ehp ENVIRONMENTAL HEALTH PERSPECTIVES
http://www.ehponline.org

An Integrated Experimental Design for the Assessment of Multiple Toxicological End Points in Rat Bioassays

Fabiana Manservisi, Clara Babot Marquillas, Annalisa Buscaroli, James Huff, Michelina Lauriola, Daniele Mandrioli, Marco Manservigi, Simona Panzacchi, Ellen K. Silbergeld, and Fiorella Belpoggi

http://dx.doi.org/10.1289/EHP419

Received: 12 October 2015
Revised: 27 April 2016
Accepted: 20 June 2016
Published: 22 July 2016

Note to readers with disabilities: *EHP* will provide a 508-conformant version of this article upon final publication. If you require a 508-conformant version before then, please contact ehp508@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

NIH National Institute of Environmental Health Sciences

Environ Health Perspect DOI: 10.1289/EHP419
Advance Publication: Not Copyedited

An Integrated Experimental Design for the Assessment of Multiple Toxicological End Points in Rat Bioassays

Fabiana Manservisi[1], Clara Babot Marquillas[4], Annalisa Buscaroli[1], James Huff[3], Michelina Lauriola[1], Daniele Mandrioli[1,2], Marco Manservigi[1], Simona Panzacchi[1], Ellen K. Silbergeld[2], and Fiorella Belpoggi[1]

[1]Cesare Maltoni Cancer Research Center, Ramazzini Institute, Bentivoglio 40010, Bologna, Italy; [2]Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD, USA 21205; [3]National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA 27709; [4]Visiting Researcher from University of Barcelona, Leonardo da Vinci Programme at the Cesare Maltoni Cancer Research Center, Ramazzini Institute, Bentivoglio 40010, Bologna, Italy

Address of institution: Cesare Maltoni Cancer Research Center, Ramazzini Institute

Via Saliceto 3, 40010 Bentivoglio, Bologna, Italy

Tel. +39 051 6640460 - Fax +39 051 6640223

e-mail: crcadmin@ramazzini.it - www.ramazzini.org

Address correspondence to:

Dr. Fiorella Belpoggi, Director, Cesare Maltoni Cancer Research Center, Ramazzini Institute

Via Saliceto, 3, 40010 Bentivoglio, Bologna, Italy

Tel. +39 051 6640460 - Fax +39 051 6640223

ED_002435_00009603-00002

e-mail: belpoggif@ramazzini.it - www.ramazzini.org

Running title: The Ramazzini Institute integrated experimental bioassay design

Acknowledgments

We thank Luana De Angelis, BS, and Laura Falcioni, DVM, PhD, of the Cesare Maltoni Cancer Research Center, Ramazzini Institute, for their suggestions and assistance in the development of this integrated model.

Competing Financial Interests Declaration: There were no funds received to support the writing or production of this paper. The authors declare no competing financial interests.

ED_002435_00009603-00003

Abstract

Background: For nearly five decades long-term studies in rodents have been the accepted benchmark for assessing chronic long-term toxic effects, particularly carcinogenicity, of chemicals. The European Food Safety Authority (EFSA) and the World Health Organization (WHO) have pointed out that the current set of internationally utilized test methods capture only some of the potential adverse effects associated with exposures to these agents over the lifetime.

Objectives: In this paper we propose the adaption of the carcinogenicity bioassay to integrate additional protocols for comprehensive long-term toxicity assessment that includes developmental exposures and long-term outcomes, capable of generating information on a broad spectrum of different endpoints.

Discussion: An integrated study design based on a stepwise process is described, that includes the priority endpoints of the OECD and NTP guidelines on carcinogenicity/toxicity and developmental/reproductive toxicity. Integrating a comprehensive set of relevant toxicological endpoints in a single protocol represents an efficient opportunity, reducing animal use in accordance with the 3Rs (replacement, reduction and refinement). This strategy has the potential to provide sufficient data on multiple windows of susceptibility of specific interest for risk assessments and public health decision making by including prenatal, lactational, neonatal exposures and evaluating outcomes over the lifespan.

Conclusion: This integrated study design is efficient in that the same generational cohort of rats used for evaluating long-term outcomes can be monitored in satellite parallel experiments to measure biomarkers and other parameters related to system-specific responses including metabolic alterations and endocrine disturbances.

ED_002435_00009603-00004

Introduction

Synthetic chemicals have been used individually and as mixtures in consumer products for over a century, gaining intense momentum beginning after World War II. Naturally occurring elements and compounds have been used for millennia. The first bioassays for identifying chemicals posing a greater and immediate danger for carcinogenicity to individuals were first developed about 100 years ago (Yamagiwa and Ichikawa, 1918). The chemical carcinogenesis revolution and testing age began when Yamagiwa and Ichikawa in 1918 showed that coal tar applied to rabbit ears caused skin carcinomas (Yamagiwa and Ichikawa, 1918). The real impetus for testing chemicals came with passage of legislation, first in the US in 1976 and then in several EU member states, requiring evaluation of industrial chemicals and especially those in the workplace and in consumer products. This led to the development of a multinational effort to harmonize testing methods through the Environment Programme of the Organization for Economic Co-operation and Development (OECD). Over the last 30 years, many test guidelines were developed within the OECD as well as the concepts for assessing risks of chemicals identified as harmful and carcinogenic in the workplace and environment (Hartung 2009; Huff 1992; Maltoni 1976; Soffritti et al. 2002; Tomatis 1979; Silbergeld et al. 2015). Rodent bioassays have been described in the OECD Test Guideline (TG) 453 (OECD 2009) and by the U.S. National Toxicology Program (NTP) (NTP 2011a), with specifications for design and conduct of studies to evaluate toxic and carcinogenic potential of chemical, biological and physical agents in laboratory animals. Recognizing that carcinogenesis is a multistep, multivariate process (Brash and Cairns 2009; Hanahan and Weinberg 2011) it may be unrealistic to expect a basic 2-year cancer study to provide all the complex data necessary for cancer risk identification, management and regulatory decisions. Current OECD Guidelines, as planned, are not aimed to monitor cancer hazards and risks of exposure on susceptible individuals such as children and elderly. For some test articles, NTP

ED_002435_00009603-00005

carcinogenicity 2 years protocol might include perinatal exposure, but these are selected only after considering patterns of human exposure (NTP 2011a; NTP 2016). Furthermore, traditional toxicity testing methods could not identify many of the endocrine-related adverse effects of some chemicals, especially subtle effects on specific developmental stages (Huff 1996; Huff et al. 1996; Manservisi et al. 2014; Melnick et al. 2002; Vandenberg et al. 2012; Bergman et al. 2015; UNEP 2012), as happened for Bisphenol A (Maffini et al. 2006; Vandenberg et al. 2009; vom Saal et al. 2007). Consistent with these considerations, both OECD and NTP have introduced new guidelines for reproductive/developmental toxicity with more functional endpoints to assess how agents affect the reproductive and endocrine status of animals (NTP 2011b; OECD 2011).

Study designs and outcomes investigated by current guidelines and our proposed protocol on carcinogenicity/toxicity and reproductive/developmental toxicity are summarized in Table 1. The OECD reference guideline for reproductive/developmental toxicity, OECD TG 443 (Extended One-Generation Reproductive Toxicity Study), provides an evaluation of reproductive and developmental effects that may occur in offspring as a result of pre- and postnatal chemical exposure as well as systemic toxicity in pregnant and lactating females (OECD 2011). In the OECD TG 443 protocol, sexually-mature male and female rodents (parental (P) generation) are exposed to graduated doses of test substances starting 2 weeks before mating and continuously through mating, gestation, lactation, and weaning of pups (F1 generation). At weaning, pups are assigned to three group for reproductive/developmental toxicity testing (cohort 1), developmental neurotoxicity testing (cohort 2), and developmental immunotoxicity testing (cohort 3). Other F1 offspring are exposed after weaning through adulthood. Clinical observations and pathology examinations are performed on all animals for signs of toxicity, with special emphasis on integrity and performance of male and female reproductive systems and health, growth, development and function of offspring. Part of cohort 1 (cohort 1B) may

ED_002435_00009603-00006

be extended to include an F2 generation; in this case, procedures for F1 animals are similar to those for the P animals. The total number of animals involved in this OECD protocol design is more than one thousand (OECD 2011).

The NTP reference guideline for reproductive/developmental toxicity, the NTP's Modified One-Generation Reproductive (MOG) (NTP 2011b), employs pregnant animals with exposures beginning at implantation with continued dosing of dams throughout gestation and lactation (Foster 2014). At weaning, offspring are administered the test substance at the same level as their respective dams and are assigned to different cohorts: a prechronic toxicity cohort (analogous to a standard 90-day study) for evaluating clinical pathology and target organ toxicity and pathology; a teratology cohort for evaluating prenatal development; another cohort to evaluate breeding and littering for potential examination of the subsequent generation. This study design involves exposure of pregnant females throughout gestation (the P generation), lifetime exposure of the F1 and generation of two cohorts of F2 animals (developmental and reproductive).

The OECD TG 443 and the NTP MOG were introduced only recently and there is still no published data comparing studies with the same substance performed according to the two guidelines. We cannot exclude the possibility that authorities such as the U.S. Environmental Protection Agency (EPA), the U.S. Food and Drug Administration (FDA), the European Chemical Agency (ECHA) and the European Food Safety Agency (EFSA) could require (or have already required) the repetition of the tests with both guidelines considering the need of empirical evidence supporting the use of either one of the two. It is our opinion that studying regularly the same substance with both the NTP MOG and OECD TG 443 represents an unnecessary repetition. The NTP's MOG is able to generate large and robust

ED_002435_00009603-00007

datasets, includes early life exposure and teratogenicity, but requires a larger number of animals than the OECD TG 443 (Schiffelers et al. 2015; Foster 2014).

Starting from the 1990's, the Cesare Maltoni Cancer Research Centre (CMCRC) of the Ramazzini Institute (RI) performed carcinogenicity studies on low doses of chemical or physical agents which may expose millions or even billions of people to potential carcinogenic risks, such as radiations and food additives (Maltoni et al. 1985; Maltoni et al. 1999; Soffritti et al. 1999; Soffritti et al. 2002; Soffritti et al. 2007; Soffritti et al. 2008), using an alternative model, more sensitive than the traditional combined toxicity/carcinogenicity two-year protocol adopted by OECD and NTP (Bucher 2002; Huff 1992; Huff and Moore 1984; Melnick et al. 2008). The CMCRC protocol includes: prolonged period of exposure/observation of experimental animals, starting exposures from the 12th day of fetal life (gestation) and continuing through lactation and weaning until at least 130 weeks or longer (Soffritti et al. 2002). In fact, human exposures to environmental agents, also at relatively low doses, most often starts prior to and during mother's gestation, continues through lactation (via breast milk) and lasts until death.In standard/typical bioassays exposure generally starts in young adulthood and last until about 2-years, which is roughly equivalent to only 65 years in humans (Maltoni et al. 1997; Haseman et al. 2001; Huff et al. 2008; Melnick et al. 2008). Group sizes in carcinogenicity studies should also be increased whenever required for sufficient statistical power and for avoiding the possibility of false negative response: bioassays involving 100 animals or more per sex per group might be necessary for identifying carcinogenic effects of low doses and weak carcinogenic activity (Maltoni 1981; McCormick 2013). More than 500 chemical-specific bioassays have been performed at CMCRC and results are used worldwide for hazard identification and human cancer risk assessments (NRC 2014a, b).

ED_002435_00009603-00008

To satisfy the need to consider multiple effects (e.g., cancer and non-cancer) across multiple life stages and to reduce the overall number of animals required for separate studies of these endpoints we propose the following experimental design that integrates traditional cancer guidelines with more recent proposals of OECD and NTP for studying reproductive and developmental toxicity. This new integrated experimental design aims to maximize the endpoints measured for each animal, thus reducing the overall number of animals produced/utilized, in accordance with the 3Rs (replacement, reduction and refinement) (EU Directives 2010).

The central aim of the proposed Integrated Long-Term Toxicity/Carcinogenicity Study methodology is to maximize the breadth of outcomes assessed and to increase the sensitivity of testing beyond that in commonly used protocols to give more reliable and inclusive information on many important endpoints (Figure 1).

Our proposal: an integrated experimental design

The integrated experimental design proposed by the CMCRC/RI is outlined in Figure 1 and more details on each specific section of the protocol are available as "supplemental material". The study design is largely based on OECD TG 453 (modified only for duration of the experiment), OECD TG 443, NTP Guidelines. The study comprises:

a) *Toxicity/Carcinogenicity study*: animals are treated from fetal life (dams, 12th day of pregnancy) until 104 weeks of age, then observed (with or without continuous exposure, depending on chemical) until 130 weeks of age (30 months). Interim kills are included to provide information on progression/regression of non-neoplastic/neoplastic changes and mechanistic information (e.g. gene expression, serum biomarkers of inflammation, cell proliferation, etc). Animals included for

ED_002435_00009603-00009

interim evaluation are also exposed from fetal life (dams, 12th day of pregnancy) until 26-52-78 and 104 weeks of age following OECD guidelines (OECD 2009);

b) Reproductive/Developmental Toxicity: different windows of susceptibility (WOS) related to reproductive/developmental and other non-cancer effects are studied. The possible adverse effects of the substances are studied in prenatal, neonatal, prepubertal, pubertal, adult parous and nulliparous WOS, and compared among them, or with the possible long-term carcinogenicity effect.

Animal model

The laboratory rat has served as the traditional animal model of choice for research and regulatory developmental and reproductive toxicity testing conducted to support human health hazard identification and risk assessment. The rat has been used extensively for developmental and reproductive physiology and endocrinology research, and has been more thoroughly characterized in these research fields than other species, likewise for identifying likely human carcinogens (Gray et al. 2004; Maltoni et al. 1999; Teitelbaum et al. 2014).

Our proposal to use Sprague-Dawley rats is based on the evidence that they are adequately sensitive, have a long historical base, and are also recommended by the OECD (OECD 2011) (OECD 2009) the NTP (King-Herbert and Thayer 2006; King-Herbert et al. 2010) and used by many other Universities and Organizations (Manservisi et al. 2014). SD rats are a known and accepted human equivalent model for cancer (Teitelbaum et al. 2014) (Soffritti et al. 2006). The proposed protocol uses Sprague Dawley rat strains that meet the requirement of the OECD 443 and 453 guidelines: "strains with low fecundity or a well-known high incidence of spontaneous developmental defects should not be used" (OECD

ED_002435_00009603-00010

2011) and "using a strain of animal that has a acceptable survival rate for the long-term study" (OECD 2009)

There are known limits for this animal model for individual cancer endpoints: for example Sprague Dawley rats represent an optimal model for breast cancer (Teitelbaum, 2014), while the high prevalence of benign tumors of the pituitary gland and pheocromocitoma of the subrenal gland make Sprague Dawley an inappropriate model for tumors of these organs (Dinse et al. 2010).

Numbers of animals

There is widespread agreement that a the relatively small numbers of animals used in most standard toxicity tests is a serious issue in terms of sensitivity and reliability. On the other hand, there are social and ethical concerns about the numbers of animal used in these tests (Hartung and Rovida 2009). Inadequate tests are a main driver of additional testing, such that it can be argued that utilizing robust methods, with increased numbers of animals per test, will reduce overall animal testing. Current guidelines recommend study designs which encompass at least three treatment groups plus control. For the OECD TG 453 carcinogenicity/chronic toxicity protocol the minimal number of animals is 480; for the OECD TG443 the minimal number is 1760 and for the NTP Modified One- Generation Reproduction Study 3200 animals (Table 1). But because only a limited number of endpoints are assessed in each of these tests, more animals are expected to be required to empower a broad-based toxicological evaluation for hazard and risk assessment. Performing these studies separately, as is current practice, would require up to 3680 animals (Table 1).

In our proposal, breeders (virgin males and females) of about 10-15 weeks of age are matched in a single outbred mating, in a number adequate to obtain sufficient animals for the study. The objective of breeding is to generate animals in order to have no more than 1 sister and 1 brother for each control and

ED_002435_00009603-00011

exposed group (2 sisters and 2 brothers in the carcinogenicity arm) in order to avoid any bias due to familial relationship.

Studying at least three exposure groups plus controls, the number for a comprehensive human equivalent hazard identification study is 1720 animals (Fig. 1 and Table 1). A higher number of exposed and control animals included in the studies better guarantees higher sensitivity of the model, sufficient statistical power and overall saving animals that would be sacrificed in unnecessary repetition of the studies or performing uninformative underpowered studies (Hooijmans et al. 2010).

In compliance with the 3R, we suggest, whenever possible, to avoid the use of culling and use all the pups generated during the experiment, avoiding unnecessary sacrifice of animals. It is our opinion that avoiding culling also would permit generally a more rigorous measure of litter mortality and simulate a human equivalent scenario, with more genetic variability and avoiding possible selection bias (for example selecting only healthy animals with higher birth weight). Nevertheless, the use of culling might be appropriate for studying suspect endocrine disruptor substances, as litter size can impact pups weights and rate of growth, affecting puberty timing. Puberty timing regulates other end points, so that the change in body weight from not equalizing litter size early on might have an inadvertent impact on the study.

Dose ranges

Under current testing procedures (Maronpot 2004), when toxicology studies are performed, relatively high doses of a chemical are given to animals, generally higher than the doses humans are exposed. However this is not always the case, especially for various workplaces/occupations and high dose drug and cancer chemo-therapies. Toxicity testing is typically carried out with maximum tolerated dose (MTD), previously determined in shorter-term exposures experiments of 28-90 days. Higher dose

ED_002435_00009603-00012

toxicology studies show that a chemical can be lethal (and need to be avoided), or block/disrupt pregnancies, or induce birth defects. These high doses effects may not always be observed at lower doses, which is why some assume that these are safe exposures, but there may be other endpoints affected, that can't be detected by typical methods of a standard bioassay (Teitelbaum et al. 2014), Non monotonic dose response curves reveal such unexpected effects, especially for EDCs (plasticizers, pesticides, and other industrial chemicals) as shown by several toxicological and epidemiologic studies on non-cancer endpoints that are relevant to metabolic disease (Thayer et al. 2005; Thayer et al. 2006; Vandenberg et al. 2012). In the multitude of chemicals that have never been tested adequately at low doses, but were already tested for carcinogenicity at high doses, we suggest testing doses in the range of actual highest human exposure, setting LOAEL (lowest-observed-adverse-effect level) from traditional toxicological studies as the highest dose, particularly in experiments designed to test endocrine-sensitive endpoints. For chemicals never tested for long-term carcinogenic effects, at least one high dose group near the MTD should be included, obviating the problem of unnecessary repetition of the bioassay if the low dose protocol is not carcinogenic.

Estimation of daily intake of a test substance depends on knowledge of the toxicokinetics, including route of administration, distribution, metabolism and excretion, which are not all readily available from the literature (Soeborg et al. 2014). If a range of doses is unavailable or fully unknown, we propose that a Dose-Range Finding (DRF) should be performed before starting the experimental protocol in order to determine an optimal exposure concentration for each chemical selected as close as possible to the estimated human exposure; in particular, when "novel food" or similar test compounds are studied, nutritional aspects and other relevant methodological aspects related to exposure should be studied (EFSA 2013a). When conducting exposure studies with low doses (many orders of magnitude lower than the NOAEL), a systematic dose-calibration study should be performed in an appropriate rodent

ED_002435_00009603-00013

model in order to identify the administered oral dose of the test substance that result in biomarker concentrations (e.g. urine, serum) comparable to the ones observed in human population. (Teitelbaum et al. 2015). Of course other higher doses must be chosen to adequately challenge biological systems and to provide some observable indication of toxicity, without jeopardizing the health, well-being, or body weights and survival of exposed animals, as well as being optimally sensitive to adequately evaluate the potential carcinogenicity (Bucher 2000; Huff 1999; Melnick et al. 2008). Higher doses also increase a priori statistical power to detect non-cancer effects using a relatively small number of animals, although remarkable exceptions exist particularly for endocrine effects (Vandenberg 2012).

Timing of exposure

Adult exposure to some chemicals is certainly an important factor in adverse health outcomes; however, increased focus on the fetus and/or neonate is of primary concern since developing organisms are extremely sensitive to perturbations by chemicals, especially those with hormone-like activity. Certain types of adverse effects may be more severe in developing organisms and occur at chemical concentrations that are in some instances below levels that would be considered harmful in adults (Tabb and Blumberg 2006). Few guidelines for testing environmental chemicals include prenatal or early life exposures, and thus often do not provide information on risks of carcinogens related to early life exposure (Rudel et al. 2011; Tabb and Blumberg 2006). Based on results of long-term carcinogenicity bioassays testing chemical and physical agents using rodents, there is ample evidence demonstrating that exposures during early developmental phases, produce an overall increase of malignant tumors and increases of specific organ site neoplasms related to exposures to specific carcinogens as in the case of vinyl chloride and benzene (Maltoni et al. 1989; Maltoni et al. 1981; Huff et al. 2008; Soffritti et al. 2008). Early exposure to chemicals is particularly important in study designs

ED_002435_00009603-00014

if there is reason to believe human exposures begin in utero and that susceptibility may be greater during growth and early developmental stages (Rice et al. 1989).

For a clear understanding of this protocol it should be considered that 16 weeks of age in adult rats roughly correspond to 10 years in human (Sengupta 2013). In our proposal, animals belonging to the toxicity/carcinogenicity arm are observed until 130 weeks of age (corresponding to about 75-80 years of age in humans), starting exposure during fetal life (dams, 12th day of pregnancy), whereas OECD guidelines stipulate that animals should be killed and examined at 104 weeks of treatment (corresponding to about 60-65 years of age in humans)(Huff et al 2008). Interim kills are also planned following the OECD TG 453, to provide information on progression/regression of non-neoplastic events and neoplastic changes and mechanistic information.

The Reproductive/Developmental Toxicity arm mimics human exposure during critical windows of development, and include (Fig. 1): a) prenatal (F1): animals treated during embryonic life and sacrificed at postnatal day (PND 21; b) postnatal (F1): animals treated through lactation, starting from birth (PND 1) and sacrificed at PND 21; c) pre-pubertal (F1): animals treated from PND 21 to PND 42; d) pubertal (F1): animals treated from PND 42 to PND 63; e) adult parous and nulliparous (F1): female animals treated from PND 1 through lactation, until PND 181. Once adult at 10-15 weeks, parous group rats are mated (outbred) and chemical treatment continues through pregnancy, delivery of pups (F2) and lactation. At the time of sacrifice of parous rats on PND 181, F2 pups had completed weaning.

In order to verify or elucidate effects in second generation, F2 offspring generated from F1 adult parous female rats are examined and sacrificed on PND 28.

ED_002435_00009603-00015

During necropsy, frozen target tissues (including blood) and organs, together with paraffin-embedded tissues, are stored for histopathology and molecular biology studies, EDCs effects, neurotoxicity, biochemical/biohaematological changes (metabolism), and for toxic and preneoplastic lesions.

Additional endpoints and adverse effects of the test compound

The aim of our integrated experimental design is to investigate all or a majority of possible health effects related to exposure to the studied agent/compound, and minimizing the unnecessary use of experimental animals. This also avoids wasted time when doing sequential endpoint studies. Endpoints assessed in traditional toxicology/carcinogenicity testing are feed and water consumption, chemical exposure, weight loss/gain, clinical pathology, survival/mortality, changes in organ weight, preneoplastic and neoplastic diseases with histopathological analyses. However, many examined chemicals have shown to also cause complex effects in animals, affecting organ development, functional and behavioral changes (Vandenberg et al. 2012).To best evaluate these fundamental endpoints, we included in our protocol design several of the NTP MOG and OECD TG 443 endpoints for immunotoxicity, neurotoxicity and developmental/reproductive toxicity." It should be noted that this protocol is easily scalable (for example additional groups can be added if appropriate, or specific arms can be amended if previously investigated) and simple changes are feasible and would permit to target specific endpoints or tissues (for example sperm aneuploidy) that are not described in this proposal.

Discussion

In our proposed lifetime experimental design, we assess a range of adverse outcomes of interest using a relatively large population of animals (sufficient power), born at the same time after mating of outbred breeders, randomized and studied for dose related effects, with the lowest possible risk of bias

ED_002435_00009603-00016

(blinding of assessors of outcomes, randomization, blinded assessment of pathological lesions by a minimum of two assessors). Typically, for studying all the previous-mentioned parameters (WOS, fertility, development, toxicity, carcinogenicity), approximately 10-20 studies are performed, using more animals, in different laboratories, with different procedures. Our experimental model and design overcomes these deficiencies and permits more information to be gathered on toxic/mechanistic/biological parameters, using the same but fewer overall animals in a large but unique experiment. In fact, in our experimental design, rats from the same generation are used for studying chronic toxicity/carcinogenicity outcomes and distributed in satellite parallel experiments (WOS), thus minimizing variables between different arms of the multi-endpoint investigation, for detecting also reproductive/developmental toxicity.

Our integrated experimental protocol requires 1720 animals, with a reduction up to 53 percent in animal use as compared to using separate test protocols (Table 1), representing an opportunity for investigating multiple toxicological endpoints at once, sparing animal lives in accordance with the 3Rs (replacement, reduction and refinement). We also expect an important reduction in terms of time, because the realization of a single integrated experiment would take shorter time for design, approval, performance and analysis if compared with multiple and sequential ones, with benefits in terms of costs and rapid availability of data for risk assessment.

The protocol we suggest addresses several important issues in the application of toxicological research to human health risk assessment: including information on different toxicological outcomes of exposures and health hazards of importance to human populations that are currently not completely covered by standard test protocols; earlier initiation and longer duration of exposure and observation of animals (130 weeks of age instead of 110) for a more comprehensive analysis of potential effects of

ED_002435_00009603-00017

chemical exposures and outcome assessment; enabling interim analyses and other strategies to examine specific outcomes over the lifespan. For increased efficiency, results of these tests can be shared among laboratories. Ideally the *in vivo* biophase should be responsibility of one laboratory in order to favor consistency of methods and quality of long-term animal studies (Gift et al. 2013). After the biophase, various endpoints, parameters, findings, and information on each category might be evaluated by different topic-expert scientists/laboratories. Frozen tissues samples from target organs are stored in order to study mechanistic aspects of the toxic process. Other relevant evidence, including cellular and molecular analyses related to mechanisms can be included in experimental designs, as has been proposed for forthcoming OECD or NTP integrated guidelines regarding long term *in vivo* studies (Darzynkiewicz et al. 2011; Kissling et al. 2007; Recio et al. 2010; Witt et al. 2008).

Conclusions

This protocol represents a proposal to regulatory scientists and the scientific community in general.

Compared to other OECD and NTP guidelines, this protocol has the unique feature of integrating carcinogenicity, toxicity and reproductive/developmental toxicity endpoints in a single protocol, with animals of the same generation, exploring windows of susceptibility that are currently not addressed in the other guidelines design. The design and protocol discussed here requires validation in order to demonstrate that the combined test is feasible and is at least as good as the separate tests (OECD 2005). Experience in the application of this proposal will be required in order to reach the same level of confidence that has been achieved for the standard carcinogenicity bioassays (Huff 2010). A priori establishment of criteria and consensus on relevant endpoints of interest is also a good starting point for evidence-based evaluations and following systematic review of obtained results (Birnbaum et al. 2013; Mandrioli and Silbergeld 2016). This is clearly needed for example for testing endocrine active

ED_002435_00009603-00018

substances with multiple endpoints as wells as modes and mechanisms of action, as the most reliably predictive animal model has yet to be identified. With this protocol, we aim to produce robust datasets that could also support the validation and discrimination of consensus criteria for evaluating non-cancer outcomes, such as "endocrine disruption".

We propose that conducting such integrated bioassays could enhance and expand scientific evidence available for risk assessments, gathering sufficient and rapid information on several adverse effects in a unique protocol for protecting public health (Claire 2012).

ED_002435_00009603-00019

References

Bergman Å, Becher G, Blumberg B, Bjerregaard P, Bornman R, Brandt I, Casey SC, Frouin H, Giudice LC, Heindel JJ, Iguchi T, Jobling S, Kidd KA, Kortenkamp A, Lind PM, Muir D, Ochieng R, Ropstad E, Ross PS, Skakkebaek NE, Toppari J, Vandenberg LN, Woodruff TJ, Zoeller RT. 2015. Manufacturing doubt about endocrine disrupter science - A rebuttal of industry-sponsored critical comments on the UNEP/WHO report "State of the Science of Endocrine Disrupting Chemicals 2012". Regul Toxicol Pharmacol 2015 Jul 31.

Birnbaum LS. 2013. State of the science of endocrine disruptors. Environ Health Perspect 121:a107-a107.

Birnbaum LS, Thayer KA, Bucher JR, Wolfe MS. 2013. Implementing systematic review at the national toxicology program: Status and next steps. Environ Health Perspect 121:A108-109.

Brash D, Cairns J. 2009. The mysterious steps in carcinogenesis. British journal of cancer 101:379-380.

Bucher JR. 2000. Doses in rodent cancer studies: Sorting fact from fiction. Drug Metabolism Reviews 32:153-163.

Bucher JR. 2002. The national toxicology program rodent bioassay: Designs, interpretations, and scientific contributions. Annals of the New York Academy of Sciences 982:198-207.

Chhabra RS, Huff JE, Schwetz BS, Selkirk J. 1990. An overview of prechronic and chronic toxicity/carcinogenicity experimental study designs and criteria used by the national toxicology program. Environ Health Perspect 86:313-321.

Claire R. 2012. Time to bring risk assessment into the real world. Research Europe 01/2012.

Dinse GE, Peddada SD, Harris SF, Elmore SA. 2010. Comparison of NTP historical control tumor incidence rates in female Harlan Sprague Dawley and Fischer 344/N Rats. Toxicol Pathol. 2010 Aug;38(5):765-75.

EFSA. 2013a. Scientific report of EFSA: Considerations on the applicability of OECD TG 453 to whole food/feed testing. European Food Safety Authority. EFSA Journal 2013;11(7):3347.

EU Directives. 2010. Diective 2010/63/eu of the European Parliament and of the council of 22 september 2010 on the protection of animals used for scientific purposes. Official Journal of the European Union 276/33.

ED_002435_00009603-00020

Fawcett TW, Frankenhuis WE. 2015. Adaptive explanations for sensitive windows in development. Front Zool. 2015 Aug 24;12 Suppl 1:S3.

Foster PM. 2014. Regulatory forum opinion piece: New testing paradigms for reproductive and developmental toxicity--the ntp modified one generation study and OECD 443. Toxicologic pathology 42:1165-1167.

Gray LE, Jr., Wilson V, Noriega N, Lambright C, Furr J, Stoker TE, et al. 2004. Use of the laboratory rat as a model in endocrine disruptor screening and testing. ILAR journal / National Research Council, Institute of Laboratory Animal Resources 45:425-437.

Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: The next generation. Cell 144:646-674.

Hartung T. 2009. Toxicology for the twenty-first century. Nature 460:208-212.

Hartung T, Rovida C. 2009. Chemical regulators have overreached. Nature 460:1080-1081.

Haseman J, Melnick R, Tomatis L, Huff J. 2001. Carcinogenesis bioassays: Study duration and biological relevance. Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association 39:739-744.

Hooijmans CR, Leenaars M, Ritskes-Hoitinga M. 2010. A gold standard publication checklist to improve the quality of animal studies, to fully integrate the three rs, and to make systematic reviews more feasible. ATLA Alternatives to Laboratory Animals 38:167-182.

Huff J. 1992. Design strategies, results and evaluations of long-term chemical carcinogenesis studies. Scandinavian journal of work, environment & health 18 Suppl 1:31-37.

Huff J. 1996. Chemically induced cancers in hormonal organs of laboratory animals and of humans. Progress in clinical and biological research 394:77-102.

Huff J, Boyd J, Barrett JC. 1996. Hormonal carcinogenesis and environmental influences: Background and overview. Progress in clinical and biological research 394:3-23.

Huff J. 1999. Animal and human carcinogens. Environ Health Perspect 107:A341-342.

Huff J, Jacobson MF, Davis DL. 2008. The limits of two-year bioassay exposure regimens for identifying chemical carcinogens. Environ Health Perspect 116:1439-1442.

Huff J. 2010. Predicting chemicals causing cancer in animals as human carcinogens. Occupational and environmental medicine 67:720.

King-Herbert A, Thayer K. 2006. Ntp workshop: Animal models for the ntp rodent cancer bioassay: Stocks and strains--should we switch? Toxicologic pathology 34:802-805.

ED_002435_00009603-00021

King-Herbert AP, Sills RC, Bucher JR. 2010. Commentary: Update on animal models for ntp studies. Toxicologic pathology 38:180-181.

Maffini MV, Rubin BS, Sonnenschein C, Soto AM. 2006. Endocrine disruptors and reproductive health: The case of bisphenol-a. Molecular and cellular endocrinology 254-255:179-186.

Maltoni C. 1976. Occupational carcinogenesis. Predictive value of carcinogenesis bioassays. Annals of the New York Academy of Sciences 271:431-443.

Maltoni C, Lefemine G, Ciliberti A, Cotti G, Carretti D. 1981. Carcinogenicity bioassays of vinyl chloride monomer: a model of risk assessment on an experimental basis. Environ Health Perspect. Oct;41:3-29.

Maltoni C, Conti B, Cotti G, Belpoggi F. 1985. Experimental studies on benzene carcinogenicity at the bologna institute of oncology: Current results and ongoing research. American journal of industrial medicine 7:415-446.

Maltoni C, Ciliberti A, Cotti G, Conti B, Belpoggi F. 1989. Benzene, an experimental multipotential carcinogen: Results of the long-term bioassays performed at the bologna institute of oncology. Environ Health Perspect 82:109-124.

Maltoni C, Ciliberti A, Pinto C, Soffritti M, Belpoggi F, Menarini L. 1997. Results of long-term experimental carcinogenicity studies of the effects of gasoline, correlated fuels, and major gasoline aromatics on rats. Ann N Y Acad Sci. 1997 Dec 26;837:15-52.

Maltoni C, Soffritti M, Belpoggi F. 1999. The scientific and methodological bases of experimental studies for detecting and quantifying carcinogenic risks. Annals of the New York Academy of Sciences 895:10-26.

Mandrioli D, Silbergeld EK. 2016. Evidence from toxicology: The most essential science for prevention. Environmental health perspectives. Environ Health Perspect Jan;124(1):6-11.

Manservisi F, Gopalakrishnan K, Tibaldi E, Hysi A, Iezzi M, Lambertini L, et al. 2014. Effect of maternal exposure to endocrine disrupting chemicals on reproduction and mammary gland development in female sprague-dawley rats. Reprod Toxicol. 2015 Jul;54:110-9.

Maronpot RR, Flake G, Huff J. 2004. Relevance of animal carcinogenesis findings to human cancer predictions and prevention.Toxicol Pathol. Mar-Apr;32 Suppl 1:40-8.

McCormick DL. 2013. Preclinical Evaluation of Carcinogenicity using the Rodent Two-Year Bioassay. In: A Comprehensive Guide to Toxicology in Preclinical Drug Development, Chapter 17, Elsevier 423–436.

ED_002435_00009603-00022

Melnick R, Lucier G, Wolfe M, Hall R, Stancel G, Prins G, et al. 2002. Summary of the national toxicology program's report of the endocrine disruptors low-dose peer review. Environ Health Perspect 110:427-431.

Melnick RL. 1999. Introduction--workshop on characterizing the effects of endocrine disruptors on human health at environmental exposure levels. Environ Health Perspect 107 Suppl 4:603-604.

Melnick RL, Thayer KA, Bucher JR. 2008. Conflicting views on chemical carcinogenesis arising from the design and evaluation of rodent carcinogenicity studies. Environ Health Perspect 116:130-135.

NRC. 2014a. Review of the styrene assessment in the national toxicology program 12th report on carcinogens. Washington, DC:The National Academies Press.

NRC. 2014b. Review of the formaldehyde assessment in the national toxicology program 12th report on carcinogens. Washington, DC:The National Academies Press.

NTP. 2011a. Specifications for the conduct of studies to evaluate the toxic and carcinogenic potential of chemical, biological and physical agents in laboratory animals for the National Toxicology Program (NTP). Research Triangle Park NN, ed.

NTP. 2011b. NTP's modified one-generation reproduction study. Available: https://ntp.niehs.nih.gov/testing/types/mog/index.html [accessed 10 February 2016].

NTP. 2016. Toxicology/Carcinogenicity. Available: http://ntp.niehs.nih.gov/testing/types/cartox/index.html [accessed 10 February 2016]

OECD. 2005. Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. OECD Series On Testing And Assessment 34, OECD Publishing.

OECD. 2009. Test No. 453: Combined Chronic Toxicity/Carcinogenicity Studies. OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing.

OECD. 2011. Test No. 443: Extended One-Generation Reproductive Toxicity Study. OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing.

Rice JM, Rehm S, Donovan PJ, Perantoni AO. 1989. Comparative transplacental carcinogenesis by directly acting and metabolism-dependent alkylating agents in rodents and nonhuman primates. IARC scientific publications:17-34.

Rietjens IM, Alink GM. 2006. Future of toxicology--low-dose toxicology and risk--benefit analysis. Chemical research in toxicology 19:977-981.

ED_002435_00009603-00023

Rudel RA, Fenton SE, Ackerman JM, Euling SY, Makris SL. 2011. Environmental exposures and mammary gland development: State of the science, public health implications, and research recommendations. Environ Health Perspect 119:1053-1061.

Schiffelers MJ, Blaauboer BJ, Bakker WE, Hendriksen CF, Krul C. 2015. Regulatory acceptance and use of the Extended One Generation Reproductive Toxicity Study within Europe. Regul Toxicol Pharmacol. 2015 Feb;71(1):114-24. Sengupta P. 2013. The laboratory rat: Relating its age with human's. International journal of preventive medicine 4:624-630.

Silbergeld EK, Mandrioli D, Cranor CF. 2015. Regulating chemicals: Law, science, and the unbearable burdens of regulation. Annu Rev Public Health 36:175-191.

Soeborg T, Frederiksen H, Andersson AM. 2014. Considerations for estimating daily intake values of nonpersistent environmental endocrine disruptors based on urinary biomonitoring data. Reproduction 147:455-463.

Soffritti M, Belpoggi F, Minardi F, Bua L, Maltoni C. 1999. Mega-experiments to identify and assess diffuse carcinogenic risks. Annals of the New York Academy of Sciences 895:34-55.

Soffritti M, Belpoggi F, Minardi F, Maltoni C. 2002. Ramazzini foundation cancer program: History and major projects, life-span carcinogenicity bioassay design, chemicals studied, and results. Annals of the New York Academy of Sciences 982:26-45.

Soffritti M, Belpoggi F, Degli Esposti D. 2006. Cancer prevention: the lesson from the lab. In: Cancer Medicine at the Dawn of the 21st Century: The View from Bologna (Biasco G, Tanneberger S, eds). Bologna, Italy:Bononia University Press, 49–64. Available: http://www.ramazzini.org/wp-content/uploads/2008/03/Cancer-Prevention_the-lesson-from-the-lab_2006.pdf [accessed 18 April 2016].

Soffritti M, Belpoggi F, Tibaldi E, Esposti DD, Lauriola M. 2007. Life-span exposure to low doses of aspartame beginning during prenatal life increases cancer effects in rats. Environ Health Perspect 115:1293-1297.

Soffritti M, Belpoggi F, Esposti DD, Falcioni L, Bua L. 2008. Consequences of exposure to carcinogens beginning during developmental life. Basic & clinical pharmacology & toxicology 102:118-124.

Sontag JM, Page NP, Saffiotti U. 1976. Guidelines for carcinogen bioassay in small rodents. Bethesda MD.

ED_002435_00009603-00024

Tabb MM, Blumberg B. 2006. New modes of action for endocrine-disrupting chemicals. Molecular endocrinology 20:475-482.

Teitelbaum SL, Belpoggi F, Reinlib L. 2015. Advancing research on endocrine disrupting chemicals in breast cancer: Expert panel recommendations. Reprod Toxicol. 2015 Jul;54:141-7.

Teitelbaum SL LQ, Lambertini L, Belpoggi F, Manservisi F, Falcioni L, Bua L, Silva MJ, Ye X, Calafat A, Chen J. 2016. Paired serum and urine concentrations of biomarkers of diethyl phthalate, methyl paraben and triclosan in rats. Environ Health Perspect Jan;124(1):39-45.

Thayer KA, Melnick R, Burns K, Davis D, Huff J. 2005. Fundamental flaws of hormesis for public health decisions. Environ Health Perspect 113:1271-1276.

Thayer KA, Melnick R, Huff J, Burns K, Davis D. 2006. Hormesis: A new religion? Environ Health Perspect 114:A632-633.

Thayer KA, Foster PM. 2007. Workgroup report: National toxicology program workshop on hormonally induced reproductive tumors - relevance of rodent bioassays. Environ Health Perspect 115:1351-1356.

Tomatis L. 1979. The predictive value of rodent carcinogenicity tests in the evaluation of human risks. Annual review of pharmacology and toxicology 19:511-530.

Tomatis L, Aitio A, Wilbourn J, Shuker L. 1989. Human carcinogens so far identified. Japanese journal of cancer research : Gann 80:795-807.

Vandenberg LN, Maffini MV, Sonnenschein C, Rubin BS, Soto AM. 2009. Bisphenol-a and the great divide: A review of controversies in the field of endocrine disruption. Endocrine reviews 30:75-95.

Vandenberg LN, Colborn T, Hayes TB, Heindel JJ, Jacobs DR, Jr., Lee DH, et al. 2012. Hormones and endocrine-disrupting chemicals: Low-dose effects and nonmonotonic dose responses. Endocrine reviews 33:378-455.

vom Saal FS, Akingbemi BT, Belcher SM, Birnbaum LS, Crain DA, Eriksen M, et al. 2007. Chapel hill bisphenol a expert panel consensus statement: Integration of mechanisms, effects in animals and potential to impact human health at current levels of exposure. Reproductive toxicology 24:131-138.

Weisburger EK. 1983. History of the bioassay program of the national cancer institute. Progress in experimental tumor research 26:187-201.

ED_002435_00009603-00025

WHO. 2012. State of the Science of endocrine disrupting chemical - 2012: An assessment of the state of the science of endocrine disruptors prepared by a group of experts far the United Nations Environment Programme (UNEP) and WHO. Available: http://unep.org/pdf/9789241505031_eng.pdf [accessed 10th of April 2016]

Yamagiwa K, Ichikawa K. 1918. Experimental Study of the Pathogenesis of Carcinoma. J Cancer Res January Jan 1;3(1):1-29.

ED_002435_00009603-00026

Table 1. Comparison between existing NTP MOG and OECD guidelines and the RI proposed study design.

Source material	Number of animals	Issue												
		WOS/Cohort	**Start of treatment**	**End of treatment (weeks)**	**Age at necropsy (weeks)**	**Generation**	**DRF**	**Chronic toxicity/ carcinogenicity**	**Sub-chronic toxicity**	**Reprod/ Develop**	**Neuro-toxicity**	**Neuro-behavioral**	**Immuno-toxicity**	**Teratology**
OECD TG 453	480	Chronic toxicity/ carcinogenicity	6-8 weeks	104	108	F1	X	X	X	–	–	–	–	–
OECD TG 443	1760*	Reproductive (1A)	2 weeks PB	13	13	F0, F1	X	–	–	X	X	X	X	–
		Reproductive (1B)	2 weeks PB	14 or 20-25 if triggered*	14 or 20-25 if triggered*	F0, F1, F2 if triggered*								
		Neuro-behavioral (2A)	2 weeks PB	11-12	11-12	F0, F1								
		Neuro-toxicity (2B)	2 weeks PB	3	3	F0, F1								
		Immuno-toxicity (3)	2 weeks PB	8	8	F0, F1								
NTP MOG	3200*	Reproduction	GD 6	22	22	F0, F1, F2	X	–	X	X	X	X	X	X
		Prenatal/Teratology	GD 6	18	18	F1, F2								
		13-week	GD 6	18	18	F1								
		Developmental/ Neurotoxicity	GD 6	11	11	F1								
		Developmental/ Immunotoxicity	GD 6	8	8	F1								
Total Animals: 2240 - 3680 (OECD 453+OECD 443 or NTP MOG)														
RI	1720*	Chronic toxicity/ carcinogenicity	GD 12	104	130 (final sacrifice)	F1	X	X	X	X	X	X	X	–
		Prenatal	Mating	birth	3	F0, F1								
		Postnatal	PND 1	3	3	F0, F1								
		Pre-Pubertal	3 weeks	6	6	F0, F1								
		Pubertal	6 weeks	9	9	F1								
		Adulthood	PND 1	26	26	F0, F1, F2								

*Considering 15pups/litter in F2 generation (OECD TG 443 generates F2 only if triggered, while NTP MOG and RI include F2 generation by default)
Key: DRF=Dose range finding; F0=parental animals; F1=litters generated by F0 animals; F2= litters generated by F1 animals GD=gestation day, PB=pre-breed; PND=Post natal day

ED_002435_00009603-00027

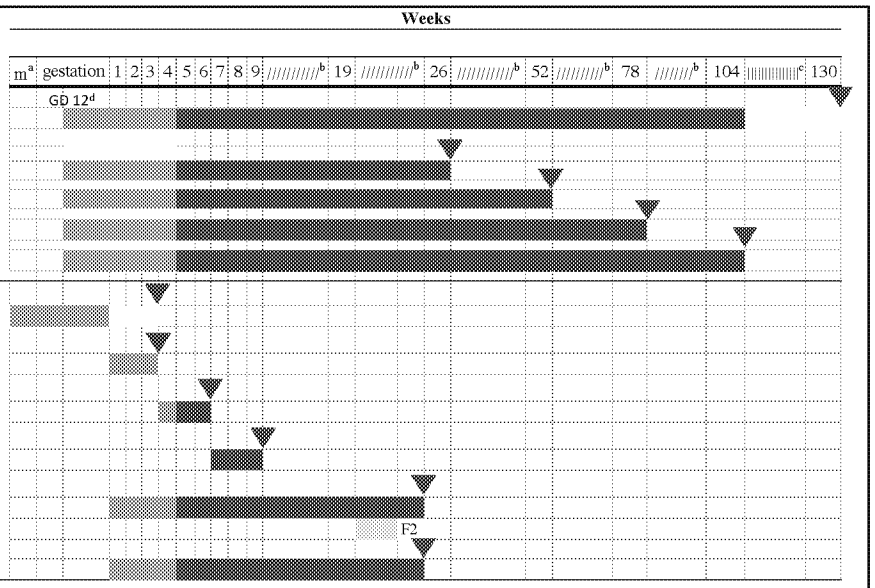
FIGURE LEGENDS

Figure 1. Integrated Long-Term Toxicity/Carcinogenicity Study experimental design. Schedule for treatment/duration for each group.

ED_002435_00009603-00028

Figure 1.

Arm	WOS/Cohort	Number of animals M	F	M+F
Chronic toxicity/ Carcinogenicity	Carcinogenicity	50	50	100
	Chronic toxicity (Interim kills)	10	10	20
		10	10	20
		10	10	20
		10	10	20
Reproductive/ Developmental Toxicity	Prenatal	10	10	20
	Neonatal	10	10	20
	Prepubertal	10	10	20
	Pubertal	10	10	20
	Adult Parous	-	10	10
	Adult Nullip	-	10	10
TOTAL F1 animals/group		**130**	**150**	**280**
TOTAL F2 animals/group		**10 litters of 15 pups**		**150**
TOTAL animals/group*				**430**



[a]: m= mating

[b]: //// = continuous treatment

[c]: IIIII = no treatment (period without dosing)

[d]: GD = Gestation Day

*: Studying at least three exposure groups plus controls, the number for a comprehensive human equivalent hazard identification study is 1720 animals

= White bars represent non dosing period

= Orange bars represent period of Parental (F0) exposure during in life phase (biophase)

= Red bars represent period of F1 exposure during in life phase (biophase)

= Yellow bars represent period of F2 exposure during in life phase (biophase)

= age (weeks) at sacrifice