5/sex/exposure group) for cholinesterase activity analyses.

Genetic Toxicology

Bacterial mutagenicity test: Samples of TPHP were incubated with *Salmonella typhimurium* tester strains, and histidine-independent mutant colonies arising on these plates were counted after incubation for 2 days.

Peripheral blood micronucleus assay (neurobehavioral and subchronic cohorts): At study termination, blood was collected from the retroorbital plexus of rats for the peripheral blood micronucleus assay.

GD = gestation day; LD = lactation day; PND = postnatal day; AGD = anogenital distance; BPS = balanopreputial separation; VO = vaginal opening; A/G = albumin/globulin; HDL = high-density lipoprotein; LDL = low-density lipoprotein; TPHP = triphenyl phosphate.

Statistical Methods

Statistical methods were chosen with consideration of distributional assumptions and the need to incorporate within-litter correlation among animals (for the perinatal rat studies). Unless specifically mentioned, all endpoints were tested for a trend across exposure groups, followed by pairwise tests for each exposed group against the control group. Significance of all trend and pairwise tests is determined by a p value of ≤0.05 and is reported at both the 0.05 and 0.01 levels.

Calculation and Analysis of Nonneoplastic Lesion Incidences

The incidences of nonneoplastic lesions are presented as numbers of animals bearing such lesions at a specific anatomic site and the numbers of animals with that site examined microscopically. Fisher's exact test,[10] a procedure that uses the overall proportion of affected animals, was used to determine statistical significance between exposed and vehicle control animals, and the Cochran-Armitage trend test was used to test for significant trends.[11] P values for these analyses are one-sided.

Analysis of Continuous Variables

Before statistical analysis, outliers identified using the Dixon and Massey test[12] for small samples ($n < 20$) and Tukey's[13] outer fences method for large samples ($n \geq 20$) were examined by DTT personnel, and biologically implausible values (that were likely due to experimental error) were eliminated from the analysis.

In most instances, no considerations for litter effects were necessary in the analysis of the continuous data, which was the case for all of the F_0 rat data and the F_1 rat data in which the cohort consisted of only one animal per litter. In these cases, two approaches were employed to assess the significance of pairwise comparisons between dosed and control groups in the analysis of continuous variables. Organ weight, body weight, and brain dimension data, which historically have approximately normal distributions, were analyzed with the parametric multiple comparison procedures of Dunnett[14] and Williams.[15; 16] Gestational length, implantations, litter sizes, postimplantation loss, pup survival, number of areolae/nipples per pup, hematology, cholinesterase, special chemistries, clinical chemistry, internal concentration data, feed consumption, spermatid, and epididymal spermatozoal data, which have typically skewed distributions, were analyzed using the nonparametric multiple comparison methods of Shirley[17] (as modified by Williams[18]) and Dunn.[19] The Jonckheere test[20] was used to assess the significance of the dose-related trends and to determine at the 0.01 level of significance whether a trend-sensitive test (the Williams or Shirley test) was more appropriate for pairwise comparisons than a test that does not assume a monotonic dose-related trend (the Dunnett or Dunn test). P values for these analyses are two-sided. For plasma concentration data, when individual concentration values were provided as "below limit of detection," one-half the LOD was used as a substitute value. However, if 80% or more of the values in the control group were below the LOD, the mean was reported as "BD" to indicate the values were "below detection" and no statistical analysis was performed on the endpoint.

When litter effects were present, body weight endpoints were analyzed using linear mixed models, with litters as a random effect. To adjust for multiple comparisons, a Dunnett-Hsu adjustment was used.[21] Pup weights were adjusted for litter size, as described in the Body

Weight Adjustments section below. AGD was adjusted for the body weight of the pup taken on the day of AGD measurement. The adjusted AGDs were analyzed as normal variates with litter effects using a linear mixed model.

For the F_1 PND 28 interim cohort clinical chemistry data, litter effects were present for non-normally distributed endpoints, so the trend across exposure groups was analyzed using a permutation test that was based on the Jonckheere trend test implemented by randomly permuting whole litters across exposure groups and bootstrapping within the litters (see, for example, Davison[22]). Pairwise comparisons were made using a modified Wilcoxon test that incorporated litter effects.[23] The Hommel procedure was used to adjust for multiple comparisons.[24]

Analysis of Gestational and Fertility Indices

Cochran-Armitage trend tests were used to test the significance of trends in gestational and fertility indices across exposure groups. Fisher's exact test was used to conduct pairwise comparisons of each exposed group with the control group. P values for these analyses are two-sided.

Body Weight Adjustments

Preweaning pup body weights were adjusted for live litter size as follows: A linear model was fit to body weights as a function of exposure and litter size. The estimated coefficient of litter size was then applied to adjust each pup body weight based on the difference between its litter size and the mean litter size. Prestandardization PND 4 body weights were adjusted for PND 1 litter size, and body weights measured between PND 4 poststandardization and PND 28 were adjusted for PND 4 poststandardization litter size. After adjustment, mean body weights were analyzed with a linear mixed model with a random litter effect.

Areola/Nipple Data

Cochran-Armitage trend tests were used to test the significance of trends for number of litters with areolae/nipples. Fisher's exact test was used to conduct pairwise comparisons of each exposed group with the control group. P values for these analyses are two-sided.

The tendency of pups from the same litter to respond more similarly than pups from different litters has been referred to as the "litter effect"[25] and reflects littermates' similarities in genetics and in utero experiences. Failure to account for correlation within litters leads to underestimates of variance in statistical tests, resulting in higher probabilities of Type I errors ("false positives"). Therefore, the number of pups with areola/nipples uses a modified Cochran-Armitage test to accommodate litter effects using the Rao-Scott approach.[26] The Rao-Scott approach accounts for litter effects by estimating the ratio of the variance in the presence of litter effects to the variance in the absence of litter effects. This ratio is then used to adjust the sample size downward to yield the estimated variance in the presence of litter effects. The Rao-Scott approach was implemented in the Cochran-Armitage test as recommended by Fung et al.,[27] formula $\overline{T}_{RS2}$.

The continuous endpoint, number of areolae/nipples per pup, was analyzed using the Jonckheere (trend) and Shirley or Dunn (pairwise) tests, as described earlier.

Analysis of Time-to-event Data

Time-to-event endpoints, such as day of attainment of BPS and VO, have several features that require careful model selection: non-normality of distributions and censored values when attainment was not observed before the end of the observation period. Further, growth retardation, reflected in the weaning weight, is an important covariate in the case of BPS and VO given the relationship between the normal day of expected attainment and body weight.

When attainment times were approximately normally distributed and attainment was observed for all or most animals, two approaches for modeling discrete developmental endpoints were taken. First, a linear model was fit to attainment day as a function of exposure concentration. A second linear model was fit to attainment day as a function of exposure concentration and weaning weight. Dunnett-Hsu adjustments were used to account for multiple comparisons.[21]

If censored observations were observed, survival analysis methods were used. In this case, a Cox proportional hazards model[28] was fit with exposure concentration and weaning weight as covariates and a Hommel adjustment for multiple comparisons.[24]

To calculate mean attainment values adjusted for weaning weight, a linear model was fit to attainment day as a function of exposure concentration and weaning weight. The estimated coefficient of weaning weight was then applied to adjust each attainment day based on the difference between the measured weaning weight and the mean weaning weight.

Analysis of Vaginal Cytology Data

Vaginal cytology data consist of daily observations of estrous cycle stages over a 16-day period. Differences from the control group for cycle length and the number of cycles were analyzed using the Shirley and Dunn tests, as described above.

To identify disruptions in estrous cyclicity, a continuous-time Markov chain model (multistate model) was fit using a maximum likelihood approach,[29] producing estimates of stage lengths for each exposure group. Confidence intervals for these estimates were obtained from bootstrap sampling of the individual animal cycle sequences. Stage lengths that were significantly different from the control group were identified using permutation testing with a Hommel adjustment for multiple comparisons.[24] When the number of observations for a particular stage was low, the corresponding stage length was estimated using a profile likelihood approach, and no confidence intervals were calculated.

Analysis of Neurobehavioral Data

Spontaneous motor activity endpoints (basic movements, rearing, and calculated total activity) captured in 5-minute epochs were analyzed separately for each sex using repeated measures analysis of variance (RMANOVA) with exposure concentration and epoch as factors and autoregressive (AR(1)) error structure reflecting temporal correlation. Rearing counts were analyzed using a Poisson mixed model with AR(1) error structure. Activity measures summed over the first 30 minutes (juvenile) or 45 minutes (adults) of the session, representing total session activity, were analyzed for each sex independently by one-way ANOVA with exposure concentration as the factor, followed by post hoc Dunnett's multiple-comparisons tests.

For acoustic startle response, the 120 dB amplitude responses for the habituation phase (trials 2-20 and the post-habituation phase (trials 21-50) were analyzed separately for each sex using RMANOVA with exposure concentration and trial as factors and autoregressive (AR(1)) error structure reflecting temporal correlation to demonstrate startle level and habituation. The 120 dB response amplitude on the initial trial 1 was analyzed by a Kruskal-Wallis test.[30] If significant, pairwise comparisons to the vehicle control (0 ppm) were conducted using post-hoc Dunn's multiple comparisons test.

For MWM-cued (visual platform), spatial acquisition (hidden platform), and reversal (hidden platform) learning, log-transformed median daily latencies and distance traveled were calculated over the four trials conducted during each daily session and analyzed across days using RMANOVA with an AR(1) error structure. For the probe test of reference memory, distance traveled and time spent in the goal quadrant were analyzed for 90 seconds as well as for the three 30-second epochs using Kruskal-Wallis tests followed by post hoc Dunn's multiple-comparisons tests. Swim speed over all quadrants was analyzed over the 90 seconds using Kruskal-Wallis tests followed by post hoc Dunn's multiple-comparison tests.

For all RMANOVA models, as well as for the Poisson mixed model for rearing, model selection proceeded as follows. An interaction model for each endpoint was fit with exposure group, time factor (5-minute epoch for motor activity, trial for startle response, and day for MWM), and their interaction. If the interaction term was significant, relevant pairwise comparisons of each exposed group to the control group were assessed at each level of the time factor. If the interaction was not significant, an additive model was fit with exposure group and time factor, followed by pairwise comparisons of exposed groups to the control group. For all RMANOVA models, Dunnett-Hsu adjustments were used to account for multiple comparisons, and normality and homogeneity of variance assumptions were evaluated using visual examination of model residual plots and the Shapiro-Wilk[31] test. The visible and hidden platform and reversal learning MWM endpoints were log-transformed to ensure normality and variance homogeneity.

Quality Assurance Methods

The 3-month study was conducted in compliance with U.S. Food and Drug Administration Good Laboratory Practice Regulations.[32] In addition, a draft of this document was audited retrospectively by an independent QA contractor against study records submitted to the NTP Archives. Separate audits covered the completeness and accuracy of the peer-reviewed pathology data. Audit procedures and findings are presented in the reports and are on file at NIEHS. The audit findings were reviewed and assessed by NIEHS/DTT staff, and all comments were resolved or otherwise addressed during the preparation of this document.

Genetic Toxicology

The genetic toxicity of TPHP was assessed by testing for mutagenic activity in the bacterial reverse mutation assay (also known as the Ames assay) and for induction of chromosome damage in a rat peripheral blood micronucleus assay.

Bacterial Mutagenicity Test Protocol

Testing procedures were modified from those originally reported by Zeiger et al.[33] Coded samples of TPHP were incubated with *Salmonella typhimurium* tester strains (TA98, TA100, TA1535, TA1537) either in buffer or S9 mix (metabolic activation enzymes and cofactors from Aroclor 1254-induced male Sprague Dawley rat liver or similarly induced male Syrian golden hamster liver S9 mix) for 20 minutes at 37°C. Top agar supplemented with *L*-histidine and *d*-biotin was added, and the contents of the tubes were mixed and poured onto the surfaces of minimal glucose agar plates. Histidine-independent mutant colonies arising on these plates were counted after incubation for 2 days at 37°C. Each trial consisted of triplicate plates of concurrent positive and negative controls and five doses of TPHP. TPHP was tested up to a dose of 10,000 μg/plate in each tester strain, greater than the current Organisation for Economic Co-operation and Development test guideline assay limit dose of 5,000 μg/plate. All trials were repeated.

In this assay, a positive response is defined as a reproducible, dose-related increase in histidine-independent (revertant) colonies in any one strain/activation combination. An equivocal response is defined as an increase in revertants that is not dose-related, is not reproducible, or is not of sufficient magnitude to support a determination of mutagenicity. A negative response is obtained when no increase in revertant colonies is observed after chemical treatment. No minimum percentage or fold increase is required for a chemical to be judged positive or weakly positive, although positive calls are typically reserved for increases in mutant colonies that are at least twofold over background.

Peripheral Blood Micronucleus Test Protocol

At the end of the 3-month study, blood samples (approximately 200 μL) were collected from male and female rats, placed in K_3 EDTA-coated tubes, and shipped overnight to the testing laboratory. Upon arrival, blood samples were fixed in ultracold methanol using a MicroFlowPLUS Kit (Litron Laboratories, Rochester, NY) according to the manufacturer's instructions. Fixed samples were stored in a −80°C freezer until analysis. Thawed blood samples were analyzed for the frequency of micronucleated immature erythrocytes (i.e., reticulocytes or polychromatic erythrocytes [PCEs]) and mature erythrocytes (i.e., normochromatic erythrocytes [NCEs]) using a flow cytometer[34]; both the mature and immature erythrocyte populations could be analyzed separately by employing special cell surface markers to differentiate the two cell types. Approximately 20,000 reticulocytes and 1×10^6 mature erythrocytes were analyzed per animal for frequency of micronucleated cells, and the percentage of reticulocytes (% RET) was calculated as a measure of bone marrow toxicity resulting from chemical exposure.

For evaluation of in vivo micronucleus data, the non-parametric statistical tests selected for trend and for pairwise comparisons with the control group do not make any assumptions about the underlying distribution of measurements and do not require equal variances among the groups. The Jonckheere test was used to test for trend, and the Dunn test was used for pairwise comparisons of each exposed group with the control group. To correct for multiple pairwise comparisons, the p value for each comparison with the control group was multiplied by the number of comparisons made. If the product was greater than 1.00, it was replaced with 1.00.

In the micronucleus test, it is preferable to base a positive result on the presence of both a significant trend test and at least one significantly elevated exposed group compared with the

corresponding control group. If two or more exposed groups are significantly elevated over the corresponding control group in the absence of a significant trend test, this result may also be considered positive. A response is considered equivocal if only the trend test is significant or if only a single exposed group is significantly increased over the control group. The absence of both a significant trend test and any significant differences between exposed groups and the control group results in a negative call. Because two calls can be made for a nonnegative effect, positive and equivocal, the original threshold values for each test (trend and pairwise) are halved to ensure that the false positive rate for the equivocal call is less than 0.05. Therefore, to maintain the overall significance level at 0.05 for positive and equivocal calls, the trend and pairwise differences are considered statistically significant if the one-sided p value is ≤ 0.025 (0.05/2).

In addition, historical control data are used to evaluate the biological significance of any observed response. Both statistical significance and biological significance are considered when arriving at a call. Ultimately, the scientific staff determines the final call after considering the results of statistical analyses, reproducibility of any effects observed (in acute studies), and the magnitudes of those effects.

References

1. Witchey SK, Sutherland V, Collins B, Roberts G, Shockley KR, Vallant M, Krause J, Cunny H, Waidyanatha S, Mylchreest E, et al. Reproductive and developmental toxicity following exposure to organophosphate ester flame retardants and plasticizers, triphenyl phosphate and isopropylated phenyl phosphate, in Sprague Dawley rats. Toxicol Sci. 2023; 191(2):374-386. https://doi.org/10.1093/toxsci/kfac135

2. National Institute of Environmental Health Sciences (NIEHS). Materials and methods and study data: perinatal and three-month toxicity study of triphenyl phosphate (CASRN 115-86-6) administered in feed to Sprague Dawley (Hsd:Sprague Dawley(R) SD(R)) rats. Research Triangle Park, NC: NIEHS Division of Translational Toxicology; 2026. https://doi.org/10.22427/DATA-500-104-003-000-3

3. National Institute of Advanced Industrial Science and Technology (AIST). Spectral Database for Organic Compounds: SDBS-2915: triphenyl phosphate. Tokyo, Japan: National Institute of Advanced Industrial Science and Technology; 1999. [Accessed: March 4, 2026]. https://sdbs.db.aist.go.jp/CompoundLanding.aspx?sdbsno=2915

4. National Institute of Standards and Technology (NIST). NIST chemistry webbook, SRD 69: triphenyl phosphate. Gaithersburg, MD: U.S. Department of Commerce, National Institute of Standards and Technology; 2026. [Accessed: March 4, 2026]. https://webbook.nist.gov/cgi/cbook.cgi?ID=C115866&Mask=400

5. Pouchert CJ, Behnke J, Aldrich Chemical Company. Spectrum B, triphenyl phosphate. In: The Aldrich Library of ^{13}C and ^{1}H FT NMR Spectra, Volume 2. Milwaukee, WI: Aldrich Chemical Company; 1993. p. 1694.

6. O'Neil MJ, Heckelman PE, Koch CB, Roman KJ. Monograph 9742: triphenyl phosphate. In: The Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals. 14 ed. Whitehouse Station, NJ: Merck & Co.; 2006. p. 1673.

7. Bolon B, Krinke G, Butt MT, Rao DB, Pardo ID, Jortner BS, Garman RH, Jensen K, Andrews-Jones L, Morrison JP, et al. STP position paper: Recommended best practices for sampling, processing, and analysis of the peripheral nervous system (nerves and somatic and autonomic ganglia) during nonclinical toxicity studies. Toxicol Pathol. 2018; 46(4):372-402. https://doi.org/10.1177/0192623318772484

8. Sills RC, Cesta MF, Willson CJ, Brix AE, Berridge BR. National Toxicology Program position statement on informed ("nonblinded") analysis in toxicologic pathology evaluation. Toxicol Pathol. 2019; 47(7):887-890. https://doi.org/10.1177/0192623319873974

9. Rao DB, Little PB, Malarkey DE, Herbert RA, Sills RC. Histopathological evaluation of the nervous system in National Toxicology Program rodent studies: a modified approach. Toxicol Pathol. 2011; 39(3):463-470. https://doi.org/10.1177/0192623311401044

10. Gart JJ, Chu KC, Tarone RE. Statistical issues in interpretation of chronic bioassay tests for carcinogenicity. J Natl Cancer Inst. 1979; 62(4):957-974. https://doi.org/10.1093/jnci/62.4.957

11. Armitage P. Statistical methods in medical research. Oxford, UK: Blackwell Scientific; 1971.

12. Dixon WJ, Massey FJ. Introduction to statistical analysis. 2nd ed. New York, NY: McGraw-Hill; 1957.

13. Tukey JW. Easy summaries--numerical and graphical. In: Exploratory Data Analysis. Reading, MA: Addison-Wesley; 1977. p. 27-56.

14. Dunnett CW. A multiple comparison procedure for comparing several treatments with a control. J Am Stat Assoc. 1955; 50(272):1096-1121. https://doi.org/10.1080/01621459.1955.10501294

15. Williams DA. A test for differences between treatment means when several dose levels are compared with a zero dose control. Biometrics. 1971; 27(1):103-117. https://doi.org/10.2307/2528930

16. Williams DA. The comparison of several dose levels with a zero dose control. Biometrics. 1972; 28(2):519-531. https://doi.org/10.2307/2556164

17. Shirley E. A non-parametric equivalent of Williams' test for contrasting increasing dose levels of a treatment. Biometrics. 1977; 33(2):386-389. https://doi.org/10.2307/2529789

18. Williams DA. A note on Shirley's nonparametric test for comparing several dose levels with a zero-dose control. Biometrics. 1986; 42(1):183-186. https://doi.org/10.2307/2531254

19. Dunn OJ. Multiple comparisons using rank sums. Technometrics. 1964; 6(3):241-252. https://doi.org/10.1080/00401706.1964.10490181

20. Jonckheere AR. A distribution-free k-sample test against ordered alternatives. Biometrika. 1954; 41(1-2):133-145. https://doi.org/10.1093/biomet/41.1-2.133

21. Hsu JC. The factor analytic approach to simultaneous inference in the general linear model. J Comput Graph Stat. 1992; 1(2):151-168. https://doi.org/10.1080/10618600.1992.10477011

22. Davison AC, Hinkley DV. Bootstrap methods and their application. Cambridge, UK: Cambridge University Press; 1997.

23. Datta S, Satten GA. Rank-sum tests for clustered data. J Am Stat Assoc. 2005; 100(471):908-915. https://doi.org/10.1198/016214504000001583

24. Hommel G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika. 1988; 75(2):383-386. https://doi.org/10.1093/biomet/75.2.383

25. Kupper LL, Portier C, Hogan MD, Yamamoto E. The impact of litter effects on dose-response modeling in teratology. Biometrics. 1986; 42(1):85-98. https://doi.org/10.2307/2531245

26. Rao JNK, Scott AJ. A simple method for the analysis of clustered binary data. Biometrics. 1992; 48(2):577-585. https://doi.org/10.2307/2532311

27. Fung KY, Krewski D, Rao JNK, Scott AJ. Tests for trend in developmental toxicity experiments with correlated binary data. Risk Anal. 1994; 14(4):639-648. https://doi.org/10.1111/j.1539-6924.1994.tb00277.x

28. Cox DR. Regression models and life-tables. J R Stat Soc Series B Stat Methodol. 1972; 34(2):187-202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

29. Kalbfleisch JD, Lawless JF. The analysis of panel data under a Markov assumption. J Am Stat Assoc. 1985; 80(392):863-871. https://doi.org/10.1080/01621459.1985.10478195

30. Kruskal W, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc. 1952; 47(260):583-621. 10.1080/01621459.1952.10483441

31. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965; 52(3-4):591-611. https://doi.org/10.1093/biomet/52.3-4.591

32. U.S. Food and Drug Administration (FDA). 21 CFR Part 58. https://www.ecfr.gov/current/title-21/chapter-I/subchapter-A/part-58

33. Zeiger E, Anderson B, Haworth S, Lawlor T, Mortelmans K. Salmonella mutagenicity tests: V. Results from the testing of 311 chemicals. Environ Mol Mutagen. 1992; 19 Suppl 21:2-141. https://doi.org/10.1002/em.2850190603

34. Witt KL, Livanos E, Kissling GE, Torous DK, Caspary W, Tice RR, Recio L. Comparison of flow cytometry- and microscopy-based methods for measuring micronucleated reticulocyte frequencies in rodents treated with nongenotoxic and genotoxic chemicals. Mutat Res. 2008; 649(1-2):101-113. https://doi.org/10.1016/j.mrgentox.2007.08.004