

Chemical and Biological Test and Evaluation—Detector Agent Simulant Relationship

Charlie Holman, Ph.D.

Army Test and Evaluation Command, Alexandria, Virginia, and Biology and Mathematics Departments,
Claflin University, Orangeburg, South Carolina

Andrew G. Loerch, Ph.D.

Systems Engineering and Operations Research Department,
George Mason University, Fairfax, Virginia

Realistic testing of chemical and biological defense systems requires an actual warfare agent. But use of such an agent is restricted to laboratory containment chambers, which are not realistic. This state of affairs has driven the chemical and biological defense community to integrate developmental testing and operational testing. Systems are challenged with both agent and simulant in laboratory containment chambers during developmental testing. A simulant is a substance that resembles the agent from the perspective of the system under test. A three-step procedure is described in this article to relate performance when challenged with simulant during operational testing to performance when challenged with agent. The procedure is based on classical logistic regression and judgment. If there is no statistical difference in performance between the agent and the simulant, then the results of the field test with the simulant can be used to predict agent performance. If there is statistical difference in performance between the agent and the simulant, but that difference is small and the system under test performs better when challenged with the agent than with the simulant, then the simulant performance is a lower bound to agent performance. What is defined as small difference is a matter of judgment. A graphical method is provided to provide insight as to the magnitude of the difference. In all other cases, the logistic regression can be used to predict performance based on operational test challenge concentrations and other parameters from the operational test.

Key words: ALO; chemical and biological defense systems; detector; evaluation; logistic regression; simulant.

An Operational Test (OT) is intended to be a realistic representation of how the system under test will be used by its intended operators in the intended operating environment. An OT includes actual warfighters executing combat missions and using the system under test in the same manner that they would use it in combat. Realistic testing of chemical and biological defense systems requires the use of an actual warfare agent. However, because of treaties, public laws, and a desire not to harm test participants, testers, the general public, or the environment, neither chemical warfare agents nor biological warfare agents are released during operational tests or any field test. Testing with an actual warfare agent is restricted to the laboratory in containment chambers.

Unfortunately, these containment chambers are not realistic environments. This state of affairs has driven the chemical and biological defense community to integrate agent chamber Developmental Testing (DT) with OT (Holman and Berkowitz 2009).

There are three methods by which the chemical and biological test and evaluation community combines or integrates the realism of actual biological or chemical agent chamber testing with the realism of actual warfighters executing missions in combat like environments. These three methods are (a) conducting DT with systems before and after OT, (b) modeling and simulation, and (c) developing agent–simulant relationships (Holman and Berkowitz 2009). A simulant is a relatively harmless substance that has some of the properties of agents and can be released into the environment.

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 2010		2. REPORT TYPE		3. DATES COVERED 00-00-2010 to 00-00-2010	
4. TITLE AND SUBTITLE Chemical and Biological Test and Evaluation-Detector Agent Simulant Relationship				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Army Test and Evaluation Command, Chemical and Biological Defense Evaluation Division, Alexandria, VA, 22302				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 6	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

Conducting agent DT with systems before and after OT can provide keen insight into determining whether using a system in the operational environment will degrade its performance. This type of testing has been used most extensively with protective garments. New Joint Service Lightweight Integrated Suit Technology (JSLIST) protective garments and JSLIST garments that went through 15, 30, 45, and 60 days of OT wear were tested in DT. The DT included swatch tests with liquid and vapor chemical warfare agent and whole system tests with simulant. As a result of this testing, curves were developed that predicted degradation in protection based on the amount of wear (Musgrave et al. 1997).

Modeling and simulation were used to integrate developmental agent chamber tests with simulant OTs for the Joint Service Lightweight Standoff Chemical Agent Detector (JSLSCAD). The JSLSCAD performance was modeled with a hierarchy of three models: (a) a vapor cloud model, (b) a scanning model, and (c) the JSLSCAD model. During the validation and verification process, the model accurately predicted performance of the JSLSCAD when challenged with simulant in open air field tests. The modeling and simulation effort was the backbone of the JSLSCAD performance evaluation (Holman et al. 2007).

Modeling and simulation was also used to evaluate the Joint Biological Standoff Detector System (JBSDS). In this effort, field measurements of the cross-sectional infrared back scatter, ultraviolet backscatter, and ultraviolet florescence of simulant were replaced with laboratory measurements for actual agent and were played back in the system software using the other parameters that were recorded in the system software during simulant release (Shirakawa et al. 2008).

Early efforts at developing an agent–simulant relationship were simply to bound a detector's performance against an agent with its performance against two simulants (Musgrave et al. 1997, 2000). Fitch et al. (2004) recommended developing both better methods to perform an agent–simulant relationship and better biological simulants. He proposed using simulants that are phylogenetically similar to the agents. These Agents of Like Origin (ALO) include the vaccine strains.

This article describes an approach based on logistic regression and judgment to develop an agent–simulant relationship and combine chamber agent test results with OT results, so that an operationally relevant evaluation can be made on chemical warfare and biological warfare agent detectors. This approach was used and is currently being used to evaluate the Joint Biological Point Detection System (JBPDs) (Holman et al. 2008; Moe et al. 2010). Biological warfare agent LE and its ALO-killed simulant are used as an example throughout this article.

Concentration

At some high concentration of an agent, a detector will always detect that agent. This high concentration is above the detection threshold, and the probability of detection is unity. At some low concentration of an agent, a detector will never detect that agent. This low concentration is below the detection threshold, and the probability of detection is zero. As the concentration of agent increases from a level that is undetectable, the probability of detection increases. The probability of detection as a function of concentration tends to be s-shaped or a sigmoid as depicted in *Figure 1*. There are many different sigmoid functions, but the logistic regression model is especially useful to model detection performance (Holman and Berkowitz 2009).

Concentration is the independent variable that has the most pronounced effect on detector performance (Holman and Berkowitz 2009).

As a general rule of thumb, the sigmoid curve is steeper (or vertical) in the laboratory than in the field. This is likely because chamber air when filtered lacks many of the impurities found in the environment. The impurities increase the variability in the detector performance. In addition, there is less measurement error, and hence less variability of response in a chamber than in the field environment.

Agent–simulant relationship procedure

The procedure described here involves testing the detector with an agent and a simulant in a chamber at various concentrations, so that a logistic regression model can be developed. The procedure then consists of three steps:

- Step 1: test of hypothesis – Test to see if there is any statistical difference between the performance of the detector when challenged with a simulant or agent in the laboratory. Ensure that sample sizes are sufficient to adequately control error. If there is no statistical difference in the performance of the detector challenged with agent or simulant, then use the simulant to predict detector performance without a transformation.
- Step 2: analysis of the difference – If step 1 demonstrates that detector performance when challenged with agent is statistically different from its performance when challenged with simulant, determine both the directionality and magnitude of the difference. If detector performance for an agent is always better than performance against a simulant, and if the difference is judged not to be too great, then field performance against a simulant can be used to form a lower bound of performance. If the

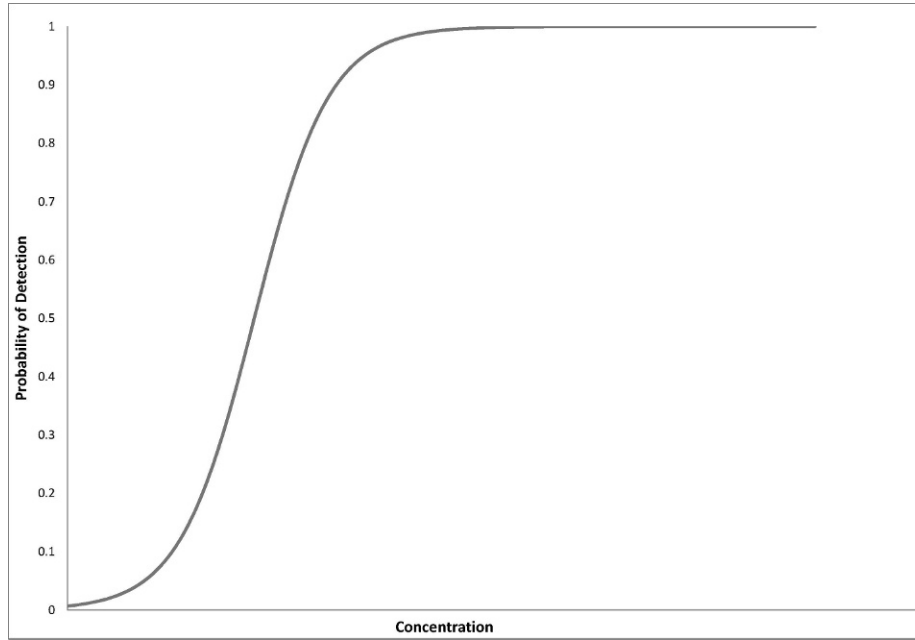


Figure 1. S-shaped or sigmoid curve depicting the relationship between agent detection and agent concentration.

detector performs well enough against this lower bound, we know that the detector will perform better against the agent.

- Step 3: For all other cases, use the logistic regression model to predict performance.

Step 1: test of hypothesis

For the JBPDS LE example, the hypothesis is as follows:

- H_0 : JBPDS performance is the same with either killed LE ALO or live LE agent.
- H_a : JBPDS performance with killed LE ALO is different from its performance with live LE agent.

A classical logistic regression statistical model was constructed for the probability of detection as a function of concentration to determine if the JBPDS detection performance differed between the agent LE and the killed LE ALO simulant. The random component is binary 0 or 1 for no detection or detection (also no identification or identification), respectively. The explanatory variables for this model are agent or simulant concentration, and agent or simulant. The Detection Model is as follows (Allison 1999; Agresti 1996; Hosmer and Lemeshow 1989):

$$\text{logit}(\pi) = \log(\pi/(1-\pi)) = \alpha + \beta_1 S + \beta_2 x$$

$$P(\text{detect}|x, S) = e^{\alpha + \beta_1 S + \beta_2 x} / (1 + e^{\alpha + \beta_1 S + \beta_2 x}),$$

where π = probability of detection; α = shift

parameter; S = 1 if live agent, 0 if killed ALO; β_1 = agent flag shape parameter; x = concentration; and β_2 = concentration shape parameter.

For this model, hypothesis is now equivalent to

- H_0 : $\beta_1 = 0$
- H_a : $\beta_1 \neq 0$

The test statistic is the likelihood-ratio test statistic: $-2 \log(L_0/L_1) = -2(L_0 - L_1)$, where L_0 is the likelihood function without β_1 , and L_1 is likelihood function of the full model. This test statistic is chi-squared with degrees of freedom equal to the difference in the number of parameters between the two models.

As can be seen in Table 1, JBPDS detection performance when challenged with a live LE biological warfare agent is statistically different from its detection performance when challenged with killed LE ALO simulant (P value = .0437). Also, as would be expected, detection performance is a function of concentration (P value = .0161) (Table 1). The Maximum rescaled R -squared is 0.8077 for this model. Live LE and killed LE ALO detection results are based on 62 challenges at various concentrations. The Hosmer and Lemeshow goodness-of-fit test chi-square value is 0.3962 with 6 degrees of freedom, which produces a P value of .99. The deviance goodness-of-fit statistic is 14.50 with 56 degrees of freedom and a P value of .99. Neither goodness-of-fit test is statistically significant, which suggests that the model is a reasonable fit.

It is interesting to note, that the difference in detector performance between the LE agent and killed

Table 1. LE versus killed LE agents of like origin analysis of maximum likelihood estimates.

Parameter	DF	Wald	Pr >
		Chi-square	Chi-square
Intercept	1	5.8085	0.0159
Natural log of concentration	1	5.7941	0.0161
Live LE or killed LE ALO indicator	1	4.0665	0.0437

DF, degrees of freedom; Pr, probability; ALO, agents of like origin.

LE ALO simulant is caused by an inherent difference in the detection of the LE agent and LE ALO and is not caused by the killing process. There is no significant statistical difference in how JBPDS detects live LE agent or killed LE agent (P value = .4564). Nor is there any significant statistical difference in how JBPDS detects live LE ALO or killed LE ALO (P value = .6447). There is, however, a significant statistical difference in detector performance between live LE agent and live LE ALO (P value = .0335).

Since detector performance when challenged with agent is statistically different from its performance when challenged with simulant, we proceed to step 2 to determine both the directionality and magnitude of the difference. Actually, regardless of the outcome of the statistical test, step 2 provides insight as to the nature of the agent-simulant relationship.

Step 2: analysis of the difference

Since the dependent variable is binary, detect or fail to detect, many of the traditional plots used to provide insight into linear regression are of minimal benefit.

Keen insight may be provided by creating a function that is the difference between the predicted probability of the detecting agent-given concentration and the predicted probability of the detecting simulant-given concentration and plotting that function against concentration. In our LE example, we create the following function:

$$\text{LE_DIF} = P(\text{Detect LE}|\text{Concentration}) - P(\text{Detect Killed LE ALO}|\text{Concentration}).$$

Figure 2 depicts a plot of LE_DIF and concentration. The X axis on this chart has been shifted to create an unclassified figure.

From this plot the following can be determined:

- The simulant-killed LE ALO accurately predicts detector performance for LE agent at high and low concentrations.
- The maximum difference in expected detection performance between challenges of LE and killed LE ALO is 0.62.

- The difference in the probability of detection between live LE and killed LE ALO
 - exceeds 0.60 over a range of 5 Agent Containing Particles per Liter of Air (ACPLA),
 - exceeds 0.20 over a range of 23 ACPLA.
- Detection performance when challenged with agent LE is greater than when challenged with LE ALO at the same concentration.

It is not surprising that the simulant-killed LE ALO accurately predicts detector performance for LE agent at high and low concentrations. At some low concentration, the JBPDS can detect neither killed LE ALO nor LE agent; hence the difference is zero. At some high concentration, the JBPDS always detects both the killed LE ALO and LE agent; hence the difference is zero.

The maximum difference in expected detection performance between challenges of LE and killed LE ALO is 0.62. Since the maximum value of a probability is unity, 0.62 is quite large.

The difference in the probability of detection between live LE and killed LE ALO that exceeds 0.60 occurs over a concentration range of 5 ACPLA. The difference in the probability of detection between live LE and killed LE ALO exceeds 0.20 occurs over a concentration range of 23 ACPLA. Both of these concentrations are quite small. A difference of 5 ACPLA is in the noise of measurement error. For field trials, concentration typically ranges from 1 to 16,000 ACPLA. Hence, the magnitude of the difference in detection performance is actually quite small.

The function LE_DIF is formed by subtracting the expected probability of detection of the killed LE ALO given concentration from the expected probability of detection of the live LE agent given concentration. Since this function is always zero or positive, it is clear that the JBPDS detects LE agent at a particular concentration at least as well as it detects killed LE ALO. Hence, the performance when challenged with the simulant-killed LE ALO is a lower bound on what the performance would be if challenged with actual LE agent. If the system performs well enough against killed LE ALO, then we know that it will perform better when challenged with actual LE agent.

If the difference in performance between the agent and the simulant is relatively small, and if the system detects agent better than it detects simulant, then the simulant performance in the field can be used as a lower bound of the performance when challenged with agent.

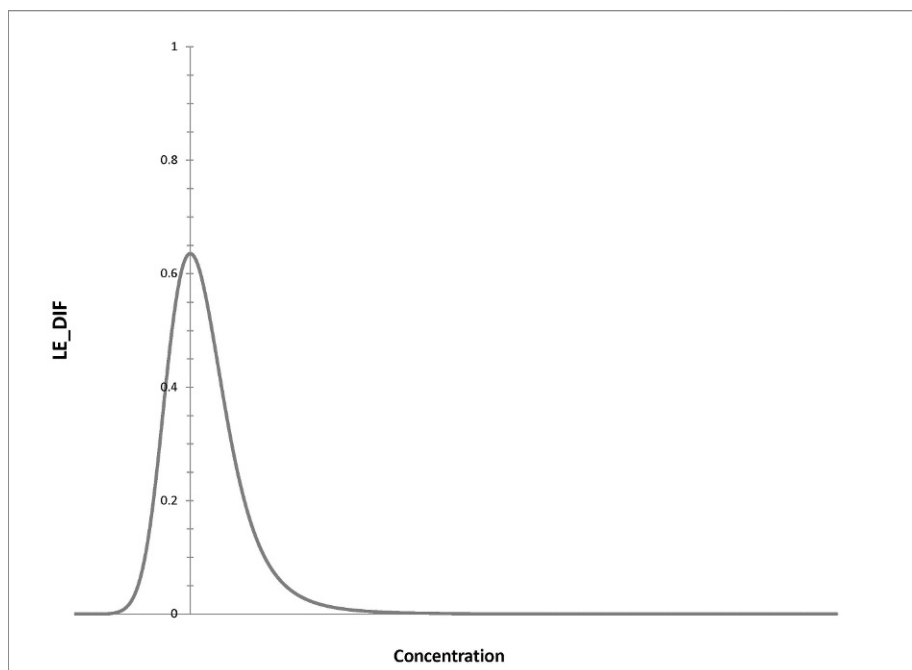


Figure 2. Joint Biological Point Detection System detection performance. In this plot, $LE_DIF = P(\text{Detect } LE|Concentration) - P(\text{Detect killed } LE \text{ ALO}|Concentration)$. DIF = difference and ALO = agents of like origin. (Concentration has been shifted and values left off to create an unclassified figure.)

Step 3: use the logistic regression model to predict performance

The logistic regression model follows and is described above. $P(\text{detect}|x,S) = e^{\alpha + \beta_1 S + \beta_2 x} / (1 + e^{\alpha + \beta_1 S + \beta_2 x})$ can always be used to predict detector performance against agent given concentration. Soldier performance can be incorporated by factoring in releases that would have been missed as a result of maintenance or soldier inattention. As a means of validation, the equation can also be used to predict performance against simulant. The predicted results against simulant can then be compared with the actual simulant performance.

There are two limitations with step 3. First, test results are being estimated by an equation based on concentration as opposed to being measured. Second, since field testing is limited to simulant and no agent, it is being estimated by extrapolation as opposed to interpolation.

Conclusion

The procedure defined in this article is useful in predicting biological warfare agent and chemical warfare agent detector performance against agent in the operational environment based on testing with both agent and simulant in the laboratory during developmental testing and on testing with simulant in the field during operational testing. This method has been used to predict the performance of the Joint

Biological Point Detection System and is currently being used on developmental detectors. $\square$

CHARLIE E. HOLMAN, Ph.D., is Chief of the Chemical and Biological Defense Evaluation Division at the U.S. Army Test and Evaluation Command in Alexandria, Virginia. He has 20 years of experience in evaluating chemical and biological defense systems. He received a bachelor of science degree in biology from Pennsylvania State University, University Park, Pennsylvania; a master of science in biology from Bloomsburg State University, Bloomsburg, Pennsylvania; a master of mathematics from the University of South Carolina, Columbia, South Carolina; and a doctor of philosophy in biodefense from George Mason University, Fairfax, Virginia. He is a member of both the International Test and Evaluation Association and Military Operations Research Society. On August 16, he became a member of the faculty at Claflin University, Orangeburg, South Carolina. E-mail: holmancharlie@earthlink.net

ANDREW LOERCH, Ph.D., is an Associate Professor and the Associate Chair of the Department of Systems Engineering and Operations Research at George Mason University. He holds a master of science in operations research from the Naval Postgraduate School, Monterey, California, and a doctor of philosophy in operations research from Cornell University, Ithaca, New York. He is also

a retired Army Colonel with 26 years of active federal service of which 15 years was spent as a military operations research analyst. He is a Past President and Fellow of the Military Operations Research Society. He is an associate editor of *Military Operations Research*, and is the editor of the MORS newest publication, *Methods for Conducting Military Operational Analysis*. Dr. Loerch directs the track in Military Applications of Operations Research in the master's program in Operations Research at George Mason University. He presently receives funding through the Joint IED Defeat Organization. He also holds a black belt in Tae Kwon Do, is the principle bassoonist in the NOVA Manassas Symphony Orchestra, and is a huge fan of the New York Yankees. E-mail: aloerch@gmu.edu

References

- Agresti, A. 1996. An introduction to categorical data analysis. New York: John Wiley & Sons.
- Allison, P. D. 1999. *Logistic regression using SAS*. Cary, NC: SAS Institute and Wiley.
- Fitch, J. P., B. P. Mark, W. Colclasure, M. Coleman, P. E. Coyle, H. H. Hill, N. B. Jackson, et al. 2004. *Review of test and evaluation methodology for biological point detectors*. Washington, D.C.: The National Academies Press.
- Holman, C. E., and R. Berkowitz. 2009. Chemical and biological defense test and evaluation—A hallmark of integrating DT and OT. *ITEA Journal* 30: 381–385.
- Holman, C. E., M. Moe, C. T. Russell, C. Jennings, R. Bartholomew, and D. Anderson. 2008. "Technical note: Whole System Live Agent Test (WSLAT) methodology and performance assessment." U.S. Army Test and Evaluation publication September 2008.
- Holman, C., I. Kuria, N. Dunn, J. Timmerman, S. Colegrove, P. Serguievski, J. Kleimeyer, et al. 2007. "Report of the modeling and simulation of the Joint Service Lightweight Standoff Chemical Agent Detector (JSLSCAD)." U.S. Army Test and Evaluation publication March 2007. (Classified at the secret level; information used in this article was only unclassified.)
- Hosmer, D. W., and S. Lemeshow. 1989. *Applied logistic regression*. New York: John Wiley & Sons.
- Moe, M., C. E. Holman, D. Kao, E. Heaps, C. Jennings, and A. Thomas. 2010. "Abbreviated system evaluation plan for the Whole System Live Agent Test (WSLAT) phase II of the Joint Biological Point Detection System (JBPDS)." U.S. Army Test and Evaluation publication April 2010.
- Musgrave, D., and C. Holman. 1997. "Biological Integrated Detection System (BIDS) pre-planned product improvement program IOTE test and System Evaluation Report Evaluation (SER)." U.S. Army ATEC Report, Alexandria, Virginia.
- Musgrave, D., C. Holman, S. Tackett, K. Finanger, J. Fuller, R. Jernigan, T. Hillard, and C. Sleeper. 2000. "System Evaluation Report (SER) for the Biological Integrated Detection System (BIDS) Pre-Planned Product Improvement (P3I) System Evaluation Report (SER)." U.S. Army Test and Evaluation publication May 2000.
- Musgrave, D., N. Dunn, F. Gasiorowski, J. Winters, S. Tackett, and C. Holman. 1997. "System evaluation report of the Joint Service Lightweight Integrated Suit Technology I (JSLIST I) Developmental and Operational Tests (DTs and OTs)." U.S. Army Test and Evaluation publication April 1997.
- Shirakawa, K., C.-T. Russell, and C. E. Holman. 2008. *Estimating Performance of a Standoff Biological Detector*. MOR SS June 10–12, 2008, New London, Connecticut: U.S. Coast Guard Academy.