Panel Classification of Self-Reported Exposure Histories: A Useful Exposure Index After a Mass-Casualty Event

David Van Sickle, PhD[a,b]
Mary Anne Wenck, DVM, MPH[a,c]
Amy Belflower, MSPH[c]
Dan Drociuk, MT[c]
Jill Ferdinands, PhD[b]

SYNOPSIS

Objective. Although rapid epidemiologic investigations of toxic exposures require estimates of individual exposure levels, objective measures of exposure are often unavailable. We investigated whether self-reported exposure histories, when reviewed and classified by a panel of raters, provided a useful exposure metric.

Methods. A panel reviewed exposure histories as reported by people who experienced a chlorine release. The panelists received no information about health-care requirements or specific health effects. To each exposure case, each panelist assigned one of five possible exposure severity ratings. When assigned ratings were not in initial agreement, the panelists discussed the case and assigned a consensus rating. Percent agreement and kappa statistics assessed agreement among panelists, Kendall's W measured agreement among panelists in their overall ordering of the exposure histories, and Spearman's rho compared the resultant rankings with individual health outcome.

Results. In 48% of the cases, the panelists' initial ratings agreed completely. Overall, initial ratings for a given case matched the consensus rating 69% to 89% of the time. Pair-wise comparisons revealed 85% to 95% agreement among panelists, with weighted kappa statistics between 0.69 and 0.83. In their overall ranking of the exposure histories, the panelists reached significant agreement (W=0.90, $p<0.0001$). Disagreement arose most frequently regarding probable chlorine concentration and duration of exposure. This disagreement was most common when panelists differentiated between adjacent categories of intermediate exposure. Panel-assigned exposure ratings significantly correlated with health outcome (Spearman's rho=0.56; $p<0.0001$).

Conclusion. Epidemiologists and public health practitioners can elicit and review self-reported exposure histories and assign exposure severity ratings that predict medical outcome. When objective markers of exposure are unavailable, panel-assigned exposure ratings may be useful for rapid epidemiologic investigations.

[a]Epidemic Intelligence Service, Centers for Disease Control and Prevention, Atlanta, GA

[b]Air Pollution and Respiratory Health Branch, Division of Environmental Hazards and Health Effects, Centers for Disease Control and Prevention, Atlanta, GA

[c]Division of Acute Disease Epidemiology, South Carolina Department of Health and Environmental Control, Columbia, SC

Address correspondence to: David Van Sickle, PhD, Department of Population Health Sciences, University of Wisconsin, 610 Walnut St., 707 WARF, Madison, WI 53726; tel. 608-261-1036; fax 608-263-2820; e-mail <vansickle@wisc.edu>.

On January 6, 2005, a freight train carrying three tanker cars of chlorine was inadvertently switched onto an industrial spur, where it collided with a parked locomotive. The train derailed and one of the chlorine tankers was breached, releasing approximately 46 tons of chlorine immediately, and an additional 14 tons over the next three days until a patch could be applied. The incident occurred on the premises of a large textile mill, where approximately 180 employees were working, and in the center of a residential town with a population of 7,000.

As part of a rapid epidemiologic assessment, the South Carolina Department of Health and Environmental Control (SC DHEC) immediately began documenting the health effects on area workers and residents. Although the assessment required estimating the level of chlorine exposure to people who sought medical care, no validated laboratory test was available to quantify chlorine exposure in biologic specimens. In addition, during the primary exposure period, little objective information was available about chlorine levels. The collision occurred in an area where routine monitoring of air quality is not performed, and where—at the time—no air monitors capable of detecting chlorine were in place. Before air-monitoring equipment was brought in to measure ambient concentrations of chlorine, state and local officials evacuated a one-mile zone around the accident site.

Without biomarkers or air-monitoring data to measure individual exposure to chlorine,[1] modeled concentrations based on distance and direction from the accident site offer the best available estimates of personal exposure.[2] However, some individuals treated at area hospitals reported chlorine exposure at locations outside the initial plume models. The length of time required to develop accurate exposure models—and the difficulty in allowing for the many personal-level factors that played important roles in the dose received—limited their potential application to a rapid epidemiologic assessment. In addition, ambient measurements have been shown to underestimate the actual exposures of individuals as determined by personal monitors.[3,4]

These obstacles highlighted the need for an alternative exposure measure to guide the subsequent epidemiological investigation. As detailed information about exposure became available relatively quickly—from interviews completed with people who sought medical care—the team developed a method for using these reported exposure histories to generate an estimate of exposure.

This article examines whether such self-reported exposure histories—when reviewed and classified by a panel of public health responders—can be used as an interim index of exposure. The objectives of the study were to (1) assess agreement in rating the severity of exposure using information obtained during interviews and (2) examine the degree to which panel-assigned severity of exposure ratings correlated with health outcomes.

METHODS

Interviews

To identify people affected by the chlorine exposure, SC DHEC began active case finding among physicians and health-care facilities in the surrounding areas. The department instituted mandatory reporting of people treated for chlorine-related symptoms. It also obtained and reviewed hospital emergency department logs, which included both contact and basic demographic information.

SC DHEC developed a questionnaire to obtain exposure information, symptoms experienced, details about decontamination and transport to medical care, psychosocial impact, preexisting cardiac or pulmonary medical conditions, and effects on pets. Two days after the collision, SC DHEC met with and administered the questionnaire to a small group of people who were still hospitalized or were in a Red Cross shelter. Others who had sought medical care for chlorine exposure were questioned by telephone. Those who could not be reached after three attempts by telephone were mailed letters requesting that recipients contact an answering service at a toll-free number. The majority of interviews were conducted by telephone. Although a total of 19 interviewers conducted interviews, five interviewers conducted the majority of the interviews. In most cases, only one person per household was interviewed.

During the interview, interviewees provided their exposure narratives. On a prepared interview form, the interviewers recorded salient details and solicited answers to a set of questions designed to estimate proximity to the chlorine release and duration of exposure (Figure 1).

The degree of medical care the person required was systematically coded on a nine-point scale, using information obtained from a review of medical records, such as the need for and length of hospitalization or of intensive care support.[5] For example, people who died as a result of the chlorine exposure were assigned a 1, and people who required intubation and intensive care were categorized as a 2. Additional details about the characteristics and distribution of health outcomes appear in the companion report describing the rapid epidemiologic assessment.[5]

Figure 1. Excerpt from questionnaire and interview guide

Please describe your exposure.

Ask additional questions to clarify points in the narrative:
  Where were you exposed to chlorine?
  Who else was present at this location?
  What time did you become aware of exposure?

Ask additional questions to obtain additional details about items described in the narrative when needed:
  Method of evacuation (car, on foot, etc.)
  Sheltering or other methods of protecting themselves
  Route of travel

To supplement the exposure categorization, ask if he/she had the following exposures and, if so, to estimate the duration:
  In the "cloud" of chlorine (yellow/green fog—too dense to see well)
  Not in the "cloud," but chlorine levels high enough to cause coughing, or burning or stinging eyes
  Able to smell chlorine
  No known exposure

Panel session

The five health-care professionals who conducted the majority of the interviews met as a panel to assign severity-of-exposure ratings. These five professionals represented a typical spectrum of public health practitioners, with a wide range of clinical and epidemiological training. None of the panel members were experts in exposure assessment, industrial hygiene, or chlorine toxicity. The team member who conducted the interview read aloud the exposure narrative and the answers to the exposure questions. If the interviewer was not a panel member, the reader was selected at random.

Only the exposure narrative and responses to questions about the duration and proximity to the chlorine release were read aloud. The panelists received no information about health-care requirements or specific health effects. In some cases, interviewees provided information about symptoms experienced during their exposure. If deemed part of the narrative, this information was not excluded from the descriptions provided to the panelists.

After listening to the exposure history, the remaining four panelists simultaneously displayed a card representing, on a five-point scale, their assessment of the exposure: 1 = no exposure, 2 = mild exposure, 3 = moderate exposure, 4 = high exposure, and 5 = extreme exposure. Each panel member's score was recorded. In the event of disagreement, to reach a consensus rating, the group could ask questions of the interviewer to clarify what had been read and discuss the exposure history. Panelists were not, however,

allowed to modify their initial rating. Additionally, panelists were not permitted to inquire about symptoms or health outcomes. The general topics of these questions and discussions were recorded.

Once the group reached consensus and assigned an exposure classification to the exposure history, that number was recorded as the consensus rating. Unless the exposure history given for household members differed significantly, the team assigned only one exposure rating per household—regardless of the number of affected occupants. However, all people in a household were included in the subsequent analysis to examine correlation of the exposure ratings with health outcome. When household members reported different exposure histories, the team assessed each independently and anonymously.

Statistical analysis

Agreement. We used the kappa statistic to measure agreement among all possible pairs of raters about assignment of people into a specific exposure category.[6] Kappa is scaled to be 0 when the amount of agreement is what would be expected to be observed by chance, and 1 when there is complete agreement. Kappa was measured between all possible pairs of raters using standard weights given by $1-|i-j|/(k-1)$, where i and j index the rows and columns of ratings by the pair of raters, and k is the maximum number of ratings. The higher the kappa statistic, the better the agreement between raters. We used the guidelines described by Landis and Koch[7] to interpret kappa statistics. Kendall's W, the coefficient of concordance, measured agreement among the raters in their overall ranking of the entire sample of exposure histories; this statistic ranges between 0 (no agreement) and 1 (perfect agreement).[8]

Correlation between panel rating and health outcomes. One objective was to assess the strength of the relationship between the severity of exposure rating and health outcomes. To accomplish this, the entire sample of people was ranked from least exposed to most exposed, using the sum of the ratings provided by the panel members. Ties were assigned their average rank.

We used the nine-point scale mentioned previously to rank people according to health outcome. The relationship between the sum of the panel-assigned exposure ratings and health outcome was analyzed by comparing the overall rankings of the sample for each variable (i.e., the sum of exposure ratings and health outcome) using Spearman's rank correlation coefficient.[8] This coefficient, which measures the relationship between two sets of ranked data, assumes a value

between −1 and +1. A positive correlation indicates that the ranks of both variables increase together, while a negative correlation indicates that as the rank of one variable increases, the other one decreases. A correlation close to zero indicates no relationship between the ranks. Correlations near −1 or +1 suggest that the two variables are trending together, and that information about one variable can provide some information about the other.

Panel-assigned ratings were analyzed using Stata Version 8.2[9] and SAS Version 9.[10]

RESULTS

The public health response and demographic characteristics of people who sought medical care for chlorine exposure are described in the companion report.[5]

Agreement

Over a period of 5 hours and 35 minutes, a panel of four raters assigned ratings to 213 exposure histories. As a given exposure history could relate to more than one person (e.g., in the case of a household), the 213 exposure histories represented a total of 232 people. An additional 60 people were interviewed after the panel was convened; as they were classified by a different group of raters, they are not included in this article.

Figure 2 shows the distribution of the initial ratings within each of the five final consensus ratings. For histories rated by consensus as having no exposure, the majority of panelists (87%) initially rated that history as having no exposure, while 13% suspected mild exposure, and none were suspicious of a more severe exposure. For histories rated by consensus as mild exposure, most panelists initially rated the history as mild exposure; 6% of raters, however, suspected no exposure, while 7% suspected moderate exposure. For histories rated by consensus as moderate exposure, a relatively larger fraction of panelists initially rated in adjacent categories: 17% initially categorized the history as mild exposure, while 9% initially suspected a high exposure.

Histories rated by consensus as having high exposure displayed the most variation in initial ratings. A total of 18% initially rated the exposure as moderate, while 12% initially classified it as extreme. The least variability was observed in histories rated by consensus as extreme exposure. A total of 1% initially suspected only a moderate exposure, while 10% initially suspected high exposure. Overall, panelists rated people into the subsequent consensus category 69% to 89% of the time, while discrepancies in ratings were largely limited to adjacent categories.
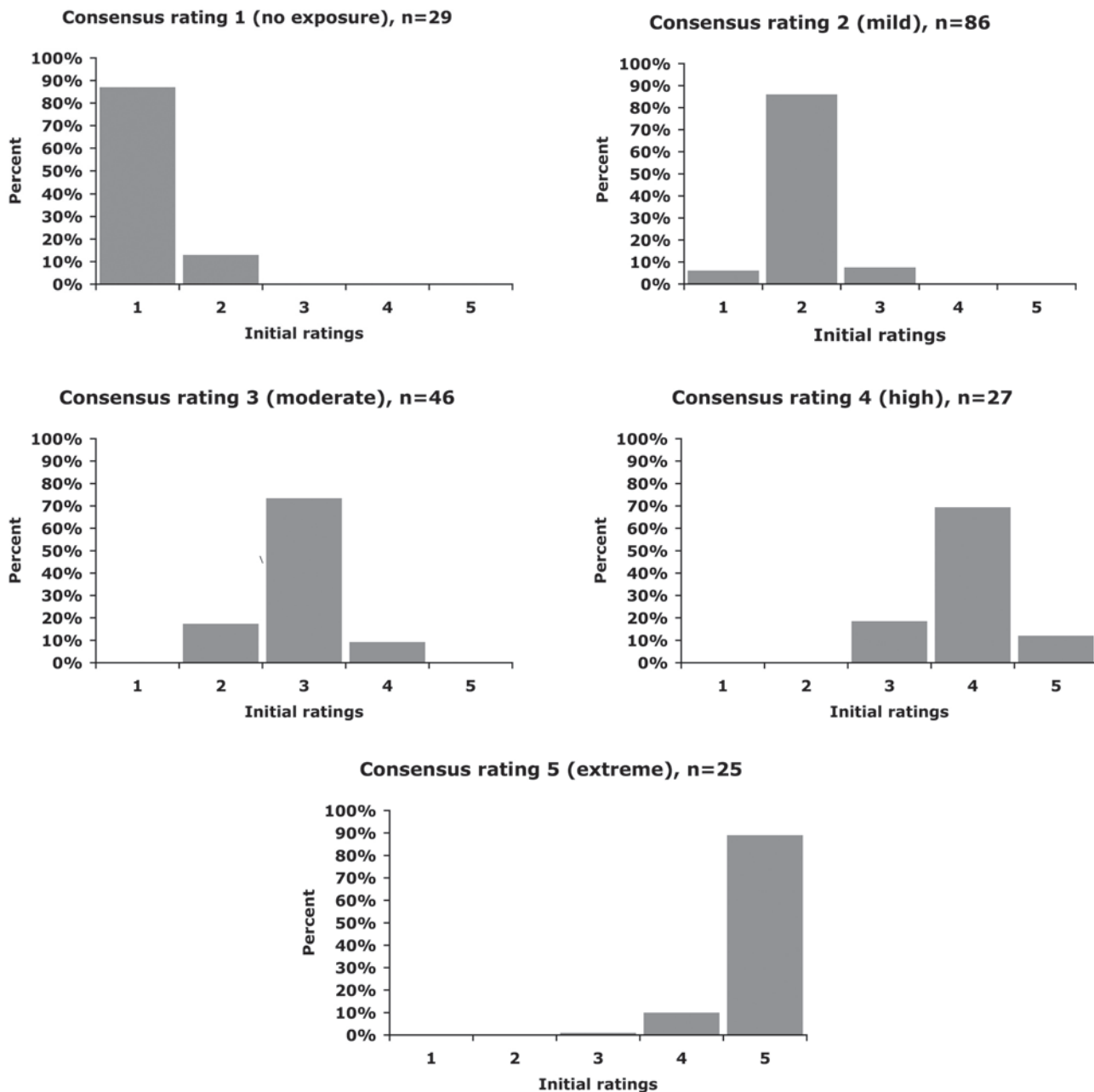
Table 1 summarizes the consensus exposure ratings assigned to the exposure histories and the frequency of initial agreement among panelists. On 102 occasions—48% of the time—the initial ratings of the panelists agreed. The frequency of complete agreement varied across consensus ratings, ranging from a low of 11% among the 27 cases receiving a consensus rating of 4 (high exposure), to a maximum of 19 (66%) among the 29 cases receiving a consensus rating of 1 (no exposure). Agreement between all possible pairs of raters ranged from 89% to 95%, while kappa statistics for these comparisons ranged from 0.69 to 0.83 (Table 2).

Table 3 presents the percentage of exposure histories in each consensus rating by their distance from the collision site. While people in category 1 or 2 reported a range of distances, the majority of people in categories 3, 4, or 5 were located within one-half mile of the collision site. A map showing the locations where individuals reported being exposed to chlorine is provided in the companion report.[5]

The topics discussed by the panel members when their individual ratings did not initially agree ($n=111$) are shown in Table 4. Overall, the most common topics were the likely concentration of chlorine gas present at the time and place of reported exposure, followed by questions about proximity and duration of exposure. Discussion and review occurred most often for people ultimately rated category 2 (55%), followed by those rated category 3 (54%) and 4 (43%). Discussions about concentration were often initiated to differentiate mild exposure from no exposure, and high exposure from extreme exposure. By contrast, questions about duration of exposure and whether the person might have been protected or sheltered from exposure were raised during deliberations on appropriate intermediate classification (i.e., between ratings 2 and 3, or 3 and 4).

The panelists showed significant agreement in their ordering of the entire exposure history sample. The value of Kendall's W (W=0.90; $p<0.0001$) indicates significant agreement between the raters in their overall ranking of all reported exposure histories, from most exposed to least exposed.

Correlation between panel rating and health outcome. To estimate the relative utility of the exposure rating, we examined how well the sum of the panelists' ratings for each person correlated with that person's health outcome. The sum of the four panelists' ratings for a person correlated significantly with that person's ranking by health outcome category (for $n=197$, Spearman's rho=0.55; $p<0.0001$). The findings were similar when comparing consensus severity ratings rather than the sum of panelist scores as the indicator variable.

Figure 2. Distribution of initial exposure ratings (percent) among histories with consensus exposure ratings 1–5

Consensus rating 1 (no exposure), n=29

Consensus rating 2 (mild), n=86

Consensus rating 3 (moderate), n=46

Consensus rating 4 (high), n=27

Consensus rating 5 (extreme), n=25

DISCUSSION

This article describes a method for assigning a severity-of-exposure rating to people affected by a toxic gas event when other exposure metrics are unavailable. We observed significant agreement among the panel of raters in (*1*) assigning ratings to self-reported exposure histories elicited from personal interviews and (*2*) in the overall ranking of the severity of the affected person's exposures. As assessed by kappa statistics, the level of agreement among all possible pairs of raters, by one commonly accepted interpretation, ranged from substantial to almost perfect.[7] These ratings significantly correlated with health outcome. Among raters, questions about exposure concentration were

Table 1. Frequency and distribution of severity-of-exposure ratings across the panelists

Panelist/rater	Severity-of-exposure rating N (percent)					Total rated histories
	1 = no exposure	2 = mild exposure	3 = moderate exposure	4 = high exposure	5 = extreme exposure	
Rater 1	23 (14)	68 (41)	41 (24)	22 (13)	14 (8)	168 (79)
Rater 2	31 (16)	74 (39)	40 (21)	20 (11)	23 (12)	188 (88)
Rater 3	15 (9)	75 (45)	34 (20)	23 (13)	20 (12)	167 (78)
Rater 4	29 (17)	64 (39)	33 (20)	15 (9)	25 (15)	166 (78)
Rater 5	24 (15)	62 (38)	34 (21)	23 (14)	20 (12)	163 (76)
Consensus	29 (14)	86 (40)	46 (22)	27 (13)	25 (12)	213[a]
Number (percent) with all panelists in agreement	19 (65)	53 (62)	12 (26)	3 (11)	15 (60)	102 (49)

[a]Total percent is >100% due to rounding.

the most common source of disagreement, followed by questions about the proximity of the person to the accident and his or her duration of exposure.

In a toxic gas release, biomarkers of exposure to the chemicals of concern may not exist, or may be too invasive, too expensive, or too time-sensitive to use. In the absence of these objective markers, and to estimate exposure, public health researchers and responders often rely on mathematical models or questionnaires and interviews.[2,11–13] The results suggest that investigators can obtain useful information about the severity of exposure by using an assignment of a consensus rating to classify reported personal exposure narratives. This investigation also demonstrates that individual experts in environmental health or exposure assessment do not necessarily need to make such determinations; a panel of public health practitioners with a diversity of training and experience can more than adequately assign exposure classifications.

Table 2. Percent agreement and kappa statistics[a] between all possible pairs of raters and consensus rating

	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5
Rater 2	93.5 0.78	—			
Rater 3	95.3 0.83	92.6 0.77	—		
Rater 4	92.1 0.75	89.4 0.69	90.2 0.72	—	
Rater 5	92.4 0.75	92.4 0.78	92.1 0.77	90.5 0.75	—
Consensus rating	97.2 0.91	95.3 0.86	96.4 0.89	92.6 0.79	95.2 0.86

[a]P-value for all kappa statistics is $p < 0.0001$.

To improve the validity and reliability of these ratings, epidemiologists and public health practitioners investigating a mass chemical exposure should attempt to elicit an exposure history that includes as much detail as possible about estimated proximity to the event and duration of exposure. For example, a question that provides panelists with specific information about the length of time a person spent inside the evacuation zone can be very helpful when estimating the severity of exposure.

Exposure ratings informed the subsequent epidemiological investigation and public health response in a variety of ways. In combination with information about the health effects and medical care requirements of affected people, exposure ratings were used to identify groups and geographic areas particularly affected by the chlorine release; to determine the need for and scope of subsequent community health assessments; and to target campaigns to deliver additional public health assistance, including influenza vaccinations and mental health services.

Limitations
The results of this study are limited in several ways. First, obtaining an exposure history that did not contain references to physical symptoms was not always possible; knowledge of these symptoms could have influenced the panelist ratings of individual exposure severity. Second, this study does not establish the validity or reliability of the panel-assigned severity-of-exposure ratings in comparison with biological markers of exposure or objectively measured physical outcomes, such as physical examinations or pulmonary function testing.

Third, the results are based on a survey of people who sought medical treatment for exposure to chlorine,

Table 3. Distance from collision site of people, by consensus exposure rating

Distance from collision site, in miles	Percent of people with consensus exposure rating				
	1 = no exposure	2 = mild exposure	3 = moderate exposure	4 = high exposure	5 = extreme exposure
<0.25	3	13	50	23	40
0.25–0.50	48	27	43	73	60
0.51–0.75	7	10	—	4	
0.76–1.00	3	10	5		
1.01–1.25	—	10	2		
1.26–1.50	—	1			
1.51–1.75	—	7			
1.76–2.00	7	2			
>2.00	31	20			

but not all eligible people participated in the survey. As a result, people who did participate in the survey may have differed systematically in the severity of their exposures or in the reporting of their exposures compared with nonparticipants. Finally, self-reported exposure histories may be less useful in the case of exposure to chemicals that, unlike chlorine gas, do not have good warning properties (detectable in air through odor or mild irritation at a level below which they are toxic).

CONCLUSION

Narrative exposure histories provide useful information and important details about personal exposure to toxic gas releases, particularly when objective measures of exposure are unavailable. When reviewed and classified by a panel of raters, these subjective reports may provide a useful exposure index to support rapid public health responses and epidemiologic surveillance.[14,15] Future research should focus on the validity and reliability of these exposure ratings in comparison with objective exposure markers.

REFERENCES

1. Belson MG, Schier JG, Patel MM. Case definitions for chemical poisoning. MMWR Recomm Rep 2005; 54(RR-1);1-24.
2. Williams FL, Ogston SA. Identifying populations at risk from environmental contamination from point sources. Occup Environ Med 2002;59:2-8.
3. Sexton K, Adgate JL, Mongin SJ, Pratt GC, Ramachandran G, Stock TH, et al. Evaluating differences between measured personal exposures to volatile organic compounds and concentrations in outdoor and indoor air. Environ Sci Technol 2004;38:2593-602.
4. Sexton K, Adgate JL, Ramachandran G, Pratt GC, Mongin SJ, Stock TH, et al. Comparison of personal, indoor, and outdoor exposures to hazardous air pollutants in three urban communities. Environ Sci Technol 2004;38:423-30.
5. Wenck MA, Van Sickle D, Drociuk D, Belflower A, Youngblood C, Whisnant MD, et al. Rapid assessment of exposure to chlorine released from a train derailment and resulting health impact. Public Health Rep 2007;122:784-92.

Table 4. Topics discussed to reach consensus when raters were not initially in agreement about exposure rating (n=111), by consensus rating

Topic	Consensus exposure rating N (percent)					Total rated (percent)[a]
	1 = no exposure	2 = mild exposure	3 = moderate exposure	4 = high exposure	5 = extreme exposure	
Concentration	8 (14)	19 (33)	9 (16)	15 (26)	6 (11)	57 (51)
Proximity	3 (6)	22 (41)	17 (31)	9 (17)	3 (6)	54 (49)
Duration	3 (6)	10 (20)	18 (35)	14 (27)	6 (12)	51 (46)
Protection	1 (3)	9 (28)	12 (37)	8 (25)	2 (6)	32 (29)
Other	—	1 (12)	4 (50)	2 (25)	1 (12)	8 (7)
Total	15 (13)	61 (55)	60 (54)	48 (43)	18 (16)	

[a]Exceeds 100% due to consideration of multiple topics for some exposure histories

6. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37-46.
7. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
8. Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991.
9. StataCorp. Stata: Version 8.2. College Station (TX); StataCorp.; 2005.
10. SAS Institute Inc. SAS: Version 9. Cary (NC): SAS Institute Inc.; 2005.
11. Bauer U, Berg D, Kohn MA, Meriwether RA, Nickle RA. Acute effects of nitrogen dioxide after accidental release. Public Health Rep 1998;113:62-70.
12. Dayal HH, Baranowski T, Li YH, Morris R. Hazardous chemicals: psychological dimensions of the health sequelae of a community exposure in Texas. J Epidemiol Community Health 1994;48:560-8.
13. Dayal HH, Brodwick M, Morris R, Baranowski T, Trieff N, Harrison JA, et al. A community-based epidemiologic study of health sequelae of exposure to hydrofluoric acid. Ann Epidemiol 1992;2:213-30.
14. Kinra S, Lewendon G, Nelder R, Herriott N, Mohan R, Hort M, et al. Evacuation decisions in a chemical air pollution incident: cross sectional survey. BMJ 2005;330:1471.
15. Baxter PJ. Evacuation decisions in chemical incidents benefit from expert health advice. BMJ 2005;330:1474-5.