Journal of Traumatic Stress, Vol. 19, No. 2, April 2006, pp. 173–184 (© 2006)

Disaster Research Methods: Past Progress and Future Directions

Fran H. Norris
National Center for PTSD and Dartmouth Medical School

Published results for 225 disaster studies were coded on methodological variables, severity of effects, and event year. Methods varied greatly, but cross-sectional, after-only designs, convenience sampling, and small samples were modal. Samples that were assessed before the disaster, selected for reasons of convenience, or were large tended to show less severe effects than other samples. Developing countries were underrepresented overall, but not in recent years. Certain desirable study characteristics (longitudinal designs, representative samples) have been decreasing in prevalence over time, whereas others (early first assessment) have been increasing. Innovations such as latent trajectory modeling or hierarchical linear modeling might advance the field's ability to capture the complexity of disasters, but the field still needs to attend to the fundamentals of sound epidemiologic research.

Regardless of whether they are natural in origin or human-caused, disasters are extremely complex events. Disasters generate an array of individually and collectively experienced stressors of varying degrees of intensity that interact with multiple characteristics of the person and environment to produce diverse outcomes that evolve over time. Although we may speak of disasters abstractly, specific studies are designed to capture the effects of a particular event on a particular population at a particular time. Thus, the elements of time and population are fundamental to the research plan and may serve to organize the primary methodological challenges that disaster researchers face.

Time, of course, is a critical variable in disaster research. The aftermath of a disaster is a motion picture, its effects a moving target. Answering important questions about onset and duration of effects can be exceptionally challenging for disaster researchers. The unpredictability of disasters makes it difficult for researchers to enter the field quickly and to plan for repeated assessments. Postdisaster intervention research must control for the natural course of recovery and deal with changing needs over time.

The element of population is important for disaster research in two ways. First, although defining the population of interest is not always straightforward, representative samples nonetheless allow the researcher to derive more accurate estimates of the prevalence and distribution of disorders in disaster-stricken communities than do other approaches. Such estimates are essential for building knowledge regarding the conditions under which disasters affect mental health. Second, there is also the issue of how well the entire body of disaster research represents the entire population of disaster victims. Most striking is the

The preparation of this article was supported by NIMH Grant R25 MH068298 awarded to Fran Norris. The Department of Veterans Affairs also provided support. Appreciation is extended to Cindy Elrod and Peggy Willoughby for their assistance in updating the literature review. A complete list of references for articles included in the review may be found at www.redmh.org.

Correspondence concerning this article should be addressed to: Fran Norris, NCPTSD, Veterans Affairs Medical Center, 215 North Main Street, White River Junction, VT 05009. E-mail: fran.norris@dartmouth.edu.

 Published online in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/jts.20109

mismatch between the locations in which disasters are most likely to occur (developing countries) and the locations in which disaster research is primarily done (the United States and other developed countries) (De Girolamo & McFarlane, 1996).

To evaluate the state of the art in disaster research, I asked four primary questions. How—and how well—has the field captured the elements of time and population? Do research methods influence results observed? What have been the trends over time in the use of good methods? What emerging methods may help the field to do a better job of capturing the elements of time and population and, more broadly, to capture the complexity of disasters?

THE DATA

The data for this review were obtained as part of a larger review of the disaster literature. A complete list of references for articles included in the review may be found at www.redmh.org. The database of 160 samples (102 distinct events) described by Norris et al. (2002; see also, Norris, Friedman, & Watson, 2002) was updated for this purpose and now is composed of over 225 distinct samples composed of over 85,000 individuals who experienced 132 distinct events. The data were captured from quantitative articles, chapters, and books published, in English, between 1981 and 2004. The included works mostly are those that were identified by the authors as relevant by their use of the word, *disaster(s)*, in their titles, abstracts, or key words or according to a variety of thesaurus terms in the PILOTS database, an electronic database of traumatic stress literature produced by the National Center for Post-Traumatic Stress Disorder. Although usually self-evident, what exactly constitutes a disaster is not always clear at the boundaries (McFarlane & Norris, in press). The focus here is on acute, collectively experienced events with sudden onset, thereby excluding research on chronic hazards (e.g., living near a landfill) and dislocation (e.g., Cambodian refugees) and terrorism (e.g., Israeli–Palestine conflict) that occurs within the context of ongoing political conflicts or war. Traumatic events occurring to individuals or families (house fire, single-car accident) were excluded, whereas similar but collectively experienced events (multiple-unit fire, massive traffic accidents) were included. Research that relied solely on archival or social indicator data or that focused solely on geographically distant, indirect, or anticipated experiences was also excluded. For example, although I included several studies that examined the effects of the September 11, 2001 terrorist attacks on samples in New York and Washington, DC, I excluded several studies that looked at the effects of terrorism on the nation as a whole or on samples outside of New York or Washington, DC. Because they were too few in number to review validly, studies of preschool age children and bioterrorism were also excluded. The boundaries for this review should not be interpreted to imply that the excluded events are less important, only that they did not meet the criteria for inclusion.

Variables Coded

I coded the data myself. All codes were checked once; many were double-checked. Although the coding of key features was enacted to gain some control over subjectivity, salience, and memory, the results are nonetheless solely the work of the author, and interrater reliability is unknown.

Sample type, disaster location, and disaster type. As shown in Table 1, each sample was coded on three primary descriptors: sample type (youth, adult survivor, rescue/recovery), disaster location (United States, other developed country, developing country), and disaster type (natural, technological, mass violence). Altogether, the review included 157 adult survivor samples (70%) and 35 school-aged youth samples (16%). Of these samples, 66% were predominantly 6 to 13 years old and 29% were predominantly 13 to 19 years old; 6% included both of these age groups. There were, in addition, 33 rescue/recovery samples (15%), such as firefighters, body handlers, and family-assistance counselors. The 34 countries represented in the database were grouped into three sets composed of the United States and its territories (116; 52%), other developed countries, composed primarily of samples from

Table 1. Samples in the Database by Disaster Type, Sample Type, and Disaster Location (n)

Disaster location Disaster type	Youth survivor	Adult survivor	Recovery worker
United States			
Natural	11	50	2
Technological	3	13	8
Mass violence	5	19	5
Other developed country			
Natural	4	17	2
Technological	4	19	11
Mass violence	0	5	1
Developing country			
Natural	7	28	3
Technological	0	3	1
Mass violence	1	3	0

the United Kingdom, Australia, western Europe, and Japan (63; 28%), and developing countries, composed of samples from the former Soviet Union, Asia other than Japan, and the Americas, other than the United States or Canada (46; 20%). Of the 225 samples, 124 (55%) experienced natural disasters (e.g., earthquakes, hurricanes, floods, wildfires), 62 (28%) experienced technological disasters (e.g., transportation accidents, industrial accidents, nuclear accidents, chemical spills, dam collapses), and 39 (17%) experienced mass violence (e.g., peacetime terrorist attacks, shooting sprees, civil disturbances).

***Severity of effects*.** The overall severity of effects observed in each sample was scored on a 4-point scale from *minimal* to *very severe*. Samples that exhibited minimal, highly specific, or transient effects were assigned a value of 1. Samples that exhibited moderate effects received a value of 2. Samples were assigned this score if they showed (a) elevations in symptoms over nonpatient norms or significant correlations between severity of exposure and psychological outcomes, and (b) rates of psychopathology below 25% in absolute terms. This category covers a wide range of actual effects. Samples that yielded rates of postdisaster psychopathology between 25% and 50% were assigned a value of 3, and those that yielded rates of psychopathology greater than 50% were assigned a value of 4. The classification relied on psychopathology as defined by the authors of the original articles and most often referred to posttraumatic stress disorder, major depressive disorder, generalized anxiety disorder, and panic disorder, in that order of frequency. The judgment was based on all disorders described by the authors in the results. Studies that examined two or more disorders almost always presented results for any disorder as well as for specific disorders. Quite often, these assignments were made based on investigators' reports of percentages above scale cut-points rather than according to diagnostic criteria; nonetheless, the last two results are relatively more severe than the first two. This strategy rather than a formal meta-analysis was used because the results of many descriptive studies did not lend themselves to derivation of effect sizes. Of the 225 samples, 20 (9%) received scores of 1 (minimal), 114 (51%) received scores of 2 (moderate), 53 (24%) received scores of 3 (severe), and 38 (17%) received scores of 4 (very severe).

***Methodological variables*.** Methodological variables included two design variables (*single postdisaster assessment* vs. *two or more postdisaster assessments; after-only* vs. *pre–post* design), timing of first and last points of data collection, coded in months, N, and sampling strategy. A range of sampling strategies emerged in the data. Least representative of disaster victims generally were *clinical* samples, composed primarily of litigants referred for clinical evaluation or of persons hospitalized at the time the data were collected. Regardless of whether the clinical samples were selected carefully to represent a total clinical population or more haphazardly, they were classified as clinical for the purposes of this review. Many samples were *convenience* samples, in which participants from nonclinical populations were self-selected or chosen because they were easiest to access. Some other samples were drawn by using *purposive* or quasi-random sampling techniques and were generally, if not precisely, representative of the population of interest. For example, researchers sometimes selected particular neighborhoods or schools for their studies to provide a cross-section of areas of impact. *Random* samples were drawn with known probability of selection and have a high probability, given high-response rates, of being highly representative of their populations. Occasionally, the size of the affected population was small enough to make sampling unnecessary, and the sample was a *census* of the population.

Event eras. The four event eras represented quartiles of the distribution: 24% of the samples experienced events that occurred before 1988, 22% events between 1988 and 1991, 28% events between 1992 and 1995, and 25% events after 1995.

DESCRIPTIVE ANALYSES OF METHODS IN DISASTER RESEARCH

Design

Most of these samples (72%, $n = 162$) were assessed once after the disaster, and the remaining (28%, $n = 63$) were assessed two or more times after the disaster. Of these, three used a successive cohort design, in which successive random samples were surveyed at various postdisaster intervals. The use of longitudinal designs did not vary according to sample type or disaster type, but it did vary by country type.[1] Longitudinal designs were less common in developing countries (15%) than in the United States and other developed countries (30%), $\chi^2(1, N = 222) = 4.49$, $p < .05$, $OR = 0.42$ (95% $CI = 0.18$–0.99).

Ten studies (4.4%) had true predisaster measures. The prevalence of pre–post designs did not differ by sample type. None of the studies conducted in developing countries and only one of the studies conducted in other developed countries had premeasures. Similarly, none of the studies conducted after incidents of mass violence and only one of the studies conducted after technological accidents had premeasures. These distributions did not lend themselves to statistical tests.

[1] Because of the large number of tests in this section, I present only the results of statistically significant, single *df* χ^2 tests. Single *df* tests were conducted only when the overall χ^2 test (e.g., design type by disaster location) was significant.

Timing of Assessment

Researchers have provided a substantial amount of data about short-term disaster effects. Although these samples were first assessed at any time from immediately to 7 years postdisaster, 61% were assessed within 6 months, 28% within 2 months. The timing of first assessment did not vary across sample type, disaster location, or disaster type.

Data about very long-term effects of disasters appear to be rare. Samples participating in longitudinal studies were interviewed as late as 17 years postdisaster, but 48% of the longitudinal samples gave their last interview within 1 year postevent.

Sampling Strategy

At 31%, convenience sampling was the mode. Otherwise, the distribution was 6% clinical, 17% purposive, 19% probability, and 27% census. Sampling strategy covaried strongly with sample type (see Table 2). Purposive sampling was disproportionately common in studies of youth (37% youth, 13% others), $\chi^2(1, N = 224) = 10.72$, $p < .001$, $OR = 4.06$ (95% CI $= 1.81$–9.12). It best described methods used in many school-based studies. Random sampling was most common in studies of adult survivors, (24% adult survivors, 6% others) $\chi^2(1, N = 224) = 12.55$, $p < .001$, $OR = 5.15$ (95% $CI = 1.76$–15.09), because population surveys primarily recruit adults. A large proportion of rescue workers (70%), but relatively small proportions of survivors (19%), were studied using the census method, $\chi^2(1, N = 224) = 32.08$, $p < .001$, $OR = 9.57$ (95% $CI = 4.20$–21.84). In these studies, all workers of a given profession (e.g., body handlers) who worked a particular event were recruited. It was not always clear whether the sample truly represented a census or was selected for reasons of convenience.

Sampling strategy also covaried with disaster location and disaster type. Census sampling was used more often in other developed countries (47%) than in the United States or developing countries (19% combined), $\chi^2(1, N = 224) = 16.48$, $p < .001$, $OR = 3.71$ (95% $CI = 1.97$–7.00). Purposive sampling was used more

Table 2. Methodological Variables by Substantive Variables in the Research (Column Percentages Except Where Noted)

	Sample type			Disaster location			Disaster type			
Variable value	Child survivor ($n = 35$)	Adult survivor ($n = 157$)	Rescue/ recovery ($n = 33$)	USA ($n = 116$)	Other developed country ($n = 63$)	Developing country ($n = 46$)	Natural disaster ($n = 124$)	Tech accident ($n = 62$)	Mass violence ($n = 39$)	Row % Severe or very severe
Design 1										
1 post time-point	77.1	74.0	63.6	69.0	71.4	84.8	76.4	65.6	73.7	40.5
2+ post time-points	22.9	26.0	36.4	31.0	28.6	15.2	23.6	34.4	26.3	40.7
Design 2										
After only	94.3	95.5	97.0	92.2	98.4	100.0	92.7	98.4	100.0	41.9
Pre-post	5.7	4.5	3.0	7.8	1.6	0.0	7.3	1.6	0.0	0.0
Timing of first assessment										
Within 2 months	17.1	27.7	39.4	31.9	23.0	23.9	27.6	29.5	25.6	33.3
3–6 months	40.0	31.0	36.4	32.8	32.8	34.8	32.5	29.5	41.0	43.8
7–12 months	25.7	25.2	9.1	24.1	21.3	21.7	23.6	21.3	23.1	42.0
>12 months	17.1	16.1	15.2	11.2	23.0	19.6	16.3	19.7	10.3	40.0
Sampling method										
Clinical	2.9	7.7	3.0	3.4	11.3	6.5	0.8	11.3	15.4	85.7
Convenience	25.7	35.9	18.2	32.8	27.4	34.8	40.7	17.7	25.6	31.0
Purposive	37.1	14.1	6.1	14.7	8.1	32.6	22.8	4.8	15.4	54.1
Random	8.6	24.4	3.0	25.9	6.5	17.4	25.2	8.1	15.4	23.8
Census	25.7	17.9	69.7	23.3	46.8	8.7	10.6	58.1	28.2	45.0
Sample size										
1–100	25.7	32.5	51.5	31.9	49.2	19.6	24.2	46.8	46.2	44.0
101–400	45.7	43.9	36.4	44.0	38.1	47.8	50.8	41.9	20.5	39.4
401–1000	8.6	16.6	9.1	13.8	7.9	23.9	17.7	6.5	15.4	46.9
>1000	20.0	7.0	3.0	10.3	4.8	8.7	7.3	4.8	17.9	15.8

often in developing countries (33%) than in developed countries (12%), $\chi^2(1, N = 224) = 9.54$, $p < .01$, $OR = 3.43$ (95% $CI = 1.60–7.34$). Clinical samples were used less often in studies of natural disasters (1%) than in studies of human-caused disasters (13% combined), $\chi^2(1, N = 224) = 15.57$, $p < .001$, $OR = 0.06$ (95% $CI = 0.01–0.43$). However, convenience samples were used more often in studies of natural disasters (41%) than in studies of human-caused disasters (21% combined), $\chi^2(1, N = 224) = 10.36$, $p < .001$, $OR = 2.61$ (95% $CI = 1.43–4.76$). Census sampling was used much more often after technological disasters (58%) than after others (16%), $\chi^2(1, N = 224) = 40.10$, $p < .001$, $OR = 7.96$ (95% $CI = 4.09–15.48$). Sample type and disaster type were confounded, with 32% of technological-accident samples being composed of rescue/recovery workers, compared to only 8% of other disaster types, $\chi^2(1, N = 225) = 18.94$, $p < .001$, $OR = 5.50$ (95% $CI = 2.53–11.96$). However, in a logistic regression predicting census sampling (vs. all other types), independent effects were observed for both rescue/recovery vs. other sample types, $OR = 6.50$ (95% $CI = 2.65–15.93$), $p < .001$, and technological accident vs. other disaster types, $OR = 6.12$ (95% $CI = 3.02–12.39$), $p < .001$.

Sample Size

When possible, larger sample sizes (*N*s) are generally preferred because smaller *N*s yield error and low power and limit the ability to examine important subgroups. The size of these samples varied from very small (11) to very large (5,687). The median size was 150. Approximately one third (34%) of the samples were composed of fewer than 100 persons, and only 23% were composed of more than 400 persons. As shown in Table 2, the distribution of sample size varied with sample type. Small samples ($\leq$100) were disproportionately common in studies of rescue/recovery workers (52%) compared to survivors (31%), $\chi^2(1, N = 225) = 4.91$, $p < .05$, $OR = 2.34$ (95% $CI = 1.11–4.94$), and large samples ($N > 1{,}000$) were disproportionately common in studies of youth (20%) compared to adults (6%), $\chi^2(1, N = 225) = 5.73$, $p < .05$, $OR = 3.71$ (95% $CI = 1.35–10.22$), probably because of the use of school-based studies.

Sample sizes also varied with disaster location and disaster type. Small samples were more common in other developed countries (49%) than in either the United States or developing countries (28% combined), $\chi^2(1, N = 225) = 8.49$, $p < .01$, $OR = 2.44$ (95% $CI = 1.34–4.45$), and more common in studies of technological accidents (47%) than in studies of other disaster types (29%), $\chi^2(1, N = 225) = 5.84$, $p < .05$, $OR = 2.11$ (95% $CI = 1.15–3.84$).

Relations Between Method Variables

Do researchers make trade-offs such that certain desirable features make undesirable features more likely? Timing of assessment was related to other methodological choices. Compared to samples first assessed after 2 months, rapidly assessed samples were more likely to have been selected for reasons of convenience (44% of rapidly assessed samples vs. 27% of those first assessed later), $\chi^2(1, N = 222) = 5.58$, $p < .05$, $OR = 2.10$ (95% $CI = 1.14–3.87$), and less likely to have been selected randomly (8% of rapidly assessed vs. 23% of assessed later samples), $\chi^2(1, N = 222) = 7.55$, $p < .05$, $OR = 0.29$ (95% $CI = 0.11–0.78$). In addition, rapidly assessed samples were disproportionately likely to be small in size ($N < 100$), (47% of rapidly assessed vs. 29% of assessed later samples), $\chi^2(1, N = 223) = 6.47$, $p < .01$, $OR = 2.20$ (95% $CI = 1.20–4.02$).

Other relations between method variables were observed as well. Sampling strategy was strongly related to sample size, scored categorically, $\chi^2(12, N = 224) = 75.55$, $p < .001$ (see Table 2). Clinical samples were disproportionately likely to be small in size (71% vs. 31% of others had Ns < 100), $\chi^2(1, N = 224) = 8.78$, $p < .01$, $OR = 5.46$ (95% $CI = 1.65–18.03$), as were census samples (50% vs. 28%), $\chi^2(1, N = 224) = 9.15$, $p < .01$, $OR = 2.57$ (95% $CI = 1.39–4.72$). Random samples were disproportionately likely to be very large (19% vs. 6% of others had Ns $> 1{,}000$), $\chi^2(1, N = 224) = 6.14$, $p < .05$, $OR = 3.66$ (95% $CI = 1.37–9.77$), as were purposive samples (19% vs. 6% of others), $\chi^2(1, N = 224) = 5.08$,

$p < .05$, $OR = 3.40$ (95% $CI = 1.24$–9.34). The sampling types also differed when N was scored continuously, $F(4,223) = 9.85$, $p < .001$. In post hoc tests, purposive and random samples did not differ from each other; and clinical, convenience, and census samples did not differ from each other. The first set of samples, however, had larger sizes than the second set.

IMPLICATIONS OF METHODS FOR EFFECTS OBSERVED

Bivariate Tests

As also shown in Table 2, studies that had one postdisaster assessment were no more likely to find severe or very severe effects than were studies that had two or more postdisaster assessments.[1] A significant difference was not expected even though effects do dissipate over time (Norris et al., 2002). In longitudinal studies, the magnitude of effects was coded according to the maximum severity observed. For example, a sample that showed moderate effects at time 1 and minimal effects at time 2 would be coded as showing moderate effects. Studies using pre–post designs were much less likely to find severe (or very severe) effects than were studies using after-only designs, Fishers exact test, $p < .01$. The timing of the first postdisaster assessment was unrelated to the severity of effects observed. It is common for protocols to ask about symptoms and reactions experienced since the disaster, which may explain why the timing of when those questions were asked did not influence the overall severity of the effects observed.

Sampling method was also related to the severity of observed effects. Severe effects were disproportionately common in clinical samples (86% vs. 38% of others combined), $\chi^2(1, N = 224) = 13.01$, $p < .001$, $OR = 9.95$ (95% $CI = 2.17$–45.62) and disproportionately uncommon in both convenience samples (31% vs. 45% of others combined), $\chi^2(1, N = 224) = 4.08$, $p < .05$, $OR = 0.55$ (95% $CI = 0.30$–0.99), and random samples, (24% vs. 45% of others combined), $\chi^2(1, N = 224) = 6.40$, $p < .05$, $OR = 0.39$ (95% $CI = 0.18$–0.84). Sample size showed a trend toward influencing the severity of effects observed, $\chi^2(3, N = 225) = 6.85$, $p < .08$. Only 16% of very large samples ($N > 1{,}000$) showed severe effects compared to 43% of samples with $N \leq 1000$, $\chi^2(1, N = 225) = 5.88$, $p < .05$, $OR = 0.25$ (95% $CI = 0.07$–0.89).

Multivariate Tests

The implications of methods for results were examined more fully in an ordinal regression equation predicting the overall severity of effects (range 1–4) from the methodological variables, with sample type, disaster location, and disaster type controlled. The advantage of this method is that all effects were independent of the effects of the other variables in the equation. Because rescue/recovery samples differed in so many ways from others, this analysis was limited to survivor samples ($n = 189$). Independent variables were either coded as dummy variables (design, sampling strategy) or treated as continuous various (time of first assessment, N). Of the two design variables, only *pre–post* (scored 1) versus *after-only* (scored 0) was included in the equation. Because the timing of first assessment was highly skewed, it was recoded into categories of month 1–2, month 3–6, month 7–12, 2nd year, 3rd year, 4th year, 5th year, and 6th year or later. Because sample size was likewise highly skewed, it was recoded into categories of 1–100, 101–200, up to 1000, then 1001–2000, 2001–3000, and >3000. Sampling strategy was scored as four dummy variables for the categories clinical, convenience, purposive, and random, with census as the reference variable. (Combined, 19% of the survivor samples were censuses.)

The overall fit of the model was significant, $\chi^2(12, N = 189) = 69.15$, $p < .001$. Estimates for the method variables are shown in Table 3. Design, sampling strategy, and sample size were all independently related to severity of observed effects. Less serious effects were associated with the inclusion of pre-event measures, convenience sampling or random sampling, and a larger N.

Table 3. Results From Ordinal Regression Analysis Predicting Severity of Effects From Methods

Variable	B	$SE\,B$	Wald
Pre-post design	−2.18	0.77	7.97**
Timing of first assessment	−0.04	0.06	0.29
Clinical	0.49	0.67	0.54
Convenience	−1.22	0.46	7.17**
Purposive	−0.26	0.52	0.26
Random	−1.12	0.55	4.13*
N	−0.11	0.05	4.93*

Note. Census sample was the reference category for the four dummy variables representing sampling strategy (clinical, convenience, purposive, random).
*$p < .05$. **$p < .01$.

TRENDS IN THE RESEARCH

Table 4 shows the distributions of sample type, disaster location, and disaster type by event year. The proportions of samples composed of youth, adult survivors, or rescue/recovery workers has not changed over time.[1] However, the distribution of disaster locations has changed. Compared to earlier eras, a larger proportion of studies in the most recent era were conducted in developing countries, χ^2 (1, $N = 223$) = 24.29, $p < .001$, $OR = 6.96$ (95% $CI = 3.44$–14.11). Two 1999 earthquakes in Turkey and Taiwan accounted for approximately half of these samples, creating some uncertainty about whether or not this finding represents the beginning of a trend. The types of disasters studied have also been changing. The proportion of studies focused on technological accidents were lower in the most recent era than in previous eras, χ^2(1, $N = 223$) = 9.31, $p < .001$, $OR = 0.30$ (95% $CI = 0.13$–0.70), whereas the proportion of studies focused on mass violence was higher, χ^2(1, $N = 223$) = 5.60, $p < .01$, $OR = 2.43$ (95% $CI = 1.17$–5.02).

Many things influence the quality of a study other than design, sample size, and sampling strategy. Nonetheless, other things being equal, most disaster researchers would probably agree that quality increases with use of longitudinal designs, early initiation, representative samples, and larger Ns. Table 4 shows the distributions of the method variables over the four event eras included in the review. Whereas one indicator of methodological progress in the field would be for the use of longitudinal designs to have increased over time, this has not been the case. In fact, the proportion of studies with two or more postdisaster assessments was lower in the most recent era than it was earlier, χ^2(1, $N = 220$) = 5.12, $p < .05$, $OR = 0.42$ (95% CI = 0.19–0.93). However, compared to earlier studies, a higher percentage of studies of events occurring in or after 1992 have begun within 6 months of the disaster, χ^2(1, $N = 221$) = 7.55, $p < .01$, $OR = 2.15$ (95% CI = 1.24–3.73).

Sampling strategies also varied by event era. The proportion of samples that were high in representativeness (random or census) was lower in the most recent era (26%) than it was in earlier eras (51%), χ^2(1, $N = 222$) = 11.04, $p < .001$, $OR = 0.34$ (95% $CI = 0.18$–0.66). There was no relation between sample size in categories and event era. In fact, the correlation between sample N and event year approached 0, $r = .01$.

WHAT INNOVATIONS MAY APPLY?

Because the possibilities are vast, I focused on two selected aspects of this speculative question: How can we better study variations in the course of recovery? And how can we better capture interdependence in population-level studies? Keeping in mind the goal of capturing the complexity of disasters, I argue here that advanced regression modeling approaches hold promise for advancing the state of the art in disaster research. In this section, I describe a selection of these approaches and present examples from trauma research.

Capturing Time: Latent Trajectory Modeling

Traditionally, effects of time have been studied by using repeated measures ANOVA or linear regression. In these methods, there is one intercept and one slope for the sample. A more recent advancement is latent trajectory modeling (LTM; also known as *growth curve analysis*), which emphasizes individual trajectories. In this method, there

Table 4. Study Variables by Event Year (Column Percentages)

Variable value	Before 1988 ($n = 54$)	1988–1991 ($n = 50$)	1992–1995 ($n = 62$)	After 1995 ($n = 57$)
Sample type				
Child survivor	11.1	20.0	17.7	12.3
Adult survivor	75.9	62.0	67.7	75.4
Rescue/recovery	13.0	18.0	14.5	12.3
Disaster location				
USA	51.9	56.0	67.7	29.8
Other developed country	33.3	30.0	25.8	22.8
Developing country	14.8	14.0	6.5	47.4
Disaster type				
Natural disaster	50.0	48.0	62.9	59.6
Technological accident	42.6	40.0	16.1	12.3
Mass violence	7.4	12.0	21.0	28.1
Design 1				
1 post time-point	59.6	72.0	74.2	83.9
2+ post time-points	40.4	28.0	25.8	16.1
Design 2				
After only	94.4	96.0	91.9	100.0
Pre-post	5.6	4.0	8.1	0.0
Timing of first assessment				
Within 2 months	24.1	26.5	32.8	28.1
3–6 months	22.2	30.6	37.7	40.4
7–12 months	31.5	22.4	18.0	21.1
>12 months	22.2	20.4	11.5	10.5
Sampling method				
Clinical	9.4	8.0	1.6	7.0
Convenience	18.9	26.0	41.9	38.6
Purposive	5.7	16.0	16.1	28.1
Random	24.4	12.0	14.5	17.5
Census	34.0	38.0	25.8	8.8
Sample size				
1–100	25.9	42.0	40.3	28.1
101–400	50.0	34.0	45.2	42.1
401–1000	18.5	14.0	4.8	21.1
>1000	5.6	10.0	9.7	8.8

is one intercept and one slope for each participant. Because slopes vary, they can be studied. In this approach, repeated measures serve as multiple indicators on two latent factors representing intercept and slope. Although I did not locate an example from disaster research, examples are beginning to appear in trauma research. Murphy, Johnson, Chung, and Beaton (2003) used this method to predict posttraumatic stress disorder (PTSD) symptoms in parents 4 months to 5 years after the death of a child. Female gender was significantly associated with higher initial PTSD symptoms (the intercept) and with greater improvements over time (the slope). Repressive coping was positively associated with initial symptoms but unrelated to improvements. Social support, on the other hand, was unrelated to initial symptoms but positively associated with improvements. This analysis was informative for showing that resources that offer protection against the development of PTSD may not be the same as the resources that promote recovery from PTSD.

Latent trajectory modeling may have important applications to postdisaster intervention research. Many behaviors targeted by postdisaster interventions are changing naturally over time, and thus it may be less important to change the attribute at a particular point than it is to change the developmental trajectory. In an example from another area of research, Curran and Muthén (1999) first estimated the normal trajectory of the targeted behavior in a control group then estimated a second growth factor unique to the treatment group, and showed how the growth curve was altered as a function of treatment. This approach may be highly relevant to evaluating public health interventions after disasters where the goal is to hasten or facilitate the community's recovery and where it is not possible to use an experimental design.

Capturing Interdependence: Hierarchical Linear Modeling

A variety of methods for analyzing population-level data exists. Of these, hierarchical linear modeling (HLM; Bryk & Raudenbush, 1992) seems to hold particular promise for disaster research because it might promote a better understanding of the transactions of individual, family, and community recovery. I did not identify use of HLM within past disaster research, but Perkins and Taylor's (1996) study of community disorder and fear of crime is a relevant and good example (i.e., it is easy to see how the approach would transfer). In this study, the investigators randomly selected 50 blocks in 50 different Baltimore neighborhoods, 12 households each. They used three different methods for measuring community-level constructs: (a) a survey of participant perceptions of crime, physical disorder, social disorder; (b) observations (e.g., graffiti, litter) by trained raters recorded on a structured inventory; and (c) content analysis of newspaper articles (e.g., reports on crime). The investigators aggregated data to the block level. Perkins and Taylor analyzed the data by using HLM, designed for analyzing data where individuals are nested in larger units. Level I tests whether there are effects of individual-level predictors. Level II tests whether, in addition, there are effects of community-level predictors. For example, individual-level fear of crime, the dependent variable, was predicted by both the individual's perception of physical disorder (level I) and the observational data on physical disorder (level II).

This research suggests that it would be feasible to develop measures of conditions and processes at family and community levels after disaster—for example, losses, disruption, and extent of recovery in the victims shared built, natural, social, and economic environments. If so, it would also be feasible to test whether individuals' postdisaster outcomes are influenced by their personal losses and resources, their families' losses and resources, and their communities' losses and resources. Such an approach might be particularly useful in examining whether efforts to strengthen community recovery can, in fact, influence individual recovery.

SUMMARY AND CONCLUSIONS

To evaluate the state of the art in disaster research, I asked four primary questions. The first question concerned where we stand now. Undeniably, the field as a whole has not done very well in capturing the element of time. Less than a third of these samples were assessed more than once after the disaster, and very few were assessed before the disaster. Researchers have provided a substantial amount of data about short- and intermediate-term disaster effects but relatively little data about very long-term effects, as half of the longitudinal samples gave their last interview within 1 year postevent. Overall, the field has performed better than anticipated with regard to capturing the element of population. Although convenience sampling was the mode (31%), over half of the samples were highly or at least moderately representative of their respective populations. On the other hand, the median sample size was 150, and less than a fourth of these samples were composed of more than 400 persons. Thus, generally speaking, these studies were not powerful ones and were limited in their abilities to explore effects for important subpopulations.

The second question explored in this review was whether the research methods employed have influenced the results observed. This did appear to be the case; certain method variables were associated with sample-level severity of effects. Controlling for sample type, disaster location, and disaster type, outcomes were less severe when the disaster was examined by using pre–post designs (vs. after-only designs), large samples, or convenience samples. The finding for pre–post design may indicate, on the one hand, that controlling for predisaster symptoms makes for a more conservative test that lessens confounding of post-disaster mental health with predisaster mental health. In an earlier meta-analysis of 52 controlled disaster studies, Rubonis and Bickman (1991) showed that psychopathology effect-size estimates were greater when control data were collected retrospectively than when they were collected prospectively. On the other hand, there are a few examples in this literature of samples that were only minimally exposed but studied precisely because there were predisaster data available for the research.

That both random selection and larger samples were associated with less severe effects likely indicate that large sample surveys show weaker effects than do other types of disaster research. As Galea, Nandi, and Vlahov (2005) pointed out, general populations should not be expected to show the severity of impact that victim groups do. A sizable proportion of persons making up most general population samples were only indirectly exposed to the focal event. Disasters do appear to have indirect effects on secondary victims or the community-at-large, but these effects are generally less severe than are the effects of direct exposure on primary victims (e.g., Norris, Phifer, & Kaniasty, 1994). However, not all large samples were probability samples. Many of the largest samples were purposive samples, drawn from schools. These samples also often include substantial numbers of students with indirect forms of exposure to the disaster.

Of the various findings, the most intriguing is the weaker effects shown by convenience samples. This finding contradicts Rubonis and Bickman's (1991) earlier result but is based upon a much larger number of studies, including many that would not have met their criteria for inclusion. Moreover, in Rubonis and Bickman's analysis nonrepresentative samples included clinical/litigant, as well as convenience samples. Although the assumption may be that convenience samples are composed of disproportionately exposed or distressed persons, in reality this is not always the case. In fact, sometimes samples are chosen not because they are the most important but solely because they are accessible. These samples may do a disservice to the research base as a whole.

The third theme explored in this review was whether we are making progress. What have been the trends over time in the use of good methods? The number of disaster studies has increased steadily over the past 25 years, but the quality of disaster research has not kept pace with the quantity of disaster research. Samples sizes have remained the same, and the use of longitudinal designs and representative samples has actually decreased over time. However, there were some positive trends as well. The timing of first assessment showed improvement over time, with over two thirds of recent studies beginning within 6 months and more than a fourth beginning within 2 months. This trend should facilitate a better understanding of the acute postdisaster period. On the other hand, one might ask whether we are trading rapidity for quality because samples that were assessed rapidly (within 2 months) were disproportionately likely to be small and selected for reasons of convenience. A second positive finding was that developing countries are far better represented in research that is more recent. Two events accounted for the majority of these samples, and thus it would be premature to conclude that greater global representativeness is indeed a trend. Moreover, there are broader issues in international and/or crosscultural research that are not captured by simply counting studies. Most quantitative studies have done little that explicitly expands our knowledge of how culture shapes the experience of disaster stress. Another trend was a steady decrease in the proportion of studies focused on technological accidents coupled with a steady increase in the proportion of studies focused on mass violence. This trend is neither positive nor negative in and of itself, but it is responsive to the concerns of the public sector over the impact on survivors and the public of intentional disasters, such as those caused by terrorist and sniper attacks.

The fourth and final aim of this review was to consider what recent methodological innovations would be beneficial to the field. Methodological innovations have not been priorities perhaps because the field still struggles to produce studies of consistently sound quality (Norris, Galea, Friedman, & Watson, in press). Yet numerous innovations are on the horizon that would serve this field well by allowing researchers to model the complexity that is a fundamental feature of disaster. Latent trajectory modeling could greatly advance our understanding of factors that influence the course of recovery, and HLM could enable us to better portray the ecology of disaster recovery. Yet the utility of these methods rests on the fundamentals of study design. These innovative methods require sample sizes greater than are the norm, more waves of data than are the norm, and/or more complicated sampling strategies than are the norm. Use of methodological innovations in disaster research will be impeded as much by systemic barriers as by a lack of investigator awareness of them. Innovative policies and targeted research initiatives may be required.

REFERENCES

Bryk, A., & Raudenbush, S. (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage.

Curran, P., & Muthén, B. (1999). The application of latent curve analysis to testing developmental theories in intervention research. American Journal of Community Psychology, 27, 567–595.

De Girolamo, G., & McFarlane, A. (1996). The epidemiology of PTSD: A comprehensive review of the international literature. In A. Marsella, M. Friedman, Gerrity, E., & R. Surfield (Eds.), Ethnocultural aspects of posttraumatic stress disorder: Issues, research, and clinical applications (pp. 33–85). Washington, DC: American Psychological Association.

Galea, S., Nandi, A., & Vlahov, D. (2005). The epidemiology of post-traumatic stress disorder after disaster. Epidemiologic Reviews, 27, 1–14.

McFarlane, A. C., & Norris, F. (in press). Definitions and concepts in disaster research. In F. Norris, S. Galea, M. Friedman, & P. Watson (Eds.), Methods for disaster mental health research. New York: Guilford Press.

Murphy, S., Johnson, L. C., Chung, I., & Beaton, R. (2003). The prevalence of PTSD following the violent death of a child and predictors of change 5 years later. Journal of Traumatic Stress, 16, 17–25.

Norris, F., Friedman, M., Watson, P., Byrne, C., 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., Friedman, M., & Watson, P. (2002). 60,000 disaster victims speak, Part II: Summary and implications of the disaster mental health research. Psychiatry, 65, 240–260.

Norris, F., Galea, S., Friedman, M., & Watson, P. (Eds.) (in press). Methods for disaster mental health research. New York: Guilford Press.

Norris, F., Phifer, J., & Kaniasty, K. (1994). Individual and community reactions to the Kentucky floods: Findings from a longitudinal study of older adults. In R. Ursano, B. McCaughey, & C. Fullerton (Eds.), Individual and community responses to trauma and disaster (pp. 378–400). Cambridge: Cambridge University Press.

Perkins, D., & Taylor, R. (1996). Ecological assessments of community disorder: Their relationship to fear of crime and theoretical implications. American Journal of Community Psychology, 24, 1–3.

Rubonis, A., & Bickman, L. (1991). Psychological impairment in the wake of disaster: The disaster–psychopathology relationship. Psychological Bulletin, 109, 384–399.