



NTP

National Toxicology Program

U.S. Department of Health and Human Services

NTP RESEARCH REPORT ON
BASELINE CHARACTERISTICS
OF DIVERSITY OUTBRED
(J:DO) MICE RELEVANT TO
TOXICOLOGY STUDIES

NTP RR 6

JUNE 2018

NTP Research Report on Baseline Characteristics of Diversity Outbred (J:DO) Mice Relevant to Toxicology Studies

Research Report 6
National Toxicology Program

June 2018

National Toxicology Program
Public Health Service
U.S. Department of Health and Human Services
ISSN: 2473-4756

Table of Contents

About This Report.....	iii
Peer Review	v
Publication Details	vi
Abstract.....	vii
Introduction.....	1
Methods.....	3
Animals and Treatment	3
In-Life Resting-State Glucose.....	3
Terminal Blood and Tissue Collection	4
Histopathology	4
Clinical Chemistry	5
Statistical Analyses	5
Power Calculations	5
Results.....	7
Body Weight and Body Weight Gain	7
Organ Weights	7
Epididymal Sperm Evaluation	7
Clinical Chemistry	7
Insulin and Leptin	8
Histopathology	8
Power Calculations	8
Discussion.....	10
References.....	14

Tables

Table 1. Terminal Body and Organ Weights in Male DO Mice.....	18
Table 2. Sperm Counts and Motility Measurements in Male DO Mice	19
Table 3. Reference Ranges for Clinical Chemistry Parameters.....	20
Table 4. Serum Insulin and Leptin at Terminal Sacrifice in Male DO Mice.....	21
Table 5. Histopathology Findings in the Liver and Kidney of Male DO Mice	22
Table 6. Available Comparator Data and Power Calculations for Common Endpoints in Mice from Five NTP Studies.....	23
Table 7. Power Calculations for Organ and Body Weights in Mice	24
Table 8. Power Calculations for Clinical Chemistry Endpoints in Rats.....	25

About This Report

Authors

Alison H. Harrill¹, Susan Borghoff², Leah Zorrilla³, Chad Blystone¹, Grace E. Kissling⁴, Dave Malarkey¹, Keith Shockley⁴, Greg Travlos¹, Michael J. DeVito¹

¹Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

²Integrated Laboratory Systems Inc., Research Triangle Park, North Carolina, USA/ToxStrategies, Research Triangle Park, North Carolina, USA

³Integrated Laboratory Systems Inc., Research Triangle Park, North Carolina, USA/Bayer Crop-Sciences, Research Triangle Park, North Carolina, USA

⁴Division of Intramural Research, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

Clinical chemistry data analysis, manuscript writing

Alison H. Harrill, Ph.D.

Oversight coordination of animal studies, study design and interpretation

Chad Blystone, Ph.D., DABT

Histopathological review

Dave Malarkey, D.V.M., Ph.D.

Greg Travlos, D.V.M., DACVP

Study design, oversight for study conduct and data analysis, manuscript writing

Michael J. DeVito, Ph.D.

Integrated Laboratory Systems Inc., Research Triangle Park, North Carolina, USA/ToxStrategies, Research Triangle Park, North Carolina, USA

Study design, coordination and implementation of in-life study and sample collection

Susan Borghoff, Ph.D., DABT

Integrated Laboratory Systems Inc., Research Triangle Park, North Carolina, USA/Bayer Crop-Sciences, Research Triangle Park, North Carolina, USA

Study design, coordination and implementation of life study and sample collection

Leah Zorrilla, Ph.D.

Division of Intramural Research, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

Statistical analysis and data quality control

Grace E. Kissling, Ph.D.

Keith Shockley, Ph.D.

Contributors

Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

Management of peer review of draft document

Mary S. Wolfe, Ph.D.

ICF, Durham, North Carolina, USA

Technical editing and document production, peer-review support

Natalie Blanton, M.P.H.

Penelope Kellar, M.S.

Whitney Mitchell, B.S.

Blake C. Riley, B.S.

Anna Stamatogiannakis, B.S.

Peer Review

The draft *NTP Research Report on Baseline Characteristics of Diversity Outbred (J:DO) Mice Relevant to Toxicology Studies* was evaluated by the reviewers listed below. These reviewers served as independent scientists, not as representatives of any institution, company, or governmental agency. In this capacity, reviewers determined if the design and conditions of these NTP studies were appropriate and ensured that this NTP Research Report presented the experimental results and conclusions fully and clearly.

Peer Reviewers

Cory Brayton, D.V.M., DACVP, DACLAM

Director, Phenotyping Core

Associate Professor, Molecular and Comparative Pathology

Johns Hopkins University, School of Medicine

Baltimore, MD

John Cullen, V.M.D., Ph.D., DACVP, FIATP

Professor, Anatomic Pathology

North Carolina State University, College of Veterinary Medicine

Raleigh, NC

Publication Details

Publisher: National Toxicology Program

Publishing Location: Research Triangle Park, North Carolina

ISSN: 2473-4756

DOI: <https://doi.org/10.22427/NTP-RR-6>

Report Series: NTP Research Report Series

Report Series Number: 6

Official citation: Harrill, AH, Borghoff, S, Zorrilla, L, Blystone, C, Kissling, GE, Malarkey, DE, Shockley, K, Travlos, G, DeVito, MJ. 2017. *NTP Research Report on Baseline Characteristics of Diversity Outbred (J:DO) Mice Relevant to Toxicology Studies*. NTP RR 6. Research Triangle Park, NC: National Toxicology Program (6): 1-32.

Abstract

Genetic variability in humans is a key characteristic underlying susceptibility potential for responses to environmental chemicals, yet animal testing paradigms typically do not capture the breadth of human diversity. Diversity Outbred (J:DO) mice are a population of outbred animals with well-randomized genetic sequence polymorphisms and a genetic architecture similar to that of human populations. This study was designed to evaluate baseline characteristics of J:DO mice that could serve as a foundation for future toxicological studies using this population-based model. Adult male DO mice ($n = 75$) were fed a control diet and monitored for 13 weeks. Clinical chemistry values and calculated reference ranges, organ and body weights, and background histopathology for the liver and kidney are provided. Measurements of sperm parameters and levels of leptin and insulin are included to assess variability in reproductive and endocrine parameters. Power calculations are provided for endpoints having comparator data from prior recent NTP studies, allowing researchers to determine the number of DO mice needed—per endpoint—to observe a comparable statistical difference to that using conventional rodent stocks (80% power, $p = 0.05$). The data suggest sample sizes for guideline toxicological studies when using DO mice would need to be increased to at least 16–34 mice for most conventional endpoints. In contrast, hormonal and reproductive endpoints would require >100 mice per endpoint to achieve the same power for determining a difference in experimental groups comparable to that of conventional studies, which may impact study design. The results of this initial characterization of DO mice suggest an inherent variability in the underlying biology that might provide utility for evaluating population dynamics in responses relevant to genetically diverse human populations.

Key Words: J:DO, Diversity Outbred, population, variability, reference range, power calculation

Introduction

Human variability is a key characteristic that underlies the potential for and severity of a response to environmental chemicals. Characterizing uncertainty and variability in human health risk assessments is, therefore, an important consideration for assessors¹. A recognized opportunity for informing risk assessments is to identify methods of adjusting animal points of departure (PODs) to human PODs, which will require methods to quantitatively characterize population variability as it relates to pharmacokinetics and pharmacodynamics.

Central to characterizing population dynamics is understanding sources of variability in response to chemical exposures. Sources of variability can be characterized as extrinsic or intrinsic. Extrinsic sources of variability include co-exposures in the subjects' environment, food sources and diet, and psychosocial stressors. Intrinsic factors are numerous, and effective at multiple biological scales, ranging from cellular diversity (e.g., genetic sequence variation, epigenetic variability) to whole organs (e.g., co-morbidities, co-exposures) to the overall status of the organism (e.g., gender, life stage, aging, circadian rhythm/chronotype)². Genetic differences have been shown to have significant impacts on human health, including susceptibility to adverse effects caused by exposure to both drugs and chemicals [e.g., apolipoprotein E (APOE)-4/4 allele carriers respond poorly to conventional Alzheimer's disease treatments³; HLA-B*5701 genotyping reduces the risk of hypersensitivity reactions to the drug abacavir⁴].

Although characterizing all sources of variability in an experimental setting for each exposure might be intractable, new tools are available that enable exploration of the contribution of genetic variation to population dynamics in response to chemicals assessed in toxicological studies. One such tool of interest to the National Toxicology Program (NTP) is the Diversity Outbred (DO) mouse population, which originally was created by the Complex Trait Consortium as a tool for understanding the contribution of gene variants to disease outcomes, as well as baseline characteristics that vary across human populations, such as exercise tolerance and propensity toward obesity⁵.

DO mice are an outbred stock derived from a set of recombinant inbred lines called the Collaborative Cross (CC) that was created by intercrossing eight genetic co-founder strains. The founder strains of the DO and CC mice include 129S1/SvImJ, A/J, C57BL/6J, CAST/EiJ, NOD/ShiLtJ, NZO/HILtJ, PWK/PhJ, and WSB/EiJ^{6, 7}. DO mice are robust and breed efficiently, with average first litters comprising seven pups (± 2.4 SD)⁸. In the DO mouse population, each individual mouse is genetically unique and heterozygous for many genetic alleles, similar to human populations. Owing to the highly structured breeding program from which DO mice were derived, this population is highly genetically diverse with balanced allelic variation that is evenly spread across the genome. The frequency of minor (rare) genetic alleles within the population averages 12.5% (with equal contribution from the eight founder strains). Toxicological studies in the DO mice offer advantages over human epidemiological studies, in part because minor allele frequencies in humans average below 5%. More humans would therefore be needed to power an investigation into population dynamics adequately than if DO mice were used⁹. DO mice have been used to identify genetic sequence variants that underlie a wide variety of complex traits, such as pain sensitivity⁹, development of atherosclerosis¹⁰, and neurobehavioral characteristics¹¹. Similarly, DO mice afford an opportunity to study toxicologically relevant traits, such as adverse responses to drugs and chemicals.

Several proof-of-principle experiments have been conducted that engender enthusiasm for using DO mice in toxicological investigations. French et al. (2015)¹² evaluated the micronucleus response to benzene inhalation in a 28-day study in DO mice and found that DO mice were 10 times more sensitive than B6C3F1 mice to benzene toxicity. Consequently, toxicity evaluations using the DO would have yielded a benchmark dose closer to the dose at which humans have been shown to have adverse health effects due to occupational benzene exposure¹³. With an appropriate study design, genetic risk factors for toxicity can also be identified, if desired. For example, linkage mapping on peripheral blood micronucleated reticulocytes implicated a quantitative trait locus (QTL) on Chromosome 10 associated with resistance to benzene toxicity. One gene within the QTL support interval was sulfotransferase 3a1 (*Sult3a1*), which is involved in benzene detoxification and for which an inherited copy number expansion conferred toxicity protection¹². Similarly, DO mice have been used to investigate idiosyncratic (rare), but severe, liver injury that occurred in consumers taking herbal supplements containing green tea extract. The primary phytochemical in green tea extract with purported health benefits and antioxidant capabilities is epigallocatechin gallate (EGCG). Church et al. (2015)¹⁴ identified a subset of DO mice that was extremely sensitive to EGCG-induced hepatotoxicity (analogous to the small subset of human individuals that experience EGCG-induced liver toxicity) and used genetic mapping to identify translatable genetic risk factors for susceptibility. A clinical study found that variants in human orthologs of three of the candidate risk genes first identified in mice were enriched in individuals who experienced EGCG-induced liver toxicity. Such information could be useful in causality assessments in the clinical setting of liver injury suspected to be caused by green tea extract-containing herbal supplements.

These new murine models and approaches to understanding population variability have promise for understanding population dynamics of adverse chemical responses. Previous studies^{12; 14} demonstrate the strength of DO mice to identify human-relevant responses, including those that occur in only a subset of genetically sensitive humans. The genetic diversity inherent to DO mice confers the potential for better representing variation in responses that could occur in human populations, and incorporating that variation into risk assessment¹⁵. To best design such studies, one would need to understand the variability in the chosen animal model and how that compares to the variability in the traditional models¹⁶.

An important first step toward evaluating the DO as a population-based model for assessing chemical risks is an understanding of baseline variability in characteristics that are commonly evaluated in subchronic toxicity studies, such as clinical pathology and organ weights. An understanding of baseline population variability in a variety of contexts and in comparison to the current standard—B6C3F1/N mice—will enable power calculations for detection of significant changes in biological parameters during toxicological evaluation. To this end, we report here clinical chemistry, organ and body weights, and background histopathology findings for adult DO mice. Where possible, we report clinical reference ranges for common analytes to form a basis of comparison for future investigations using DO mice. We provide a quantitative assessment of variability within each parameter and report estimates of sample sizes that would be necessary to detect chemical-induced biological perturbations in a variety of commonly measured endpoints.

Methods

Animals and Treatment

Seventy-five male Diversity Outbred (J:DO; The Jackson Laboratory) mice (6 weeks old) were procured and singly housed in polycarbonate microisolator cages with irradiated absorbent heat-treated hardwood bedding (Northeastern Products Corporation, Warrensburg, NY) and cotton fiber Nestlets™ (Ancare, Bellmore, NY) throughout the 14 weeks of the study. All animals were housed in the Integrated Laboratory Systems AAALAC-accredited animal facility that was maintained at temperatures between 20 and 25°C, with a relative humidity of 30–70%.

Additionally, animal room light controls were set to maintain a 12-hour light/dark cycle. All animal procedures were in compliance with NIH and NTP guidance, with animals handled and treated according to the Guide for the Care and Use of Laboratory Animals and in compliance with all applicable federal, state, and local regulations. Study and procedures were approved by the Institutional Animal Care and Use Committee at Integrated Laboratory Systems.

Following a 1-week acclimation period, animals were placed on an ad libitum control diet (D12450J; 10% kcal from fat, 20% kcal from protein, and 70% kcal from carbohydrates; Research Diets, New Brunswick, NJ; n = 75), except during designated fasting periods. Consistent with standard NTP protocol, animals had ad libitum access to reverse-osmosis-treated tap water from glass bottles. Retrospective annual analysis for the quality of drinking water from the City of Durham, NC was reviewed and showed no potential contaminants in the drinking water that might have affected the study outcome. Results of comprehensive chemical analyses of reverse osmosis water by National Testing Labs, Inc. (Cleveland, OH) were reviewed and likewise showed no potential contaminants that might have affected the study outcome.

Due to the large number of animals in this study, the mice were divided into 4 replicate cohorts consisting of 18 or 19 animals per cohort. To maintain time consistency of sample collection, the study start days were staggered for the cohorts and one cohort was started on each of 4 consecutive days. Clinical observations were performed during the acclimation period (within 2 days of arrival) and weekly throughout the study. The initial week 1 assessment coincided with beginning of feed consumption measurements for each cohort. Feed consumption (g/kg BW/day) was measured for each animal during the initial acclimation period and weekly until study termination (data not shown). Body weights were recorded at the start of the study and weekly through study termination at week 14.

One day prior to the start of the study and on study termination, each mouse was anesthetized via isoflurane inhalation, and blood was collected from the retro-orbital (RO) sinus; approximately 100–150 µL of blood was collected per animal (no more than 1% of body weight). The blood was collected into a Microtainer™ serum separator tube (Becton Dickinson and Company, Franklin Lakes, NJ) and allowed to clot for approximately 1 hour. The blood was then centrifuged; the serum was aliquoted into 1.2-mL cryogenic vials and stored at or below –70°C for subsequent glucose analyses.

In-Life Resting-State Glucose

In-life blood glucose evaluations were performed on all mice at week 14, 1 day prior to study termination. Briefly, mice were transferred to clean cages with access to water but without access

to feed beginning at 0700 hours. The mice were fasted 4–5 hours, and resting-state glucose levels then were measured using a handheld Blood Glucose Monitoring System Meter (AlphaTRAK 2, Abbott Laboratories, Chicago, IL). Blood samples were obtained by tail prick; approximately 1 μ L of blood was collected directly onto the glucometer test strip for immediate analysis. Following resting-state glucose evaluation, the mice were returned to their original housing and provided ad libitum access to feed.

Terminal Blood and Tissue Collection

At the scheduled day of study termination and at a consistent time of day, mice were weighed and each mouse then anesthetized via isoflurane inhalation. Up to 1,400 μ L of blood was collected from the RO sinus into serum separator tubes (2 tubes per animal, up to 700 μ L per tube). The anesthetized mouse was then euthanized by cervical dislocation and submitted for necropsy. The blood was allowed to clot for approximately 1 hour and then centrifuged. The serum was harvested and aliquoted into up to four 1.2-mL cryogenic vials (50 μ L, 200 μ L, 200 μ L, and remainder) and stored at or below -70°C until analyzed.

At necropsy, the liver, brain and left and right kidneys (weighed individually), testes (left weighed only), and epididymides (weighed individually) were removed and weighed. The left lobe of the liver was cut in half (transverse cut) with one half placed into 10% neutral buffered formalin (NBF). The right kidney (longitudinal cut) was placed into 10% NBF; the right testis was collected in modified Davidson's solution. The right epididymis was used for sperm motility and concentration evaluations. Epididymal sperm motility and concentration parameters were measured by computer-assisted sperm analysis (CASA, Hamilton-Thorne, Beverly, MA) per NTP specifications. The instrument parameters were minimum contrast, 30; minimum cell size, 4 pixels; straightness threshold, 50.0%; path velocity cutoff, 10.0 $\mu\text{m/s}$; progressive minimum path velocity, 50.0 $\mu\text{m/s}$; static head size, 0.13 to 2.43; static head intensity, 0.10 to 1.52; and static elongation, 5 to 80.

Hyperactivated sperm were evaluated by allowing the sperm to disperse into 1.0 mL phosphate-buffered saline (PBS) for 5 min at 37°C . Debris was removed by a brief low-speed centrifugation ($100 \times g$ for 15 sec at room temperature), and the sperm were then centrifuged at $500 \times g$ for 8 min at 4°C and resuspended in 2 mL of M16 medium (Sigma). Aliquots were either processed as below or incubated at 37°C in 5% $\text{CO}_2/95\%$ air for 1.5 hours. Hyperactivated sperm were measured using the following sorting parameters: track speed $>170 \mu\text{m/s}$, amplitude of lateral head displacement $>9.0 \mu\text{m/s}$, and linearity $<30\%$ consistent with Ghanayem et al. (2010)¹⁷.

Histopathology

Tissues in 10% NBF [liver, right kidney (longitudinal section)] were transferred to 70% histology-grade alcohol 18 to 24 hours post-necropsy. The fixed tissues were histologically processed and paraffin-embedded within 5 days after transfer to alcohol. A pathologist, without knowledge of animal identity, evaluated hematoxylin and eosin-stained liver and kidney tissue sections for morphological alterations. Gradable lesions were graded on an ascending scale of 1 to 5, depending on severity; nongradable lesions were noted as present. Grading was based on a five-point scale as follows: 1 – minimal, 2 – mild, 3 – moderate, 4 – marked, and 5 – severe. Minimal changes included few or small to nearly undetectable changes, Mild changes were slightly increased over minimal, Moderate changes were obvious but not overwhelming, Marked

changes were many in number or large, Severe changes included extensive increase in size or number or degeneration that is often associated with adverse effect on the tissue or whole animal. The histopathology evaluation was performed with a low or no threshold, such that any lesion was recorded regardless of significance.

Clinical Chemistry

Clinical chemistry evaluations were performed on all serum samples collected at study termination on an Olympus AU400e chemistry analyzer (Beckman Coulter, Inc., Irving, TX) using commercially available reagents and standard methods. Biochemical markers included serum concentrations of glucose, triglycerides, cholesterol, high-density lipoprotein (HDL), low-density lipoprotein (LDL), non-esterified free fatty acids (NEFA), blood urea nitrogen (BUN), creatinine, calcium (Ca), inorganic phosphorus (iPhos), sodium (Na), potassium (K), chloride (Cl), total carbon dioxide (CO₂), anion gap, total protein, albumin, and globulin. Serum activities of alanine aminotransferase (ALT), alkaline phosphatase (ALP), sorbitol dehydrogenase (SDH), and creatine kinase (CK) also were determined. Serum insulin and leptin concentrations were determined by electrochemiluminescent immunoassay methods using a Sector Imager 2400 and mouse-specific reagents from Meso Scale Discovery in blood samples collected at necropsy and at the end of week 1 (MSD; Gaithersburg, MD).

Statistical Analyses

Data are presented as the mean \pm standard deviation or as the result from the individual animal. Comparisons of metrics across time points were made using a two-sided Student's t-test where the data were normally distributed. When the variances between treatment groups were not equal, a Welch corrected t-test was used. P-values are two-sided and considered significant if $p < 0.05$.

Normal reference ranges for each analyte in male DO mice were calculated according to standard practice for setting reference intervals set by the American Society for Veterinary Clinical Pathology, which mirror the 2008 Clinical Laboratory and Standards Institute recommendations for human patient ranges¹⁸. Outlier values were first identified and removed using the robust regression and outlier removal (ROUT) analysis in Graphpad Prism (version 7.00; La Jolla, CA), which identifies outliers using nonlinear regression ($Q = 0.1\%$)¹⁹. Next, the data were tested for normality using the D'Agostino-Pearson omnibus normality test (Graphpad). For analytes with a normal distribution, the reference range was set at the mean \pm 2 SD. When the lower-bound value (-2 SD) was below zero, the minimum value was set to zero (for sperm concentration and sperm percent hyperactivity). For non-normally distributed endpoint data, analyte values were rank ordered; the minimum value of the reference range was equivalent to the ranked value corresponding to the 2.5th percentile and the maximum value was equivalent to the ranked value at the 97.5th percentile. Outlier values were removed only for the purposes of setting the reference range; these outlier values were included in all other analyses, including power calculations.

Power Calculations

One goal of the study was to determine how many DO mice per group would be needed to achieve the same statistical power as 10 B6C3F1/N mice per group typically used in 13-week

NTP studies. Data on selected endpoints from male control B6C3F1/N mice were summarized from five recent NTP subchronic studies that used either drinking water or gavage exposures²⁰⁻²⁴. For each selected endpoint, the magnitude of change that could be detected using 10 B6C3F1 mice per group with 80% power at two-sided $p = 0.05$ was determined assuming normally distributed endpoints. The number of DO animals required to detect this difference with 80% power at two-sided $p = 0.05$ then was calculated from the standard deviations observed in the male DO mice, using standard formulas²⁵(PROC POWER in SAS 9.3). Endpoints evaluated included organ weights, organ-to-body weight ratios, body weight change, epididymal sperm counts, and % sperm motility. Clinical chemistry analyses are not performed on mice in subchronic studies at NTP but are performed in rats. As an approximation, standard deviations from male control F344/N (four studies) and F344/NTac (one study) rats in recent NTP subchronic studies that used either drinking water or gavage exposures²⁰⁻²⁴ were used, as described above, to determine the clinical chemistry differences that could be detected with 80% power at the two-sided 0.05 level of significance. Then, based on the standard deviations observed in male DO mice, the number of DO mice required to detect the same difference with 80% power at the two-sided 0.05 level of significance was determined for each endpoint.

Results

Body Weight and Body Weight Gain

The mean of the initial body weights of the 75 DO animals was 23.6 ± 2.9 g (median 23.8 g). NTP study guidelines for subchronic and chronic toxicological studies dictate that body weights for individual animals at study start be no more than $\pm 20\%$ of the mean. Therefore, the acceptable body weight range for this study would be 18.9–28.3 g; four mice (5% of total) had body weights greater than the upper boundary, and two mice (3%) had body weights less than the lower boundary. These animals would not conform to the current NTP specifications set for standard B6C3F1 mouse studies. Final body weights at the end of the 14-week study were 33.7 ± 5.0 g (median 32.7 g). Over the course of the study, the animals gained 10.1 ± 3.8 g on average (median 9.5 g) from starting weights on week 1.

No abnormal clinical signs were observed over the course of the study, except for three animals exhibiting mild or moderate alopecia. All animals survived until the end of the study, with the exception of one animal, which was found dead on week 3 with no abnormal findings observed at necropsy.

Organ Weights

Organ weights from the DO mice at week 14 are presented in Table 1. The average coefficient of variation (CV) across organs tested was 18%, with the largest CVs for the left and the right epididymis (26% and 27%, respectively). The smallest CV observed was in the brain weights (8%), although the CV was larger (16%) for the brain/body weight ratio (Table 1). Liver weight/body weight ratios averaged approximately 4% ($4.10 \pm 0.49\%$) with a CV of 20%.

Epididymal Sperm Evaluation

Epididymal sperm concentrations and motility were evaluated on the day of necropsy (Table 2). The CV for the various measurements of sperm counts and quality varied from 8.6% to 80%. The percent hyperactive sperm ranged from 0.12% to 9.07%, with a CV of 80. Sperm counts were the second most variable sperm measurement and ranged 66-fold from 0.5×10^6 to 33.1×10^6 sperm/mL with a CV of 58%. Measurements related to sperm quality varied similarly to organ weights with CVs from 8.6% to 31%.

Clinical Chemistry

Several terminal serum clinical chemistry measurements were analyzed in these mice. Some parameters, such as albumin, Ca, Na, Cl, and total protein concentrations varied little across animals, with CVs less than 7% (data not shown). Reference ranges for each analyte are presented in Table 3. Where data were available, reference ranges for both B6C3F1 mice tested previously at NTP and common laboratory mouse outbred stock CD-1 were compared to ranges calculated for the DO stock²⁶⁻²⁹. These analytes exhibited a similar range across mouse models. Reference ranges for albumin, BUN, fasting glucose, K, and Na concentrations and ALT and ALP activities were equivalent between the DO and the B6C3F1 mice. Reference ranges for cholesterol, creatinine, and total protein concentrations in the DO mouse were more tightly

constrained than for the B6C3F1 mouse, with creatinine concentration exhibiting the tightest range (0.1–0.2 mg/dL in the DO as compared to 0.2–0.8 mg/dL in B6C3F1 mice).

Insulin and Leptin

Serum insulin and leptin concentrations had the largest variability of the endpoints evaluated in this study (Table 4). Serum insulin at terminal sacrifice demonstrated large variability and ranged from 71.8 to 63,225 I.U./mL with a CV of 282%. Serum leptin averaged 5,904 I.U./mL and ranged from 219 to 4,415 I.U./mL with a CV of 85%.

Histopathology

Microscopic lesions observed in the liver are presented in Table 5. These findings were typical for background lesions observed in mice and consisted of hepatocellular fatty change, glycogen accumulation, single cell necrosis, and increased mitosis; basophilic focus; chronic inflammation; periportal mononuclear cell infiltrates; and megakaryocytosis. With the exception of hepatocellular fatty change and glycogen accumulation, all findings were minimal (grade 1) in severity. Hepatocellular glycogen accumulation was the most common finding (incidence, 72/75; 96%). Hepatocellular fatty change also was observed in 25% of the DO mice (19/75). Evidence of chronic inflammation was observed in 44% of the animals (33/75). Chronic inflammation was noted almost exclusively in animals having minimal to mild fatty change.

Microscopic lesions also were observed in the kidney (Table 5). These findings included cortical tubule regeneration (31/75; 41%), vacuolization (15/75; 20%), dilation (10/75; 13%), and hypertrophy (5/75; 7%); papillary and cortical tubule mineralization (13/75; 17% and 2/75; 3%, respectively); mononuclear cell infiltrate (52/75; 69%); and, more rarely, chronic progressive nephropathy and artery inflammation (each in 1/75; 1%). Nearly all changes were minimal to mild, and a few occurred only in single animals.

Power Calculations

NTP evaluates the subchronic toxicity of chemicals in male and female B6C3F1 mice in 90-day exposure studies consisting of 3–5 dose concentration groups and controls with 10 mice/sex/group. To compare the variation in common study parameters between B6C3F1 mice and DO mice, the data from the male control B6C3F1 mice from the five recent subchronic toxicity studies in which oral gavage or drinking water was used as the route of exposure were summarized. Table 6 shows group means and standard deviations of the individual NTP studies compared to the DO mice in the current study. The means of the DO mice for the various parameters were similar to those of B6C3F1 mice; however, the standard deviations were two to eight times larger in the DO mice (Table 4).

Table 7 shows the absolute and percent differences from control detectable with a power of 80% and two-sided $p = 0.05$ for organ weights, relative organ weights, and sperm counts and motility, based on the data from the control B6C3F1 mice from the NTP studies. The typical 90-day subchronic toxicity study has 80% power ($p = 0.05$) to detect differences between control and treated groups of 4–7% in body weight, liver, kidney and epididymal weights, and sperm motility (Table 6). Epididymal weights and sperm counts are more variable, and the NTP 90-day studies in B6C3F1 mice can detect a 15% or more difference between control and treated animals with a power of 80% (at $p = 0.05$). Having equivalent statistical power to detect the

same differences between control and treated animals in a DO mouse study would require approximately 16–34 mice/treatment group for all endpoints except sperm counts (Table 7). Due to the large variance in sperm counts in the DO mouse, 150 DO mice/treatment group would be required to have the same power as the standard 90-day subchronic toxicity study with 10 B6C3F1 mice/treatment group. In contrast, the variation in clinical chemistry parameters in the DO is similar to that for the rats used at NTP (Table 8); to attain the same power as a typical NTP subchronic study in rats evaluating the same clinical chemistry parameters would require no more than 16 DO mice per treatment group (Table 8).

As shown in Table 4, due to wide variability of leptin and insulin concentrations within the population, 288 DO mice would be needed to achieve 80% power to detect a 20% difference in serum leptin levels from the controls. Similarly, 3,138 DO mice would be needed to detect a change within these parameters for insulin levels. This observation demonstrates a remarkable capacity of the DO mice for population-wide variability and inherent, wide population dynamics for hormonal traits related to metabolism.

Discussion

Toxicologists have struggled to balance experimental designs that are efficient and sensitive to small chemical-induced changes with those that provide insight into population variability. Although traditional inbred strains and F1 hybrids offer an advantage in economy of scale, because all of the animals are genetically identical, potential human-relevant hazards might be missed by conventional screens. The current study provides a foundation for toxicological assessments using a population-based approach that encompasses inherent genetic diversity.

One approach to informing study designs that use newer population-based models is to estimate the number of animals required for the population-based models to have equivalent statistical power with the traditional model. At NTP, subchronic toxicity studies use 10 B6C3F1 mice per treatment group; at the end of the study, body weight, body weight gains, organ weights, hematology, clinical chemistry, and histopathology of approximately 40 tissues are evaluated for treatment-related effects. In addition, males are examined for sperm counts and sperm motility, while vaginal cytology is evaluated in females. In the current study, we calculated the absolute and percent differences that could be statistically detected with 80% power at two-sided $p = 0.05$ using control data on B6C3F1 male mice from the five most recent studies published by NTP²⁰⁻²⁴ that used either drinking water or diet as the route of exposure. These routes of exposure were used to avoid confounding of the different vehicles used in gavage studies.

A typical NTP subchronic study has 80% power to detect between a 4% and 7% difference from controls for body weight and organ weights at the two-sided 0.05 level of significance. The exception is testes and relative testes weights in which the typical NTP subchronic toxicity study can only distinguish a 13% and 15% difference from controls, respectively. Power calculations for clinical chemistry measurements in rats indicated 80% power to detect between a 2% and 5% difference between treated and control groups. Power calculations for sperm measurements in B6C3F1 mice indicate the typical NTP subchronic study has 80% power to detect at least a 16% difference in sperm counts between control and treated groups.

The variability in the DO mouse for the endpoints typically evaluated in an NTP subchronic toxicity study is greater than that observed for the currently used rodent models. Power calculations indicate that using 16–34 DO mice per group results in the same statistical power as using 10 B6C3F1 mice per group for organ weights and body weight changes. Surprisingly, clinical chemistry parameters typically measured in NTP studies require no more DO mice than NTP subchronic toxicity studies with 10 rats per group, with the exception of albumin and BUN concentrations, and ALP activity, which would need between 12 and 20 DO mice per group to achieve the same power. These findings suggest that, for most endpoints, tripling the size of the study when using DO mice could provide the same sensitivity to identify a chemical hazard as a typical study using 10 B6C3F1 mice per treatment group, while at the same time obtaining a better understanding of population variability.

For routine screening use, a DO mouse study could have limitations for identifying male testicular toxicants. Based on our calculations, studies to examine male mouse reproductive toxicants could require up to 150 DO mice per treatment group to obtain the same power as 10 B6C3F1 mice/treatment group, due to the large variability in sperm counts across DO mice. Of note is that reports in rats have indicated inhaled anesthetic used prior to collection could

impact observed motile sperm counts in the vas deferens³⁰. In this study, isoflurane was used, which is a departure from typical NTP guidance that prescribes CO₂ euthanasia. Although choice of anesthetic in this study could impact overall concentrations, the Campion et al. (2012)³⁰ study found coefficients of variation in sperm counts between anesthetic types were unaffected and the expected impacts on animal-to-animal variability and power calculations presented here are anticipated to be minimal.

In the male DO mice, body weights at study start ranged from 17.2 g to 30.1 g. Harrill et al. (2016)¹⁶ measured body weights in young adult female mice of 34 inbred strains on standard rodent chow, and the body weight averages per strain ranged from 11.7 g to 30.5 g. The results from Harrill et al. (2016)¹⁶ suggest that variability in the DO mice is consistent with the variability observed across a wide variety of mouse strains. A wide dynamic range in many of the endpoints measured in this study, however, is not surprising, given the variability in the eight genetic co-founder strains of the population. For example, two of these strains are well-known models for non-obese diabetes, obesity, or metabolic syndrome: NOD/ShiLtJ³¹ and NZO/HILtJ³². In some instances, this background might provide advantages for assessing risk in western societies, where obesity is highly prevalent and the average body mass index is increasing³³. Although the variability of the DO mouse provides unique opportunities to relate phenotypic responses to specific genetic variants, it can pose challenges for toxicology. The variability in sperm counts measured in this study suggests the use of the DO mice for screening chemicals that cause reproductive toxicity could require over 150 animals per dose group, which is unlikely to be feasible for most regulatory toxicological studies.

A high degree of variability was observed in hormonal endpoints (insulin and leptin concentrations). Variability in insulin measurements might be linked to variability that has been observed in the non-fasted state. In part, this wide degree of variability in measurements across DO mice also could reflect population variability observed for these parameters within humans. This variability also might be due to genetic combinations arising from the characteristics of the eight founder strains of the DO population, as they relate to obesity and diabetes. For example, two of the DO founder strains, NOD/ShiLtJ and NZO/HILtJ, are well known for their non-obese diabetic and obese body traits, respectively. Similarly, C57BL/6J and CAST/EiJ are known to be susceptible or resistant to diet-induced obesity, respectively³⁴. Because each DO mouse contains genetic material from all of these strains in a unique and randomized pattern, the expectation is that an admixture of alleles results in a subset of individuals that exhibit extreme phenotypes of metabolic parameters. These characteristics are important to consider when designing a study aimed at understanding metabolic syndrome or obesity. NTP is currently investigating effects of a 13-week high fat diet study in DO mice, the findings of which will be the subject of a subsequent publication.

In this study, reference ranges were calculated for common clinical chemistry parameters to provide a resource for future toxicological studies that seek to distinguish an “abnormal” value that falls outside of the range. Importantly, many of the analytes tested in the study exhibited reference ranges that are consistent with B6C3F1 mice or mice evaluated by veterinary pathology laboratories that are independent of strain or stock (Table 3). That reference intervals for serum cholesterol and creatinine concentrations and SDH activity were narrower than those identified for B6C3F1 mice was surprising. This discrepancy could reflect either a greater biological constraint on values in DO mice or simply a difference in instrumentation used to assay the clinical chemistry analytes between the DO and historical B6C3F1 data. Of note is that

current guidelines typically recommend using at least 120 animals for setting a reference range; here we base our analysis on values from 75 animals, which can be expected to produce relatively robust reference limits¹⁸. We propose that reference ranges set in this study provide a reasonable starting point, acknowledging that further refinements to the ranges as more DO mouse data are collected would be desirable for setting the definitive range for each analyte. That each clinical pathology lab establish reference ranges using the instrumentation used for test samples with unknown values is generally accepted. This practice often is not done, however, with investigators instead relying on transference of intralaboratory reference ranges calculated elsewhere³⁵. The reference ranges presented in this report thus provide a convenient tool to establish a range of baseline values that can be used in future studies to, for example, determine a toxicological effect of a chemical agent.

In the current study, we performed histopathological analyses of the liver and kidneys in a manner consistent with the approach NTP uses in its subchronic toxicity studies. In the livers of the DO mice, hepatocellular fatty change, glycogen accumulation, single-cell necrosis, and increased mitosis; basophilic focus; chronic inflammation; periportal mononuclear cell infiltrates; and megakaryocytosis were noted. The incidence and severity of these liver findings in the control animals were considered within the normal physiological range and consistent with the incidence and severity observed in control B6C3F1 mice in NTP subchronic studies³⁶. Evaluation of the kidneys indicated a background incidence of cortical tubule regeneration, vacuolization, dilatation, and hypertrophy; papillary and cortical mineralization; mononuclear cell infiltrates; chronic progressive nephropathy; and artery inflammation. Similar to the liver findings, nearly all these changes were minimal to mild, and a few types of lesions occurred only in single animals, consistent with the incidence and severity observed in control B6C3F1 mice in NTP subchronic studies. Although these lesions have been occasionally described as pathological responses to toxicants, they often are found as mild or minimal spontaneous changes in untreated animals³⁷. The histopathological findings indicate that unexposed DO mice are no different from the typical mouse strains used in NTP studies and that in a screening bioassay, sample sizes of 16–34 DO mice/treatment group might be adequate for detecting most kidney or liver toxicants via histopathological assessment.

Additional studies using common NTP testing parameters would be desirable for increasing the utility of the DO mouse population for toxicological studies. First, repeating the studies using female mice would be prudent to account for potential sex differences in the clinical chemistry and histopathology parameters reported here. Also noteworthy is that departures in the current study from current NTP testing guidelines should be taken into account in subsequent studies. For example, the control diet used in this study (D12450J; 4.3% fat, 19.2% protein, and 67.3% carbohydrate, by weight) is a departure from the typical rodent diet used in long-term toxicological studies (NTP-2000; 8.2% fat, 14.6% protein, 52% carbohydrate, by weight). Additional studies are necessary to determine the overall effect of diet composition on baseline parameters; however, that reference ranges for DO mice calculated in this study are in line with veterinary reporting estimates for mice on a variety of diets (Table 3) is encouraging. This observation might imply that minor differences in standard control diets will have smaller impacts on baseline clinical chemistry measurements as compared to the effect of inter-individual variability inherent to including diverse genetic backgrounds. Similarly, investigating commonly measured endpoints would be desirable, such as hematology and pathology of additional organs, which were not collected during the course of the current investigation.

Using DO mice for toxicological studies would provide insight into population variability in responsiveness to the test chemical. Using a population-based screen over isogenic strains/hybrids or other genetically limited outbred stocks has several advantages. One such advantage is the ability to detect toxicity responses that are relatively rare, yet occur in the human population in susceptible individuals¹⁴. A better understanding of the dynamics in population responses might provide key information that enables determination of extrapolation factors in benchmark doses to diverse human populations¹⁵. Because this current study examined relatively few endpoints, further experiments in DO mice to evaluate the variability of a particular endpoint of concern would be advisable. Based on the data presented here, use of DO mice in typical guideline toxicity studies would require modification to accommodate the need for larger sample sizes in each group. Taken together, the current dataset represents an important first step toward understanding background variation in commonly measured parameters, the reference range of key clinical chemistry markers, and discernment of sample sizes needed to adequately power a toxicological study using DO mice. These data are expected to form a foundation for advanced study designs to query population variability in toxicity responses.

References

1. Rodricks JV, Levy JI Science and decisions: advancing toxicology to advance risk assessment. *Toxicol Sci.* 2013; 131(1):1-8. <http://dx.doi.org/10.1093/toxsci/kfs246>
2. Zeise L, Bois FY, Chiu WA, Hattis D, Rusyn I, Guyton KZ Addressing human variability in next-generation human health risk assessments of environmental chemicals. *Environ Health Perspect.* 2013; 121(1):23-31. <http://doi.org/10.1289/ehp.1205687>
3. Cacabelos R Pharmacogenetic basis for therapeutic optimization in Alzheimer's disease. *Mol Diagn Ther.* 2007; 11(6):385-405. *Mol Diagn Ther.* <http://dx.doi.org/10.1007/BF03256262>
4. Mallal S, Phillips E, Carosi G, Molina JM, Workman C, Tomazic J, Jagel-Guedes E, Rugina S, Kozyrev O, Cid JF et al. HLA-B*5701 screening for hypersensitivity to abacavir. *N Engl J Med.* 2008; 358(6):568-579. *N Engl J Med.* <http://dx.doi.org/10.1056/NEJMoa0706135>
5. Gatti DM, Svenson KL, Shabalin A, Wu LY, Valdar W, Simecek P, Goodwin N, Cheng R, Pomp D, Palmer A et al. Quantitative trait locus mapping methods for diversity outbred mice. *G3 (Bethesda).* 2014; 4(9):1623-1633. <http://dx.doi.org/10.1534/g3.114.013748>
6. Collaborative Cross Consortium. The genome architecture of the Collaborative Cross mouse genetic reference population. *Genetics.* 2012; 190(2):389-401. <http://dx.doi.org/10.1534/genetics.111.132639>
7. Philip VM, Sokoloff G, Ackert-Bicknell CL, Striz M, Branstetter L, Beckmann MA, Spence JS, Jackson BL, Galloway LD, Barker P et al. Genetic analysis in the Collaborative Cross breeding population. *Genome Res.* 2011; 21(8):1223-1238. <http://dx.doi.org/10.1101/gr.113886.110>
8. Churchill GA, Gatti DM, Munger SC, Svenson KL The Diversity Outbred mouse population. *Mamm Genome.* 2012; 23(9-10):713-718. <http://dx.doi.org/10.1007/s00335-012-9414-2>
9. Recla JM, Robledo RF, Gatti DM, Bult CJ, Churchill GA, Chesler EJ Precise genetic mapping and integrative bioinformatics in Diversity Outbred mice reveals Hydin as a novel pain gene. *Mamm Genome.* 2014; 25(5-6):211-222. <http://dx.doi.org/10.1007/s00335-014-9508-0>
10. Smallwood TL, Gatti DM, Quizon P, Weinstock GM, Jung KC, Zhao L, Hua K, Pomp D, Bennett BJ High-resolution genetic mapping in the diversity outbred mouse population identifies Apobec1 as a candidate gene for atherosclerosis. *G3 (Bethesda).* 2014; 4(12):2353-2363. <http://dx.doi.org/10.1534/g3.114.014704>
11. Logan RW, Robledo RF, Recla JM, Philip VM, Bubier JA, Jay JJ, Harwood C, Wilcox T, Gatti DM, Bult CJ et al. High-precision genetic mapping of behavioral traits in the diversity outbred mouse population. *Genes Brain Behav.* 2013; 12(4):424-437. <http://dx.doi.org/10.1111/gbb.12029>
12. French JE, Gatti DM, Morgan DL, Kissling GE, Shockley KR, Knudsen GA, Shepard KG, Price HC, King D, Witt KL et al. Diversity outbred mice identify population-based exposure thresholds and genetic factors that influence benzene-induced genotoxicity. *Environ Health Perspect.* 2015; 123(3):237-245. <http://dx.doi.org/10.1289/ehp.1408202>

13. Gatti D, French JE, Schughart K QTL mapping and identification of candidate genes in DO mice: A use case model derived from a benzene toxicity experiment. *Methods Mol Biol.* 2017; 1488:265-281. http://dx.doi.org/10.1007/978-1-4939-6427-7_12
14. Church RJ, Gatti DM, Urban TJ, Long N, Yang X, Shi Q, Eaddy JS, Mosedale M, Ballard S, Churchill GA et al. Sensitivity to hepatotoxicity due to epigallocatechin gallate is affected by genetic background in diversity outbred mice. *Food Chem Toxicol.* 2015; 76:19-26. <http://dx.doi.org/10.1016/j.fct.2014.11.008>
15. Harrill AH, McAllister KA New rodent population models may inform human health risk assessment and identification of genetic susceptibility to environmental exposures. *Environ Health Perspect.* 2017; 125(8):086002. Environmental health perspectives. <https://doi.org/10.1289/ehp1274>
16. Harrill AH. Mouse population-based toxicology for personalized medicine and improved safety prediction In: Will Y, McDuffie JE, Jeffy BD, Olaharski AJ, editors. *Drug discovery toxicology: From target assessment to translational biomarkers.* Hoboken, NJ: John Wiley and Sons, Inc.; 2016. p. 314-329. <http://dx.doi.org/10.1002/9781119053248.ch20>
17. Ghanayem BI, Bai R, Kissling GE, Travlos G, Hoffler U Diet-induced obesity in male mice is associated with reduced fertility and potentiation of acrylamide-induced reproductive toxicity. *Biol Reprod.* 2010; 82(1):96-104. *Biol Reprod.* <http://dx.doi.org/10.1095/biolreprod.109.078915>
18. Friedrichs KR, Harr KE, Freeman KP, Szladovits B, Walton RM, Barnhart KF, Blanco-Chavez J, American Society for Veterinary Clinical Pathology ASVCP reference interval guidelines: determination of de novo reference intervals in veterinary species and other related topics. *Vet Clin Pathol.* 2012; 41(4):441-453. <http://dx.doi.org/10.1111/vcp.12006>
19. Motulsky HJ, Brown RE Detecting outliers when fitting data with nonlinear regression - a new method based on robust nonlinear regression and the false discovery rate. *BMC Bioinformatics.* 2006; 7:123. <http://dx.doi.org/10.1186/1471-2105-7-123>
20. NTP (National Toxicology Program). 2012. Toxicology and carcinogenesis studies of N,N-dimethyl-p-toluidine (CAS No. 99-97-8) in F344/N rats and B6C3F1/N mice (gavage studies). <https://www.ncbi.nlm.nih.gov/pubmed/23027199>.
21. NTP (National Toxicology Program). 2014. Toxicology and carcinogenesis studies of β -picoline (CAS No. 108-99-6) in F344/N rats and B6C3F1/N mice (drinking water studies). https://ntp.niehs.nih.gov/ntp/htdocs/lt_rpts/tr580_508.pdf.
22. NTP (National Toxicology Program). 2014. Toxicology studies of tetrabromobisphenol A (CAS No. 79-94-7) in F344/N rats and B6C3F1/N mice and toxicology and carcinogenesis studies of tetrabromobisphenol A in Wistar Han [CrI:WI(Han)] rats and B6C3F1/N mice (gavage studies). https://ntp.niehs.nih.gov/ntp/htdocs/lt_rpts/tr587_508.pdf.
23. NTP (National Toxicology Program). Draft, 2014. Toxicology studies of green tea extract in F344/N rats and B6C3F1/N mice and toxicology and carcinogenesis studies of green tea extract in Wistar Han [CrI:WI(Han)] rats and B6C3F1/N mice (gavage studies).

24. NTP (National Toxicology Program). Draft, 2014. Toxicology studies of bromodichloroacetic acid (CAS No. 71133-14-7) in F344/N rats and B6C3F1/N mice and toxicology and carcinogenesis studies of bromodichloroacetic acid in F344/NTac rats and B6C3F1/N mice (drinking water studies).
25. Snedecor GW, Cochran WG. Statistical methods. Ames,; Iowa State University Press; 1967.
26. Derelanko MJ, Auletta C. Handbook of toxicology. Boca Raton: CRC Press, Taylor & Francis Group; 2014.
27. Mouse Biochemistry. 2017. https://en.wikivet.net/Mouse_Biochemistry.
28. Reference values for laboratory animals. 2017. <https://www.researchservices.umn.edu/services-name/research-animal-resources>.
29. Rodent reference ranges. University of Pennsylvania School of Medicine; 2002. <http://cal.vet.upenn.edu/projects/ssclinic/refdesk/rodentrr.htm>.
30. Champion SN, Cappon GD, Chapin RE, Jamon RT, Winton TR, Nowland WS Isoflurane reduces motile sperm counts in the Sprague-Dawley rat. Drug Chem Toxicol. 2012; 35(1):20-24. Drug Chem Toxicol. <http://dx.doi.org/10.3109/01480545.2011.564182>
31. Mathews CE, Xue S, Posgai A, Lightfoot YL, Li X, Lin A, Wasserfall C, Haller MJ, Schatz D, Atkinson MA Acute versus progressive onset of diabetes in NOD mice: Potential implications for therapeutic interventions in type 1 diabetes. Diabetes. 2015; 64(11):3885-3890. <http://dx.doi.org/10.2337/db15-0449>
32. Joost HG, Schurmann A The genetic basis of obesity-associated type 2 diabetes (diabesity) in polygenic mouse models. Mamm Genome. 2014; 25(9-10):401-412. <http://dx.doi.org/10.1007/s00335-014-9514-2>
33. Finucane MM, Stevens GA, Cowan MJ, Danaei G, Lin JK, Paciorek CJ, Singh GM, Gutierrez HR, Lu Y, Bahalim AN et al. National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9.1 million participants. Lancet. 2011; 377(9765):557-567. [http://dx.doi.org/10.1016/S0140-6736\(10\)62037-5](http://dx.doi.org/10.1016/S0140-6736(10)62037-5)
34. Kreznar JH, Keller MP, Traeger LL, Rabaglia ME, Schueler KL, Stapleton DS, Zhao W, Vivas EI, Yandell BS, Broman AT et al. Host genotype and gut microbiome modulate insulin secretion and diet-induced metabolic phenotypes. Cell Rep. 2017; 18(7):1739-1750. <http://dx.doi.org/10.1016/j.celrep.2017.01.062>
35. Geffre A, Friedrichs K, Harr K, Concordet D, Trumel C, Braun JP Reference values: a review. Vet Clin Pathol. 2009; 38(3):288-298. <http://dx.doi.org/10.1111/j.1939-165X.2009.00179.x>
36. Harada T, Enomoto A, Boorman GA, Maronpot RR. Liver and gallbladder In: Maronpot RR, BG, Gaul BW, editor. Pathology of the Mouse: Reference and Atlas Vienna, IL; 1999. p. 119-183.

37. Creasy D, Bube A, de Rijk E, Kandori H, Kuwahara M, Masson R, Nolte T, Reams R, Regan K, Rehm S et al. Proliferative and nonproliferative lesions of the rat and mouse male reproductive system. *Toxicol Pathol.* 2012; 40(6 _suppl):40S-121S.
<http://dx.doi.org/10.1177/0192623312454337>
38. Serfilippi LM, Pallman DR, Russell B Serum clinical chemistry and hematology reference values in outbred stocks of albino mice from three commonly used vendors and two inbred strains of albino mice. *Contemp Top Lab Anim Sci.* 2003; 42(3):46-52.
39. *Handbook of Toxicology.* Boca Raton, FL: CRC Press, Taylor & Francis Group; 2014.

Baseline Characteristics of Diversity Outbred (J:DO) Mice Relevant to Toxicology Studies

Table 1. Terminal Body and Organ Weights in Male DO Mice

Weights	Count	Mean	SD	SEM	CV	Median	Min	Max
Body (g)	71	33.66	4.99	0.59	15%	32.70	24.90	49.40
Brain (g)	74	0.47	0.04	0.00	8%	0.47	0.39	0.57
Liver (g)	74	1.39	0.28	0.03	20%	1.36	0.80	2.11
Left kidney (g)	74	0.21	0.04	0.00	18%	0.21	0.14	0.32
Right kidney (g)	74	0.22	0.04	0.00	17%	0.21	0.15	0.32
Total kidney (g)	74	0.43	0.07	0.01	17%	0.42	0.28	0.64
Left testis (g)	74	0.10	0.02	0.00	15%	0.10	0.07	0.14
Left epididymis (g)	74	0.05	0.01	0.00	26%	0.05	0.03	0.10
Right epididymis (g)	74	0.05	0.01	0.00	27%	0.05	0.03	0.12
<u>Brain</u> ^a	71	1.43	0.22	0.03	16%	1.40	0.92	1.94
Body (%)								
<u>Liver</u>	71	4.10	0.48	0.05	12%	4.11	2.59	5.31
Body (%)								
<u>Kidney</u>	71	1.29	0.21	0.03	17%	1.28	0.74	1.88
Body (%)								
<u>Left kidney</u>	71	0.63	0.10	0.01	17%	0.63	0.36	0.93
Body (%)								
<u>Right kidney</u>	71	0.66	0.11	0.01	17%	0.64	0.37	0.95
Body (%)								
<u>Left testis</u>	71	0.31	0.05	0.01	18%	0.31	0.19	0.53
Body (%)								
<u>Left epididymis</u>	71	0.15	0.04	0.00	28%	0.14	0.08	0.30
Body (%)								
<u>Right epididymis</u>	71	0.15	0.05	0.01	30%	0.14	0.10	0.38
Body (%)								

^aBody (%) is tissue weight in g/body weight × 100. SD = standard deviation, SEM = standard error of the mean, CV = coefficient of variation, Min = minimum value, Max = maximum value.

Baseline Characteristics of Diversity Outbred (J:DO) Mice Relevant to Toxicology Studies

Table 2. Sperm Counts and Motility Measurements in Male DO Mice

Measure	Count	Mean	SD	SEM	CV	Median	Min	Max
Sperm concentration	73	12.94	7.47	0.87	58%	13.2	0.50	33.1
Sperm motility	63	49.56	15.22	1.92	31	51.2	11.00	81.00
Path velocity	63	129.44	27.58	3.22	21	129.45	73.20	195.50
Progressive velocity	63	109.88	29.79	3.84	27	110.45	43.7	179.7
Track speed	63	228.99	35.74	4.18	16	228.95	157.40	310.20
Lateral amplitude	63	10.69	1.19	.14	11	10.60	8.4	14.1
Beat frequency	63	35.52	3.65	.42	10	35.65	26.3	44.1
Straightness	63	79.24	6.83	.80	8.6	81.30	54.00	89.40
Linearity	63	46.73	8.63	1.01	18	45.80	25.3	65.7
% Hyperactivity	40	3.1	2.49	0.39	80	2.34	0.12	9.07

SD = standard deviation, SEM = standard error of the mean, CV = coefficient of variation, Min = minimum value, Max = maximum value.

Baseline Characteristics of Diversity Outbred (J:DO) Mice Relevant to Toxicology Studies

Table 3. Reference Ranges for Clinical Chemistry Parameters

Analyte	DO Reference Range	Hsd:ICR (CD-1) Outbred Reference Range ^a	B6C3F1 Reference Range ^b	Analyte ^c	DO Reference Range
Albumin (g/dL)	2.4–3.2	2.6–3.7	2.5–4.2	Glucose AUC	31,283–64,065
ALP (U/L)	35–97	29–78	20–85	Glucose AUC/mg	461.1–1076.1
ALT (U/L)	11–46	26–59	20–50	Fasting T0 glucose	69–157
Anion gap	8.8–30.8	NA	NA	Glucose T0/T180	0.33–1.23
BUN (mg/dL)	16–39	18–29	12–34		
Ca (mg/dL)	8.6–9.8	9.6–10.7	NA		
Cholesterol (mg/dL)	72–96	98–186	80–130	Wk 1 insulin (ng/mL)	0.112–3.19
CK (U/L)	24–270	47–354	NA	Wk 14 insulin (ng/mL)	0.0727–3.49
Cl (mEq/dL)	108–118	110–116	NA	Wk 1 leptin (ng/mL)	0.350–5.20
CO ₂ (mEq/L)	13–33	NA	NA	Wk 14 leptin (ng/mL)	0.415–17.20
Creatinine (mg/dL)	0.1–0.2	0.2–0.4	0.2–0.8		
Fasting glucose (mg/dL)	69–157	NA	81–165		
Glob (g/dL)	1.6–2.7	1.7–2.2	NA		
HDL (mg/dL)	47–113	NA	NA		
IPhos (mg/dL)	4.8–9.8	6.8–11.3	NA		
K (mEq/dL)	4.2–7.4	7.9–11.2	3.6–7.3		
LDL (mg/dL)	6–22	NA	NA		
Na (mEq/dL)	145–155	152–156	147–163		
NEFA (mEq/dL)	0.8–2.1	NA	NA		
SDH (U/L)	9.9–32.9	NA	18–57		
Total bile acids (μMol/L)	0.4–4.2	NA	NA		
Total protein (mg/dL)	4.2–5.3	4.8–5.8	4.0–6.0		
Triglycerides (mg/dL)	69–388	81–280	NA		

^aCD-1 reference ranges for male 18-week-old mice³⁸.

^bB6C3F1 reference ranges for 18- to 20-week-old animals in NIEHS 13-week subchronic studies³⁹.

^cHistorical reference ranges for alternative strains and stocks were not available for glucose, insulin, or leptin concentrations. Analytes were measured using conventional clinical chemistry procedures.

NA indicates that reference ranges were not available for comparison.

Abbreviations: ALP = alkaline phosphatase; ALT = alanine aminotransferase; AUC = area under the curve; BUN = blood urea nitrogen; Ca = calcium; CK = creatine kinase; Cl = chlorine; CO₂ = carbon dioxide; DO = Diversity Outbred; Glob = globulin; HDL = high density lipoprotein; IPhos = inorganic phosphate; K = potassium; LDL = low density lipoprotein; Na = sodium; NEFA = non-esterified fatty acid; SDH = sorbitol dehydrogenase; T0 – time zero minutes; T180 = time 180 minutes; Wk = study week.

Baseline Characteristics of Diversity Outbred (J:DO) Mice Relevant to Toxicology Studies

Table 4. Serum Insulin and Leptin at Terminal Sacrifice in Male DO Mice

Measure	Count	Mean	SD	SEM	CV	Median	Min	Max	Number needed to detect a 20% difference
Serum insulin	72	3143	8863	1052	282	664	72	63225	3,138
Serum leptin	72	5905	5042	598	81	4415	219	18557	288

SD = standard deviation, SEM = standard error of the mean, CV = coefficient of variation, Min = minimum value, Max = maximum value.

Table 5. Histopathology Findings in the Liver and Kidney of Male DO Mice

Liver Findings		Number Affected	Kidney Findings		Number Affected
Number examined		75	Number examined		74
Nothing adverse discovered		0	Nothing adverse discovered		4
Basophilic focus		1	Arterial inflammation		1
Hepatocyte fatty change		19	Cortical tubule dilation		10
Grade 1		14	Grade 2		1
Grade 2		4	Grade 1		2
Grade 3		1	Grade 2		7
Hepatocyte; glycogen accumulation		72	Grade 3		1
Grade 1		26	Cortical tubule hypertrophy		5
Grade 2		35	Grade 1		5
Grade 3		9	Cortical tubule mineralization		2
Grade 4		2	Grade 1		2
Hepatocyte single-cell necrosis		5	Cortical tubule regeneration		31
Grade 1		5	Grade 1		29
Chronic inflammation		33	Grade 2		2
Grade 1		33	Cortical tubule vacuolization		15
Megakaryocytosis		2	Grade 1		8
Grade 1		2	Grade 2		6
Periportal mononuclear cell infiltrate		9	Grade 3		1
Grade 1		9	Grade 4		1
			Chronic progressive nephropathy		1
			Grade 2		1
			Mononuclear cell infiltrate		52
			Grade 1		47
			Grade 2		3
			Grade 3		2
			Papilla mineralization		13
			Grade 1		11
			Grade 2		2

Grading was based on a five-point scale as follows: 1 – minimal, 2 – mild, 3 – moderate, 4 – marked, and 5 – severe. A kidney sample was missing from the archive, resulting in 74 kidney samples examined.

Baseline Characteristics of Diversity Outbred (J:DO) Mice Relevant to Toxicology Studies

Table 6. Available Comparator Data and Power Calculations for Common Endpoints in Mice from Five NTP Studies

Endpoint	DO Mouse	TR 587²²	TR 585²³	TR 583²⁴	TR 580²¹	TR 579²⁰
Body weight at start of study (g)	23.7 ± 2.8	22.6 ± 0.4	24.4 ± 0.3	22.3 ± 0.3	21.8 ± 0.4	23.1 ± 0.3
Body weight at end of study (g)	33.7 ± 5.0	37.4 ± 0.9	40.7 ± 1.0	39.3 ± 1.0	38.5 ± 1.3	33.6 ± 1.4
<i>Organ weight in grams (relative organ weight)</i>						
Liver (g)	1.4 ± 0.27	1.44 ± 0.01	1.79 ± 0.06	1.53 ± 0.07	1.79 ± 0.05	1.43 ± 0.07
(relative g/g × 100)	(4.1 ± 4.8)	(38.33 ± 0.74)	(49.91 ± 0.54)	(38.60 ± 1.04)	(46.37 ± 0.82)	(42.52 ± 0.81)
Right kidney (g)	0.22 ± 0.04	0.28 ± 0.00	0.31 ± 0.01	0.29 ± 0.01	0.30 ± 0.01	0.27 ± 0.01
	(6.3 ± 1.0)	(7.4 ± 0.15)	(7.53 ± 0.14)	(7.28 ± 0.07)	(7.92 ± 0.26)	(7.98 ± 0.24)
Right testes (g)	0.10 ± 0.02	0.11 ± 0.00	0.12 ± 0.01	0.12 ± 0.00	0.12 ± 0.00	0.12 ± 0.00
	(3.1 ± 0.55)	(3.2 ± 0.15)	(2.93 ± 0.009)	(3.14 ± 0.07)	(3.02 ± 0.13)	(3.47 ± 0.08)
Left epididymis (g)	0.049 ± 0.04	0.28 ± 0.00	0.31 ± 0.01	0.29 ± 0.01	0.30 ± 0.01	0.27 ± 0.01
	(1.5 ± 0.41)	(7.4 ± 0.15)	(7.53 ± 0.14)	(7.28 ± 0.07)	(7.92 ± 0.26)	(7.98 ± 0.24)
Sperm concentration (10 ⁶ /cauda epididymis)	0.22 ± 0.04	0.28 ± 0.00	0.31 ± 0.01	0.29 ± 0.01	0.30 ± 0.01	0.27 ± 0.01
	(6.3 ± 1.0)	(7.4 ± 0.15)	(7.53 ± 0.14)	(7.28 ± 0.07)	(7.92 ± 0.26)	(7.98 ± 0.24)

Baseline Characteristics of Diversity Outbred (J:DO) Mice Relevant to Toxicology Studies

Table 7. Power Calculations for Organ and Body Weights in Mice

Endpoint	Absolute difference between two groups that can be detected with 80% power at p = 0.05 with 10 B6C3F1 mice per group	Percent difference between two groups that can be detected with 80% power at p = 0.05 with 10 B6C3F1 mice per group	Number of DO mice to achieve the same power as n = 10 B6C3F1 mice
Body weight (g)	2.2	5.6	27
Liver weight (g)	0.11	7.2	33
Kidney weight (g)	0.02	7	22
Testis weight (g)	0.005	14.8	32
Epididymis weight (g)	0.009	4.1	16
Relative liver weight (%)	0.23	6	23
Relative kidney weight (%)	0.05	6.9	20
Relative testis weight (%)	0.02	12.9	29
Relative epididymis weight (%)	0.02	6.4	34
Sperm/cauda epididymis	3	16.3	150
% Sperm motility	2.8	4	32

Table 8. Power Calculations for Clinical Chemistry Endpoints in Rats

Serum Analyte	Absolute difference between two groups that can be detected with 80% power at p = 0.05 with 10 F344/N or F344/NTac rats per group	Percent difference between two groups that can be detected with 80% power at p = 0.05 with 10 F344/N or F344/NTac rats per group	Number of DO mice per group to achieve the same power as n = 10 F344/N and F344/NTac rats
Albumin	0.1	2.3	19
Alkaline phosphatase	11.9	6	12
Alanine aminotransferase	33.6	25.9	5
Total bile acids	6	25.7	6
Blood urea nitrogen	3.3	19.4	20
Creatine kinase	111	39.2	7
Creatinine	0.06	8.8	5
Total protein	1.7	5.9	— ^a
Sorbitol dehydrogenase	9.7	28.1	6

^aNumbers are too small for a reliable estimate.



National Toxicology Program

NTP Central Data Management, MD K2-05

National Institute of Environmental Health Sciences

P.O. Box 12233

Research Triangle Park, NC 27709

<http://ntp.niehs.nih.gov>

ISSN 2473-4756