National Institute of Environmental Health Sciences

United States Environmental Protection Agency

Data Extraction Challenge for Systematic Review

A Joint NIEHS-EPA Initiative

Charles Schmitt - charles.Schmitt@nih.gov
National Institute for Environmental Health Sciences (NIEHS)

The Team

- John Bucher (NIEHS)
- Nicole Kleinstreuer (NIEHS)
- Andy Rooney (NIEHS)
- Charles Schmitt (NIEHS)
- Andy Shapiro (NIEHS)
- Vickie Walker (NIEHS)
- Ashley Williams (ICF)
- Mary Wolfe (NIEHS)
- Alicia Frame (EPA)
- Kristan Markey (EPA)
- Seema Schappelle (EPA)
- Kris Thayer (EPA)
- Michelle Taylor (EPA)

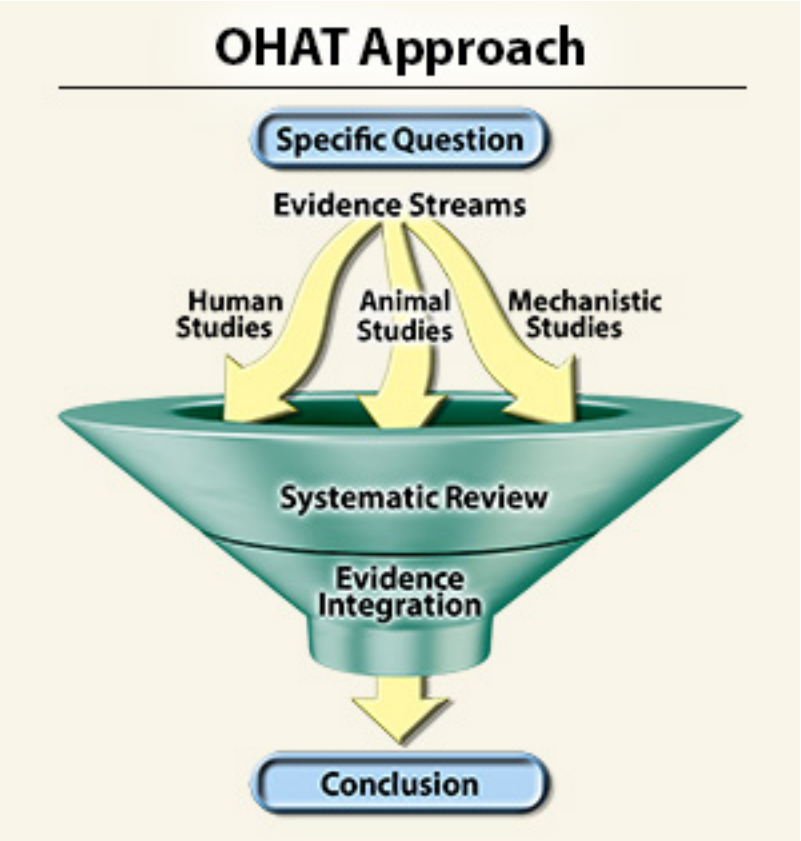
Special thanks to:
Ian Soboroff (NIST)
Hoa Dang (NIST)

And a number of others we've been mining for knowledge on challenges

Background

What Is Systematic Review?

- Systematic review is a predetermined, multistep process used to *identify*, *select*, critically *assess*, and *synthesize* evidence from scientific studies to reach a conclusion.
- NTP and EPA use the systematic review process to conduct literature-based health evaluations to assess whether exposure to environmental substances (e.g., chemicals) has adverse effects on health or to determine the state of the science.



Systematic Review Example

- *What detrimental impacts on neurobehavior does fluoride exposure cause?*

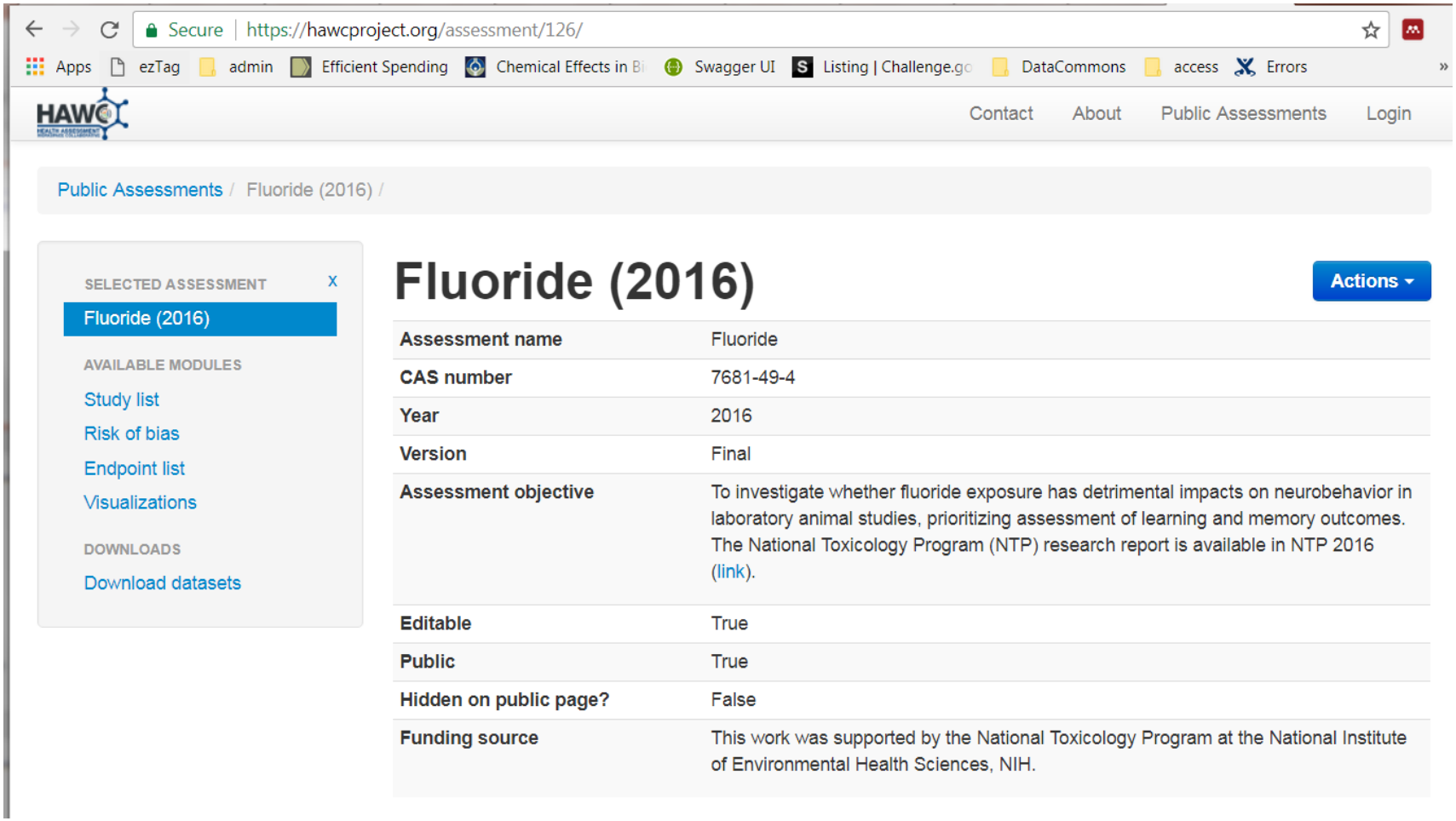
Systematic Review Example

- *What detrimental impacts on neurobehavior does fluoride exposure cause?*
- Simplified Study:
  - Expose 3 groups of animals to increasing doses of test article
  - Expose 4th group to negative control substance
  - Expose 5th group to positive control substance
  - Measure effect for one or more endpoints
    - 3-chamber assay to test socialization
    - Pathology assay to determine neural tissue damage
  - Analyze dose-response against positive and negative controls
    - Determines statistics, e.g., lowest effect level

Systematic Review Pipeline

- *What detrimental impacts on neurobehavior does fluoride exposure cause?*
  - Formulate review question
  - Define criteria to include/exclude articles
  - Locate articles (1000s)
  - Select articles (100s)
  - Assess study quality, determine risk of bias
  - ***Extract data from studies***
  - Meta-analysis and synthesis of studies
  - Interpret results in light of review question

Example Reviews



HAWC: https://hawcproject.org/assessment/126/

National Institute of Environmental Health Sciences

United States Environmental Protection Agency

Need – A Tool for Machine Assisted Data Extraction

the positive control wells treated with natural ligands (1 nM of 17β-estradiol) ordinary showed maximum response and it showed well reproducibility. Description of PC50 and PC10 is illustrated in Fig. 1.

2.6. Animals

Crj:CD (SD) rats at post-natal day (pnd) 10 and dams were purchased from Charles River Japan, Inc. (Shiga, Japan). Dams and pups were kept in polycarbonate pens until weaning. All rats were weaned at pnd 17 and then housed individually in stainless steel, wire-mesh cages during the study. The immature rats were weighed, weight-ranked and assigned randomly to each of the treatment and control groups. Each group consisted of six rats. Body weights and clinical signs were recorded on a daily basis throughout the study. Rats were provided with tap water and a commercial diet (CRF-1, Oriental Yeast Co., Tokyo, Japan) ad libitum before weaning and with water automatically and a commercial diet (MF, Oriental Yeast Co.) ad libitum after weaning. The animal room was maintained at a temperature of 23 ± 2 °C, a relative humidity of 55 ± 5% and was artificially illuminated with fluorescent light on a 12-h light/dark cycle (06:00–18:00 h). All animals were cared for according to the principles outlined in the guide for animal experimentation prepared by the Japanese Association for Laboratory Animal Science.

2.7. Animal study design

The 21 chemicals, i.e. all of those mentioned above except for dibutyl phthalate and ethynyl estradiol, were injected subcutaneously on the dorsal surface at doses of 2, 20 and 200 mg/kg from pnd 20 to pnd 22, i.e. for 3 days. The high dose was selected on the basis of the previous uterotrophic assay using bisphenol A, in which the uterine response was clearly detected at a dose of 160 mg/kg per day injected subcutaneously (Yamasaki et al., 2000). On the other hand, doses of dibutyl phthalate or ethynyl estradiol were 0, 40, 200 and 1000 mg/kg per day or 0, 0.2, 2 and 20 μg/kg per day, respectively. These doses were based on the results of preliminary studies. The concentration and stability of each chemical was confirmed. The volume of olive oil contained in each chemical solution was 4 ml/kg for subcutaneous injection. A vehicle control group given only olive oil was also established. The animals were killed approximately 24 h after the last ad-

Select DE Modules

Test Subject Module
Species: Rat — Ok | Reject | Edit
Strain: Crj:CD — Ok | Reject | Edit
Source: Charles River Japan, Inc — Ok | Reject | Edit

Experiment Group Module
Route of Admin: sub. inj. — Ok | Reject | Edit

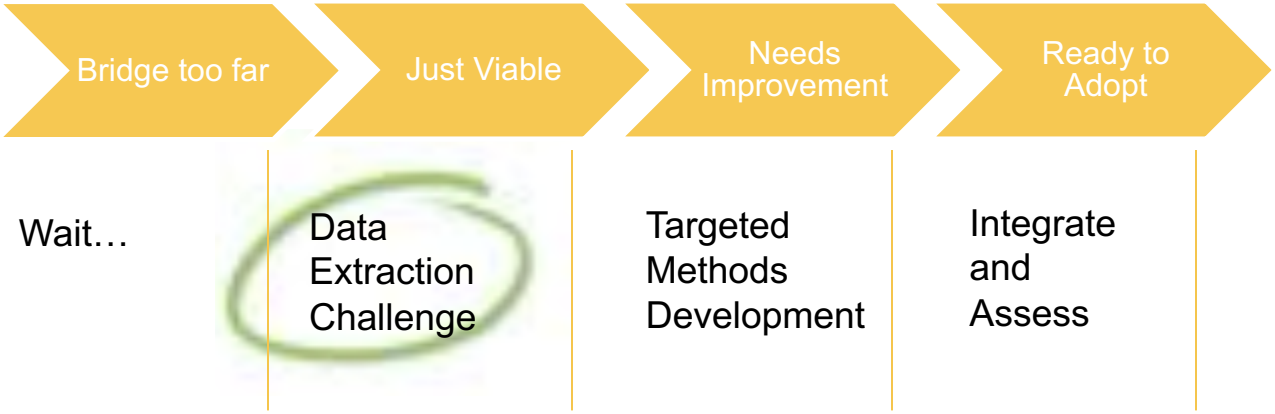
DE Module 3…

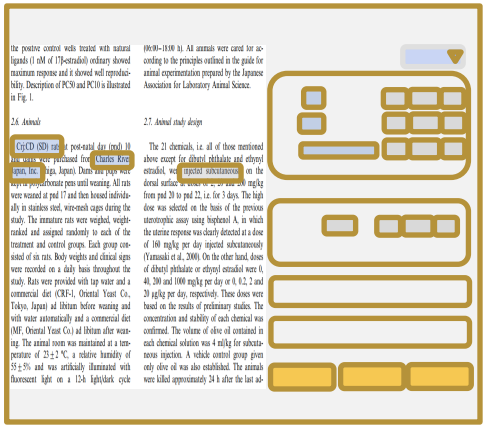
DE Module 4…

Export to Clipboard | Export to App 1 | Export to App 2

Incorporating Automated Data Extraction (DE)

DE methods development pipeline





** For some DE tasks determining where we are on the pipeline is fairly clear (e.g., gene name extraction), other tasks (e.g., risk of bias) are not as obvious*

National Institute of Environmental Health Sciences

United States Environmental Protection Agency

2018 TAC Challenge

Focus - Animal Studies & Animal Treatment Groups

With, pilot of Measures & Endpoints

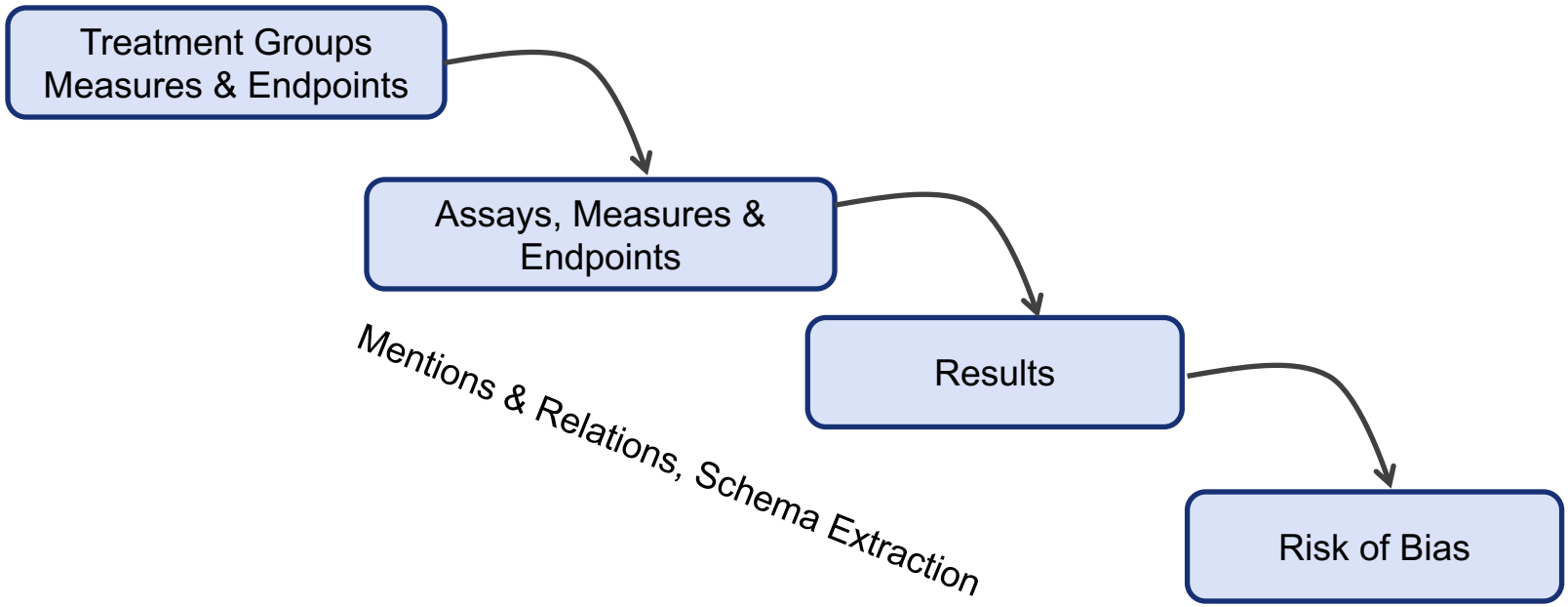
Conceptual Schema for Animal Studies

- Journal Article
  - Studies
    - Experiments
      - Treatment/Animal Groups
        - Type
        - Animal Information
        - Exposures
        - Doses
        - Measures
        - Endpoints
        - Assays
      - Results
      - Risk of Bias

Can we extract these items and relations?

Challenge Series – Not a one time challenge

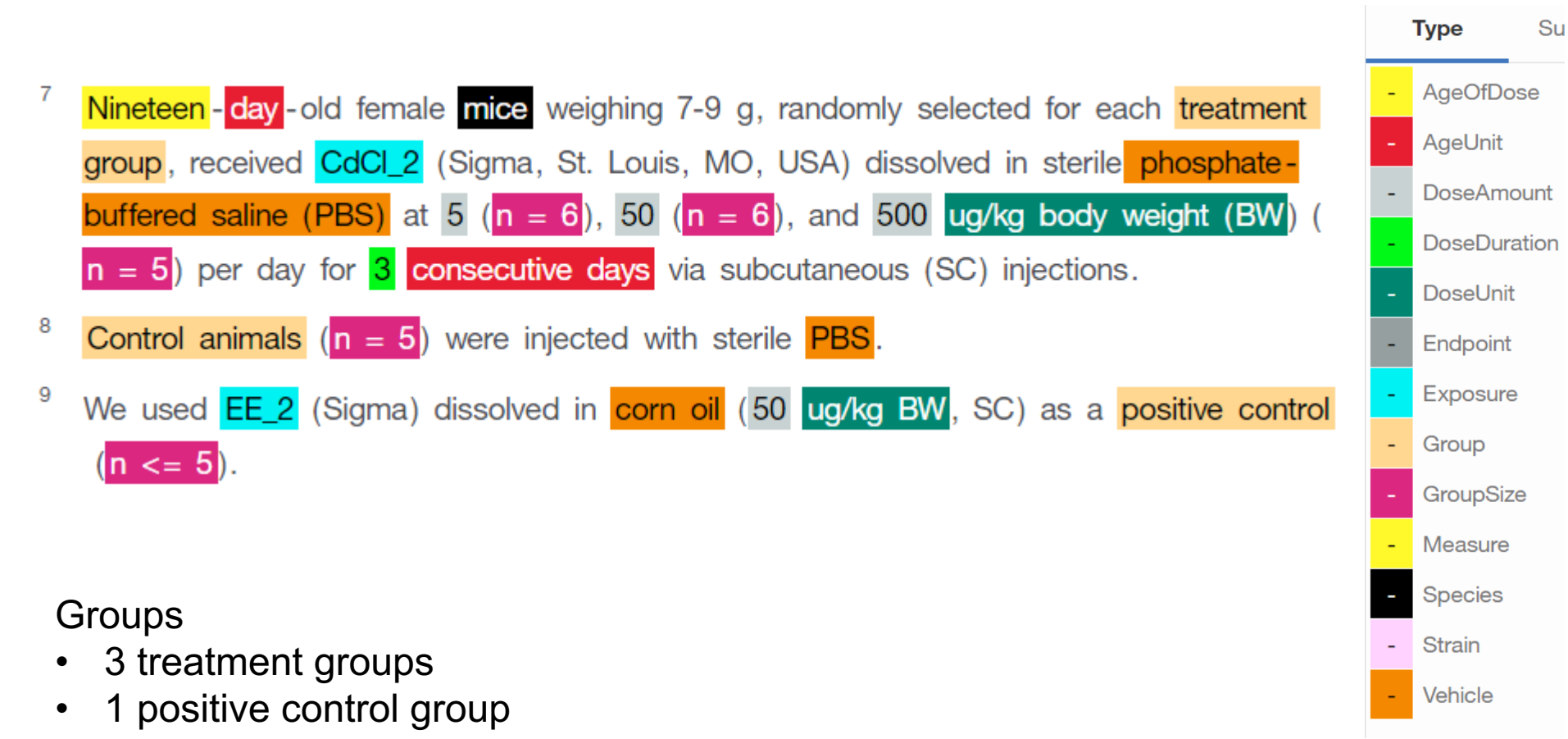
Our goal is to close the gaps thorough a coordinated series of challenges



DRAFT

Annotation Example

Entity annotation – Treatment Groups



Groups

- 3 treatment groups
- 1 positive control group
- 1 negative control group

This is a one of the nicer example in that there is minimal variation across groups

Entity Annotation – False positives

2 Treatment groups (four rats per group) were 0 - 0.03 - 0.1 - 0.3 - 1 - 10 ug/kg bw EE2 (Steraloids, Newport, RI) and 0 - 0.03 - 0.1 - 0.3 - 1 - 10 mg/kg bw zearalenone (ZEA) (Sigma, Zwijndrecht, The Netherlands).

3 The control group received only the vehicle ethanol (70%, 0.3 ul/g bw) (Merck, Roden, The Netherlands).

4 EE2 and ZEA were dissolved in ethanol and mixed with 1 ml of custard.

5 Animals were housed individually during feeding, in order to monitor whether the entire test compound was ingested.

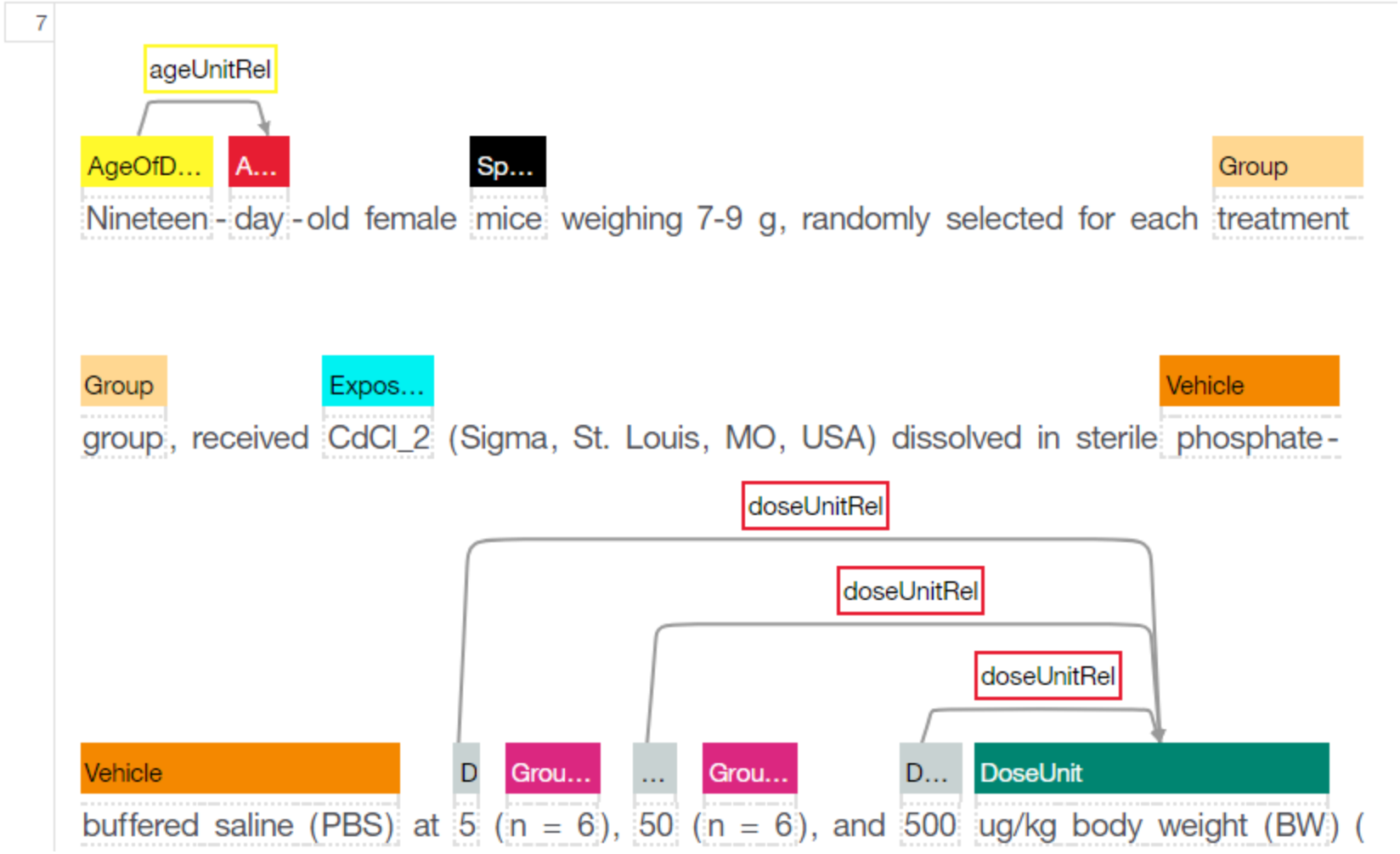
6 Exposure started at postnatal day 21 and lasted for 3 consecutive days.

7 At day 4, animals were weighed and sacrificed by bleeding under isoflurane anesthesia.

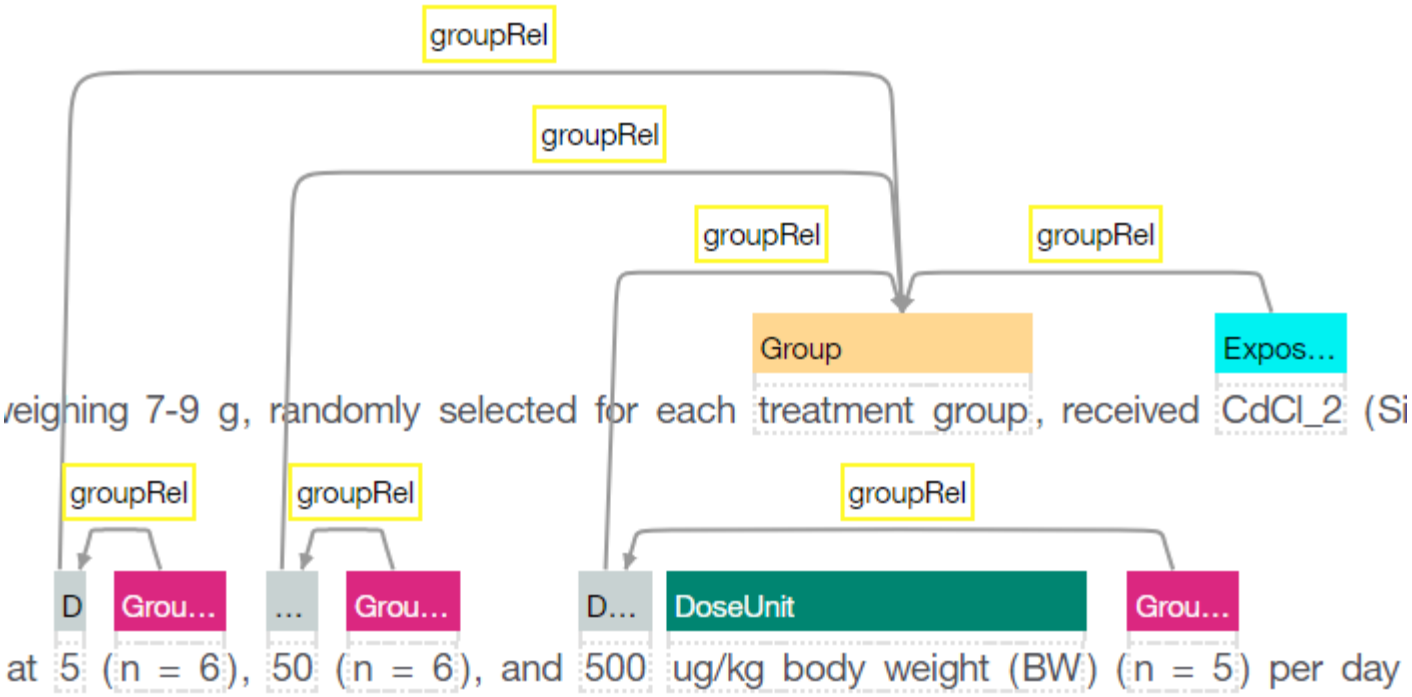
8 The doses used in this study were chosen based on previously carried out dose-finding experiments and a yeast-based reporter gene assay.

9 Furthermore, initial dose-range finding experiments showed that concentrations above 10 ug/kg bw for EE2 and 10 mg/kg bw for ZEA did not result in higher relative uterus weights (results not shown).

Relation annotation – simpler cases



Relation annotation – treatment groups



Relationship structure: Entities to a Group anchor

Treatment Groups

1 For each individual animal, doses were calculated daily.

2 Treatment groups (four rats per group) were 0 - 0.03 - 0.1 - 0.3 - 1 - 10 ug/kg bw EE2 (Steraloids, Newport, RI) and 0 - 0.03 - 0.1 - 0.3 - 1 - 10 mg/kg bw zearalenone (ZEA) (Sigma, Zwijndrecht, The Netherlands).

3 The control group received only the vehicle ethanol (70%, 0.3 ul/g bw) (Merck, Roden, The Netherlands).

Relationship structure: Dose Amount defines anchor for groups

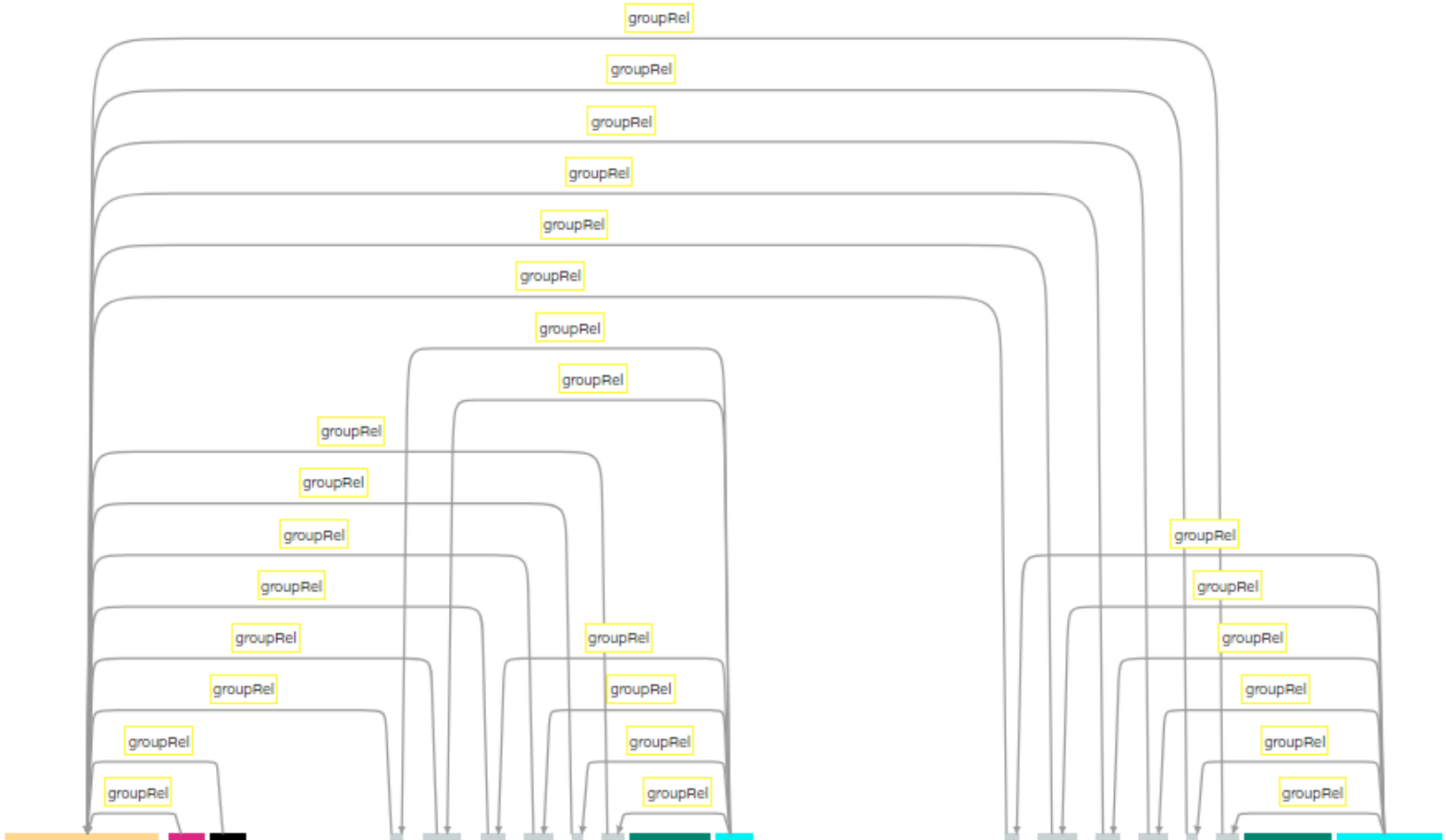
12 treatment groups

6 dose levels, 2 exposures, 2 dose units, same species/group size

1 control group

11509742_methods.txt

1 For each individual animal, doses were calculated daily.

2 

Treatment groups (four rats per group) were 0 - 0.03 - 0.1 - 0.3 - 1 - 10 ug/kg bw EE2 (Steraloids, Newport, RI) and 0 - 0.03 - 0.1 - 0.3 - 1 - 10 mg/kg bw zearalenone (ZEA) (Sigma, Zwijndrecht,

Annotations - Mentions

- **Group**: an indicator of a treatment group or positive/negative control group
- **Group Size**: number of animals in a test or control group
- **Exposure**: the treatment, positive control, or negative control substance
  - including dose and unit
- **Vehicle**: the solution the exposure is in
  - Possibly including dose and unit
- **Animal Species & Strain**: the scientific species and strain names

Annotations - Mentions

- **Age at First/Last Exposure**: the age at which the first and last doses are given
  - Including time unit (e.g., PND – post natal days)
- **Duration of Exposures**: number of days from when the first dose is given to when the last dose is given.
- **Measure**: the experimental variable being measured as part of an assay
- **Endpoint**: the experimental condition of interest.

Annotations - Relations

- **AgeUnitRel**: a relationship between age of exposure value and age of exposure unit
- **DoseUnitRel**: a relationship between dose value and dose unit
- **ExposureRel**: a relationship between the exposure substance and the vehicle
- **SpeciesRel**: a relationship between strain and species
- **GroupRel**: a relationship between two mentions where one of the mentions is a ‘grouping’ entity

Tasks

- **Task 1**: Extract mentions (Group Size, Group Type, Species, Strain, etc) except for measures/endpoint
  - This is similar to NLP Named Entity Recognition (NER) evaluations.
  -
- **Task 2**: Identify the relations between mentions from Task 1
  - This is similar to many NLP relation identification evaluations.
- **Task 3**: Extract meansure & endpoint mentions and identify relations between measures, endpoints and treatment group
  - This is similar to Tasks 1& 2 but focused on measures and endpoints.

Training & Test Data

- 100-200 articles pulled from prior systematic reviews
- Additional set of un-annotated articles
  - E.g., for embeddings
- Finalizing set of articles
  - Balancing open access, breadth of journals, date of articles, single studies versus multiple study articles
- Train/Test split will be determined after annotation is completed

Other Aspects

- **Following procedures already in place for FDA adverse event challenge**
  - Evaluation:
    - Precision/Recall/F1 measures on mention and relationship level annotations with and without mention/relation type
  - 3 separate submissions
  - Rejection of submissions that don't meet XML standards
  - Registration procedures
  - …

Draft Timeline

Time frame	Milestone
Nov, Dec 2017	Pilot Annotations
Jan 2018	Annotations Guidelines
May 2018	Registration deadlines
Mid Sep 2018	Submissions due
Early Oct 2018	Results to participants
Mid Oct 2018	Workshop proposals due
Mid-late Oct 2018	Notification of acceptance
Early Nov 2018	Workshop papers due
Mid Nov 2018	TAC 2018 workshop

We welcome any and all feedback

charles.schmitt@nih.gov