

## **Mental health research in the aftermath of disasters: using the right methods to ask the right questions**

Sandro Galea

### **Introduction**

Several research teams have documented the consequences of the September 11, 2001, terrorist attacks in New York City and throughout the country. The chapters in this part bring together reports from three of these research teams and summarize some of the key findings from each of their studies. Two commentaries offer perspectives on the challenges that this research faces and what these challenges suggest for post-disaster research in general. There are several key methodologic points that emerge from these preceding chapters. In this discussion I will synthesize the key methodologic issues that emerge both from the empiric papers and from the accompanying commentaries. Some of my comments are congruent with those already articulated in the preceding six chapters. I restate those comments here to reflect their importance and to present them as a part of a broader reflection on this area of research.

Overall, we can fruitfully consider the key methodologic issues at hand along two principal lines, namely issues pertaining to nature of the sample and issues that pertain to assessment methods used in the research.

### **Population sampling**

#### **Choosing the right sample to ask the right questions**

We can consider that there are three principal types of samples that have, appropriately, been the focus of most post-disaster research. These are: samples of persons who were directly affected by a disaster (frequently referred to as "victims" in the literature), samples of rescue works (including police, fire-fighters and others), and general population samples. As noted by North *et al.* (this volume) in their commentary, studies that are implemented in each of these samples are not comparable and indeed neither should they be. We can readily imagine that persons in

each of these three samples may be affected by a particular disaster in different ways. For example, the relation between intensity of traumatic event exposure and risk of post-traumatic stress disorder (PTSD) is well established (Brewin *et al.*, 2000; Galea *et al.*, 2005). It is to be expected that samples of those directly affected by a disaster will have higher prevalence of psychopathology than either of the other two samples. Similarly, the event exposures of general population samples are likely to be immensely heterogeneous, ranging from direct personal exposure (i.e., some persons in the general population will have been direct victims) to no exposure whatsoever, and as such, the overall prevalence of psychopathology in general population samples would be expected to be far lower than that in victim groups.

These different types of samples present researchers with an opportunity to ask different questions in the aftermath of disasters. Samples of disaster victims allow inquiry about individual-level mechanisms that explain the associations between risk factors and risk of psychopathology. For example, recent work using such samples has explored both the role of social support in shaping post-disaster resilience (Tucker *et al.*, 2000) and the physiologic precursors of PTSD (Goenjian *et al.*, 2003).

Samples of rescue workers are marked by heterogeneity in exposures that is typically greater than that in victim samples. Rescue workers may include, for example, both fire-fighters who were directly involved in the extraction of persons from a disaster site and construction workers who were involved in cleanup operations. In many disaster instances, the September 11, 2001, terrorist attacks among them, rescue workers are exposed to disaster sites for a prolonged period after the disaster. Therefore, these samples are particularly suited to questions that pertain to the relations between nature of exposure and subsequent psychopathologic outcomes. In addition, rescue workers may be exposed to occupational hazards ranging from fine particulate matter that may result in respiratory disease to direct physical injury from unsafe work sites. Occupational and environmental health researchers then have a unique opportunity to better inform our understanding of the aftermath of disasters, and how the consequences of these events can be minimized both in magnitude and in duration.

Finally, general population samples introduce opportunities to consider the population burden of disaster consequences, to understand the determinants of population rates of disease, and to understand disasters (and their consequences) within multivariate and multilevel frameworks that can facilitate our understanding of the factors *beyond* the individual that influence post-disaster outcomes. Direct victim samples do little to elucidate the overall burden of mental health problems. This may be a particular concern in large disasters such as the September 11, 2001, terrorist attacks where, in the months following the event, state and federal officials needed population estimates of disease burden in order to implement adequate mental health relief resources. Population-based samples

can assess both those who were directly affected as well as those who were less directly affected by the disaster, hence providing such estimates. Importantly, research using population-based samples needs to ensure that there is a well-conceived rationale as to why a particular general population sample may be affected by a disaster. As noted in some of the preceding chapters, there may be conceptual reasons why PTSD, which nosologically requires linkage to exposure, may not be plausibly present in distant populations (e.g., general US samples) after a disaster in New York City but may well be plausible in areas closer to the disaster site.

Epidemiologically, the determinants of population rates of disease may be different than the determinants of individual risk (Rose, 1992). Factors that have only small effects on individual risk of pathology may, by virtue of their ubiquity, be sentinel determinants of population rates. Population-based samples are then essential to identify such factors. For example, although controversial, there have been some provocative findings published recently which suggest that television exposure may be linked to a greater risk of psychopathology among direct victims of a disaster (Ahern *et al.*, 2004). This observation was facilitated by the heterogeneity of a population-based sample; samples that studied only victims would not have been likely to isolate television viewing as a risk factor. The ubiquity of modern television exposure to disasters unfolding in real time suggests that television viewing may be a critical determinant of population rates of psychopathology after future disasters.

Although there has been a longstanding appreciation in the scientific literature that factors which influence post-disaster outcomes include community-level factors there is very little empiric research about factors at levels beyond the individual that may affect population prevalence of post-disaster pathology (Norris *et al.*, 2002; Galea *et al.*, 2005). The social context undoubtedly modifies the individual experience of disasters and elucidating the role of social context in these circumstances is emerging as one of the most promising areas of disaster research. In addition, community-level factors such as social cohesion have no individual analog and must be studied *at* the group level often necessitating a representative, large, population-based sample to test specific hypotheses.

In sum, there is much room for growth in disaster research, particularly in moving beyond the victim-based samples that have been the traditional focus of most of this research, to innovative use of other samples asking questions that extend what we understand in the area.

### **The relevance of different sampling strategies**

A smaller, but important, consideration about choosing samples to inform disaster research relates to the methods that are used to collect these samples and the implications these methods have for inference from this data. Perhaps an easy example

of this concern arises in the use of random representative samples and the so-called "convenience" samples. Although randomly selected representative population-based samples may provide us with an opportunity to generalize to populations at large, in contrast, convenience samples embed biases about reasons for participation that make it difficult to draw inference from such samples to other populations.

More subtle differences in sample selection may also make generalization from studies employing different sampling strategies challenging. For example, although both the studies conducted by Galea *et al.* and Silver *et al.* discussed in chapters in this part of the book, recruited population representative samples, the former used random-digit-dialing techniques for sample selection while the latter employed an existing Internet-based cohort that was originally recruited through random-digit-dialing techniques (Galea *et al.*, this volume; Silver *et al.*, this volume). Many factors, including computer literacy, personal need for ongoing incentives, and amount of free time, may influence ongoing participation in an Internet-based cohort. The epidemiologic experience with these methods is limited and it is difficult to fully know how to account for the potential factors that influence ongoing participant involvement in this cohort, and as a result, to draw firm conclusions about the comparability of these two samples, even though both may mirror their target population socio-demographically. This issue may be a particular challenge when considering sampling strategies employed across countries. For example, in a series of studies that have assessed the consequences of the Chernobyl nuclear meltdown in the Ukraine, Bromet *et al.* (2000) recruited samples that represented the populations of interest in the Ukraine. However, the incentives to participate in such a study within a rigidly regimented centrally controlled system are undoubtedly different than the incentives that motivate volunteers in the USA to participate in disaster research. The challenges in obtaining representative samples, and the specific implications of the decisions made by the investigators to obtain such samples need to be understood and carefully considered when interpreting results of post-disaster research. An upcoming volume that concerns itself with methodologic issues in post-disaster research addresses these issues in substantially more detail (Norris *et al.*, 2006).

#### Comparing data across studies

Inherent in this discussion is the fact that data obtained from different sampling frames is frequently not directly comparable and as such cross-study inference can only be made with judicious attention to the nature of the sampling frame and its components. In some ways this observation is self-evident. For example, it is clearly unwise to use data from samples of children, such as those presented by Hoven *et al.* in this volume, to extrapolate to potential findings in adults, or vice versa (Hoven *et al.*, this volume). Other cross-sample comparisons may be more

tempting to make and may well lead to erroneous conclusions. This may be particularly relevant when considering general population samples given these samples' heterogeneity. For example, reviews of the empiric epidemiologic literature have long suggested that the psychological consequences of human-made disasters may be more pronounced than the consequences of natural disasters (Norris *et al.*, 2002). However, as has been previously noted (Galea *et al.*, 2005), the persons who have typically been enrolled in studies of human-made disasters (e.g., explosions) tend to be more directly affected by the events than studies of natural disasters (e.g., tornadoes), where the samples typically include persons in the general community whose exposure is more heterogeneous. It is likely that the reason for lower estimates of psychopathology in the latter samples is simply that these samples include more persons whose exposure was less marked than those in the former samples.

Limitations of cross-sample comparisons have somewhat limited the field. In some ways all disasters are unique and as such observations drawn from one disaster study and one particular sample may require replication in several subsequent disasters before we can draw generalizations confidently. It is worth noting though, that it is often possible to compare specific subsamples within larger studies. For example, in the series of studies after the September 11, 2001, terrorist attacks conducted by Galea *et al.* (and discussed in part in an earlier chapter in this volume), the prevalence of PTSD among the subset of persons in the general population sample who actually were in the World Trade Center complex during the attacks was 34% (Galea *et al.*, 2003), virtually identical to the prevalence of PTSD documented among victims of the Oklahoma City bombing (most of whom were in the Murrah Federal building during the bombing) studied by North *et al.* (1999). Such comparisons can increase our confidence in replicability of these epidemiologic findings and can be immensely helpful to public health planners.

#### **Comparing data within studies**

A parallel issue to the above, and one that has been the subject of some debate in research related to this issue, is the issue of within-sample comparison. As Havenaar and Bromet suggest in their chapter in this part (Havenaar & Bromet, this volume), identifying control communities that are exactly comparable to those affected by disasters is next to impossible. This becomes even more apparent when we consider the potential contributions of community-level factors to population rates of disease. The search for control communities that are comparable to affected communities both on individual socio-demographic parameters and on key contextual characteristics is almost certainly futile. General population samples then present a particularly important opportunity for within-study comparison of persons who were differently exposed to a disaster but all of whom may plausibly have been at risk for psychopathology after the event. For example, in the series of studies

conducted by Galea *et al.* after September 11, 2001, a sampling frame that included New York City and the New York metropolitan area allowed the investigators to compare persons who were geographically closer or further from the disaster focus (as discussed in a chapter in this book). This sample also allowed for comparison between those who were highly exposed (e.g., close to the World Trade Center) and those who were at most only indirectly affected by the event. This then permits comparison to other studies that have sampled only persons who were direct victims of the event (as noted earlier in comparison to North *et al.*, chapter) and also allows for intra-study comparisons between exposure groups, testing hypotheses both about risk factors for individual psychopathology and population-level determinants of disease rates.

## Assessment methods

### Rigorous adherence to DSM-IV nosology

There is little question that the field of mental health epidemiology has been immeasurably advanced by the establishment of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM), and in particular since the DSM-III which was published in 1980. The DSM provides diagnostic algorithms and, as a result, allows for comparability of diagnosis across clinicians. North *et al.* in their chapter, discuss the implications of the DSM nosology for PTSD (North *et al.*, this volume). I concur with North *et al.* that the DSM should provide the basis for assessment of psychopathology in disaster research. However, it is also worth noting that diagnosable disorders and diagnostic criteria within psychiatry have not been subjected to extensive validation. Several authors (e.g., see Rounsaville *et al.*, 2002) have addressed the uncertainties of validating psychiatric diagnoses as established in the DSM. In particular, the DSM assumes that "psychiatric disorders are discrete biomedical entities with clear phenotypic boundaries" (p. 8) (Rounsaville *et al.*, 2002; Pfefferbaum *et al.*, in press). The DSM itself acknowledges this issue and notes that diagnostic criteria are "meant to serve as guidelines to be informed by clinical judgment." (American Psychiatric Association, 2000, p. xxii). Epidemiologic studies attempt to obtain data from participants in a systematic and invariant manner. As a result, numerous structured instruments have been developed that allow researchers to systematically collect psychiatric data from participants in a reproducible fashion. However, clearly, none of these instruments allow for clinical judgment and *none* of them can make diagnoses. Probably the most broadly accepted full structured interviews today are the CIDI, Composite International Diagnostic Interview and the clinician-administered PTSD scale (CAPS) (a discussion of the properties of the CIDI is provided in the Havenaar and Bromet chapter in this

volume). However, these instruments are too lengthy for use in most general population surveys, where the researcher must balance issues of volunteer burden and costs, and obtain valid information as efficiently as possible. As a result, general population studies usually rely on brief instruments that have been validated, typically in studies that have compared their discriminatory ability to clinician diagnosis. Beyond such validation, the choice of instrument used then must depend on the particular study exigencies, the questions that are being asked, and the methods that are being used. It is in turn the responsibility of the researchers to make sure their decisions and instruments are presented clearly and of the consumers of the research to evaluate the validity of these findings accordingly.

There are two final points worth making in this regard. First, while I concur wholeheartedly with North *et al.* discussion (in this volume) about the need for semantic caution in presenting findings about psychopathology, I would suggest that since no diagnosis can be established absent clinical assessment, the essence of careful research interpretation lies not in nomenclature used for mental health status detected (i.e., whether the appropriate term is "probable PTSD" or not), but rather in judicious interpretation of the findings of studies with careful attention paid to what the measures used in a particular studies were and how they were implemented. Second, regardless of our potential concern for the nosologic implications of findings from general survey research in the general population, the fact remains that several studies have shown that a substantial proportion of persons in the general population reported psychopathologic symptoms after the September 11, 2001, terrorist attacks, and that a proportion of those persons reported that these symptoms reduced their functioning (Galea & Resnick, 2005). It seems to this author self-evident that we must then consider the implications of this observation, both for the purposes of public mental health planning, and as an occasion for critical reflection on our current DSM-IV nosology and its potential limitations.

#### **Choice of screening instrument**

A corollary, but important point to this thinking is that the choice of assessment instrument, dictated, as I note above, primarily by what the key questions of interest are and what the best methods might be to address these methods. Use of different instruments across studies inevitably further complicates cross-study comparison. The use of different assessments instruments is critical because structural differences between assessment methodologies may lead to meaningfully different population-based prevalence estimates and small differences in diagnostic prevalence can produce significant underestimates or overestimates of the post-disaster needs of a community. To illustrate, in the 1–2 months following the September 11 terrorist attacks, Schlenger *et al.* (2002), using the PTSD Check List (PCL) as an assessment instrument, estimated that 11.2% of New York City area residents met criteria for

probable PTSD. This estimate was 50% higher than of the 7.5% prevalence of PTSD estimated by Galea *et al.* (2002) in a sample of Manhattan residents using the National Women's Study (NWS) as an assessment instrument. Applying each of these prevalences to the population of over 8 million New York City residents, one would estimate that approximately 900,000 (Schlenger *et al.*, 2002) vs. 600,000 (Galea *et al.*, 2002) persons met criteria for probable PTSD, a difference of 300,000 people. Clearly, a discrepancy of 3.7% points can have significant implications for public mental health planning and resource allocations. Subsequent analyses have shown that the NWS instrument is more specific than the PCL and it is likely that these differences between these two population representative studies are primarily a function of the assessment instrument used (Ruggiero *et al.*, in press). It is then important that studies conducted after disasters explicitly state the psychometric properties of the assessment measures used to enable consumers of the research to adequately assess how to evaluate results from individual studies and how to compare results across studies.

### Conclusion

As discussed throughout this section, and indeed in the whole book, post-disaster research presents particular challenges that must be addressed in order for this work to adequately document and understand the consequences of these events. While some of these challenges are endemic to all work carried out after disasters and mass traumas, others are particular to specific events. Clearly, researchers interested in the aftermath of disasters need to consider these issues, develop and implement study designs that best address them, and derive the best-possible inference from their studies. Equally as important, however, is the careful interpretation of post-disaster research. Unfortunately, both consumers of research and researchers themselves often fail to consider the methodologic and conceptual nuances that are guiding a particular study. Results from studies conducted in New York City are frequently conflated with studies conducted across the USA. Clearly, persons in each of these sampling frames may be at risk for different consequences of the September 11 terrorist attacks for different reasons. Scrupulous attention to decisions and assumptions embedded in particular study designs would go a long way to judicious inference drawn from post-disaster research. Conversely, researchers have a responsibility to present their work clearly, in such a way that consumers of the research can understand both its limitations and the extent to which inference can be drawn from their work. As the field expands, and as particular attention to methodologic rigor in these areas becomes paramount, the contribution of the field will undoubtedly grow and we will come closer to the goal of minimizing human suffering after mass traumas and terrorism.



## REFERENCES

- Ahern, J., Galea, S., Resnick, H. & Vlahov, D. (2004). Television images and probable PTSD after September 11: the role of background characteristics, event exposures and peri-event panic. *Journal of Nervous and Mental Disease*, 192(3), 217–226.
- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: American Psychiatric Association.
- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, (4th ed. Revised). Washington, DC: American Psychiatric Association.
- Brewin, C.R., Andrews, B. & Valentine, J.D. (2000). Meta-analysis of risk factors for post-traumatic stress disorder in trauma-exposed adults. *Journal of Consulting and Clinical Psychology*, 68, 746–766.
- Bromet, E.J., Goldgaber, D., Carlson, G., Panina, N., Golovakha, E., Gluzman, S.F., Gilbert, T., Gluzman, D., Lyubsky, S. & Schwartz, J.E. (2000). Children's well-being 11 years after the Chernobyl catastrophe. *Archives of General Psychiatry*, 57, 563–571.
- Galea, S., Ahern, J., Resnick, H., Kilpatrick, D., Bucuvalas, M., Gold, J. & Vlahov, D. (2002). Psychological sequelae of the September 11 terrorist attacks in New York City. *New England Journal of Medicine*, 346, 982–987.
- Galea, S., Vlahov, D., Resnick, H., Ahern, J., Susser, E., Gold, J., Bucuvalas, M. & Kilpatrick, D. (2003). Trends in probable posttraumatic stress disorder in New York City after the September 11 terrorist attacks. *American Journal of Epidemiology*, 158(6), 514–524.
- Galea, S. & Resnick, H. (2005). Post-traumatic stress disorder in the general population after mass terrorist incidents: considerations about the nature of exposure. *CNS Spectrums*, 10(2), 107–115.
- Galea, S., Nandi, A. & Vlahov, D. (2005). The epidemiology of post-traumatic stress disorder after disasters. *Epidemiologic Reviews*, 27, 78–91.
- Galea, S., Ahern, J., Resnick, H. & Vlahov, D. (this volume). Post-traumatic stress symptoms in the general population after a disaster: implications for public health. In *9/11: Mental Health in the Wake of a Terrorist Attacks*, eds. Y. Neria, R. Gross & R. Marshall. Cambridge, UK: Cambridge University Press.
- Goenjian, A.K., Pynoos, R.S., Steinberg, A.M., Endres, D., Abraham, K., Geffner, M.E. & Fairbanks, L.A. (2003). Hypothalamic–pituitary–adrenal activity among Armenian adolescents with PTSD symptoms. *Journal of Traumatic Stress*, 16(4), 319–323.
- Havenaar, J.M. & Bromet, E.J. (this volume). Capturing the impact of large scale events through epidemiological research. In *9/11: Mental Health in the Wake of a Terrorist Attacks*, eds. Y. Neria, R. Gross & R. Marshall. Cambridge, UK: Cambridge University Press.
- Hoven, C.W., Mandell, D.J., Duarte, C.S., Wu, P. & Giordano, V. (this volume). An epidemiological response to disaster: the New York City Board of Education's post 9/11 needs assessment. In *9/11: Mental Health in the Wake of a Terrorist Attacks*, eds. Y. Neria, R. Gross & R. Marshall. Cambridge, UK: Cambridge University Press.
- Norris, F.H., Friedman, M.J., Watson, P.J., Byrne, C.M., Diaz, E. & Kaniasty, K. (2002). 60,000 disaster victims speak. Part I: an empirical review of the empirical literature, 1981–2001. *Psychiatry*, 65, 207–239.

- Norris, F., Galea, S., Friedman, M., Watson, P. (2006). *Research Methods for Studying Mental Health after Disasters and Terrorism* (forthcoming from Guilford Press).
- North, C.S., Nixon, S.J., Shariat, S., Mallonee, S., McMillen, J.C., Spitznagel, E.L. & Smith, E.M. (1999). Psychiatric disorders among survivors of the Oklahoma City bombing. *Journal of the American Medical Association*, 282, 755–762.
- North, C.S., Pfefferbaum, B. & Hong, B. (in press). Historical perspective and future directions in research on psychiatric consequences of terrorism and other disasters. In *9/11: Mental Health in The Wake of a Terrorist Attacks*, eds. Y. Neria, R. Gross & R. Marshall. Cambridge, UK: Cambridge University Press.
- Pfefferbaum, B., Stuber, J., Fairbrother, G. & Galea, S. (in press). Panic reactions to the September 11 attacks in adolescents. *Journal of Traumatic Stress*.
- Rose, G. (1992). *The Strategy of Preventive Medicine*. New York: Oxford University Press.
- Rounsaville, B.J., Alarcon, R.D., Andrews, G., Jackson, J.S., Kendell, R.E. & Kendler, K. (2002). Basic nomenclature issues for DSM-V. In *A Research Agenda for DSM-V*, eds. D.J. Kupfer, M.B. First & D.A. Regier. Washington DC: American Psychiatric Association, pp. 1–29.
- Ruggiero, K., Rheingold, A.A., Resnick, H., Kilpatrick, D.G. & Galea, S. (in Press). Comparison between two widely-used PTSD screening instruments: implications for public mental health planning. *Journal of Traumatic Stress*.
- Schlenger, W.E., Caddell, J.M., Ebert, L., Jordan, B.K., Rourke, K.M., Wilson, D., Thalji, L., Dennis, J.M., Fairbank, J.A. & Kulka, R.A. (2002). Psychological reactions to terrorist attacks: findings from the National Study of Americans' reactions to September 11. *Journal of the American Medical Association*, 288, 581–588.
- Silver, R.C., Holman, E.A., McIntosh, D.N., Poulin, M., Gil-Rivas, V. & Pizarro, J. (in press). Coping with a national trauma: a nationwide longitudinal study of responses to the terrorist attacks of September 11th. In *9/11: Mental Health in The Wake of a Terrorist Attacks*, eds. Y. Neria, R. Gross & R. Marshall. Cambridge, UK: Cambridge University Press.
- Tucker, P., Pfefferbaum, B. & Nixon, S J. (2000). Predictors of post-traumatic stress symptoms in Oklahoma City: exposure, social support, peri-traumatic responses. *Journal of Behavioral Health Services and Research*, 27, 406–415.