

DISCUSSION GROUP 2

Data and Information Necessary for Risk Assessment Applications

Chair

Paul S. Fischbeck, *Carnegie Mellon University*

Panelists

Martha Grabowski, *Lemoyne College and Rensselaer Polytechnic Institute*

Paul Green, *Washington State Ferries*

Sonja Haber, *Human Performance Analysis Corporation*

Najmedin Meshkati, *University of Southern California*

John (Jay) S. Niederhauser, *Puget Sound Pilots*

Stephen Ricks, *Clean Bay*

Discussion Paper

Connell, Linda J. Aviation Safety Incident Reporting: The NASA Aviation Safety Reporting System. (See page 24.)

Description/Objectives

Reliable data about a range of identified risk factors are needed to support complete risk analyses. However, there are considerable difficulties with data to support risk analyses: data sets and information sources can be incomplete, inconsistent, and of different degrees of accuracy and utility for risk assessment. Data difficulties— incompleteness, inaccuracy, inconsistency, unreliability,

unavailability—and the need to integrate data from different sources in order to perform risk assessments are common topics in many risk domains, including the maritime world. This session

- Provides an overview of issues associated with data and information necessary for risk assessment applications,
- Focuses on approaches to overcoming data and information limitations, and
- Summarizes lessons learned from maritime and other domains.

A glimpse of future data needs, sources, and problems is also provided.

SUMMARY OF DISCUSSION

Presented by Paul S. Fischbeck

We looked at data and information necessary for risk assessment applications. As you can see, we had a variety of people from industry, government, and academia. It was a very successful panel.

Why Collect Data?

One point is, if better data are the answer, and we kept hearing that yesterday, we need better data. If that is the answer, what is really the question? What are we really

collecting the data for? Why collect the data? What are we going to use the data for?

We started our discussion with why collect data? One of the first key points that emerged was that collecting data is good business. You can make better business decisions if you have good data. Data have many applications other than risk analysis. If you understand what is going on, then you can improve schedules and so forth. With this demonstration, you also get upper management buy-in. They understand the value of collecting data.

Then, once you have data, you can build models, which we just heard about, and those models have applications aside from just doing risk assessment. They can help you prevent accidents and reduce injuries and impacts from accidents; they can also do things such as justify a government program or show why a regulation is a good idea or not a good idea. They can save you money. By having a good model of operations, you can, in fact, save money. You can also demonstrate use scenario generation and do "what if" analyses not only about risk but about other applications as well. So, you get this buy-in and there is more application to data than just risk analysis.

Types of Data

We talked about the types of data we may want to collect (Exhibit 1). There are accident data and there are near-miss or incident data, which we talked about in detail yesterday. But we also need to know about normal operations. We need to know what happens every day. This is often overlooked. A discussion that came up several times was about how it takes a lot of effort and we have to merge four or five databases. Let's find out what nor-

EXHIBIT 1 Types of Data

- Accident data
- Near-miss (incident) data
- Normal operations
- Baseline (preaccident) conditions
- Real-time ship-specific data
- Reliability of mechanical systems
- Human factors performance
- Environmental conditions
- Organizational norms and procedures
- Causal data
- Need leading indicators (cholesterol) more than deaths (heart attack rates)
- Note: marine data are not aviation data
 - Variety of platforms
 - Controlling agency/organization

mal operations look like. We need to find that out so we can compare how unusual accidents are and how unusual those precursor events are. For instance, for environmental reasons, we also may want to find out what baseline is appropriate. To determine the impact of an accident, we may have to determine the conditions before the accident to know what actually happened.

There was also a need that was brought up for real-time, ship-specific data. If you have a pilot that is going out to a ship, wouldn't it be nice to know what has happened recently, such as how the engine has been performing in the past 24 to 48 hours? What about the crew? Is there any way to retrieve that kind of information? It would certainly help the pilot make better decisions. The information may not be year-end summary data, but at the same time, getting real-time information could be critical.

Other data types include, for example, the reliability of mechanical systems. Where is that information collected? How is that collected? Who is collecting it? And what about human performance, environmental conditions, and organizational norms and procedures? You have to understand the whole thing. You can't just focus on, for instance, human performance. That has been a focus of a lot of discussion, but without the context of other things it is useless; understanding the ships norms in operating for the ship must be part of the context.

There is a need for causal data. Someone mentioned not wanting to know heart attack rates. I want to know the relationship between cholesterol and heart attack rates. That is what is going to be useful to me to make my decisions. So, reporting in year-end summary data will not be what is useful.

Another thing that came up was that maritime industry and maritime data are not the same as aviation industry and aviation data. They are not the same as nuclear power industry and nuclear power data. These are very different environments. You cannot assume that what works well for aviation will work well for the marine industry.

In particular, the variety of platforms makes the marine industry different from other industries. Linda Connell talked about her NASA database. They were dealing with 200 types of platforms. Two hundred types of aircraft configurations. Well, there are 200 different types of ships that pull into the Los Angeles Harbor every day. So the variety and the variability between platforms and the crews that operate them and the way they are maintained is far, far more variable than it is in the aviation and nuclear power industries. This has to be understood and recognized up-front.

Also, if you look at the overarching controlling agencies, when it comes to aviation there tends to be a very nice structure on top. This is the same for nuclear power. There are some tightly controlled regulatory bodies that sit on top. When it comes to the marine industry, they

don't exist at the international or national level. You don't have quite the same control. So, that also makes data collection more important. Other questions arise; for instance, how do you collect data from foreign nationals operating in your waters?

How to Collect Data

How do you collect these data? One thing that came up was that there are a lot of data out there that have already been collected in various ways, shapes, and forms.

But we have to be creative in order to find it. Unfortunately, there is no one-stop shopping. Different people collect different things for different reasons, and to be able to pull it all together and make sense of it is always a very difficult task. The ability to go through and perhaps use the carrots we talked about—the fact that data collection is good business—may be a way to encourage more data collection. At the same time, if you are collecting data from a variety of sources, you must be aware of built-in biases that exist and sometimes hide behind the data.

The data can be everything from low-level data to pre-processed data. There is a hierarchy. When you do census collection, you get every single data point in the population. Then you go to survey data; there was a comment that this is not being done enough. People are always pushing for that complete census. Maybe survey data can be very helpful. Then you fall back on models and then simulation outputs, which we have seen used in the Prince William Sound study. Then there are expert opinions. So, be aware of this continuum of types of data collection.

As was just mentioned, data collection should be rated based on model needs. You don't want to go out and collect data because it is easy and the data are under the lamp post—you have the data, that is great, but does that help you make the decision you're interested in? You have to keep that in focus. There is a technique called the value of information, which is part of decision analysis, that allows you to focus on your data needs.

Then here comes the big guy, which is the incident reporting data. Yesterday Linda Connell talked about the ASRS database and how, hopefully, they will have a real opportunity to get people to report, to put up an immunity deal that allows them to report incidents. So, we can start to capture the concerns that people have on ships. This has to be one of the key ways to do that. If we can't get trust and buy-in from international sources, then we have a real problem; it has to be done. Trust is critical. A system was modeled after ASRS that was started in 1980 for the marine industry. It was a flop. Trust was violated and it was a day of disaster. We are still living with the repercussions and stigma of that particular loss of trust. It has to be there; it has to be guaranteed; it has to be a gold standard that cannot be violated.

Data Quality

We have all different types of data and poor data can, in fact, undermine the credibility of the entire analysis. At the same time, perfect data are never going to be available. You have a battle going back and forth between what you want and what you can get. When are data good enough to include or to make a decision? That is where the modeling problem comes into play. So, it is important to state the source and accuracy of what you have. Don't hide it. Be open with it. Let people know what is going on. There are quantitative methods including probability distributions and qualitative methods for doing that. You should seek more sources. Get verification. That is a key element. Don't be afraid to show what you don't know. If you don't know something, don't say that you do. Put uncertainty on it. Allow for the uncertainty to be truly registered.

Who Has Access to What Data?

You've collected the data. Now, who gets to look at it? Different industries have tried different levels of exposure, different levels of access. Nuclear power has some great databases, but they are very tightly held for many reasons; they are very tightly held within utilities. They are shared among themselves, but there is no access to them from outside.

There was a discussion about open and closed databases and here the key was that it was believed that open was better. For this industry, openness is critical. Once again, immunity and trust are critical. Because you have an open database, you can't let people backsolve and figure out who reported what, what is going on. It has to be solid. You have to sanitize the data fields so you can prevent this backtracking. Proprietary databases have to be honored. At the same time, if you can share the carrot and the advantages of sharing data, then that may break down some of the barriers. Don't be surprised if people misuse the data. You are going to collect all these data; you are going to put the data on the web or somewhere to make them publicly accessible. People will do bad analyses. That is going to happen. But, that does not mean it shouldn't be done.

We had a small discussion, with mixed results, about whether we should charge a fee for access to the data.

How Are the Data Maintained?

You've collected the data; you've decided who has access. The question is, how do you maintain the data? Who owns the data? Is it a private or a public ownership—critical question. Is it a central location or do you have a web page that points to the different data sources? Many

current databases are in very poor condition so it takes time, effort, and money to go through and clean them up and get them into acceptable form. Who pays for that? Who can afford to pay for that? Who is going to do that? Once that is done, that is when you don't want to throw the data away. That is when you want to really make it available to a lot of other people, once you have cleaned it up. But, how do you do that? Let me turn it over to who are the maintainers. If we standardize fields, data merging is much easier. Working with the offshore platform people, having common IDs that are consistent and reliable allows you to merge data together. That is really important.

The other kicker is—who owns the data and who is liable for errors in the data? Suppose there are omissions—something is missing and you make a decision and the data should have been there. They weren't—you make the wrong decision. Who is liable? Is the data owner liable in some way, shape, or form? Lawyers become involved here. They have to come up with disclosures that say, "Here are data, if you use them, it is your own responsibility." What is legal? What is the legality of data errors and so forth?

What Can Be Done with Current Data?

There are a lot of current data out there and a lot of different sources. To be able to go through and start pulling the data we already have and seeing what is out there is an important first step. But, it is tough. Once again, because of the quality of the data, wouldn't it be nice to know all these different things—how good are those databases, what are they being used for today, what could they be used for, how can owners be encouraged to share, and so forth.

So there are a lot of questions about existing databases. We don't know all these things. It would be nice to know—here is a database and here is a taxonomy, here is a description of each of the databases, and here are the limits and strengths and weaknesses and applicability and previous applications and so forth—so we could know what is out there. Who is going to pay for that?

In summary here are some of the major points. Collecting data is good business. That was a big insight to me. I had not realized how important that was. There is a need for this accident, incident, and normal operations data. We need normal operations data. We need to prioritize collection. Let's help the modelers, let's help the decision makers, let's help the policy makers by giving them the data they need—not the data that are easy to find. There is this . . . where we have "poor data can hurt you, but perfect data is not available." You have to understand that and then be open about data limits, the biases, the quality, and the uncertainties of the data you're using. Let's open the data up to the public.

QUESTIONS AND ANSWERS AND DISCUSSION

Question: How many of these types of data are being collected? How many organizations are collecting them?

Answer: One group with a tremendous amount of data are classification societies. They hide behind the idea of liability. The sad part about that to me is that the shipping people basically own those people. We serve on the boards and so on and so forth. There are a lot of data there. I never see anybody mention what goes on in the Nautical Institute and their simple system of reporting accidents. I don't know if that is going to be included or not, but there is a system that goes on that is voluntary, and the amount of data that come in there is amazing. So there is a platform that people will, in fact, report. How far we take it, I don't know. One thing to do is to really go through and find out what data are out there and find out how big a carrot you need to put in front of that person to open it up. If you want to go through and justify, and you can show a bottom-line improvement in profitability by having better data access, I think doors will open. It can span not only risk and accident prevention but also other things—better maintenance, more reliable systems, less dead-in-the-water time.

Question: I want to get a ship owner's point of view here. Accidents and pollution incidents, and knowing what the definition of an accident is, and knowing what the definition of a pollution incident is, it is very important from our standpoint as ship owners to operators to the industry. Perception of it. So, when we start talking about the data that were collected, I think it is important to make sure that we know what the definitions are of the types of data collected. I haven't heard anything said here about the definition of an accident. We had a discussion about what a fatality is. Different people record fatalities differently. The issue of a pollution incident, and some people may think of pollution incidents as being oil in the water, as opposed to . . . as opposed to broken glass or whatever it is. So, it is important to make sure as part of the idea of collecting data that we have definitions of the data we are collecting. I think that if data are recorded as an incident, once again, if you have the specific information that describes what is going on, then different people will define things differently. You are absolutely right that you cannot merge apples with oranges. If one person defines a spill as over 10,000 gallons, and someone else says a spill is anything over 5 gallons, and you merge those databases, you have a real problem. So, going back to the various databases and having a clear idea what is already in there—there was some long discussion about how people are going through and pulling data off the web and doing quick,

ad hoc analyses through them, making bold statements, and they really don't know what the underlying data were or the definitions of what they are dealing with. This can be very problematic.

What is the alternative? The alternative is to have all the data secret or hidden or password protected, or you pay \$1.00 to look at it, and you can control access. So, there are two sides to the coin about how public and how private this should be. But you are absolutely right.

Comment: It is interesting because in the '60s and '70s, the International Chamber of Shipping had an excellent program that had fire and explosion accidents, tanker accidents, navigational accidents, many things like that. It all fell apart around 1980 because of concerns about leaks. You can't overestimate the impact of that on people being willing to even have something reported anonymously. It is a huge problem.

Comment: What you are saying is there are three problems here. One is that we don't really know the status of who has what current data, the reliability of it, and the pluses and minuses to all that is out there. Second, whether they are near-miss data or any other data using some of the existing data sets, there are legal implications on who holds or who releases the data that have to be overcome. That has to come primarily because for the data to be widely accepted and used in risk analysis and decision making, they need to be available to the public. So, there are really three somewhat separate issues. The question I have is this: Did your group discuss who, where, why, or what might be a method base that has the potential to do that sort of thing?

Answer: We talked about the need for a trustworthy organization, one that is, in fact, removed from the biases you might find. So, the IMAS people are going to come here and tell me who that contractor is who is going to maintain that database or how that will be collected. There is a very strong need to make sure the controlling agency is above reproach and that they can be trusted. They have to be shielded from the legal onslaught that is going to occur.

Comment: What I'm hearing you say is that, even if we assume we can overcome the issue to be near-miss systems, there are broader issues relative to currently available data, let alone who we get it from or if we gather from the right sources, that have to be looked at. If we can get one system to work, we have to start somewhere. If we can get the incident material in place and working, that would go a long way to showing everybody else that we can get over these hurdles and the end result is valu-

able for business, for reduction of risk, for saving the environment, for all these different reasons. You then get real big payback at the end.

Question: I was going to ask if you discussed quality management techniques as a method of incorporating this—if you are an inherently competitive industry and if each company addresses it internally as an industry-wide practice?

Comment: Good data collection is good business and improves performance, absolutely. There was some discussion about some organizations that people had witnessed that had collected no data. They were operating big things in dangerous waters and they had no data. So, there is an amazing variety. Some companies track washers and find out how many have been used. Others operate and don't collect anything.

Question: ASRS didn't start out with 34,000 inputs a year. It took 10 years spool-up time. But, if you look at the previous attempt in 1980 for the ship equivalent incident database, that was about \$10 million a year for 4 or 5 years. That was \$50 million down the tubes. Not only that, but it hurt the possibility of it ever happening again because people are still around from 1980 and they are the ones who were burned and they are still here. To get them engaged and involved again requires some real guarantees. But, without that, without one of these steps, taking one of these databases and getting it going and demonstrating the availability, the applicability, and the viability of such a database, we're going to be spinning our wheels again and we'll come back in 5 years and have the same discussion again.

Question: One useful point of input, one of the most useful databases that we've found, is in the insurance PMI direction. The investigation reporting of the accidents is good and it incorporates a great deal of the human factors information. . . .

Comment: Analysts know where the data are. But, to know what data are out there, where they are, and the quality of that data and who to call to get that data, it shouldn't be a secret handshake. It has to be opened up and made available to people.

Comment: The first lawsuit that goes after that data is yet to come. Someone is going to sue to find out the data. We have to have things in place. We have to be proactive and understand that now. This is the reality. This is the world we are working in.