Causality, covariates and consensus in ISCRAM research: towards a more robust study design in a transdisciplinary community

Zeno E. Franco*

Pacific Graduate School of Psychology
Palo Alto, CA 94303, USA
E-mail: zfranco@pgsp.edu
*Corresponding author

Nina Zumel

Quimba Software
San Francisco, CA, USA
E-mail: zumel@quimba.com

Kathy Blau, Knute Ayhens-Johnson and Larry E. Beutler

Pacific Graduate School of Psychology
Palo Alto, CA 94303, USA
E-mail: kblau@pgsp.edu
E-mail: knute@pgsp.edu
E-mail: lbeutler@pgsp.edu

Abstract: Research in disaster management encompasses a variety of academic disciplines. Yet, despite calls to expand the range of methodologies used and elaborate a nascent theory of disaster management, progress towards a transdisciplinary framework is slow. Some reasons for this are explored by focusing on the research efforts of the international community for Information Systems in Crisis Response and Management (ISCRAM). Similar to the primary disciplines it draws from, ISCRAM research is typified by case study evaluations. As a result of poorly articulated case study methodologies and the lack of alternative methods, the confidence in causal and generalisability claims remains questionable. Performance evaluation techniques may close these gaps, but several limiting factors must first be addressed – in particular, parameterising and controlling for context variables must receive more attention. The need for well-explicated covariates, such as a disaster severity index that describes the relative impact *between* incident types, is explored in some detail. The relationship connecting the context and performance assessment variables is briefly considered. Finally, we suggest that the quality of research and theory building is contingent on a deeper, transdisciplinary dialogue about the nature of scientific evidence within ISCRAM – a discussion that may also gradually inform a general theory of disaster management.

Keywords: disaster management performance assessment; major disaster impact scale; methodological pluralism.

Reference to this paper should be made as follows: Franco, Z.E., Zumel, N., Blau, K., Ayhens-Johnson, K. and Beutler, L.E. (2008) 'Causality, covariates and consensus in ISCRAM research: towards a more robust study design in a transdisciplinary community', *Int. J. Emergency Management*, Vol. 5, Nos. 1/2, pp.100–122.

Biographical notes: Zeno Franco is a PhD candidate in Clinical Psychology at the Pacific Graduate School of Psychology, USA, and a Predoctoral Fellow with the National Centre on the Psychology of Terrorism. He recently completed a three-year US Department of Homeland Security Fellowship and specialises in the behavioural science aspects of US policy responses to major disasters and terrorism.

Nina Zumel, PhD, is the Co-founder and Owner of Quimba Software, a research and specialty software development company in the San Francisco Bay Area. She holds a BS in Electrical Engineering and Computer Science and a PhD in Robotics. Her research focuses on training systems and models of behaviour in disaster management.

Kathy Blau and Knute Ayhens-Johnson are PhD students in Clinical Psychology at the Pacific Graduate School of Psychology, USA, and Research Assistants with the National Centre on the Psychology of Terrorism.

Larry E. Beutler, PhD, ABPP, is the William McInnes Distinguished Professor of Psychology and former Chair of the PhD Programmes at the Pacific Graduate School of Psychology, USA. He is a Consulting Professor of Psychiatry and Behavioural Sciences at the Stanford University School of Medicine and a former Visiting Professor of Homeland Security at the Naval Post Graduate School. He currently serves as the Director of the National Centre on the Psychology of Terrorism and the Palo Alto Medical Reserve Corps programme. He is the former Editor of the *Journal of Clinical Psychology* and the *Journal of Consulting and Clinical Psychology*. He is a Fellow of the American Psychological Association and the American Psychological Society.

1 Introduction

Research in the field of disaster management is a dauntingly complex task as it brings together a wide variety of disparate academic fields performing inquiry at the nexus between theory and practice – often gathering data in turbulent environments. The diversity of disciplines contributing to the field only add to the intricacy of the subject by introducing variations in concepts, theories, and perspectives that are discipline specific. While these complexities have been explored to some degree in general disaster research (Stallings, 2002), comparatively little theoretical work in the field of disaster management has been done to ground the research process, create a shared research agenda, or to frame this work adequately in a transdisciplinary context. To be more specific, by *disaster management* we mean the process of individuals, communities, first responders, professional emergency managers, local and regional political leaders, regional and national agencies, and, at times, the highest levels of national governments acting to control the effects of a disaster. Used in this sense, 'disaster management' is a temporally specific term and does not address pre-planning or long term recovery, but specifically the decisions and actions that characterise the acute response and recovery period.

Disaster management shares considerable conceptual overlap with ideas and structures used in the description of emergency management (Drabek, 2004). That being the case, we occasionally use the terms interchangeably here. However, we recognise that by so-doing, distinctions in terms of scale and controllability of the incident are inevitably lost. Given the overlaps and the absence of a common, grounding theoretical structure, we know of no good alternative. Indeed, it is the absence of a unifying theory and the presence of divergent conceptual/theoretical and language systems within this diverse field that stimulated us to write this paper.

Several authors, over a period of decades, have called for more attention, analysis, and research in the areas of emergency and disaster management *theory* – though progress remains slow (see for example, Lalonde, 2007; McEntire, 2004; McEntire and Fuller, 2002; Moore, 1956; Wybo and Latiers, 2006). Different disciplines even disagree in the level of organisation that they accept as characterising a 'theory' – what is a 'theory' to one, may be simply a 'concept' to another. Drabek (2004, p.5), one of the preeminent scholars in disaster research, notes that the broad perspectives drawn from a variety of disciplines, "may provide the basis for 'true' theories of emergency management and/or disaster responses. Collectively, they offer a foundation, but the house, so to speak, has yet to be built." Currently, a comprehensive theory of disaster management is nascent at best.

Disaster management is certainly informed by theories that govern disaster studies specifically (*e.g.*, Normal Accident Theory; Perrow, 1984, *etc*.). However, because disaster management lies at the interface of the disaster and human attempts to control the disaster's consequences, this area of research must also incorporate an incredibly wide array of theories of human behaviour – at the individual, group and population levels (Drabek, 2004; Franco *et al*., 2007; Moore, 1956). Stated simply, theories that seek to explain disasters are necessary but insufficient to describe and predict the human activity involved in managing these events.

In most scientific disciplines, it is thought that theories will be inextricably linked to research questions, and by extension to the specific methods used to test the theories (Wampold, 2003). However, within the behavioural sciences, variations exist in what constitutes an adequate research method to assess a theory (Borkovec and Castonguay, 2006; Hollon, 2006). In part because of the history of disaster research being largely grounded in sociology and in part because of the complexities associated with collecting data in the immediate aftermath of disasters, much of the research in this arena has relied on case studies, interviewing and naturalistic observation. In contrast, quantitatively oriented researchers express concern about the ability of the case study paradigm to adequately address the issue of causality. Ordinarily, the arguments distil to a consideration of how precisely one can control extraneous variables in order to isolate causal chains (Kazdin, 1998). Thus, controlled experiments favouring external validity are considered to be more desirable as indicators of causality followed by quasi-experimental, $n = 1$ studies, groups of case studies, or single-case studies that favour internal validity (Borkovec and Castonguay, 2006; Hollon, 2006; Kazdin, 1998). Each of these methodologies has both its advocates and detractors.

As Drabek (2002, p.121) noted, many have been skeptical of the "…qualitative, and hypothesis generating framework" used extensively by the Disaster Research Centre and other early scholarly efforts in this field. Clearly, much was gained through these processes in the first fifty years of formal disaster research, and these approaches remain valuable tools. However, Drabek also argues that in order for the field to mature further, a range of more controlled methodologies and rules of evidence must be applied in order to finally address the issue of causation. In fact, he laments that, "it is hard to believe that quasi-experimental designs have not yet been implemented routinely…" (Drabek, 2002, p.143), and goes on to assert that "as more diverse perspectives and orientations push the field in different directions, we can anticipate that alternative methods will become more pronounced" (p.145).

In part because of our own interest in the influence of technology on the activity of disaster management, and also in order to constrain the range of this discussion and make it concrete, we focus here on current efforts to improve disaster management through the application of technology and information science. In particular, we address the research stream being produced by one group, the international community for Information Systems for Crisis Response and Management (ISCRAM; Van De Walle and Turoff, 2006). This is a vibrant, interdisciplinary research community, bringing together emergency management professionals, academic specialists in a variety of hazards, and technologists interested in applying information science to effect change in the way crises are managed.

While this is an innovative research community, striving to move from an interdisciplinary footing to truly transdisciplinary one – the goal of transdisciplinarity has not quite been met (given the criteria for transdisciplinarity offered by Scholz and Marks, 2001). We argue that there are at least three primary limitations to the research being produced in the ISCRAM community. First, and foremost among these, is the lack of a deep multidisciplinary dialogue about what constitutes scientific evidence. Second, ISCRAM research continues to be dominated by a single methodology – the case study, even when researchers make implicit claims that are not best addressed by this research approach. And third, there has been comparatively little effort to analytically or inferentially generalise from the findings offered in the proceedings of ISCRAM to a theory of disaster management (Franco *et al.*, 2007).

Recently, McEntire (2004) identified ten barriers to theoretical development in the area of emergency management. For the remainder of this article, we focus on three of the barriers he specified in particular, "What paradigms should guide our field?"; "What variables should be explored in academic research?"; and "What is the proper balance for knowledge generation?" (McEntire, 2004, pp.6–8). Our goal is to further explore these three questions and to suggest a few solutions drawn from research approaches used in psychology and practitioner/clinical sciences more broadly. Further, we address the ways in which efforts to improve disaster management – again focusing particularly on technology interventions – can be objectively measured. Or, stated more openly, our goal is to begin a deeper conversation between psychology and the other disciplines represented in the ISCRAM community – with the aim of articulating the relationships between the behavioural and information sciences in the context of disaster management. The hope behind this effort is that a drive towards model specification and testing will improve the quality of both bench and transdisciplinary research produced and, by extension, that this effort will encourage further theory development in the field.

To accomplish this, the remainder of the paper is organised as follows: we first articulate some limitations of current research in ISCRAM. Next, we begin an exploration of predictor and dependent variable measurement that may improve model specification in ISCRAM research. One of the primary limiting factors in conducting high quality experimental or quasi-experimental research – the lack of a single index that allows for comparison of disaster impact *across disaster types* is discussed in detail. The role of performance assessment variable operationalisation is also explored. Some of speculations that are both worth testing and are actually *testable* are offered from this transdisciplinary perspective – following and extending Drabek's (2002) suggestions. Finally, responding to McEntire's (2004) and Drabek's (2002) admonition to relate practice and inquiry to a nascent theory of disaster management, the importance of careful generalisation of research findings is considered – both as a way of informing decisions about the adoption new technologies in disaster management and as means for more closely binding ISCRAM research to this theory building process.

2 Current limitations in Information Systems in Crisis Response and Management research

Technology holds incredible promise to improve the quality of human life across in a wide range of applications. Perhaps one of the most hopeful of these is the application of information science to bring order out of the chaos inherent to disasters. However, there are also persistent difficulties associated with the use and adoption of technology (*e.g.*, Grudin, 1988), which are almost certainly exacerbated in the context of disaster. Because of cognitive limitations, situational constraints and emotional arousal, otherwise helpful technologies may be abandoned entirely or used in unanticipated, seemingly irrational ways (Smith, 1991). Further, in the most extreme events, even hardened technologies may fail or simply be unavailable (GAO, 1993).

The ISCRAM community is dedicated to the development and evaluation of technologies designed to support a variety of tasks associated with crisis response. Yet, a recent informal review of research quality in the ISCRAM community suggested that about 70% of papers accepted for publication in the conference proceedings were based on case studies, under half of the papers had a formal methods section, just a quarter of the papers reported any type of statistics, participant pools were typically based on convenience sampling, and just 30% of the studies made relatively sophisticated generalisability claims that served to frame the research in a broader context (Franco *et al*., 2007). This despite calls for increased methodological rigor specifically in the evaluation of technology systems designed to aid in disaster response (Drabek, 2002).

2.1 Transdisciplinary – crossing the final hurdle

If we understand the term transdisciplinary to mean: a shared framework of understanding between scientists from different disciplines, which is based on the contributing traditional disciplines, but also extends them; and that is designed to address complex societal problems by engaging in a "process of mutual learning between science and society" (adapted from Scholz and Marks, 2001, p.237), then it is clear that by this definition, the area of disaster management should be profoundly transdisciplinary.

Yet most research in disaster management continues to be conducted from an interdisciplinary perspective. Let us be clear – it is not that conducting bench research within a single discipline is better or worse that doing interdisciplinary or transdisciplinary research – each approach has vital, specific contributions to make (Scholz and Marks, 2001). However, one of goals of the ISCRAM community from its outset seems to be to bring together an otherwise diffuse group of developers and researchers to work specifically on a set of societal problems. As its founders have noted, "a lot of research and development activities were dealing with hardware in computers, communications, and sensors, there did not appear to be an active community concerned with the requirements, design, development or evaluation of information systems supporting individuals and organisations as they respond to or manage a crisis" (Van De Walle and Turoff, 2006, p.365).

The community's very composition is transdisciplinary, reflecting the necessity of bringing together pure technologists, academic researchers, and interface this work with input from first responders, emergency managers, and government officials – all of whom are represented in ISCRAM's membership (Van De Walle and Turoff, 2006). It is clear that the community strives to achieve true transdisciplinarity – and the integration of academic disciplines with outside stake-holders demonstrates this commitment powerfully. In fact, in all but one dimension, ISCRAM has achieved its goal of becoming a transdisiplinary forum.

The final hurdle – and perhaps the most difficult one – is to establish a shared scientific framework and standards of evidence that are mutually acceptable to the various disciplines represented. It seems that a conversation about what constitutes supportable evidence is necessary and overdue. In the review process that this paper went through, an anonymous reviewer commented that the thrust of this argument and our emphasis on 'empiricism' was misplaced. We want to be clear that our aim is not to create a methodological hegemony driven by research standards that have been found acceptable in our own primary discipline – that is not the point at all. Moreover, we strongly support epistemological and methodological pluralism.

However, the majority of research being conducted on technology-assisted disaster management continues to rely on case studies, specification of technology design, and typically unsupported generalisability and utility claims about the technology (Franco *et al*., 2007). Case studies are acceptable for many types of research in disaster studies generally, however, they are much less convincing when the assertions offered have to do with the relationship between a specific intervention (in this case an information system or technological tool) and improvement in disaster management performance. Research questions dealing specifically with such interventions are probably better addressed through the methods developed to support evidence based practice.

In a decades long process of evaluating the quality of evidence produced in psychological research (as well as medicine, and other practitioner-oriented sciences), the relatively utility of case study design in terms of evaluating and prescribing changes to improve practitioner performance has waned (see, *e.g.*, Centre for Evidence Based Medicine, 2001; Harris *et al*., 2001; Sackett *et al*., 2000; Siwek *et al*., 2002), giving way to research approaches that improve validity, generalisability, and these approaches have been agreed upon through a series of consensus statements (see, *e.g.*, Standards for Educational and Psychological Testing, AERA, APA & NCME, 1999; 1966; APA, 1954). Let us be clear once again – this is not a value judgment against case study design, as different types of methodologies are uniquely suited to particular research questions (Wampold, 2003). Rather, as Drabek (2002) points out, while much of the evidentiary foundation of disaster research and theory is based on case studies, this method alone seems less appropriate to measure changes in practitioner performance. These questions are better addressed through controlled n = 1, *ex post facto*, or randomised controlled trials (Borkovec and Castonguay, 2006; Hollon, 2006; Kazdin, 1998). In the review of this paper it was noted that while these approaches may be desirable, they may be impractical because of the difficulty in collecting data in the midst of disasters. Again, preeminent disaster researchers have called for exactly this type of research (Drabek, 2002), and we share the belief that these forms of research are needed to close the gap in causal evidence that currently exists.

3 Laying the groundwork for increased rigor in Information Systems in Crisis Response and Management research

Currently, most studies in the ISCRAM community addresses efficacy to varying degrees, without adequately examining effectiveness (for a detailed discussion of the efficacy versus effectiveness debate, see, *e.g.*, Fishman, 2000; Clarke, 1995). However, even if we can imagine broader studies of effectiveness of ISCRAMs, the execution of either type of study remains hampered because several critical tools necessary to make causal inference claims remain underdeveloped.

3.1 What paradigms should guide our field? Identifying key dimensions of interest for a transdisciplinary research framework

Prior efforts to comprehensively address disaster management have identified a number of variables that may impact planning, mitigation, and response. For example, in addition to the skill of professional disaster managers, other factors that may contribute to the overall success or failure of a response include the cultural attitudes of the affected population; centralisation or decentralisation of government response; quality of crisis communication; political will; rumour; poverty; as well as the type, scale and scope of the disaster itself (McEntire and Fuller, 2002; Gheytanchi *et al.*, 2007).

If we attempt to cluster the large number of contributing variables logically, it follows that there are key, overarching dimensions that should at least be addressed – and if possible controlled for – in studies attempting to evaluate the effect of technology applications on disaster management (Davies, 1965). This can be accomplished either directly through the experimental process (*i.e.*, use of predictor variables, specific sampling techniques, and inferential generalisation), through the expression of limitations to the robustness of findings (*i.e.*, case to case transferability claims; another acceptable generalisation strategy), or by analytically relating the results of a study to an overarching theory (*i.e.*, analytic generalisation; Firestone, 1993). We assert that the key dimensions of interest in this area of research are the *disaster type and scale*, the *intervention* (in this case an information system), and the *characteristics of individuals or groups the intervention is intended to support* (Franco *et al.*, 2007; Markus and Robey, 1988).

To illustrate how such a framework enables the precise articulation of ISCRAM research problems, we use as an example the tremendous amount of effort currently being expended on SMS text-messaging system as a tool for crisis communication (*e.g.*, Gomez and Turoff, 2007). While this technology has clear utility in many contexts – for example, as a readily available mode of communication among (trained) citizen responders, or between citizen responders and professional responders – it may not be effective in all disasters or for all populations. Several negative, unanticipated scenarios are conceivable, if one were to attempt to extend the use of SMS technology from the specific group and processes mentioned above to a wider population, without specifically considering and addressing the differences in domain knowledge (that is, knowledge of disaster response processes), technological sophistication, and sociocultural background between the original target population (citizen responders) and the new target population (general public in a specific community). For instance, it is conceivable that given the history of distrust towards the government in African American communities in the Southern USA, a poorly worded, widely broadcast SMS message might engender rumours that the government was attempting a 'land grab' by sending an evacuation order directly to residents rather than communicating through trusted local political leaders – resulting in poorer performance than if less technologically based communication strategies were used (Allport and Postman, 1947; Freimuth *et al*., 2001; Miller, 1992).

By remaining in conversation with these three key dimensions, researchers will be discouraged from mistaking *efficacy* of a technological tool in a controlled environment (or in a specific environment) for *effectiveness* in environments that are less sterile than a laboratory, or that vary significantly (with respect to our dimensions) from the original evaluation environment. The field is replete with specialists in advanced analytic paradigms, Bayesian learning systems and the like – however, if the findings from these approaches are not tested against the contextual changes endemic to disasters, the utility of these tools remains suspect and the causal assertion that a given technology actually improves overall management performance remains weak.

3.2 What variables should be explored? Towards a major disaster impact index

Drawing once again on a transdisciplinary perspective, we are reminded of the importance of defining a consensus-based index that describes the severity of disaster impact. The lack of such a measure represents a serious problem to those interested in conducting *ex post facto* research designs, as this research is predicated on the use of a valid covariate or predictor variable (for a discussion of the use of severity indices in medical intervention evaluation research, see for example, Batchelor *et al*., 2001; Mitchell *et al*., 2007; Peitzman *et al*., 1999).

While clearly not a simple task, the development of a single index that *adequately* (not perfectly) describes the relative impact of a variety of disaster types would represent an important advance in the field of disaster management research – both for its own sake, and because it would facilitate the implementation of a wide range of experimental and quasi-experimental disaster management performance assessment designs. The requirements for such a consequence index should be examined carefully. At minimum, it should be able to summarise several key factors that are functionally related to disaster response and recovery – such as casualty rates, number of structures destroyed, and loss of key infrastructure. The scale should also consider these figures in the context of the region affected by the disaster. This might include, for example, the geographical area involved, total population in the affected zone, and relative impact in urban versus non-urban areas. Additionally, such a tool should be understandable and useful to three distinct constituencies – researchers, policymakers, and the public. Further, the measure should be distinct from and not conflated with performance metrics. Finally, such a system should also be meaningful in a multinational, cross-cultural context.

3.2.1 A review of the current state of the art

To our knowledge, there are just a few indices that seek to equate impact *between* different types of disasters (post-disaster as opposed to predictive modelling of hazards) – instead, most evaluate *within* specific disaster types, such as tornadoes, floods, and so on. One comprehensive measure is the Bradford Disaster Scale (BDS; Keller, 1989), a logarithmic scale, which uses number of fatalities as its initial foundation. More recently, this scale was extended to include evacuations caused by chemical plant accidents, reinsurance costs, and number of injuries in the evaluation of disasters (Keller *et al.*, 1997). This extension of the BDS used data from chemical and oil industry accidents derived from the Major Hazard Incident Data Service (MHIDAS), with criteria for the disaster involving one of the following:

- five or more fatalities
- damage in excess of US$1 million
- 50 or more individuals evacuated
- ten or more injuries.

Trend analyses have demonstrated the utility of the BDS as a tool for assessing performance over time. For example, these analyses have shown that the USA experienced increased financial costs related to these events, while reducing the number of fatalities and injuries over time. In contrast, costs in the UK were relatively stable over the period examined, but fatalities increased (Keller *et al.*, 1997).

The latest incarnation of the BDS is a sophisticated, multifaceted indexing tool. However, several problems exist with the system. First, it has been primarily used to address chemical plant accidents, not major disasters *per se*. While the definition of disaster is widely contested, it seems clear that the inclusion criteria for the BDS scale squarely addresses normally occurring accidents with little overlap to extraordinary catastrophes. Fatality injury rates from major disasters (excluding epidemics and famines) often begin at the level of several tens of deaths, and the median number of fatalities for the top ten disasters in 2006 was 880 (EM-DAT, 2006). Further, the recent Indian Ocean Tsunamis serve as a stark reminder that catastrophic incidents may result substantially higher casualty rates in very short periods of time. Thus, it may be more effective compare the number of fatalities in a given event to the historical absolute range of fatalities for major disasters.

A second problem with the BDS is its reliance on inputs that are tailored to the industrial accident context. As noted earlier, to effectively equate a variety of major disasters in a way that is meaningful for research on the efficacy of disaster management, the index should use inputs that are logically tied to disaster management activities. For example, the number of fatalities provides a great deal of information about what types and amounts of resources will have to be made available (*e.g.*, mortuary personnel, refrigerated trucks, *etc*.). However, reinsurance costs are less informative as the amount of insured property varies substantially from one catastrophe to the next. Using parameters that describe destruction of infrastructure in more absolute terms may more accurately represent the level of response needed to meet the challenges of a particular catastrophic event.

Third, the BDS relies on an open-ended logarithmic scale. While this approach is clearly useful in academic settings, it may be less so in public policy arenas. For example, the magnitude of an earthquake is traditionally reported using the Richter scale (M_L) in the USA, and the public has become comfortable with the term. However, contemporary seismologists typical report magnitudes based on extensions of the Richter system (*e.g.*, Moment Magnitude, M_W) and the public's actual understanding of either logarithmic scale remains poor (Miller, 1997). Other systems, such as the European Macroseismic Scale (EMS; Grünthal, 1998) and the Torino Impact Hazard Scale for asteroids (Binzel, 2000) were designed to qualitatively express the actual affect of these hazards on humans at the individual or group level, reflecting a more intuitive approach. Unfortunately, measures that are entirely qualitative and accessible to the public typically reduce the utility of these systems as quantitative research tools.

Another approach to quantifying disaster impact has been to measure the number of built structures damaged or destroyed by the event. The Damage Index (DI; Blong, 2003; 1998) is also a logarithmic scale, designed to describe the destruction of standardised house equivalents by relating a replacement ratio for these structures to a Central Damage Ratio, which uses a five-point benchmarking system to qualitatively describe damage across a range of disaster types. The DI serves as one approach to assessing the cost of a disaster that avoids the problems associated with reinsurance based measures, thus increasing the validity of the model in areas where substantial portions of the population lack property insurance. However, the DI still suffers from some of the problems associated with the BDS, especially a lack of immediate transparency to the public and policymakers.

Finally, a third approach which attempts to capture the multifaceted nature of disaster consequences has been offered (Christen *et al.*, 1994). This multi-attribute method uses fuzzy set theory to quantitatively express membership functions that may be easily quantified or qualitatively expressed. These terms are expressed as a number between 0 and 1, where 0 is taken to indicate normality and 1 represents a maximally catastrophic state. The authors hypothesised a linear relationship between the logarithmic indicator value (*e.g.*, number of fatalities or km^2 of ecosystem damage) and the Impact Value, such that for four fatalities the Impact Value = 0.02, and for 100 fatalities, the Impact Value = 0.06, *etc*. The authors note that in addition to the ability to summarise a multi-attribute event in a single value, the values obtained in examining several chemical plant accidents correspond well to intuitive assessments of the relative consequences of these accidents, and the system can be used to compare across accidents for evaluative purposes. The authors suggest that other ways of determining membership functions may be used, thus increasing the flexibility and acceptability of such a system – but also underscoring the relatively subjective nature of this approach.

3.2.2 Alternative approaches to fatality scaling

As we have noted, in order to be broadly useful, a major disaster impact scale must take into account the geographical scale of the event (Alley, 1984) – and optimally would also address temporal boundedness as well – however we will leave time out of this discussion for the moment. While basing the geographical aspect of the index on an objective measure such as km^2, the relationship between geography and disaster management response may be better expressed by tying this sub-scale to actual political boundaries. These political boundaries often delimit the types and number of resources that can be fielded as part of an organised response. Further, using such an approach makes the index more familiar to the general public and policymakers by contextualising the disaster and its response in geographical units that are commonly used in everyday life.

Table 1 Geographic impact

Ordinal ranking	*Description*
1	City
2	County
3	State
4	Region
5	National
6	Multinational
7	Continent
8	Hemisphere
9	World (non-catastrophic)
10	World-catastrophic (equivalent to a ten on the Torino asteroid impact hazard Scale)

Further, if necessary, the complete destruction of built structures in any of these geographical units may be roughly expressed as tonne-equivalents TNT using a standard blast distance (useful in equating the destructive power of earthquakes, tsunamis, hurricanes, and explosions – but less so with disasters that do not involve physical destruction, *e.g.*, dirty bombs, *etc*.).

Similarly, a ranking based on the number of absolute number of fatalities normalised against the number of deaths from a wide range of major catastrophes may serve as a functional fatality scale. Assuming that extreme events are both rare and would not warrant the type of performance evaluation we are recommending here, the maximal historical figures for number of fatalities in a major disaster appears to be 1–2 million souls (excluding temporally unbounded events such as famine, pandemic, drought, *etc*.). Arbitrarily setting the maximum to 1 million for the sake of convenience, and using exponential decay (*i.e.*, successively halving the figures), it is possible to derive an ordinal scale that functionally describes differences in fatality impact at levels in between orders of magnitude (obscured to some extent by natural log scales) with a ranking of 1–10. Expressing the number of fatalities from a major disaster as a simple ranking system that is tied to a look-up table should be more immediately understandable to the general public, thus increasing the transparency and utility of the scale in policy discussions.

A few simple transformations allow us to further normalise the scale, and express much the same information in a format that is meaningful in a research context. The median of the scale is 977 (interestingly, a close approximation of the median number of fatalities from the top ten major disasters of 2006, which was 880; EM-DAT). By normalising the absolute number of deaths to the median, and re-expressing these figures as $\log_2 n$, an open-ended logarithmic scale emerges that retains many of the properties of the initial scale.[1] Optimally, a fatality scale should express both an absolute value and a figure corrected for relative population in the geographical-political area of interest – an area for future consideration.

Table 2 General fatality scale

# Fatalities (expressed as exponential decay from maximal value)	*Example event*	*Simple ordinal fatality scale*
1 000 000+	Yellow River Flood, 1887	10+
500 000–999 999	Tangshan Earthquake, 1976	10
250 000–499 999	Kaifeng seawall destroyed, 1642	9
125 000–249 999	Indian Ocean Tsunami, 2004	8
62 500–124 999	Ashgabat Earthquake, 1948	7
31 250–62 499	Bam, Iran Earthquake, 2003	6
15 625–31 249	Nevada del Ruiz Eruption, 1976	5
7813–15 624	Hurricane Mitch, 1998	4
3906–7812	Kobe Earthquake, 1995	3
1953–3905	San Francisco Earthquake, 1906	2
977–1952	Hurricane Katrina, 2005	1

Table 3 Median normalised fatality scale

Absolute number of fatalities	*(n/median)*	*Log_2 (n/median)*
1 000 000	1024.000	10
500 000	512.000	9
250 000	256.000	8
125 000	128.000	7
62 500	64.000	6
31 250	32.000	5
15 625	16.000	4
7813	8.000	3
3906	4.000	2
1953	2.000	1
Median 977	1.000	0
488	0.500	–1
244	0.250	–2
122	0.125	–3
61	0.063	–4
31	0.031	–5
15	0.016	–6
8	0.008	–7
4	0.004	–8
2	0.002	–9
1	0.001	–10

Ultimately, the development of a single value index that expresses the multiple facets of a disaster in a functional, transparent, and fairly objective way – that can be tied to the operational aspects of disaster management is central to any effort to evaluate the relative impact of our efforts to improve disaster management. Research on performance assessment in disaster management is currently constrained by the relatively few comprehensive indices on offer, the lack of debate about their merits (for example, it may prove more useful to relate logarithmic scales such as the BDS and Blong's DI scale and then re-express this information as a descriptive, ordinal scale) and in the end, by a lack of consensus on an accepted covariate controls that will facilitate both inquiry with the research community and dialogue with a wider audience.

3.3 Shared performance assessment metrics

In order to evaluate the effectiveness of a disaster management intervention, we must link a disaster impact index scale to a well delineated set of outcome measures. That is to say, after grouping by disaster severity or controlling for it – what outcomes can we reasonably expect to be within our ability to constrain, and how do we measure these outcomes? To date, most of benchmarking efforts have involved the assessment of disaster planning and readiness (EMAP, 2004; Caudle, 2005). However, none of these approaches measures *actual* performance in disasters – profoundly hampering our ability to accurately examine outcomes in specific disasters, making it difficult to compare performance across time, and limiting the ability to relate interventions to outcomes (Franco *et al*., 2005).

3.3.1 National level performance indices

As a starting point, a number of performance assessment benchmarking systems drawn from a variety of sources (such as military, international coalition, medical student training, and flight crew performance metrics) can be applied to professional disaster managers. Each measure has its strengths and weaknesses, and should be used judiciously in an attempt to close the causal chain for the particular research question asked. What is essential to note is that all of these metrics are designed to evaluate activities that are undertaken *after* the disaster has occurred, thus removing the potential conflation between a disaster severity index and these measures. This is an important point as it is easy (as witnessed by our own internal discussions and the conversation with the reviewer of this paper) to inadvertently mix the two ideas. Temporality is one of the key ways of disambiguating a covariate from the dependent variables in this context. Several performance metrics are suggested here as potential candidates, but this is by no means a comprehensive or prescriptive list (for a detailed discussion, see Franco *et al*., 2005):

- *360-degree global assessment* – A process in which all actors in the disaster confidentially evaluate other actors with whom they interact.
- *Impact on situation* – if critical performance goals can be identified, such as the prevention of a chemical plant from being engulfed in a fire, dichotomous results can be stated.

- *Action indices* – specific activities may be viewed as central to disaster response performance. Evaluating whether or not these actions took place and if these activities were completed in the correct order can offer some insight, however this approach encourages specific behavioural patterns at the expense of improvisation.
- *Time indices* – the time it takes to initiate action, the time to complete an action, and the time to detection of a problem have all been offered as ways of assessing performance. Additionally, more complex measures such as synchronisation indices that examine the time difference between first and last team prepared to take action can provide insight into command and coordination performance.
- *Process variables* – for example, the co-location of knowledge, resources, and decision-making authority has been identified as a prerequisite to optimal performance in large international relief efforts.
- *Lessons learned overlap* – each major disaster results in several sets of after action reports and concomitant 'Lessons Learned'. However, there is little evidence that these lessons are, in fact, learned and addressed in following disasters. If performance does improve, less overlap in after action recommendations should be in evidence over time.
- *Attitudes* – governments must engender trust and provide guidance to citizens in the face of catastrophe. Measures of individual attitudes towards government crisis communication can provide important information about how successful these efforts were.

3.3.2 Population and political level performance indices

While performance metrics for professional disaster managers tell part of the story, one postulate that can be advanced is that as the scale and scope of the disaster increases, the relative impact of technologies designed to assist professional disaster managers diminishes (Franco *et al*., 2007). If this is the case, it may be possible that entirely new classes of technology can be targeted to address population level or political level actors. Thus it follows that performance metrics that address the ability of the population to be resilient in the face of disaster or the ability of political actors to make good decisions and properly execute their duties must also be developed in order to adequately test such questions.

Adaptive individual performance is necessary for optimal overall response and recovery in the event of a catastrophic incident. If even a relatively small percentage of the population behaves in unexpected ways, the ability of professional disaster managers to maintain situational control can be seriously impaired. For instance, one study demonstrated that large portions of disaster affected populations do not behave in expected ways, with about 1/3 responding maladaptively (Weisath, 1989). As the scale of the disaster increases, the absolute number of people responding poorly may simply overwhelm the professional disaster management response. Thus, the development of population level dependent variables or performance measures is fundamental to evaluating overall disaster management effectiveness. Some potential metrics follow, again these are illustrative and not exhaustive (see for example, Raphael, 2005):

- *Compliance* – the percentage of people in a given geographical area who elect to follow the crisis communication orders issued by local, state, or federal officials (*e.g.*, number evacuating *versus* number wandering aimlessly *versus* refusals to evacuate).
- *Panic* – the percentage of individuals who act in a manner that is not consonant with prior community level disaster planning, or in direct conflict with crisis communication, which is driven by emotions of fear or terror.
- *Emergent behaviour* – a number of emergent behaviours within populations may occur, some of which contribute to response and recovery, and others which impede these activities.
- *Community interoperability* – the Cuban disaster management model suggests one way professional disaster managers' jobs can be made easier is if they can expect groups from widely varied backgrounds to act in a fairly standardised manner in relationship to other communities, relief workers, and outside agencies.
- *Time to declaration* – Where clear criteria for the declaration of disaster or evacuation order exist, the time difference from when this declaration should be made, and when it is actually made serves as one index of political performance.
- *Quality of crisis communication* – effective crisis communication requires political actors to have a strong situational/operational awareness, a good grasp of how to work with the media, ability to disambiguate conflicting information, and the ability to establish trust with the public.

4 What is the proper balance for knowledge generation? Going for the Gold(standard) – reconsidering the plausibility of effectiveness research

We have argued that by engaging in a process of transdisciplinary model generation – that is, the articulation of the key dimensions, methods of controlling for these dimensions (*i.e.*, covariates such as a disaster impact scale), and the identification performance assessment metrics – some specific, testable speculations can be offered. Some propositions that follow logically from this approach are offered here (and many more are conceivable):

- users are less likely to adopt and use ISCRAM systems in the midst of a disaster as compared to users in less complex environments where similar technologies might be applied
- the impact of ISCRAMs decreases as a function of the complexity of the disaster because the role of professional disaster managers is reduced (in other words, ISCRAMs are typically designed to address the emergency management layer, and not the political or population layers)
- there is (or is not) a demonstrable improvement in overall disaster management performance when a given ISCRAMs is used
- there is (or is not) a demonstrable improvement in overall disaster management performance when a given class of ISCRAMs is adopted.

The initial speculation may be fairly easy to test using an efficacy model, but the second and third questions require evaluating effectiveness and are therefore predicated on the further specification of the predictor and dependent variables we have just described. The fourth may require careful generalisation from a number of individual studies or meta-analysis to make a best practices policy recommendation. If we consider the process of conducting large-scale quasi-experimental designs (and even the possibility of randomised controlled trials) in terms of the ISCRAM community, a proximal set of goals for this work might involve introducing a technology intervention – such as an information system designed to manage disaster victim data (for example, the SAHANA victim registry, which was deployed successfully in six major disasters from 2004–2006; Samaraweera and Corera, 2007) – into a variety of disaster situations and evaluating impact using a few well chosen performance metrics, such as proportion of missing persons found compared to proportion expected at given disaster impact scores, elapsed time from separation to reunion, *etc*.

Performing *ex-post facto* and randomised controlled trials at the scale necessary to evaluate these technologies certainly seems like a tall order. However, it should be noted that research involving fairly complex interventions designed to deliver societal level treatments using cities, counties or regions of a country as the unit of analysis has been done successfully, and can show clear, measurable changes in behaviour and outcomes (see Rogers *et al*., 1999; Singhal and Rogers, 1999). Natural experiments involving differences in disaster management approaches occur on a daily basis – for example in the USA, counties typically use one of three or four methods to request mutual aid from neighbouring communities (Aldrich, 2007) – yet little is known about the relative utility of each of these methods. Current estimates suggest that 1 million mutual aid requests for firefighting equipment and personnel occur each year in the USA alone (USFA, 2006), providing a more than adequate statistical power to detect differences in mutual aid systems if the influence of contextual variables can be adequately described and controlled for. Equally, the adoption or non-adoption of a specific technology intervention at the county or regional level may offer similar quasi-experimental designs that can begin closing the causal gap. Achieving “gold standard” research within the ISCRAM community is not out of reach.

While systems such as SAHANA have obvious and tremendous potential as ways to mitigate the problems encountered by a disaster manager, their effectiveness remains an empirical question that is not adequately addressed through descriptions of the technology or single case studies. Viewed from an evidence based practice perspective, such studies are obviously open to many threats to methodological bias that render their findings causally inconclusive (Oosterhuis *et al*., 2004). These considerations may seem like apparent negatives, but this not the case. Well conceived technology systems can and will improve disaster management. Conducting high quality research may initially call into question the utility of such systems, but over time findings from transdisciplinary or ‘oriented’ research will result in refinements of existing systems, more rapid future system development, and improved policy recommendations for the adoption of specific technological applications during catastrophic events (Häberli *et al*., 2001).

5 Discussion

The present effort does not seek to set forth a comprehensive theory of disaster management – it is too early in the process of inquiry for that. However, this is an appropriate moment for deeper reflection on the process and direction of disaster management research, and the activity of *theorising* can offer a number of benefits to the field at this juncture (Weick, 1989; 1995). In the review process for this paper, it was noted that the approach taken here represents a largely positivist, inductive stance – one emphasising model specification, standing in contrast to the philosophical and methodological underpinnings of most disaster research that has been conducted to date. We agree and disagree with this critique. We agree that, at first glance, the suggestions made here appear to be an effort to push ISCRAM research towards an inductive research model – and we argue that for some types of research questions, this model may be particularly useful and informative. However, it is easy to over simplify this argument and reduce it to an 'either…or' – *i.e.*, that we are arguing that research in the ISCRAM community must move to this model exclusively or risk being seen as scientifically inconclusive.

Instead, what we *are* asserting the generalisability of the research conducted in the ISCRAM community is a critically important feature of this effort – in this context, the generalisation process has serious theoretical, practical, and ethical implications. In other domains, inefficiencies in the research process might simply be corrected over time without any real social cost. However, within this field, the inappropriate adoption of a given technology without the careful considerations of its limitations (generalisabiliy) is an ever-present risk with potentially dramatic consequences. A principled framework for matching the capabilities of existing technologies to the sociotechnical needs of the disaster management processes will reduce the risks associated with accepting a technology that should not be adopted (a form of Type I error).

Here, the buoyant creativity of the ISCRAM community must be met with an equally cautionary stance that examines the ethics of this work (Ernst, 2001). The effective management of a disaster can save lives and prevent tremendous hardship. Just as easily, ineffective management (or management interventions) can have profound negative effect on entire populations (Gheytanchi *et al.*, 2007). If we assert that the system of disaster management that is currently in place is not sufficient and that our technologies can improve performance, our footing necessarily changes from that of the detached scholar to that of a practitioner facing an ill patient. As we increasingly take on the mantle of responsibility for improving disaster management, the need to base recommendations on solid evidentiary foundation becomes ever more acute.

Inductive and deductive approaches to generalisation suffer almost equally from inherent problems that remain unresolved, and perhaps are even philosophically unresolvable. Rosenberg (1993; as cited in Lee and Bakersville, 2003, p.225) summarises Hume's assertion that, "the deductive arguments…are no more convincing than their most controversial premises and so generate a regress, while inductive ones beg the question. Accordingly, claims that transcend available data, in particular predictions and general laws, remain unwarranted." Further, Lee and Bakersville (2003) note that some philosophers of science argue generalisation is never justified. Yet the show must go on. Disasters will not wait for us to resolve the problems of evidence endemic to science. Policy makers and emergency management practitioners must make tough decisions about which technologies to adopt and which to jettison based personal experience, good judgment, and the largely unsubstantiated claims of system developers. As researchers, we are obligated to make the best recommendations for change and improvement that we can, recognising that our suggestions will always be imperfect. Generalisation is a fact of life in this setting, and the *more* accurate our predictions of the utility and limitations of the various technology solutions offered, the better we meet our ethical duty as researchers.

There are a number of forms of generalisation and different ways of referring to these processes – *i.e.*, inferential (statistical), analytic, and case-to-case translation (Firestone, 1993); or generalising from data to description; from data to theory; from theory to description; and from concepts to theory (Lee and Bakersville, 2003). Our limited, informal review of research quality of ISCRAM proceedings (Franco *et al.*, 2007) suggests that claims regarding the generalisability of technology interventions are largely absent from published case studies, and when extrapolations were present they were often unsophisticated and did not rest on a well explicated qualitative methodology. Further, similar to other areas within information science, the case studies often appeal to inferential generalisation techniques rather than using approaches that have been established as more appropriate to interpretivist research (Lee and Baskerville, 2003). These limitations greatly diminish the reader's ability to attempt for themselves a case-to-case extrapolation based on the study's description (Firestone, 1993). Further, analytic generalisation from findings of multiple case studies to a broader, overarching theory of disaster management remains largely absent from the discussion, which seems to be a missed opportunity given the unique centrality of the ISCRAM community in this transdisciplinary research space.

We have argued that it is increasingly possible to conduct high quality, controlled research within the context of ISCRAM, and as a team predominated by psychologists, we would feel more comfortable about research quality and generalisation claims using the evidentiary rules of statistical inference. Further, we assert that this type of research might better address the fundamental research questions being asked here, which largely revolve around testing the utility of an intervention in a particular context. We believe, as do others (*e.g.*, Drabek, 2004; McEntire, 2004), that this type of research may offer important insight, unique contributions, and we have attempted here to address some of the constraints that are hindering this process.

It is easy to suggest this research approach is simply tantamount to better model specification for evaluation research in disaster management. Clearly, these activities will help with creating better parameters and models. But model building and theory building are not distinct activities – they are in fact deeply intertwined, as models can be understood as complex hypotheses designed to test a theory (Forster, 2000). As such models are tested and improved, a theory of disaster management will be variously explicated, expanded, and constrained. Especially at this stage in our collective understanding of disaster management, this must be a highly iterative processes of theorising, model specification, testing, and the interpretation of results to strengthen or weaken a network of ideas – binding the nascent theory structure to findings and conceptual analysis (Davies, 1965; Cronbach and Meehl, 1955; Machado and Silva, 2007).

It remains concerning that despite the strengths of quantitative/inductive perspective, comparatively little work using these tools is performed in the ISCRAM community. However, the positivist approach offered here does not and *should not* be taken as an attempt to preclude high quality interpretivist or qualitative research. In fact, given the multidisciplinary nature of this research, a trans-theoretical and trans-methodological approach may offer greater explanatory and predictive power than simply using one set of tools. When complementary research is possible (*i.e.*, triangulation), confidence in generalisability claims and advances in theory will be strengthened – in part because of the philosophical independence of the methods (Jick, 1979).

The management of the terrorist attacks in the USA on September 11, 2001 was complicated by communication failures, problems associated with a lack of interoperability, and cultural differences between response agencies. We run a similar risk in ISCRAM research – that our lack of a shared evidentiary foundation and theoretical vision will impair our ability to communicate within this group and to offer well reasoned advice to policymakers. The call of transdisciplinary research asks that each of us begin the process of transcending the prescriptions and prohibitions of our own disciplines and encounter the philosophical, methodological, and cultural beliefs held by our fellow scientists. One of the risks of this type of multidisciplinary research is that such communities tend to spend more time discussing what they have in common than what divides them – thus limiting the amount of problem solving power available to the group and creating a false sense of consensus (Cacioppo, 2007). Each discipline within ISCRAM contributes a unique epistemological perspective, and a deeper conversation about methodology and rules of evidence used by each constituent field may dramatically improve the quality of research produced within the ISCRAM community. We must first accept and even encourage dissensus before a true consensus on measurement and evidence can be achieved.

Acknowledgments

This material is based on work supported by the National Science Foundation under Grant No. IIP-0637999.

The authors thank the anonymous reviewer for insightful comments on an initial draft of this paper. This version reflects a ‘conversation’ with the reviewer, which we feel better illustrates some of the strengths and limitations of our arguments. Moreover, this conversation places the effort in the context of a deeper transdisciplinary dialogue.

References

Aldrich, J. (2007) *Personal Communication*, August.

Alley, W.M. (1984) ‘The Palmer drought severity index: limitations and assumptions’, *Journal of Climate and Applied Meteorology*, Vol. 23, pp.1100–1109.

Allport, G.W. and Postman, L. (1947) *The Psychology of Rumor*, New York: Holt.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA & NCME) (1966) *Standards for Educational and Psychological Testing*, Washington, DC: AERA.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA & NCME) (1999) *Standards for Educational and Psychological Testing*, Washington, DC: AERA.

American Psychological Association APA (1954) 'Technical recommendations for psychological tests and diagnostic techniques', *Psychological Bulletin*, Supplement, Vol. 51, No, 2, pp.201–238.

Batchelor, J.S., McGuinness, A. and Ryan, J. (2001) 'Adult trauma scoring models: a critical review', *Trauma*, Vol. 1, pp.175–180.

Binzel, R.P. (2000) 'The Torino impact hazard scale', *Planetary and Space Science*, Vol. 48, pp.297–303.

Blong, R.J. (1998) 'Damage scales', *Natural Hazards Quarterly*, Vol. 4, http://www.riskfrontiers.com/nhq/nhq4-3tables.htm.

Blong, R.J. (2003) 'A new damage index', *Natural Hazards*, Vol. 30, pp.1–23.

Borkovec, T.D. and Castonguay, L.G. (2006) 'Effectiveness research', in J.C. Norcross, L.E. Beutler and R.F. Levant (Eds.) *Evidence-Based Practices in Mental Health: Debate and Dialogue on the Fundamental Questions*, Washington, DC: American Psychological Association Press.

Cacioppo, J.T. (2007) 'Better interdisciplinary research through psychological science', *APS Observer*, Vol. 20, pp.3, 48.

Caudle, S. (2005) 'Homeland security: approaches to results management', *Public Performance & Management Review*, Vol. 28, pp.352–375.

Centre for Evidence Based Medicine (2001) 'Evidence based medicine – levels of evidence', Oxford: CEBM, http://www.cebm.net/index.aspx?o=1047.

Christen, P., Bohneblust, H. and Seitz, S. (1994) 'A methodology for assessing catastropic damage to the population and environment: a quantitative multi-attribute approach for risk analysis based on fuzzy set theory', *Process Safety Progress*, Vol. 13, pp.234–238.

Clarke, G.N. (1995) 'Improving the transition from basic efficacy research to effectiveness studies: methodological issues and procedures', *Journal of Consulting and Clinical Psychology*, Vol. 63, pp.718–725.

Cronbach, L. and Meehl, P. (1955) 'Construct validity in psychological tests', *Psychological Bulletin*, Vol. 52, pp.281–302.

Davies, J.T. (1965) *The Scientific Approach*, London: Academic Press.

Drabek, T.E. (2002) 'Following some dreams: recognizing opportunities, posing interesting questions, and implementing alternative methods', in R.A. Stallings (Ed.) *Methods of Disaster Research*, Xlibrus.

Drabek, T.E. (2004) 'Theories relevant to emergency management versus a theory of emergency management', *Emergency Management Higher Education Conference*, Maryland: NETC.

EM-DAT (2006) 'Disasters in numbers 2006', http://www.em-dat.net/documents/Confpress%202006.pdf.

Emergency Management Accreditation Program (EMAP) (2004) 'EMAP standard 2004', Technical Report, Lexington, KY: EMAP, September.

Ernst, R. (2001) 'The responsibility of science and scientists', in J. Klein, R. Grossenbacher-Mansuy, A.B. Häberli, R. Scholz and M. Welti (Eds.) *Transdisciplinarity: Joint Problem Solving Among Science, Technology, and Society – An Effective Way for Managing Complexity*, Basel: Birkhäuser Verlag, pp.6–22.

Firestone, W.A. (1993) 'Alternative arguments for generalizing from data as applied to qualitative research', *Educational Researcher*, Vol. 22, pp.16–23.

Fishman, D.B. (2000) 'Transcending the efficacy versus effectiveness research debate: proposal for a new, electronic', *Journal of Pragmatic Case Studies, Prevention & Treatment*, Vol. 3.

Forster, M.R. (2000) 'Key concepts in model selection: performance and generalizability', *Journal of Mathematical Psychology*, Vol. 44, pp.205–231.

Franco, Z., Joseph, L. and Beutler, L.E. (2005) 'Developing and evaluating advanced distributed learning systems for disaster training', National Center on the Psychology of Terrorism White Paper, http://www.terrorismpsychology.org/white_papers/disaster_adl.doc.

Franco, Z.E., Zumel, N. and Beutler, L.E. (2007) 'A ghost in the system: integrating conceptual and methodological considerations from the behavioral sciences into disaster technology research', in B. Van de Walle, P. Burghardt and C. Nieuwenhuis (Eds.) *Proceedings of the 4th International Conference on Information Systems for Crisis Response and Management ISCRAM2007*, Brussels: VUBPRESS, pp.115–124.

Freimuth, V.S., Quinn, S.C., Thomas, S.B., Galen, C., Zook, E. and Duncan, T. (2001) 'African American views on research and the Tuskegee Syphilis study', *Social Science and Medicine*, Vol. 52, pp.797–808.

GAO (1993) 'Disaster management: recent disasters demonstrate the need to improve nation's response strategy', *Government Accountability Office*, Washington, DC.

Gheytanchi, A., Joseph, L., Gierlach, E., Kimpara, S., Housley, J., Franco, Z.E. and Beutler, L.E. (2007) 'The dirty dozen: twelve failures of the Hurricane Katrina response and how psychology can help', *American Psychologist*, Vol. 62, pp.118–130.

Gomez, E.A. and Turoff, M. (2007) 'Interoperable communication: an analysis of SMS text-message exchange', in B. Van de Walle, P. Burghardt and C. Nieuwenhuis (Eds.) *Proceedings of the 4th International Conference on Information Systems for Crisis Response and Management ISCRAM2007*, Brussels: VUBPRESS, pp.45–50.

Grudin, J. (1988) 'Why CSCW applications fail: problems in the design and evaluation of organization of organizational interfaces', *Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work*, pp.85–93.

Grünthal, G. (Ed.) (1998) *European Macroseismic Scale 1998 EMS-98*, European Seismological Commission, Subcomm. on Engineering Seismology, Working Group Macroseismic Scales, Imprimerie Joseph Beffort, Helfent-Bertrange, Luxembourg.

Häberli, R., Bill, A., Grossenbacher-Mansuy, W., Klein, J.T., Scholz, R.W. and Welti, M. (2001) 'Synthesis', in J. Klein, W. Grossenbacher-Mansuy, R. Häberli, A. Bill, R.W. Scholz and M. Welti (Eds.) *Transdisciplinarity: Joint Problem Solving Among Science, Technology, and Society – An Effective Way for Managing Complexity*, Basel: Birkhäuser Verlag, pp.6–22.

Harris, R.P., Helfand, M., Woolf, S.H., Lohr, K.N., Mulrow, C.D., Teutsch, S.M. and Atkins, D. (2001) 'Current methods of the U.S. Preventative Service Task Force: a review of the process', *American Journal of Preventative Medicine*, Vol. 20, pp.21–35.

Hollon, S.D. (2006) 'Why I do clinical trials', in J.C. Norcross, L.E. Beutler and R.F. Levant (Eds.) *Evidence-Based Practices in Mental Health: Debate and Dialogue on the Fundamental Questions*, Washington, DC: American Psychological Association Press.

Jick, T.D. (1979) 'Mixing qualitative and quantitative methods: triangulation in action', *Administrative Science Quarterly*, Vol. 24, pp.602–611.

Kazdin, A.E. (1998) *Methodological Issues and Strategies in Clinical Research*, 2nd ed., Washington, DC: American Psychological Association.

Keller, A.Z. (1989) *The Bradford Disaster Scale*, University of Bradford, September, Nos. 12–13.

Keller, A.Z., Meniconi, M., Al-Shammari, I. and Cassidy, K. (1997) 'Analysis of fatality, injury, evacuation and cost data using the Bradford disaster scale', *Disaster Prevention and Management*, Vol. 6, pp.33–42.

Lalonde, C. (2007) 'The potential contribution of the field of organizational development to crisis management', *Journal of Contingencies and Crisis Management*, Vol. 15, pp.95–104.

Lee, A.S. and Baskerville, R.L. (2003) 'Generalizing generalizability in information systems research', *Information Systems Research*, Vol. 14, pp.221–243.

Machado, A. and Silva, F.J. (2007) 'Toward a richer view of the scientific method: the role of conceptual analysis', *American Psychologist*, Vol. 62, pp.671–681.

Markus, M.L. and Robey, D. (1988) 'Information technology and organizational change: causal structure in theory and research', *Management Science*, Vol. 34, pp.583–598.

McEntire, D.A. (2004) 'The status of emergency management theory: issues, barriers, and recommendations for improved scholarship', *FEMA Higher Education Conference*, Maryland.

McEntire, D.A. and Fuller, C. (2002) 'The need for a holistic theoretical approach: an examination from the El Niño disasters in Peru', *Disaster Prevention and Management*, Vol. 11, pp.128–140.

Miller, D.E. (1992) '"Snakes in the greens" and rumor in the inner city', *Social Science Journal*, Vol. 29, pp.381–394.

Miller, S. (1997) 'Earthquakes: prediction and the media – a case study in public understanding of science', *Geophysical Journal International*, Vol. 131, pp.530–533.

Mitchell, A.D., Tallon, J.M. and Sealy, B. (2007) 'Air versus ground transport of major trauma patients to a tertiary traum centre: a province wide comparison using TRISS analysis', *Canadian Journal of Surgery*, Vol. 50, pp.129–133.

Moore, H.E. (1956) 'Toward a theory of disaster', *American Sociological Review*, Vol. 21, pp.733–737.

Oosterhuis, W.P., Bruns, D.E., Watine, J., Sandberg, S. and Horvath, A. (2004) 'Evidence-based guidelines in laboratory medicine: principles and methods', *Clinical Chemistry*, Vol. 50, pp.806–818.

Peitzman, A.B., Courcoulas, A.P., Stinson, C., Udekwu, A.O., Billiar, T.R. and Harbrecht, B.G. (1999) 'Trauma center maturation: quantification of process and outcome', *Annals of Surgery*, Vol. 230, pp.87–94.

Perrow, C. (1984) *Normal Accidents*, New York: Basic Books.

Raphael, B. (2005) 'Crowds and other collectives: complexities of human behavior in mass emergencies: commentary on "understanding mass panic and other collective responses to threat and disaster"', *Psychiatry,* Vol. 68, pp.115–120.

Rogers, E.M., Vaughan, P.W., Swalehe, N.R., Svenkerud, P. and Sood, S. (1999) 'Effects of an entertainment-education radio soap opera on family planning behavior in Tanzania', *Studies in Family Planning*, Vol. 30, pp.193–211.

Sackett, D.L., Straus, S.E., Richardson, W.S., Rosenberg, W. and Haynes, R.B. (2000) *Evidence-Based Medicine: How to Practice and Teach EBM*, Churchill Livingstone: Edinburgh.

Samaraweera, I. and Corera, S. (2007) 'Sahana victim registries: effectively track disaster victims', in B. Van de Walle, P. Burghardt and C. Nieuwenhuis (Eds.) *Proceedings of the 4th International Conference on Information Systems for Crisis Response and Management ISCRAM2007*, Brussels: VUBPRESS, pp.17–32.

Scholz, R.W. and Marks, D. (2001) 'Learning about transdisciplinarity: where are we? Where have we been? Where should we go?', in J. Klein, W. Grossenbacher-Mansuy, R. Häberli, A. Bill, R.W. Scholz and M. Welti (Eds.) *Transdisciplinarity: Joint Problem Solving Among Science, Technology, and Society*, Berlin: Birkhäuser Verlag.

Singhal, A. and Rogers, E.M. (1999) *Entertainment-Education: A Communication Strategy for Social Change*, Mahwah, NJ: Lawrence Erlbaum.

Siwek, J., Gourlay, M.L., Slawson, D.C. and Shaughnessy, A.F. (2002) 'How to write an evidence-based clinical review article', *American Family Physician*, Vol. 251, pp.251–258.

Smith, V.L. (1991) 'Rational choice: the contrast between economics and psychology', *Journal of Political Economy*, Vol. 99, pp.877–896.

Stallings, R.A. (Ed.) (2002) *Methods of Disaster Research*, Xlibrus.

US Fire Administration (USFA) (2006) 'Fire statistics – fire departments', http://www.usfa.dhs.gov/statistics/departments/index.shtm.

Van De Walle, B. and Turoff, M. (2006) 'ISCRAM: growing a global R&D community on information systems for crisis response and management', *Int. J. Emergency Management*, Vol. 3, pp.364–369.

Wampold, B.E. (2003) 'Bashing positivism and reversing a medical model under the guise of evidence', *The Counseling Psychologist*, Vol. 31, pp.539–545.

Weick, K.E. (1989) 'Theory construction as disciplined imagination', *The Academy of Management Review*, Vol. 14, pp.516–531.

Weick, K.E. (1995) 'What theory is not, theorizing is', *Administrative Science Quarterly*, Vol. 40, pp.385–390.

Weisath, L. (1989) 'A study of behavioral responses to an industrial disaster', *Acta Psychiatrica Scandinavica Supplementum*, Vol. 355, pp.13–24.

Wybo, J. and Latiers, M. (2006) 'Exploring complex emergency situations'dynamic: theoretical, epistemological and methodological proposals', *Int. J. Emergency Management*, Vol. 3, pp.40–51.

Note

1 Note that for the publicly available scale, only positive whole numbers are expressed beginning with 1 for the median value – as the general public is unlikely to intuitively understand using negative numbers to describe a 'major disaster.'