Development of a data organization program for inhalation toxicology studies

Grady Roberts
Mentored by Dr. Benjamin Wong

Introduction

In the field of inhalation toxicology, massive amounts of real-time respiratory dynamics (RD) data are generated in studies focusing on inhalation of toxic chemicals. These data are used to understand disease mechanisms, diagnose exposure casualties, and develop countermeasures. One obstacle to achieving these goals is inefficiency in data analysis (Shields, 2014). In the situation of an airborne chemical attack, first responders must be able to identify the agent, the best course of treatment, and the severity of each potential casualty, a task that may be made easier by assessing RD parameters that are clinically relevant and easily measurable. In addition, existing assessment and treatment methods may be improved through the analysis of relevant experimental RD data, with more efficient analysis leading to better treatment. To this end, this study sought to develop a data organization and management program for increasing the efficiency of analyzing large sets of RD data.

Materials and Methods

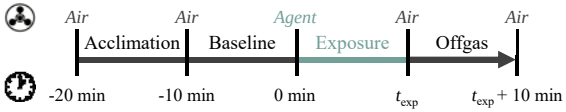


Figure 1: Diagram of inhalation exposure setup.

RD data were collected in three distinct periods, corresponding to before (baseline), during (exposure), and after (offgas) an exposure (Fig. 1). The length of the exposure phase was varied to create distinct doses, calculated as the product of the concentration of the chemical, C, and the length of exposure, t_{exp}. The organization program was designed to accommodate this structure of exposure, or any other structure consisting of three data collection phases.

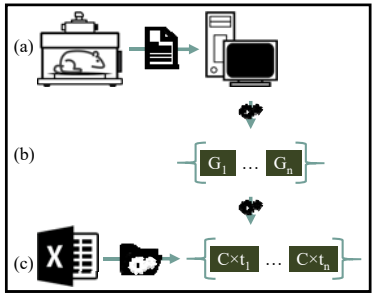


Figure 2a: Raw data were imported to the program from FinePointe (Version 2.3.1.16, Data Sciences International [DSI]).

Figure 2b: Raw data were formatted into tables separated by subject group, denoted Gn.

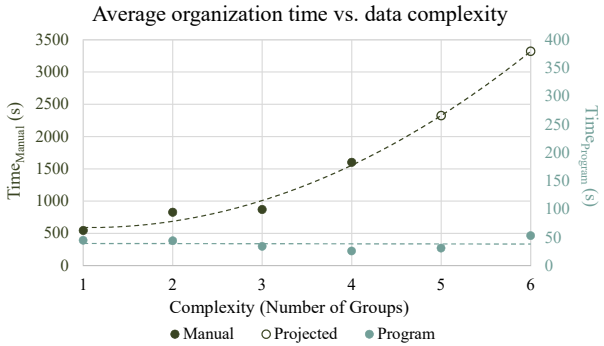
Figure 2c: Dosage information for each subject was imported to the program and data were organized by dose groups.

Materials and Methods (cont.)

The program, written in Microsoft Excel VBA (v7.0.1639), was designed to operate in three main stages. First, raw RD data for each exposure phase were imported into the program (Fig. 2a) in the form of text files exported from FinePointe (v2.3.1.16, DSI). Next, the program systematically looped through the data to format it into tables organized by exposure groups (Fig. 2b). Finally, a Microsoft Excel (v14.1.7166.5000) spreadsheet containing dosage information for each subject was imported (Fig. 2c). Subjects were cross-referenced with the dosage sheet to organize the data into groups by $C \times t_{exp}$ dose.

The efficacy of the program in increasing analysis efficiency was tested by recording the time taken to organize RD data programmatically and time taken to organize data manually in the same format outputted by the program at varied levels of complexity.

Results



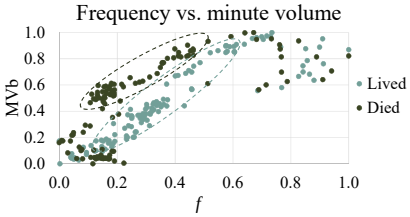
Graph 1: Graph of time taken to organize data, by method. Open points represent projected values based on second degree polynomial fit curve.

The average organization time when using the program was significantly less than the time required to perform manual organization at each complexity level (p-values of 4.59×10^{-6}, 2.39×10^{-7}, 3.10×10^{-8}, 5.79×10^{-9}, for one, two, three, and four groups, respectively), as determined by a two-sample t-test. The magnitude of the p-value decreased as the number of groups increased, implying that organizing data with the program is more effective for larger datasets. A typical group consists of at least eight subjects and an average of 866 data points per group. This finding was supported by a graph of average organization time by method (Graph 1). For all levels,

Results (cont.)

averages of manual organization times were less than those for programmatic organization. Manual organization time data was found to be proportional to the square of the data complexity while the programmatic organization exhibited essentially no relationship.

Discussion



Graph 2: Graphing two normalized parameters against each other may be able to predict patient survivability. Distinct clusters were observed (dashed ovals). Data were normalized.

The organization program significantly increased the efficiency of RD data organization, expediting analysis and fulfilling the goal of the study. The potential for human error that could affect the data, especially as a result of the mental fatigue caused by the repetitive actions involved in organization, was reduced. By organizing the data in a flexible structure, it is possible to represent the data in a manner that is not currently available within FinePointe (v2.3.1.16). For example, normalized graphs of the relationship between parameters (Graph 2) can be produced. Such graphs are important diagnostic tools because the cluster (denoted by the dashed ovals on Graph 2) the patient falls in may be able predict their survivability or outcome. An increased ability to predict where a patient is in their course will allow first responders to triage patients. Further work for this project would include development of automatic graphing capabilities and a master program to create additional instances of the program for each desired parameter, because the current program only handles one parameter.

References

Data Sciences International. FinePointe Software v2.3.1.16 [Computer software]. Available from https://www.datasci.org

Shields, R. (2014). *Combating the inefficiency in data aggregation*. Retrieved from https://www.exchangewire.com/blog/2014/01/23/combating-the-inefficiency-in-data-aggregation/

Acknowledgements

I would like to thank the entire inhalation toxicology team at ICD.