Variability of animal studies for acute toxicity, skin sensitization, and mechanistic responses

Nicole C. Kleinstreuer, PhD
NICEATM

December 17, 2019
Sloppy reporting on animal studies proves hard to change
Scientists appear to ignore guidelines adopted 7 years ago

BY MARTINA EINHORN

nearly read any paper on an animal experiment, and you’re likely to find many questions: What strain of mice was used, and what were their sex and age? Were animals randomly assigned to control and treatment groups? Was the researcher who examined outcomes blinded to what group they were in? The absence of such details partly explains why between 5% and 40% of animal studies aren’t reproducible. It may also help explain why so many treatments reported to work in animals have flopped in humans. (Science, 22 November 2013, p. 1602.) Yet it’s perplexing surprisingly hard to solve the problem.

In 2009, the UK’s National Center for the Replacement, Refinement & Reduction of Animals in Research (NC3RR) in London developed a checklist of items that any paper about its research ought to include. More than 1000 scientific journals and two dozen fundraising agencies have endorsed the so-called ARRIVE guidelines—short for Animal Research: Reporting of In Vivo Experiments. (Science has not officially endorsed them, but encourages authors to comply.) But 7 years later, studies suggest that many scientists are either unaware of the guidelines or are ignoring them.

“We just don’t seem to make much progress,” says Merle Kruger-Eckert of Radboud University Medical Center in Nijmegen, the Netherlands, who co-organized a 26-Sept- ember meeting in Philadelphia where scientists met with journal editors and funders such as the United Kingdom Medical Research Council and the Wellcome Trust to discuss ways of speeding up implementation of the guidelines. One problem may be that existing compliance can take a lot of work, both for authors and journals. The 28 items in the checklist provide a “gold standard,” says Malcolm Macleod, a neurologist at the University of Edinburgh who has studied the problems in animal experimentation. The list covers a wide range of issues, from a paper’s title and study design to how the animals were cared for, results, and conflicts of interest. But a 2016 survey showed almost no improvement in reporting in journals of Nature Publishing Group (NPG) and PLOS during the first 3 years after the guidelines were introduced, even though both publishers had endorsed ARRIVE. That study’s lead author Linda Sones of VU University Medical Center in Amsterdam, says that an expert-developed analysis shows that things weren’t much better in the 2013-16 period.

Macleod and colleagues have tested one...
Reproducibility of Animal Data: Hazard

Binary Hazard Classification

- Uterotrophic: ~74%
- Hershberger: ~72%
- Skin Sensitization: ~78%
- Acute Systemic: ~81%
- Skin Irritation: ~76%
- Eye Irritation: ~84%

Kleinstreuer et al. 2016; Browne et al. 2018; Kleinstreuer et al. 2018a; Dumont et al. 2016; Hoffmann et al. 2018; Kleinstreuer et al. 2018b; Karmaus et al. in prep; Leuchtefeld et al. 2018
Conditional probability given a previous test result: Eye Irritation

491 substances with at least two Draize studies and extractable eye irritation category in REACH registrations 2008-2014

<table>
<thead>
<tr>
<th>Prior Type</th>
<th>1</th>
<th>2A</th>
<th>2B</th>
<th>Non</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>73%</td>
<td>16.1%</td>
<td>0.4%</td>
<td>10.4%</td>
<td>46</td>
</tr>
<tr>
<td>2A</td>
<td>4.2%</td>
<td>32.9%</td>
<td>3.5%</td>
<td>59.4%</td>
<td>138</td>
</tr>
<tr>
<td>2B</td>
<td>0.2%</td>
<td>4%</td>
<td>15.5%</td>
<td>80.2%</td>
<td>86</td>
</tr>
<tr>
<td>Non</td>
<td>1.1%</td>
<td>3.5%</td>
<td>1.5%</td>
<td>93.9%</td>
<td>400</td>
</tr>
<tr>
<td>Laboratory No.</td>
<td>A</td>
<td>F</td>
<td>G</td>
<td>J</td>
<td>K</td>
</tr>
<tr>
<td>---------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>14</td>
<td>38-58</td>
<td>58-60</td>
<td>26-60</td>
<td>56-60</td>
<td>18-62</td>
</tr>
<tr>
<td>1</td>
<td>23-59</td>
<td>35-57</td>
<td>55-86</td>
<td>23-48</td>
<td>10-37</td>
</tr>
<tr>
<td>10</td>
<td>9-33</td>
<td>40-57</td>
<td>39-59</td>
<td>26-55</td>
<td>8-42</td>
</tr>
<tr>
<td>2</td>
<td>4-100</td>
<td>25-54</td>
<td>0-23</td>
<td>4-57</td>
<td>4-49</td>
</tr>
<tr>
<td>9</td>
<td>20-37</td>
<td>26-34</td>
<td>12-61</td>
<td>0-43</td>
<td>10-37</td>
</tr>
<tr>
<td>21</td>
<td>9-43</td>
<td>74-80</td>
<td>22-56</td>
<td>49-63</td>
<td>10-12</td>
</tr>
<tr>
<td>16</td>
<td>15-25</td>
<td>8-18</td>
<td>8-110</td>
<td>17-23</td>
<td>18-30</td>
</tr>
<tr>
<td>24</td>
<td>13-21</td>
<td>8-37</td>
<td>35-53</td>
<td>8-37</td>
<td>6-23</td>
</tr>
<tr>
<td>7</td>
<td>2-36</td>
<td>42-50</td>
<td>40-72</td>
<td>11-26</td>
<td>11-30</td>
</tr>
<tr>
<td>13</td>
<td>0-37</td>
<td>2-39</td>
<td>0-57</td>
<td>0-30</td>
<td>0-14</td>
</tr>
<tr>
<td>18</td>
<td>0-28</td>
<td>16-21</td>
<td>17-62</td>
<td>2-13</td>
<td>6-45</td>
</tr>
<tr>
<td>8</td>
<td>0-29</td>
<td>8-18</td>
<td>55-86</td>
<td>2-47</td>
<td>2-11</td>
</tr>
<tr>
<td>27</td>
<td>2-23</td>
<td>14-24</td>
<td>2-66</td>
<td>0-26</td>
<td>all 4</td>
</tr>
<tr>
<td>30</td>
<td>2-6</td>
<td>12-19</td>
<td>2-30</td>
<td>8-16</td>
<td>13-22</td>
</tr>
<tr>
<td>Any laboratory</td>
<td>0-100</td>
<td>2-80</td>
<td>0-110</td>
<td>0-83</td>
<td>0-99</td>
</tr>
</tbody>
</table>
• Systematic literature review to identify “guideline-like” studies
• Identify *in vivo* reference chemicals
  • Active chemicals verified in ≥2 independent studies
  • Inactive chemicals verified in ≥2 independent studies (with no positive results in any study)
Validating NAMs for Endocrine Disruptor Screening

Development of a curated Hershberger database


Evaluation of androgen assay results using a curated Hershberger database


Screening Chemicals for Estrogen Receptor Bioactivity Using a Computational Model

Patience Browne, Richard S. Judson, Warren M. Casey, Nicole C. Kleinstreuer, and Russell S. Thomas

A Curated Database of Rodent Uterotrophic Bioactivity

Nicole C. Kleinstreuer, Patricia C. Ceger, David G. Allen, Judy Strickland, Xiaoqing Chang, Jonathan T. Hamm, and Warren M. Casey

Integrated Model of Chemical Perturbations of a Biological Pathway Using 18 In Vitro High-Throughput Screening Assays for the Estrogen Receptor

Richard S. Judson, Felicia Maria Magpantay, Vijay Chickarmane

Development and Validation of a Computational Model for Androgen Receptor Activity

Nicole C. Kleinstreuer, Patricia Ceger, Eric D. Watt, Matthew Thomas Martin, Keith A. Houck, Patience Browne, Russell S. Thomas, Warren Casey, David Dix, David Allen, Shraddha Sakamuru, Menghang Xia, Rui Li, and Richard S. Judson
Ex: Acute Oral Toxicity

Bootstrapping of the standard deviations for 1120 repeat test chemicals identified a 95% confidence interval for LD50 values of
$$\pm 0.31 \log_{10}(\text{mg/kg})$$
Investigating Sources of Variability

Acute Toxicity Dataset: Chemicals Spanning EPA Hazard Categories

Variability classes based on EPA categories

Standard Deviation

Low

Medium

High
Acute Toxicity Dataset: Chemicals Spanning EPA Hazard Categories

Investigating Sources of Variability

No significant differences found in ToxPrint Chemotype enrichment
No significant differences driven by physicochemical properties
Benchmarking Alternative Models
Animal data reproducibility as threshold for performance

Skin Sensitization
Defined Approaches (AOP WoE and KE 1/3 STS) accepted by EPA based on comparison to LLNA (mouse) data

Interim Science Policy: Use of Alternative Approaches for Skin Sensitization as a Replacement for Laboratory Animal Testing
DRAFT FOR PUBLIC COMMENT
April 4, 2018
EPA’s Office of Chemical Safety and Pollution Prevention:
Office of Pesticide Programs
Office of Pollution Prevention and Toxics
Skin Sensitization: Lab Animal vs Human Data (n≈150)

**LLNA**

<table>
<thead>
<tr>
<th>Hazard</th>
<th>Potency</th>
</tr>
</thead>
<tbody>
<tr>
<td>72%-82%</td>
<td>54% - 60%</td>
</tr>
</tbody>
</table>

**GPMT / Buehler**

<table>
<thead>
<tr>
<th>Hazard</th>
<th>Potency</th>
</tr>
</thead>
<tbody>
<tr>
<td>~72%</td>
<td>~60%</td>
</tr>
</tbody>
</table>

**Reproducibility of Multiple Tests (~100 chems)**

<table>
<thead>
<tr>
<th>Hazard</th>
<th>Potency</th>
</tr>
</thead>
<tbody>
<tr>
<td>~78%</td>
<td>~62%</td>
</tr>
</tbody>
</table>

ICCVAM. 1999. NIH Publication No. 99-4494
ICCVAM. 2010. NIH Publication No. 11-7709
Urbisch et al. 2015. Reg Tox Pharm 71:337-351.
Dumont et al. 2016. Tox In Vitro 34: 220-228
Human data and human biology as the gold standard

Using the AOP framework to develop testing strategies
**Adverse Outcome Pathway (AOP) for Skin Sensitization**

**Chemical Structure & Properties**
- Electrophilic substance
- Metabolism
- Penetration

**Molecular Initiating Event**
- Covalent interaction with skin proteins

**Cellular Response**
- Activation of inflammatory cytokines
- Induction of cytoprotective genes
- Keratinocytes responses
  - Activation of inflammatory cytokines
  - Induction of cytoprotective genes

**Organ Response**
- Histocompatibility complexes presentation by DCs
- Activation of T cells
- Proliferation of activated T-cells

**Organism Response**
- Inflammation upon challenge with allergen

**Key Event 1**
- Activation of inflammatory cytokines
- Induction of cytoprotective genes

**Key Event 2**
- Induction of inflammatory cytokines and surface molecules
- Mobilisation of DCs

**Key Event 3**
- Dendritic Cells (DCs)
  - Induction of inflammatory cytokines and surface molecules
  - Mobilisation of DCs
  - Histocompatibility complexes presentation by DCs
  - Activation of T cells
  - Proliferation of activated T-cells

**Key Event 4**
- T-cell proliferation
- Inflammation upon challenge with allergen

*Adapted from illustration by D. Sailstad*

OECD (2014)
Defined Approaches (DAs) combine *in vitro* and *in silico* data using simple decision trees or machine learning algorithms to predict skin sensitization.
All non-animal AOP-based DAs evaluated perform as well or better than the animal test at predicting human skin sensitization:

Hazard: 74% (mouse) vs. 75-85% (DAs)

3-class Potency: 59% (mouse) vs. 55-69% (DAs)
Eye Irritation: Reconstructed Human Tissue Models

OECD/Guideline for the Testing of Chemicals

Reconstructed human Cornea-like Epithelium (RCE) test method for identifying chemicals not requiring classification and labelling for eye irritation or serious eye damage

INTRODUCTION

1. Serious eye damage refers to the production of tissue damage in the eye, or serious physical damage to vision, following application of a test chemical to the external surface of the eye, which is not fully reversible within 21 days of application, as defined by the United Nations Globally Harmonized System of Classification and Labeling of Chemicals (GHS) (1). According to UN GHS, eye irritation refers to the production of changes in the eye following application of a test chemical to the external surface of the eye, which are fully reversible within 21 days of application. Test chemicals inducing serious eye damage are classified as UN GHS Category 1, while those inducing eye irritation are classified as UN GHS Category 2. Test chemicals not classified for eye irritation or serious eye damage are defined as those that do not meet the requirements for classification as UN GHS Category 1 or 2 (1A or 2B), i.e., they are referred to as UN GHS No Category.

2. The assessment of serious eye damage is commonly carried out by the use of in vitro systems (OECD Test Guideline TD101), adapted in 1987 and revised in 1997, 2001 and 2017 (2). The choice of the most appropriate test method and the use of the Test Guideline should be seen in the context of the OECD Guidance Document on an Integrated Approach to Testing and Assessment (IATA) on the Testing and Assessment of the Sensitivity to Eye Damage and Eye irritation (3).

3. The Test Guideline describes an in vitro procedure allowing the classification of chemicals (substances and mixture) not requiring classification and labelling for eye irritation or serious eye damage in accordance with UN GHS. If a medium and reconstructed human corneal epithelium (RCE) which clearly assesses the histological, morphological, histopathological, and physicochemical properties of the human corneal epithelium. Four other in vitro test methods have been validated, considered scientifically valid and adopted by OECD (4-7).

You can view the full document at http://www.oecd.org/…/oe...
Mechanistic Mapping of HTS Assays

Example: Developmental Toxicity

Human Teratogenic Mechanisms

- Endocrine disruption
- Oxidative stress
- Vascular disruption
- Folate antagonism
- Neural crest cell disruption
- Specific receptor- or enzyme-mediated

Van Gelder et al. 2010; Knudsen and Kleinstreuer 2011; Saili et al. 2019
Mechanistic Mapping of HTS Assays

Example: Carcinogenicity

Hallmarks of Cancer & Characteristics of Carcinogens

- Inflammation
- Oxidative stress
- Genotoxicity/instability
- Angiogenesis
- Immortalization/proliferation
- Immunosuppression
- Invasion/metastasis
- Specific receptor- or enzyme-mediated

Hanahan & Weingberg 2011; Smith et al. 2016; Guyton et al. 2018; Chiu et al. 2018
Acknowledgments

- ILS/NICEATM group
- Sciome collaborators
- ICCVAM agencies
- ICATM partners
- Patience Browne (OECD)
- Anna Lowit (EPA)
- Richard Judson (EPA)
- Tom Knudsen (EPA)
- Tony Williams (EPA)
Challenges

• Scientific
  – Considering human population/genetic variability
  – Incorporating metabolic competence
  – Developing complex systems models
  – Reporting, collection, and curation of reference data

• Non-scientific
  – Increasing awareness, education, and training
  – Cross-sector communication
  – Funding for human-centric research and education
Automating Reference Data Identification

- Project with Oak Ridge National Labs (ORNL) and FDA CFSAN to apply text-mining (NLP) approaches & ML to identify high-quality data
- Semi-automated retrieval and evaluation of published literature (trained on uterotrophic database)
- Apply to developmental toxicity studies (with ICCVAM DARTWG)
  - Define literature search keywords, identify corpus
  - Extract/characterize study protocol details from regulatory guidelines: minimum criteria
  - Apply ML algorithms to identify high-quality studies, expert check
Integrated Chemical Environment (ICE)

Validation Studies

Databases

Published Data

Computational Models

https://ice.ntp.niehs.nih.gov/