Available online at www.sciencedirect.com
ScienceDirect
Journal of Hazardous Materials 159 (2008) 119–129
Journal of Hazardous Materials
www.elsevier.com/locate/jhazmat

The CSB Incident Screening Database: Description, summary statistics and uses

Manuel R. Gomez[a,*], Susan Casper[b,1], E. Allen Smith[c]

[a] *Chemical Safety and Hazard Investigation Board, United States*
[b] *University of North Carolina School of Public Health, United States*
[c] *US Chemical Safety Board, United States*

Received 11 June 2007; accepted 6 July 2007
Available online 7 October 2007

Abstract

This paper briefly describes the Chemical Incident Screening Database currently used by the CSB to identify and evaluate chemical incidents for possible investigations, and summarizes descriptive statistics from this database that can potentially help to estimate the number, character, and consequences of chemical incidents in the US. The report compares some of the information in the CSB database to roughly similar information available from databases operated by EPA and the Agency for Toxic Substances and Disease Registry (ATSDR), and explores the possible implications of these comparisons with regard to the dimension of the chemical incident problem. Finally, the report explores in a preliminary way whether a system modeled after the existing CSB screening database could be developed to serve as a national surveillance tool for chemical incidents.

Keywords: Chemical process safety; Incident; Accident; Incident screening; PSM; RMP; Chemical accidents; CSB

1. Introduction

The Clean Air Act Amendments of 1990 established the current national federal framework for the prevention of accidental chemical releases from establishments that produce, process, handle and store chemicals.

One outcome of this legislation was the creation of the United States Chemical Safety and Hazard Investigation Board (Chemical Safety Board or CSB for short) modeled after the National Transportation Safety Board. The purpose of the CSB is to reduce chemical releases and its consequences, and one of its basic activities is to investigate chemical accidents and issue reports of those investigations[2] (see www.csb.gov).

The legislation also mandated OSHA to promulgate regulations for process safety management (www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=STANDARDS&p_id=9760), and an EPA regulatory system – the Risk Management Program – that requires manufacturers and users of certain chemicals to implement chemical accident prevention and preparedness activities and to submit reports to EPA every 5 years that include an accident history (http://yosemite.epa.gov/oswer/ceppoweb.nsf/content/caa112r.htm). The EPA regulations apply to approximately 13,000–15,000 establishments that handle approximately 140 extremely hazardous substances above EPA-defined threshold amounts. The OSHA regulations apply to a list of relatively similar but not exactly equivalent chemicals, but there are important differences in the universe of facilities covered by

* Corresponding author. Tel.: +1 202 261 7611; fax: +1 202 974 7611.
E-mail address: manuel.gomez@csb.gov (M.R. Gomez).
[1] Formerly with the US Chemical Safety Board.
[2] This report focuses on a tool used by the CSB to select investigation targets, which is relevant to one of the CSB's major tasks, although not the only one. The legislation also mandates the CSB to pursue two other major areas of activity. One is to issue reports of more general chemical process safety issues with recommendations to "Congress, federal, state and local agencies, including the Environmental Protection Agency and the Occupational Safety and Health and Administration." These may include recommendations for regulations. The other major activity is "establish by regulation requirements binding on persons for reporting accidental chemical releases into the ambient air subject to the Board's investigatory jurisdiction." For more information see the legislative text within the CSB web page (http://www.csb.gov/legal_affairs/docs/Legislative%20Authority.pdf).

0304-3894/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhazmat.2007.07.122

the two agencies, especially regarding non-fuel retail facilities and public facilities in states with federal OSHA jurisdiction.

Despite the intent of this legislation to "reduce the likelihood and consequences of accidental releases," the frequency, character, consequences, risk factors or trends in chemical releases from fixed facilities in the US are not well characterized [1,2]. The available information allows only for a very limited understanding of the problem, especially about whether the problem is getting better or worse and why. There is substantial and relatively detailed information about the incidents in the establishments regulated by EPA's risk management program, and this information is potentially valuable from a public policy perspective. The regulated establishments are presumably the riskiest by virtue of having large amounts of the regulated extremely hazardous substances. The regulated substances and firms represent only a part of the problem, however. There is substantial evidence – to which this report contributes – that many chemical releases with important consequences occur outside of the universe of facilities that are mandated by EPA to report information, yet there is no comprehensive national system to collect information about all chemical incidents.

More than a decade after the promulgation of the 1990 amendments, therefore, there is still an urgent need to more completely and accurately define the dimension of the problem of accidental chemical incidents. This need has been discussed elsewhere [1,2]. In this context, it is useful to explore whether the currently available data can be improved to better understand and address the problem.

The CSB has developed a screening system to identify and select those chemical incidents that it should investigate with its limited resources, from among the many that occur every year. Although this system was not developed for the collection of reliable statistical information, it contains information that can shed light on the dimension and characteristics of the chemical incident problem. This report is a preliminary effort to mine this system for information that may be relevant to public policy. Because of the weaknesses of the available data, the analyses should be seen primarily as a hypothesis-generating exercise.

2. Objectives

The primary objectives of this report are to:

- Briefly describe the Chemical Incident Screening Database currently used by the CSB to identify and evaluate chemical incidents for possible investigations; and,
- summarize descriptive statistics from this database that may help to estimate the frequency, character and consequences of chemical releases across the nation.

In addition, the report:

- Compares some of the information in the CSB database to roughly similar information available from databases operated by EPA and the Agency for Toxic Substances and Disease Registry (ATSDR), and explores the possible implications of these comparisons with regard to the dimension of the problem; and,
- briefly explores whether the existing databases, in modified form, might form the basis for a national surveillance system for chemical incidents.

3. Materials

3.1. The CSB Chemical Incident Screening Database

Since 2001, the CSB has selected chemical incidents for investigation with the help of a system that collects and evaluates information about such incidents from multiple sources on a near real-time basis. The information stored from this screening mechanism is called the Chemical Incident Screening Database. The purpose of this system is to help the CSB identify those incidents that may be most deserving of CSB field investigations in a timely manner, from among the many that occur every year, so as to target limited investigation resources most effectively. The system was not designed as a surveillance instrument and does not currently attempt to comprehensively collect statistically valid data concerning chemical incidents.

The screening system has evolved from a relatively informal mechanism into a more systematic approach, especially after early 2005. Prior to 2001, incident screening was performed somewhat inconsistently in terms of methods and frequency. Between 2001 and 2004, screening methods improved steadily, and, since June 2004, incident screening has become more systematic and comprehensive. It is now performed daily using more comprehensive sources than before. Most importantly, more incidents with lower consequences are routinely identified now than in the period before 2002.

From the beginning, however, the incident screening system has obtained information about chemical incidents in fixed facilities from media sources and some government sources. The daily number is rarely more than a dozen or so, and duplicates are easily identified. The information is "scored" by CSB incident screeners with an algorithm that weights several variables relevant to the worthiness of an incident for investigation. Higher scores prompt the collection of more information, which may in time lead to the deployment of an investigative team. The agency investigates only a very small number of the incidents that come to its attention through this screening mechanism, because its resources are very limited.

The information about the incidents in the database comes from day-of-incident media or other reports, and is therefore of uncertain accuracy. CSB staff seeks additional information for only a few dozen reports each year, and only about 6–12 are eventually investigated, of the more than 600 typically identified. A 2006 CSB report describes this system and the outcomes of its application regarding incident identification and selection in more detail (http://www.csb.gov/index.cfm?folder=news_releases&page=news&NEWS_ID=277).

The CSB system collects information from three sources:

- Incidents identified from thousands of media sources, i.e., national, state, and local TV, radio and newspaper, searched for the CSB by a contractor using relevant keywords;
- Daily reports of relevant incidents from EPA's National Response Center (http://www.nrc.uscg.mil/nrchp.html); and,
- Reports from the National Transportation Safety Board for evenings, nights and weekends (http://www.ntsb.gov/).

Between 2001 and early 2005, the incident information was coded primarily for basic demographic information (date, company name, location, source of report, etc.) and the reported number of fatalities, injuries and evacuations and shelter in place orders. Some of the information was collected in free-text form. Numerous improvements were made since its early stages, and especially in early 2005, when the collection and coding of all the information became more systematic and additional information began to be collected when possible (e.g., very rough estimates of ecosystem damage, estimated property losses).

3.2. EPA incident information

The EPA maintains a database with the information that is required to be reported by establishments that are regulated by the risk management provisions of the Clean Air Act Amendments of 1990. This database goes by the name RMP*Info (see: http://yosemite.epa.gov/oswer/ceppoweb.nsf/content/RMPoverview.htm). This report used information about the reported number of incidents, fatalities and worker injuries in the period 1995–2005 from this database.

RMP*Info contains much more information about chemical incidents than is used in this report. A full description or critical evaluation of the database, however, is beyond the scope of this report. A limited number of analyses of the first reporting period for this database (1995–99) exist in the literatures [3,4], and analysis of the data from the most recent reporting period (1999–2004) is reportedly underway [5].

3.3. ATSDR hazardous substances emergency events surveillance (HSEES) system

The ATSDR has maintained the hazardous substances emergency events surveillance (HSEES) system since 1990. It is a state-based surveillance system for chemical incidents (http://www.atsdr.cdc.gov/HS/HSEES/hsees.html). Currently 15 states report data to the system using a standard protocol; this number has varied between 13 and 15 since 1998. The system's goals can be paraphrased as an effort to count the chemical "events," describe the resulting distribution of deaths and injuries, understand the risk factors associated with the events, and develop strategies to reduce their human impacts.

This report used selected information from HSEES about the reported number of incidents, fatalities and worker injuries, as well as chemicals/substances involved, for the period 1998 to the present. The information collected by HSEES includes events in both fixed facilities and in transportation, but the analysis in this report included data from fixed facilities only, excluding homes and illegal activities. Unlike the EPA database and the CSB system, the HSEES system captures many events that occur outside industrial workplaces (i.e., in offices, schools, etc.), and likely many events smaller than those identified by the other two databases as well. In other words, the definition of "event" is different from and more inclusive than the definitions of "accident" and "incident" by EPA and CSB, respectively, which are understood to refer to workplace events. Also, the HSEES definition excludes incidents that involve "only petroleum." A detailed description of HSEES, numerous reports, and a list of published analyses of the reported data can be found in the ATSDR-HSEES web page (http://www.atsdr.cdc.gov/HS/HSEES/hsees.html).

4. Methods

4.1. CSB data analysis

CSB data were compiled to summarize descriptive statistics as follows:

(i) For the period 2001–05:
   - The number of incidents, fatalities and injuries by year;
   - the number of incidents, fatalities and injuries by state, type of incident; and,
   - proportions of incidents resulting in fatalities, injuries or both.

(ii) In addition, the following information was also examined for the period March 2005–2006 (after substantial improvements were put in place):
   - The sources from which incidents were identified (to understand which sources provide most of the information and the extent of overlap among them); and,
   - selected characteristics of identified incidents (type of incident, impacts, location by state, industry, chemical/substances involved).

4.2. Comparisons of CSB data with other data sources

In addition to the descriptive information of the CSB data, the rate of incidents, fatalities and injuries by year estimated from the CSB data were compared semi-quantitatively to similar estimates from the EPA and HSEES data.

4.3. Extrapolation of HSEES data

The HSEES data covers only 15 states. It was necessary to extrapolate this data to all 50 states in order to permit comparisons to the information in the CSB database, which is national in scope, and also to allow for semi-quantitative comparisons to the EPA database, which captures incidents only in the regulated universe of firms. To extrapolate, the authors used only the data for 10 states for which data were available for the period 1998–2005, excluding six other states that have participated in the system for shorter periods of time. A ratio of yearly incidents per number of manufacturing industries for those 10 states was estimated from the reported number of incidents and the total number of manufacturing establishments in them, as reported by Census Bureau data in the year 2000 (approximately mid-

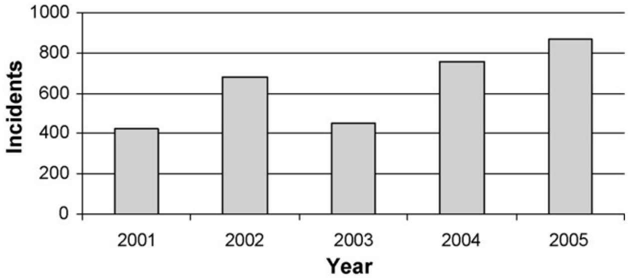

Fig. 1. Total incidents identified in CSB database 12 March 2001–2005.

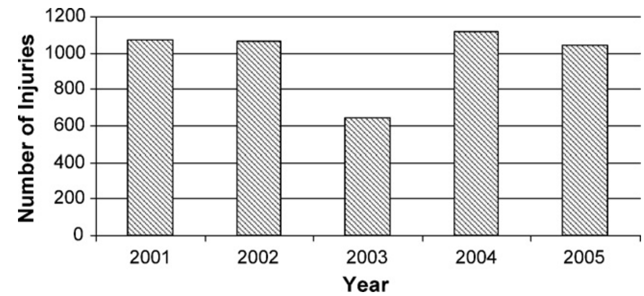

Fig. 3. Injuries identified in CSB database 12 March 2001–2005.

point of the time period in question). The resulting estimated yearly rate of incidents for the 10 states was extrapolated to the number expected for the manufacturing establishments in the entire nation. This admittedly crude extrapolation may introduce a number of biases. For example, the number of manufacturing establishments is not an accurate measure of exposure to process safety risks, of chemical-related activity, or of the number of employees involved in it. Likewise, the mix of manufacturing establishments varies across states. Despite these potential shortcomings, the extrapolation permits some useful comparisons, as discussed in Section 5. In any event, more elegant methods of extrapolation were neither available nor desirable given the uncertainties in the number of incidents (the numerator in the extrapolation).

5. Results

5.1. CSB database. Analysis of 2001–2005 and 2005–2006 data

This section of the report summarizes data that were collected by the CSB between March of 2001 and the end of 2005. The collection and coding of data changed in the early part of 2005, permitting other analyses, which are described in the next section of this report. During this 5-year period, the CSB identified a total of 3478 incidents, with 273 fatalities from 182 incidents, 5142 injuries in 1203 incidents, 96 incidents resulting in both fatality and injury and 1997 incidents without fatality or injury.

Fig. 1 summarizes the total number of incidents identified in the CSB database between March 2001 and 2005. The average number of incidents per year was 637 (adjusting for the “short” year in 2001).

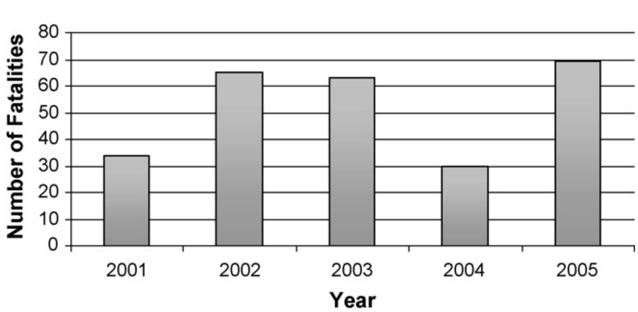

Fig. 2. Fatalities identified in CSB database 12 March 2001–2005.

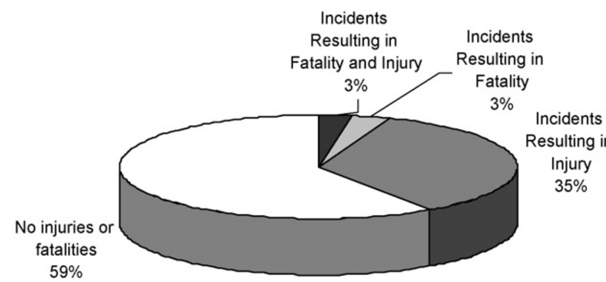

Fig. 4. Percentages of incidents resulting in fatality and/or injury 12 March 2001–2006.

Fig. 2 shows the number of fatalities per year identified in the CSB database between March 2001 and 2005. The average number of fatalities per year was 52.2. Note that the fact that figure may appear slightly suggestive of an upward trend could well be an artifact of the improved sensitivity of the CSB data collection methods, as described earlier.

Fig. 3 shows the number of injuries by year identified in the CSB database between March 2001 and 2005. The average number of injuries per year was 987.

Fig. 4 summarizes the proportions of incidents that resulted in fatalities only (3%), injuries only (30%), both (3%), or none (59%) in this time period.

Fig. 5 reports the number of identified incidents for the 15 states with the highest number in the period from 12 March 2001, to 12 March 2006. Texas accounted for the highest proportion, about 9.8% (309) of the 3140 incidents reported to CSB. California represented the second highest percentage, 6.4% (202) of the incidents, and Ohio accounted for the third highest, 5.6% (176).

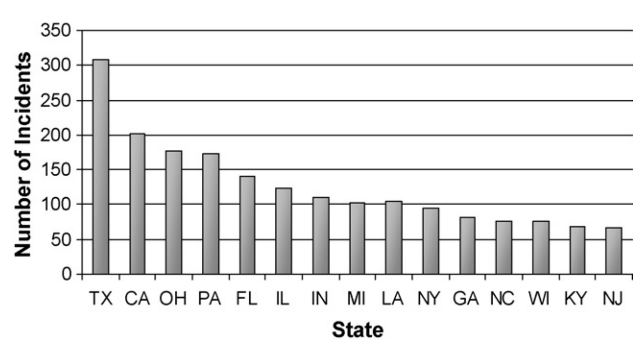

Fig. 5. Number of incidents by state in CSB database. Fifteen states with the highest number of incidents 12 March 2001–2006.

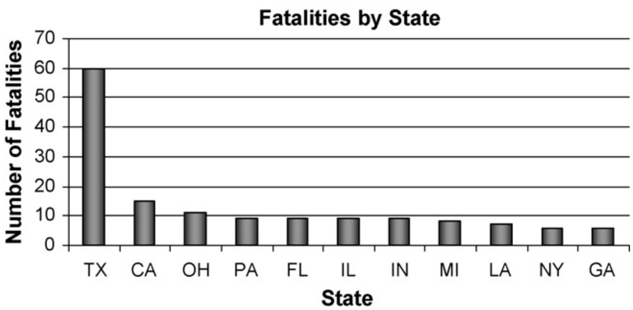

Fig. 6. Number of fatalities by state in CSB database. Includes 11 states with the highest number of fatalities 12 March 2001–2006.

Figs. 6 and 7, respectively, summarize the 11 states with the highest number of fatalities and injuries, respectively. As was the case for the total number of incidents, Texas also had the most fatalities and injuries in the CSB data.

5.2. *CSB database. Analysis of March 2005–2006 CSB data*

This section of the report summarizes the information that was collected in the 1-year period between March of 2005 and March of 2006. The collection and coding of data improved in late 2004 and early 2005 in several ways. As a result, the information is amenable to several additional types of analyses, and, likely more accurate with regard to all the variables. As described below, this 1-year period identified 714 incidents, 52 fatalities and 946 injuries.

5.2.1. *Reporting source of incidents*

CSB's incident screeners receive information about incidents from three major sources. As illustrated in Fig. 8, the majority (70%) of the incidents in this period were recorded as having been identified solely from CSB media searches. The CSB was notified of 18% of the incidents solely by the NRC, and 2% solely by the NTSB. A small proportion of the incidents were identified by more than one source. Five percent were identified by both the media and the NRC, 4% were identified by the media and NTSB, and 1% was identified by all three sources.

Figs. 9 and 10 summarize the sources of information, respectively, for fatalities and injuries identified in the CSB database. To a lesser degree than was the case for the sources of information for incidents as a whole, the majority of the fatalities (55%) and nearly half (49%) of the injuries were identified solely from media sources. These percentages are smaller than those for incidents as a whole. It is of note that NTSB was the sole source of identified fatalities in 31% of the cases, and in 20% of the injuries.

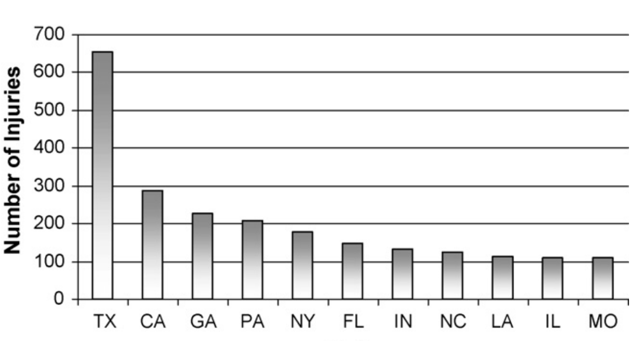

Fig. 7. Number of injuries by state in CSB database. Includes 11 States with the highest number of injuries 12 March 2001–2006.

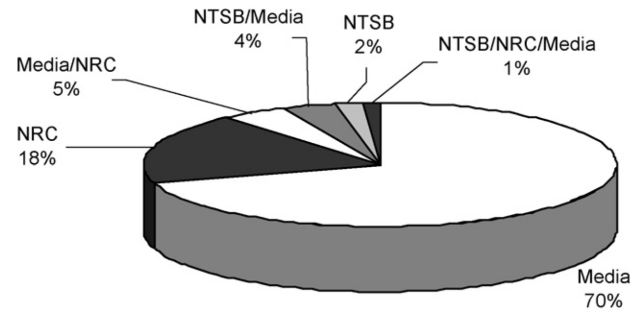

Fig. 8. Recorded incidents by source of reports in CSB database 12 March 2001–2006.

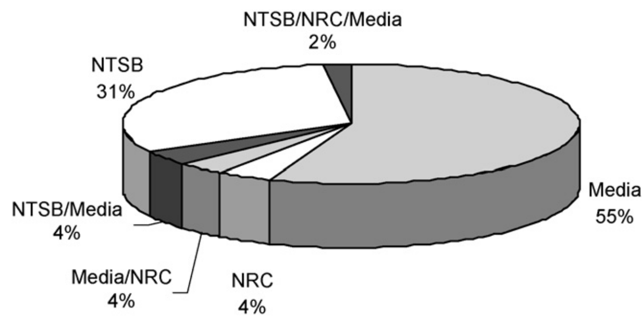

Fig. 9. Fatalities by reporting source in CSB database 12 March 2005–2006.

5.2.2. *Types of incidents*

The incidents are coded as releases, fires and/or explosions, uniquely or in combination. As Fig. 11 indicates, chemical releases (without concomitant fire or explosion) constituted 50% of the incidents identified in the period 12 March 2005–2006.

As Fig. 12 illustrates, from 12 March 2005 to 12 March 2006, 24% (171) of the incidents resulted in public impact, including shelter in place or evacuation, and 76% (712) of the incidents resulted in no public impact.

Fig. 13 illustrates that most of the recorded fatalities (69%) were associated with explosions, although explosions accounted for only 20% of the incidents. Injuries follow a similar trend,

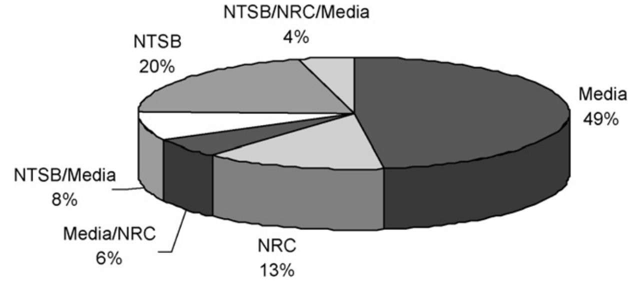

Fig. 10. Injuries by reporting source in CSB database 12 March 2005–2006.

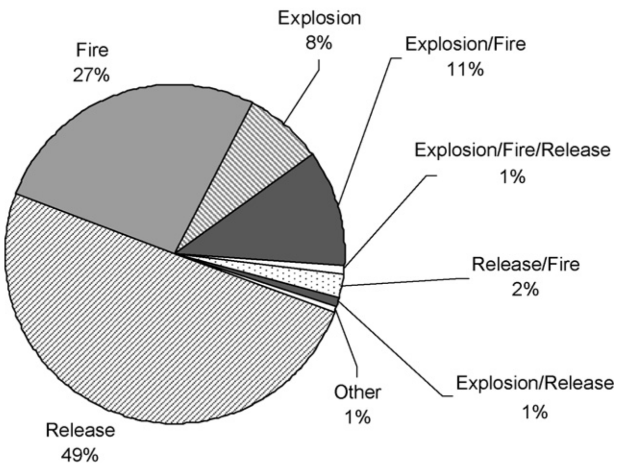

Fig. 11. Incident type 12 March 2005–2006.

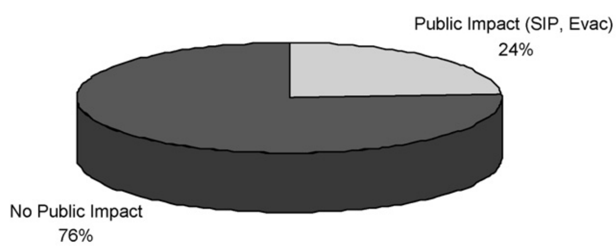

Fig. 12. Public impact of incidents 12 March 2005–2006.

as illustrated in Fig. 15. In the same vein, although releases accounted for 50% of the identified incidents (Fig. 11), only 19% (10) of the 52 total fatalities were associated with them as during that time period, as illustrated by Fig. 13 (Fig. 14).

5.2.2.1. Industry type. As illustrated in Fig. 15, manufacturing industries accounted for 63% (441) of incidents identified by the CSB from 12 March 2005, to 12 March 2006, while 36% (225) occurred in other industries.

Figs. 16 and 17 summarize the distribution of incidents within different sectors of the manufacturing and non-manufacturing sectors, respectively.

5.2.2.2. Chemicals involved. Fig. 18 summarizes the substances most frequently reported to be involved in the incidents

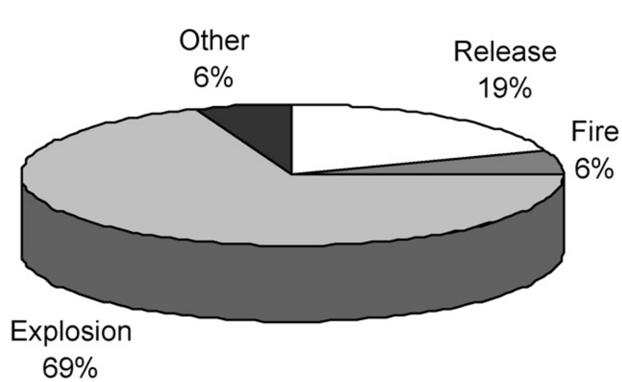

Fig. 13. Fatalities by incident type 12 March 2005–2006.

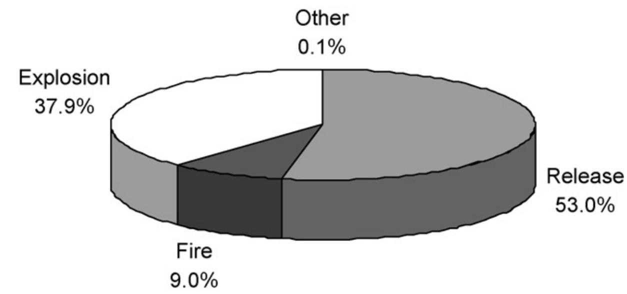

Fig. 14. Injuries by incident type 12 March 2005–2006.

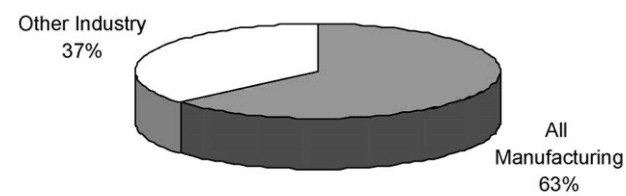

Fig. 15. Manufacturing vs. non-manufacturing incidents, 12 March 2005–2006.

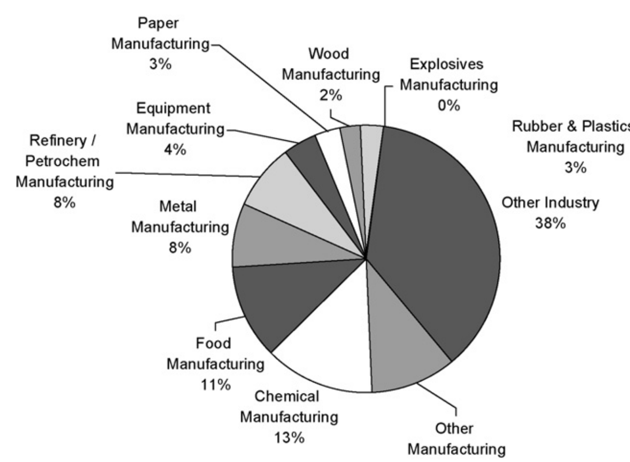

Fig. 16. Distribution of incidents identified in different sectors of manufacturing industries in CSB database 12 March 2005–2006.

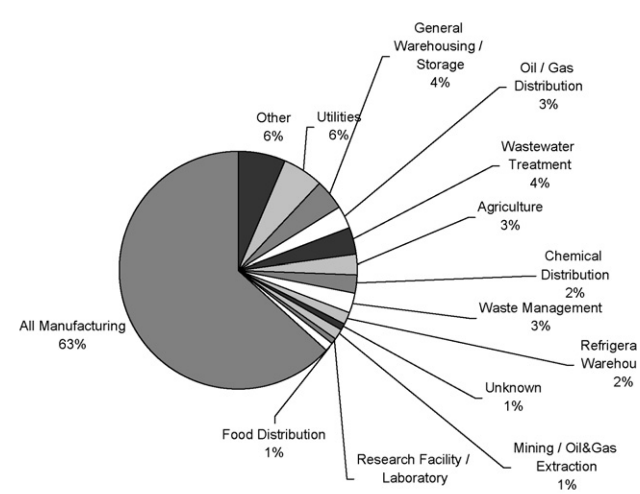

Fig. 17. Distribution of incidents identified within different sectors of the non-manufacturing industries in CSB database 12 March 2005–2006.

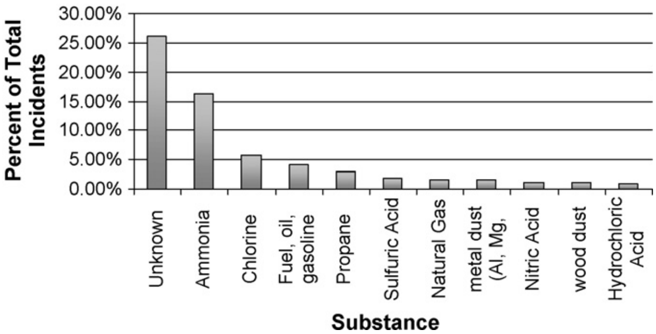

Fig. 18. Substances involved in CSB incidents 12 March 2005–2006.

during this time period in the CSB data. The substances could be identified from the reports in only 74% of the cases; 26% were unknown. Of those incidents for which the substance involved was known, ammonia, chlorine and fuel/oil/gasoline were the three top substances in reported frequency.

5.3. *Comparison of selected CSB, EPA and ATSDR information*

Some of the information from the CSB database can be usefully compared to estimates derived from similar databases at EPA and ATSDR. The comparisons are rough at best, but in the absence of better data, they may shed light on important aspects of the chemical incident problem and how to measure it. At the very least, these comparisons may suggest areas that should be the subject of future research.

Table 1 summarizes the average number of incidents per year identified by the three databases in similar but not exactly overlapping time periods. The EPA estimates include only the firms that are regulated by the Risk Management Program. The HSEES data are rough extrapolations from the data in 10 states, as described before. They also include locations such as offices and schools, and likely events that are much smaller than those in the EPA and CSB records. The CSB data presumably are national in scope.

The average number of incidents identified by the CSB is more than twice the number reported to EPA, and the estimate of incidents from the HSEES is far larger than both. Despite the different reporting periods and universes of fixed establishments involved in this comparison, these differences suggest that a substantial number of chemical incidents occur outside the universe of firms regulated by EPA.

Table 2
Estimated average worker fatalities per year

Source of data	Reporting years	Average fatalities/year (range)
CSB	2001–2005	52 (30–69)
HSEES[a]	1998–2005	47 (range not estimated)
RMP*Info	1995–2005	7 (2–18)

[a] Fixed facility events only, excludes events that include “only petroleum,” events in homes, and illegal activities. Totals reported are extrapolated from data from 10 states.

This conclusion requires at least two assumptions. First, that the “true” number of incidents did not vary substantially in the slightly different time periods covered by the three systems within 1995–2005. The periods compared overlap but are not the same, and the CSB data includes only the latter part of the decade. The second assumption concerns the impact that the different definitions of an “incident” in the three systems (“incident” as roughly defined by the search criteria of the CSB database, “accident” in RMP*Info, and “event” in HSEES) may have on the counts. Very different definitions could at least partly explain the differences. They almost certainly play a role in the huge difference between the HSEES estimates and the other two sources. As discussed earlier, the HSEES system almost certainly includes many smaller incidents than is the case for EPA and the CSB. The EPA system was presumably designed to capture only relatively serious incidents. The EPA and HSEES systems have exact written definitions (of “incident” and “event”) that are presumably followed by the reporting units and states in the two systems, respectively. The CSB’s “definition” is less precise. The identification of an incident results from the use of search words applied to media reports, so that it does not depend on *a-priori* criteria in a traditional manner. In a relatively small number of instances, there is also some judgment on the part of screeners, who may “throw out” an incident that is deemed to be too small.

Tables 2 and 3 summarize the average estimated number of yearly fatalities and injuries, respectively, in the same time periods as Table 1. A pattern similar to that observed in Table 1 can be seen. The average number of worker deaths identified in the CSB and HSEES databases is 6–8 times the number in the EPA reports; and the number of injuries is 2–12 times larger. The same limitations as described above could theoretically account for these differences (different reporting periods and different “incident” definitions). A review of the 52 identified incidents with deaths in the CSB database, however, suggested that all or nearly all would satisfy the EPA definition of “accident,” so that

Table 1
Estimated average incidents per year

Source of data	Reporting years	Average incidents/year (range)
CSB	2001–2005	637 (410–840)
HSEES[a]	1998–2005	16,727 (range not estimated)
RMP*Info	1995–2005	308 (106–476)

[a] Fixed facility events only, excludes events that include “only petroleum,” events in homes, and illegal activities. Totals reported are extrapolated from data from 10 states.

Table 3
Estimated average worker injuries per year

Source of data	Reporting years	Average injuries/year (range)
CSB	2001–2005	987(640–1600)
HSEES[a]	1998–2005	4176 (range not estimated)
RMP*Info	1995–2005	345 (207–446)

[a] Fixed facility events only, excludes events that include “only petroleum,” events in homes, and illegal activities. Totals reported are extrapolated from data from 10 states.

Table 4
Chemicals most commonly reported with incidents

HSEES [6]	HSEES [7]	EPA RMP*Info	CSB
Most frequently released chemicals (all incidents 1998–2001)	Most frequently released chemicals (chemicals & allied products, 1993–2000)	Number of accidents reported by chemical involved 1994–1999	Substances involved in incidents March 2005–2006
Volatile organic compounds (33%)	Other inorganics (21.8.7%)	Ammonia (656)	Ammonia (26%)
Other inorganics (19.7%)	Volatile organic compounds (18.6)	Chlorine (518)	Chlorine (6%)
Other/unclassified (17.3%)	Other (17.6)	Hydrogen fluoride (101)	Fuel/oil/gasoline (4%)
Acids (13.6%)	Mixtures (11%)	Flammable mixture (99)	Propane (3%)
Ammonia (5.8%)	Acids (7.8%)	Chlorine dioxide (55)	Sulfuric acid (2%)
Bases (2.2%)	Ammonia (6.2%)	Propane (54)	Natural gas (2%)
Chlorine (1.6%)	Pesticides (4.1%) multiple substances (4.1%)	Sulfur dioxide (48)	Metal dust (2%)
Pesticides (1.5%)	Polychlorinated biphenyls (0.1)	Ammonia (conc. >20%) (43)	Nitric acid (1%)
Paints & dyes (0.4%)	Bases (3.5%)	Hydrogen chloride (32)	Wood dust (1%)
Polychlorinated biphenyls (0.1)	Paints & dyes (2.3%)	Hydrogen (32)	Hydrochloric acid (1%)

they presumably would have been reportable under RMP if the establishments in which they occurred had been covered by the EPA regulations.

Table 4 illustrates that the chemicals/substances most commonly reported to be involved in chemical incidents were quite similar for the CSB and EPA databases. Some of the differences are easily understood. For example, only the CSB records "wood dust" and "metal dust" as chemical categories.

Two sets of reported HSEES figures are included to illustrate how differences in coding approaches can make it difficult to combine data from different systems for analysis or comparisons. The HSEES data are not directly comparable to the others because ATSDR codes the information in very different groupings. In addition, the HSEES percentages in this table include both fixed facility and transportation incidents. Finally, there are also other differences in the incidents recorded in the HSEES in contrast to those by EPA and the CSB, as discussed earlier. All these factors may affect the types of substances reported to be involved.

5.3.1. *Comparison of CSB data with EPA information*

Although the EPA data, especially for the latest reporting period, is not necessarily a gold standard for incidents among regulated facilities, it is not unreasonable to assume that it may be the most likely to accurately capture relatively serious incidents for the covered establishments, and especially incidents that involve deaths or injuries, because the database is made up of reports that are required by regulation. A comparison of the incidents that appear in the EPA database to those in the CSB database in similar time periods, therefore, may serve as a measure of the completeness with which the CSB media searches (and supplementary government sources) identify chemical incidents.

We carried out several comparisons of the incident information reported to the EPA to the information recorded by the CSB system during the period 2001–2005. All the incidents reported to EPA with worker fatalities in the facility and those with more than five worker injuries were searched by company name, location and date in the CSB data. We did not include incidents involving deaths of responders or the public. In addition, a randomly selected sample of approximately 4% of all the incidents reported to EPA was also searched in the CSB data.

Table 5 summarizes the number and percentage of these three categories of incidents reported to EPA that were also found in the CSB data. Approximately 75% of the incidents with worker fatalities reported to EPA were also found in the CSB data. This percentage improved to 100% in 2004 and 2005, although the numbers of incidents with deaths were very few in those years. This suggests that whatever undercount may have occurred in fatal incidents, it has been corrected under the current system. The CSB system has become more systematic over time, and particularly since early 2005. The table also shows that approximately 57% of the incidents with >5 five worker injuries, and 40% of the randomly selected incidents in the EPA data were also found in the CSB system. There was no obvious "improvement" in the percentage "missing" in the latter time periods in the CSB data, as there was for fatal incidents. Whether this apparent undercount in the CSB data is real, and why it might be occurring is not clear. It is possible, for example, that incidents in the EPA database are reportable but too small to register in media reports. Also, the EPA data may be inaccurate. Confirmation of the accuracy of the EPA data (i.e., is it a valid "gold standard" for comparisons?) and more extensive comparisons are necessary to ascertain whether CSB data reflect a true undercount.

6. Conclusions and discussion

6.1. *The dimension and characteristics of the chemical incident problem*

The CSB database yields an estimate of approximately 600–700 chemical incidents, 52 deaths and 1000 injuries per

Table 5
Percentage of incidents reported to EPA that were found in CSB data 2001–2005

Type of incident	Percentage found in CSB data
Incidents with worker fatalities	75% (12 of 16)
Incidents with >5 worker injuries	57% (16 out of 28)
Total Incidents	40% (8 of 20)

year for the period 2000–2006. Approximately 6% of the incidents resulted in fatalities, and 38% in injuries. Texas had the largest number of recorded incidents, deaths and injuries. This is true for fatalities even if one discounts 2005, when an unusually severe incident killed 15 workers in that state. For the year 12 March 2005–2006, most incidents recorded in the CSB database occurred in manufacturing (63%), leaving a substantial proportion outside of that sector. For the same period, approximately half of the incidents in the CSB database were chemical releases, 29% were fires, and 20% were explosions. Approximately half of the injuries occurred in releases, but 69% of the fatalities occurred in explosions. About one quarter of the incidents had an impact on the general public (shelter in place or evacuations). We did not attempt to compare these proportions to those observed in the EPA and ATSDR databases, but such comparisons are feasible and could prove to be informative. Given the many uncertainties about the data, all of these preliminary findings require additional research for confirmation.

There is no gold standard of information about chemical incidents, so it is impossible to know the extent to which the estimates from the CSB database or the ATSDR projections reflect the true frequency and consequences of such incidents. There are numerous potential shortcomings in both these estimates, only some of which we can mention here. The CSB system, as emphasized above, is not designed for surveillance, but to assist in the selection of incidents to investigate. The bulk of the information comes from media sources, as opposed to an organized reporting system with precise definitions of the events that should be counted and the other parameters that should be recorded. This source is obviously of uncertain accuracy. It could result in an undercount, as some of the earlier comparisons of CSB and EPA data suggest, or in false positives (incidents are included that are not really chemical incidents). In addition, the information about the nature of the events and the fatalities, injuries and other impacts is only that which was available to reporters on the day of the event. This information could also be inaccurate. The information that the CSB receives from NTSB and NRC is also limited to that available on the day of the incidents.

In the case of the projections from the HSEES data, the estimates were derived from an extrapolation of data from only 10 states, and the number of manufacturing establishments, which is at best a very crude denominator for such extrapolations. The estimates for those 10 states are also fraught with uncertainties which are discussed in the publications by ATSDR. In addition, there are large difference in estimates of the number of incidents between the CSB and HSEES data that almost certainly result from different definitions of which incidents should be recorded. Lastly, the comparisons were not made during exactly equivalent time periods.

While the estimates almost certainly do not accurately measure the dimension of the chemical incident problem, they do suggest very strongly that a sizable proportion of the chemical incident problem – at least half or likely much more – exists outside the universe of firms regulated by EPA's RMP regulations, as indicated in the comparisons between CSB, RMP, and, probably to a lesser extent, HSEES data in Tables 1–3. This conclusion is supported by several plausible arguments.

The first supporting argument is the comparison of the yearly fatalities estimated from the three systems. Fatalities are the events that are least likely to suffer from under-reporting, misclassification or other biases, yet the CSB and HSEES databases estimate from four to seven times more deaths per year, on the average, than those reported to EPA under RMP rules for the same (or closely comparable) time periods. A similar pattern was found for the average estimated number of annual injuries (3–12 times the RMP average), and for the average estimated number of identified incidents (2–27 larger than the RMP reported average). This is buttressed by the observation that some of the incidents in EPA's RMP data may have been missing from the CSB database (Table 5), although it remains to be seen whether this apparent undercount is true today, or is a result of the relatively unsystematic collection of data by the CSB in the early years or other factors.

Finally, the CSB investigation experience, while small in numbers, also tends to confirm the picture from these comparisons. Of the 28 CSB investigations for which information was available at the time this report was completed, more than half (57%) were not covered by RMP or PSM regulations. Likewise, fully 75% of the CSB cases involving fatalities occurred in establishments that were not regulated by RMP or PSM, and more than half of the deaths (54%) occurred in non-covered workplaces, despite the disproportionate impact of 15 fatalities in 2005 in a Texas refinery, which were covered by the regulations. These estimates are generous because the firms classified as "regulated" in this analysis include some where only a part of the establishment was covered by either PSM or RMP, but not necessarily the specific process where the incident took place that was investigated by the CSB.

The conclusion that existing RMP and PSM regulations do not address risks of catastrophic chemical incidents in many situations is important for public policy purposes.

6.2. Sources of information

A large proportion of the incidents in the CSB database are recorded to have come from the CSB media searches alone (70%). A smaller proportion of the deaths and injuries come solely from those CSB searches. A substantial percentage of the recorded fatalities in the year examined were identified from NTSB reports relayed to the CSB, but this was probably due to an incident with 15 fatalities in Texas during that time. A closer review of the contribution of each of these sources would be valuable for the potential development of a surveillance system relying on them.

6.3. Comparisons of EPA and CSB data

The comparisons of the CSB data to the EPA RMP data may indicate that the CSB system undercounts some incidents, possibly the ones of smaller consequence. Although the numbers are very small, there appeared to be no undercount for incidents involving worker deaths in the more recent time period for the CSB database, when the system has become more systematic.

We can only speculate whether the possible undercount in the non-fatal incidents is in fact true, and, if so, why it might be occurring. It is quite feasible that the EPA system captures some incidents that are reportable but too small to be considered important enough to be covered by the media sources which constitute the main source of the CSB data. This could be the case, for example, for incidents that involve no serious injuries, occur within a facility's fenceline, do not trigger outside emergency response actions, or occur in large metropolitan areas with an excess of other, more "worthwhile" news items. It is also possible that the EPA data are not an accurate gold standard because of errors in reporting by the regulated firms. Indeed, some of the EPA data in the first reporting time period (1995–2000) is known to have contained mistakes in the number and type of fatalities (due to inaccurate reporting), which EPA and other analysts identified and corrected, but the data as a whole has not been validated, to the authors' knowledge. The possibility of such an undercount in the CSB data deserves additional research nonetheless, both for improved screening and especially if the CSB database model is to be used in any way as a surveillance mechanism.

Another intriguing observation is that only a small proportion of the incidents were recorded as having been identified by the National Response Center (solely or together with other sources). This could be the result of incomplete coding of the data by the CSB, shortcomings in the NRC information, or other reasons. A better understanding of this observation will require additional research.

6.4. *The character of the problem*

All the databases contain information regarding the types of incidents, the industry sectors and chemicals involved, some of the risk factors that may be involved, and other parameters of interest for prevention efforts. This report only skims the possibilities of mining this information, as in the similarities and differences for some of the most frequently reported chemicals involved in incidents in the three databases (Table 4). Some of the references to this paper, and numerous others, provide additional analysis of the factors associated with incidents in both the EPA and ATSDR databases. Additional analyses and comparisons may yield valuable information for preventive purposes, despite the differences in coding between the systems.

6.5. *Prospects for developing a surveillance tool for chemical incidents*

The use of media sources as the main source of information for the CSB screening mechanism, as described here, begs the question of whether and how a similar approach may be used to develop a surveillance system, in the absence of any other mechanism becoming available in the foreseeable future. Such a system would rely on a combination of media searches supplemented by other sources, with increased and well-defined sensitivity, and together with more accurate capture of other relevant incident information.

With today's web and search capabilities, such a system does not appear far-fetched. The possible undercount of the CSB system, if it is confirmed, could theoretically be corrected through enhanced capture of relevant incidents by appropriate modification of the search parameters and better use of supplementary information (e.g., data from NRC, NTSB, EPA, HSEES and possibly others). More precise definitions of the incidents to be counted and of other relevant parameters would need to be developed and applied consistently for accurate coding. These definitions ideally would be harmonized as much as possible with those of existing data sources. Once a surveillance tool with adequate sensitivity is available, it is theoretically not difficult to follow-up the reports with inquiries to obtain additional and more accurate information than what appears in the initial media and other accounts. The CSB currently identifies only a handful of incidents every day (typically fewer than 10 on the average); follow-up for this number of incidents would not be terribly burdensome. A definition of incidents as inclusive as that which is apparently used by HSEES would require much more. The degree of effort to collect additional follow-up information can be tiered according to the severity of the incidents, in some precisely defined manner (e.g., based on factors such as fatalities, injuries, size of the release, large property loss, or community or environmental impacts). Regardless of the definitions that are adopted and used to identify incidents, systematic follow-up could reduce false positives and provide more accurate information about the factors associated with the incidents (causes, equipment involved, amounts released, economic impacts, exact number of fatalities and injuries, etc.). Lastly, numerous commercial databases can easily be automatically searched to obtain relevant information about the characteristics of the firms in which the incidents occur (size of firm, number of employees, industrial classification, parent company, etc.).

If such a surveillance database were to prove feasible, the effort and resources required would be quite beyond the CSB's current capabilities, but there is no reason why other agencies or institutions could not perform this function.

Disclaimer statement

This paper and its accompanying presentation at the Mary Kay O'Connor 2006 International Symposium represent the individual views of the authors. They are not a product of the US Chemical Safety and Hazard Investigation Board (CSB) and their contents have not been reviewed, endorsed, or approved as an official CSB document.

Acknowledgements

We thank the staff of the HSEES within ATSDR for multiple analyses they prepared at our request from the data they collect from states in their systems.

Ms. Hillary Cohen was instrumental in the validation of the CSB data against EPA data and in support of other analysis tasks. She is with the CSB Office of Congressional, Public and Board Affairs.

References

[1] 2001 Assessment of Chemical Safety in the United States, Mary Kay O'Connor Process Safety Center, Texas A&M University, College Station, TX, MKOPSC Report 2002–03, April 2002.

[2] Defining Indicators and Metrics for Measuring Improvements in Chemical Safety, Mary Kay O'Connor Process Safety Center, Texas A&M University, College Station, TX, MKOPSC Report 2002–03, April 2002.

[3] R.P. Kleindorfer, C.B. James, R.E. Michael, L. Kiwan, A.L. Robert, F. Harold, Accident epidemiology and the U.S. chemical industry: accident history and worst-case data from RMP*Info, Risk Anal. 23 (4) (2003) 865–881.

[4] R.M. Elliott, R.K. Paul, A.L. Robert, The role of hazardousness and regulatory practice in the accidental release of chemicals at U.S. industrial facilities, Risk Anal.: Int. J. 23 (5) (2003 October).

[5] Presentation by Isadore Rosenthal, Meeting of Advisory Committee of the Risk Management and Decisions Processes Center, University of Pennsylvania, Wharton School of Business, June 16, 2006.

[6] W.E. Kaye, M.F. Orr, W.A. Wattingney, Surveillance of hazardous substance emergency events: identifying areas for public health prevention, Int. J. Hyg. Environ. Health 208 (2005) 37–44.

[7] D.M. Manassaram, M.F. Orr, W.E. Kaye, Hazardous substances events associated with the manufacturing of chemicals and allied products, J. Hazard. Mater. 104 (2003) 123–135.