17th International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
9-12 May 2016, Budapest, Hungary

ARE TOXIC LOAD-BASED TOXICITY MODELS CONSISTENT WITH EXPERIMENTAL OBSERVATIONS? INDEPENDENT ANALYSIS OF TIME-VARYING EXPOSURE DATA FROM THE 2012–2013 ECBC/NAMRU-D TOXICOLOGICAL EXPERIMENTS

Alexander Slawik[1], James Silva[1, 2], Kevin Axelrod[1, 3], Jeffry T. Urban[1], Nathan Platt[1]

[1]Institute for Defense Analyses, Alexandria, Virginia, USA
[2]Department of Physics, Boston University, Boston, Massachusetts, USA
[3]Graduate Program in Biophysics, Harvard University, Cambridge, Massachusetts, USA

Abstract: The U.S. Defense Threat Reduction Agency (DTRA) sponsored a two-year set of experiments, conducted in 2012 and 2013, that were designed and executed through a collaboration between the U.S. Army's Edgewood Chemical and Biological Center (ECBC) and the Naval Medical Research Unit Dayton (NAMRU-D) to explore the effects of time-varying inhalation exposures of hydrogen cyanide (HCN) gas on rats. In this work, we explore the comparison between the observed lethality and the predicted lethality of the "toxic load" model of exposure. Our analysis confirms the conclusion published by the authors of the experiment that the casualties observed for exposures over the experiments' full range of exposures from 2.3 to 30 minutes are not consistent with the predictions of toxic load-based toxicity models. We also conclude that a single set of fitted parameters for the toxic load model (i.e., the toxic load exponent n, probit slope m, and median lethal exposure TL_{50}) accurately models the single exposure experimental data across the experiments' longer timescales of 10 to 30 minutes. However, we found that none of the toxic load models that we considered appear to fit the experimental data for the novel, time-varying exposures well, with the Average Concentration and Griffiths-Megson models providing the least inaccurate casualty predictions.

Key words: *casualty assessment, consequence assessment, Haber's law, toxic load modelling.*

INTRODUCTION

Toxic industrial chemicals and chemical warfare agents present an acute inhalation hazard to civilians and military personnel. An individual exposure to an airborne hazardous material may be highly time-dependent due to the random effects of wind meandering and atmospheric turbulence. Several toxicological models based on the "toxic load" model of exposure have been proposed to predict the casualties arising from time-dependent exposures to airborne hazardous materials, but none were developed using data from toxicological experiments that used time-varying exposure profiles. To explore this experimentally unexplored regime, the U.S. Army's Edgewood Chemical and Biological Center (ECBC) and the Naval Medical Research Unit Dayton (NAMRU-D) performed a two-year set of experiments observing the lethality of time-varying exposures of hydrogen cyanide (HCN) gas on rats.

ECBC/NAMRU-D EXPERIMENTAL DATA

To determine the suitability of toxic load models, the U.S. Army's Edgewood Chemical and Biological Center (ECBC) and the Naval Medical Research Unit Dayton (NAMRU-D) performed a two-year set of experiments. These experiments were designed to observe the lethality of hydrogen cyanide in a healthy population of male Sprague-Dawley rats. A group of ten rats are simultaneously exposed to the airborne toxin using a pressurized, nose-only inhalation apparatus (see Sweeney et al, 2014 for details).

The experimental setup allowed the concentration of the gas to be varied as a function of time; twenty-two separate exposure profiles were tested (eleven for each year or phase of the experiment). These exposure profiles were constructed to approximate square pulses of constant concentration (Sweeney et al, 2014 validates this idealization). The profiles can be grouped into three categories: a single square pulse, two square pulses with no time gap between them, and two square pulses with a

time gap between them when the rat is not exposed to any toxin. All exposure profiles are between 2.3 to 30 minutes long. Each profile was tested a number of times with a unique concentration level, denoted as a trial.

TOXIC LOAD MODELS

Toxic load models define a quantity *TL* (toxic load) which captures the propensity of a population to suffer a physiological effect from exposure to a toxic chemical. For airborne exposures with a single constant exposure, the toxic load is defined as

$$TL = C^n \Delta t, \tag{1}$$

where C is the air concentration of toxin in the exposure, Δt is the duration of the exposure, and n is the toxic load exponent, a positive number. The toxic load model holds that the susceptibility of a population, measured in fraction affected, is log-normally distributed with respect to toxic load. The toxic load corresponding to 50% of the population affected is defined as the median toxic load (denoted by TL_{50}), and the slope of the cumulative distribution function at TL_{50} is the probit slope, denoted m. The following probit model defines the relationship between toxic load and toxic effects,

$$P = \Phi(m \log_{10} TL - m \log_{10} TL_{50}), \tag{2}$$

$$\Phi(z) = \left[Erf\left(\frac{z}{\sqrt{2}}\right) + 1\right]/2\ . \tag{3}$$

Here P corresponds to the percent of a population to exhibit the response of interest (in this case rat lethality), $\log_{10}$ is the base 10 logarithm, Φ is the cumulative distribution function of the standard normal distribution, and Erf is the error function. Figure 1 portrays a generic plot of the toxic load model, with the physical significance of the parameters indicated.

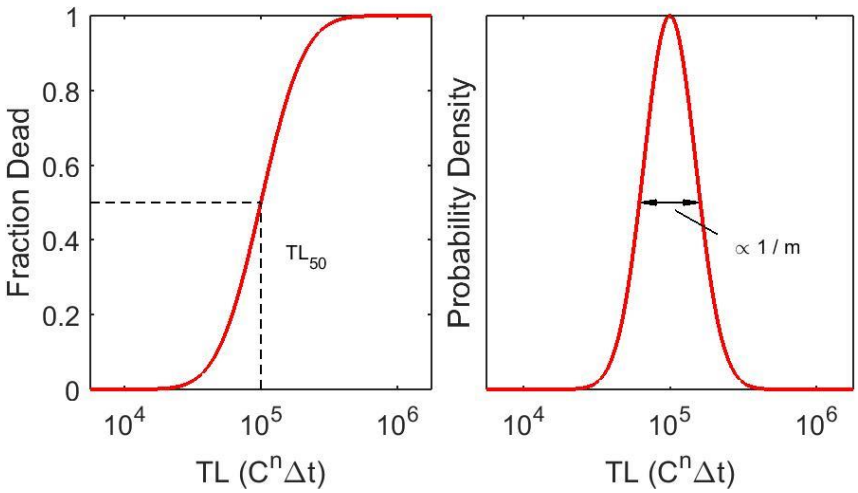

Figure 1. The relationship between toxic load and fraction dead (left) and probability density (right) for a population. The toxic load exponent n parameterizes the x-axis, weighing the relative importance of concentration and exposure time. The median toxic load TL_{50} captures the overall susceptibility of the population, while the probit slope m captures the variability of susceptibility in the population.

The toxic load model was originally defined for single square pulse exposures (ten Berge, 1983). However, as the toxic load model is phenomenological, it is unclear how to extend toxic load to a given time varying exposure. Various extensions to the toxic load model have been proposed to capture time dependence: the Ten Berge model (4), the average concentration model (5), the peak concentration model (6), the concentration intensity model (7), and the Griffiths-Megson model (8). These models define the toxic load in terms of the time dependent concentration *C(t)* and the toxic load exponent *n*:

$$TL = \int C^n(t)\, dt \qquad \text{(Ten Berge)} \qquad (4)$$

$$TL = \left(\frac{\int C(t)dt}{\Delta t}\right)^n \Delta t = \left(\overline{C(t)}\right)^n \Delta t \qquad \text{(Average Concentration)} \qquad (5)$$

$$TL = \left(\frac{\int C(t)dt}{\sup\{C(t)\}}\right) \sup\{C(t)\}^n \qquad \text{(Peak Concentration)} \qquad (6)$$

$$TL = \left(\frac{\left(\int C(t)dt\right)^{2-n}}{\left(\int C^2(t)dt\right)^{1-n}}\right) \qquad \text{(Concentration Intensity)} \qquad (7)$$

$$TL = \left(\frac{\int C(t)dt}{\Delta t_{C>0}}\right)^n \Delta t_{C>0} \qquad \text{(Griffiths-Megson)} \qquad (8)$$

The expressions for toxic load are understood to be defined only over the time interval between the onset and termination of chemical exposure. These expressions are well defined given a particular *C(t)*, exposure duration Δ*t*, and toxic load exponent *n*. The different models generally produce different values of the toxic load and significantly different casualty predictions (Czech 2011). However, in the case of constant exposures (*C(t)* = *C*), all the extensions are identical to the general toxic load model of Equation (1).

CONSTANT-CONC. EXPOSURES: ESTIMATION OF TOXIC LOAD PARAMETERS

In order to validate the proposed extensions of the toxic load model (Equations 4 – 8), we first fit the basic toxic load model described in Equation (1). Based upon our assessment of goodness of fit, we find that a toxic load model with parameters n = 1.36, TL_{50} = 5.41 x 10^4, and m = 6.17 well captures the lethality of HCN in a healthy population of male Sprague-Dawley rats resulting from constant concentration exposures of 10-30 minutes in duration. The details of the model fitting are described in a companion paper (Slawik, et al., 2016). We exclude exposures of shorter duration (2.3 and 5 minutes) because they fit the model poorly.

TIME-VARYING EXPOSURES: VALIDATION OF TOXIC LOAD EXTENSIONS

All five toxic load model extensions (Equations 4-8) can be shown to fit the 10-30 minute data poorly in plots of predicted versus observed casualties. The Griffiths-Megson and Average Concentration models (equations (8) and (4) respectively, provide the best overall casualty predictions. Figure 1 depicts the accuracy of the Average Concentration model in predicting the 10 and 30 minute time-varying exposure data with and without a time gap.

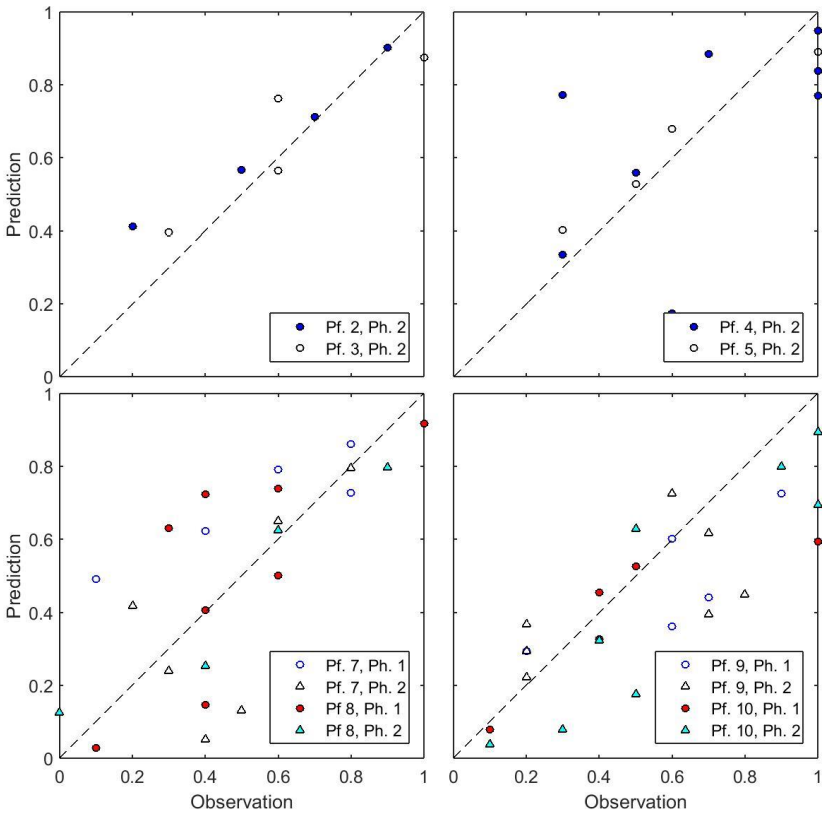

Figure 2. Predicted versus observed fractional casualties for time-varying 10 and 30 minute duration profiles using the Average Concentration model, equation (4). The time-constant 10-30 minute duration data are used to fit the toxic load parameters to generate predicted values. Each dot denotes a specific trial (10 rats). Color and shading style denotes each profile (Pf.) and phase or year of experiment (Ph.). For the 10 minute profiles, casualties are slightly over-predicted, and for the 30 minute profiles, there is scatter in the casualty comparisons. For time-varying exposures without a gap (left column), the Griffiths-Megson model, equation (8), is identical to Average Concentration model.

QUANTIFYING BIAS AND SCATTER

The accuracy of the toxic load model extensions can be assessed by comparing model predictions to observations. Careful inspection of the Average Concentration model's predictions in Figure 2 on a profile by profile basis reveals either over-prediction or under prediction bias for many profiles, and the scatter of the data points is easily visible. The average absolute error in casualty prediction is 1.6 rats, or about 16% of the total sample size. Since only ten rats are exposed for each trial, sampling error can be significant. Taking the predicted casualties as our starting point, we can determine whether the observed difference between the model and the data is larger than that expected to arise by chance alone due to sampling.

We choose the mean square error (MSE) statistic to quantify scatter and the absolute value of mean error (AME) statistic to quantify bias (ie. a tendency to over-predict or under-predict). If the rat population sampled in each trial is assumed to be perfectly described by the toxic load model extension, the observed casualties will be distributed according to a binomial distribution centered at the predicted number of casualties. We can capture the expected distribution of a statistic due to sampling variability alone using Monte Carlo simulations to sample the binomial distribution. If the observed MSE and AME for each profile is uncharacteristically large, then the toxic load model extension in question provides poor predictions. Tables 1 and 2 compare the performance of the five tested toxic load model extensions using the MSE and AME statistics respectively. Performance is measured via the p-value (the probability of obtaining as extreme a value) estimated by Monte Carlo simulation with 10,000 trials. If

the agreement between the model predictions and observations is perfect, the p-values should be evenly distributed between 0 and 1; clustering of p-values near 1 denotes poor predictions. We choose an error in the 90th percentile to denote a poor fit. Each combination of exposure profile and toxic load model extension is marked in red or green to denote "bad" or "good" fits respectively.

Table 1. Variance of Predictions: p-values of mean squared error statistic

Profile, Year	Profile type	Duration	Griffiths-Megson p-values	Ave. Conc. p-values	Ten-Berge p-values	Conc. Int. p-values	Peak Conc. p-values
Prof. 2, 2013	No gap	10 mins	0.3728	0.3728	0.9723	0.9948	0.9998
Prof 3, 2013	No gap	10 mins	0.3990	0.3990	0.8206	0.8990	0.9793
Prof. 7, 2012	No gap	30 mins	0.9746	0.9746	0.9918	0.9952	0.9997
Prof 8, 2012	No gap	30 mins	0.9934	0.9934	0.9938	0.9968	0.9996
Prof 7, 2013	No gap	30 mins	0.9800	0.9800	0.9993	0.9996	1.0000
Prof. 8, 2013	No gap	30 mins	0.4159	0.4159	0.4871	0.5881	0.8043
Prof. 4, 2013	Gap	10 mins	1.0000	0.9998	1.0000	1.0000	1.0000
Prof. 5, 2013	Gap	10 mins	0.8495	0.1781	0.9964	0.9989	0.9998
Prof. 9, 2012	Gap	30 mins	0.5639	0.8070	0.6333	0.6776	0.8840
Prof. 10, 2012	Gap	30 mins	0.9302	0.9260	0.9997	1.0000	1.0000
Prof. 9, 2013	Gap	30 mins	0.8851	0.9845	0.9647	0.9804	0.9958
Prof. 10, 2013	Gap	30 mins	0.9656	0.9989	0.8944	0.8904	0.9203

Table 2. Bias of Predictions: p-values of absolute mean error statistic

Profile, Year	Profile type	Duration	Griffiths-Megson p-values	Ave. Conc. p-values	Ten-Berge p-values	Conc. Int. p-values	Peak Conc. p-values
Prof. 2, 2013	No gap	10 mins	0.6882	0.6882	0.9964	0.9997	1.0000
Prof. 3, 2013	No gap	10 mins	0.4002	0.4002	0.9135	0.9681	0.9964
Prof. 7, 2012	No gap	30 mins	0.9864	0.9864	0.9979	0.9990	1.0000
Prof. 8, 2012	No gap	30 mins	0.9211	0.9211	0.3071	0.5740	0.9791
Prof. 7, 2013	No gap	30 mins	0.5429	0.5429	0.9924	0.9986	0.9999
Prof. 8, 2013	No gap	30 mins	0.4322	0.4322	0.6243	0.7895	0.9328
Prof. 4, 2013	Gap	10 mins	0.9314	0.2194	1.0000	1.0000	1.0000
Prof. 5, 2013	Gap	10 mins	0.9639	0.2483	0.9996	0.9999	1.0000
Prof. 9, 2012	Gap	30 mins	0.4431	0.9009	0.6398	0.7195	0.9513
Prof. 10, 2012	Gap	30 mins	0.8280	0.5758	0.9999	1.0000	1.0000
Prof. 9, 2013	Gap	30 mins	0.0179	0.9852	0.9161	0.9460	0.9972
Prof. 10, 2013	Gap	30 mins	0.9363	1.0000	0.0301	0.2928	0.7334

Table 3 summarizes the results of the Monte Carlo simulations, noting the fraction of profiles without uncommonly large scatter, bias, and scatter or bias. The Griffiths-Megson and Average Concentration model provide the least inaccurate predictions, but their predictions are still poor, failing over half of the profiles tested. The peak concentration model clearly performs poorly, and the commonly used ten-Berge model provides accurate predictions for only a quarter of the profiles.

Table 3. Toxic load model extensions' overall performance in predicting casualties

Metric	Griffiths-Megson	Average Conc.	Ten-Berge	Conc. Intensity	Peak Conc.
# profiles with acceptable scatter	6 of 12	5 of 12	4 of 12	4 of 12	2 of 12
# profiles with acceptable bias	7 of 12	7 of 12	4 of 12	4 of 12	1 of 12
# profiles with acceptable bias and scatter	5 of 12	4 of 12	3 of 12	3 of 12	0 of 12

CONCLUSIONS

The disagreement between the toxic load model predictions and the 10-30 minuet exposure data is much higher than that expected due to sampling error alone. Systematic experimental error or some physiological process in the rat not described by the models could explain this disagreement. The failure of these models to accurately describe the time-varying exposure data is troubling considering the importance modelling casualties arising from time-varying exposures in real-world airborne hazardous release incidents.

ACKNOWLEDGEMENTS: This effort was supported by the US Defense Threat Reduction Agency (through Mr. Richard J. Fry as the project monitor) and the Institute for Defense Analyses' professional development program. The views expressed in this paper are solely those of the authors.

REFERENCES

ten Berge, W.F., van Heemst, M.V., 1983: Validity and accuracy of a commonly used toxicity-assessment model in risk analysis, Fourth International Symposium on Loss Prevention and Safety Promotion in Process Industries, Vol. 1, Institute of Chemical Engineers, Rugby, United Kingdom, pp. 11-112.

Bliss, C. I. (1935). The calculation of the dosage-mortality curve. *Annals of Applied Biology*, *22*(1), 134-167.

Czech, C., Platt, N., Urban, J., Bieringer, P., Bieberbach, G., Wyszogrodzki, A., Weil, J., 2011: A comparison of hazard area predictions based on the ensemble-mean plume versus individual plume realizations using different toxic load models, Paper 2.5 at the Special Symposium on Applications of Air Pollution Meteorology, 91st Annual American Meteorological Society Meeting, Seattle, Washington, Jan. 2011.

Slawik, A.J., Axelrod, K.C., Silva, J.B., Dimitrov, I.K., Urban, J.T., Platt, N., 2016: Are toxic –load-based toxicity models consistent with experimental observations? Independent analysis of steady-exposure data from the 2012–2013 ECBC/NAMRU-D toxicological experiments, Proceedings of the 17th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, 9-12 May 2016, Budapest, Hungary

Sweeney, L. M., Sommerville, D. R., & Channel, S. R. 2014: Impact of non-constant concentration exposure on lethality of inhaled hydrogen cyanide. *Toxicological Sciences*, **138**(1), 205-216.

Sweeney, L. M., Sommerville, D. R., Channel, S. R., Sharits, B. C., Gargas, N. M., & Gut, C. P., 2015: Evaluating the validity and applicable domain of the toxic load model: Impact of concentration vs. time profile on inhalation lethality of hydrogen cyanide. *Regulatory Toxicology and Pharmacology*, **71**(3), 571-584.