The Utility of Quantile Regression in Disaster Research

Katarzyna Wyka[1,2*], Dana Sylvan[3] and JoAnn Difede[2]

[1]*Graduate School of Public Health and Health Policy, City University of New York, 55 W. 125th Street, New York, NY 10027*

[2]*Weill Cornell Medical College, 425 E 61 Street, New York, New York 10065*

[3]*Hunter College, City University of New York, 695 Park Avenue, New York, New York 10065*

Abstract: Following disasters, population-based screening programs are routinely established to assess psychological consequences of exposure. These data sets are highly skewed as only a small percentage of trauma-exposed individuals develop adverse mental health outcomes. Commonly used statistical methodology in disaster research generally involves population-averaged models, such as linear and logistic regressions. However, these models offer only a partial explanation of the complex relationships between the extent of disaster exposure, individual characteristics and adverse mental health outcomes. The aim of this report is to illustrate the benefits of using quantile regression in disaster research by analyzing the effects of a selected exposure variable, perceived threat to one's life, and education level on post-traumatic stress symptomatology among n=2960 non-rescue disaster workers exposed to the World Trade Center (WTC) disaster in New York City on 9/11. The findings of the study are in line with previous WTC research that documented the link between perceived danger associated with disaster work, low education level and elevated post-traumatic stress symptomatology. However, the use of quantile regression demonstrates the robust and differential association between these variables throughout the entire distribution of post-traumatic stress symptomatology. Specifically, we show that the effect of high perceived danger and low education level were more strongly associated with post-traumatic stress symptoms in the upper tail of the distribution, after adjusting for covariates. Quantile regression methodology has the potential to enrich disaster research by tackling research questions that were previously unanswered. This method may be particularly useful in analyzing large population-based screening programs.

Keywords: Quantile regression, Disaster research, Disaster workers, Post-traumatic stress, PTSD, World Trade Center.

INTRODUCTION

Post-disaster mental health research aims to capture the complexities of response to disasters, including natural and environmental disasters as well as terrorist attacks. Epidemiological studies based on routinely established population-based screening programs typically study the prevalence of negative mental health outcomes and their relationship with the extent of disaster exposure and individual characteristics of those affected. Disaster research collectively has shown that the most prevalent mental health consequence associated with disasters is post-traumatic stress disorder (PTSD) [1-2], an anxiety disorder characterized by re-experiencing, avoidance, arousal and reactivity as well as cognition and mood symptoms [3]. PTSD presents a significant burden for individuals, their families, and society at large and can become a chronic condition [2]. It is associated with substantial comorbidity, including depression, substance abuse and disability [4]. Subthreshold PTSD, a less severe form of post-traumatic stress disorder, is also common and associated with substantial symptomatology, comorbid conditions, and impaired functioning [5]. PTSD has been shown to affect individuals of all ages, sociodemographics and level of disaster training. Assessing which variables are most associated with the onset and persistence of post-traumatic stress symptoms is essential to efficiently offer mental health services to those most in need of treatment.

To assess risk factors associated with elevated post-traumatic stress symptoms, disaster researchers typically use multiple linear regression or logistic regression models [6]. Multiple linear regression allows to assess how the mean of the conditional distribution varies depending on the level of independent variables. Logistic regression requires that the response variable is dichotomized. Such models have several limitations in the context of disaster research. First, data obtained from post-disaster population-based screening programs are highly skewed since only a small percentage of disaster-exposed individuals develop adverse mental health outcomes, such as post-traumatic stress symptoms. Hence, linear regression models, which are based on normal theory and homogeneous distributions, have significant weakness given some of the assumptions underlying these

*Address correspondence to this author at the Graduate School of Public Health and Health Policy, City University of New York, 55 W. 125th Street, New York, NY 10027;
Tel: 646 364 0248; E-mail: kwyka@sph.cuny.edu

models are likely not satisfied. Dichotomizing the response variable, on the other hand, is typically associated with substantial loss of information and the choice of the cutoff point is often arbitrary. At best, both models offer an incomplete account of the relationships between the response and independent variables of interest. Furthermore, these models are not able to detect the partial effects of the independent variables at the different locations of the response variable distribution, *i.e*. in the tails and center of the conditional distribution. In the context of disaster research, it is plausible that the association between the extent of disaster exposure and post-traumatic stress symptomatology may be present among individuals with severe post-traumatic stress symptoms only. Likewise, while a link between education level and post-traumatic stress symptoms is well established, this association may be different among individuals with severe versus less severe post-traumatic stress symptoms.

Quantile regression methodology [7] provides an effective alternative for obtaining more sensitive insights into the relationships of interest and has the potential to yield more clinically relevant findings. Specifically, quantile regression allows to capture the effects of independent variables on the entire distribution of the response variable and assess direction and magnitude of the effects on different parts of the distribution (*e.g*. lower and upper tails, center). In this model, conditional quantiles (or percentiles) of the response variable are expressed as a linear function of the independent variables. The model is specified as follows: $y_i = \beta_0^{(q)} + \beta_i^{(q)} x_i + \varepsilon^{(q)}$ where q is a given quantile. The coefficients are interpreted as the change in a quantile q of the conditional distribution associated with one-unit difference in an independent variable x_i. Thus, the model allows to estimate whether the magnitude of the regression coefficient differs for a given independent variable at the various quantiles of interest. Standard errors and confidence limits for the quantile regression coefficient estimates can be obtained with asymptotic and bootstrapping methods. Both methods provide robust results [8], with the bootstrap method preferred as more practical [9]. A very clear and well-illustrated presentation of quantile regression methodology for social and behavioral research is given in [10, 11]. Quantile regression methodology is implemented in most professional statistical software packages. For comprehensive descriptions we refer to [12-14].

While the utility of quantile regression in biomedical research has been demonstrated in several recent publications such as [15-21] and the references therein, there is a paucity in the use of this methodology in disaster research. However, the field can benefit from modeling the effects of high order quantiles rather than mean effects, as high severity of negative mental health outcomes yield more devastating outcomes. The aim of this report is to illustrate the benefits of using quantile regression in disaster research by analyzing the effect of a selected exposure variable, perceived threat to one's life and education level on post-traumatic stress symptomatology among non-rescue disaster workers exposed to World Trade Center (WTC) disaster in New York City on 9/11. We conclude with a brief discussion where we make some suggestions for potential future avenues of research.

MATERIALS AND METHODS

The data used in this paper is part of an archival dataset collected through the Weill Cornell 9/11 Mental Health Screening Program (see [22-23] for program details). Participants were World Trade Center (WTC) non-rescue disaster workers who were deployed to work at the WTC site as part of their occupational duties (n=2960). The interviews were piggybacked onto their fitness-for-duty evaluations and were conducted by independent doctoral level psychologists. The confidentiality of all individuals was preserved by utilizing identification numbers, rather than worker names. The employer received only overall statistics on the prevalence of PTSD symptoms and related psychopathology. The Weill Cornell Medical College and the City University of New York Institutional Review Boards approved use of this data for research purposes.

The response variable of interest is the post-traumatic stress symptoms severity score assessed using the Clinician-Administered PTSD Scale (CAPS) [24], the standard measure in the assessment of PTSD. Symptom scores were summed to create a total score indicating the overall severity of the symptoms (possible range 0-136). The independent variables analyzed in this study are perceived threat to one's life while working at the site (yes versus no; exposure variable) and education level (high school diploma or less versus at least some college). Both variables were shown to be strongly associated with post-traumatic stress symptomatology in multiple studies [*e.g*. 22, 23]. The analyses are adjusted for occupational exposure

(*e.g*. duration of work, areas of operation), personal exposure (*e.g*. injury of death of a family member, friend, or colleague), lifetime trauma and psychiatric history, age, gender, race/ethnicity and marital status.

Quantile regression is used to examine the relationship between the severity of post-traumatic stress symptomatology (CAPS) and perceived threat to one's life while working at the site as well as education level. Quantiles of the response variable are defined based on the distribution of CAPS scores, where 50^{th}, 75^{th}, 90^{th} and 95^{th} quantiles are used. These quantiles correspond to minimal (CAPS=7), low-mild (CAPS=18), high-mild (CAPS=37) and high-moderate (CAPS= 50) severity levels. Quantile regression coefficients represent the estimated difference in the CAPS score for the quantiles of interest when comparing workers who reported perceived threat to one's life while working at the site versus those who did not and workers with at least some college education versus those with high school diploma or less, adjusting for all other individual characteristics listed above. Linear multiple regression estimates are also included to facilitate comparisons.

Analyses are conducted in SAS (PROC QUANT REG) [25] and replicated in R (package quantreg) [12].

RESULTS

A comprehensive description of the data can be found in [22, 23]. In summary, workers were mostly white (65.7%), male (96.9%), and married/cohabitating (75.4%), and many had completed at least some college (52.6%). The mean age was 45.2 years (SD = 9.6). Approximately one third of workers reported perceived threat to one's life while working at the site. Eight percent of participants had symptoms consistent with PTSD and another 9.3% had symptoms consistent with subthreshold PTSD. CAPS severity scores were non-normally distributed with M=13.07 (SD=16.08, range 0-95) (Figure **1**). Mean CAPS scores were higher among workers reporting perceived threat to one's life while working at the site and lower among those who completed at least some college.

Quantile regression results are reported in Table **1**. Not surprisingly, perceived threat to one's life while working at the site is a significant predictor of CAPS severity, however it is differentially associated with CAPS severity score at different location of the distribution. Specifically, the estimated difference in the CAPS severity score between workers reporting perceived threat to one's life while working at the site versus those who did not is 3.94 at the 50^{th} CAPS quantile, then it rises substantially across the quantiles to 19.21 at the 95^{th} CAPS quantile, suggesting that the effect is more pronounced among workers with more severe post-traumatic symptomatology. In contrast, linear regression results show that the estimated average difference in the CAPS severity score between those who reported perceived threat to one's life while working at the site versus those who did not is about 7 points.

Likewise, lack of at least some college education is associated with elevated CAPS scores, particularly among workers with more severe post-traumatic stress symptoms. The estimated between group difference is -.56, -1.40, -1.83 and -2.68 at the 50^{th}, 75^{th}, 90^{th} and 95^{th} quantiles, respectively. This result enriches the

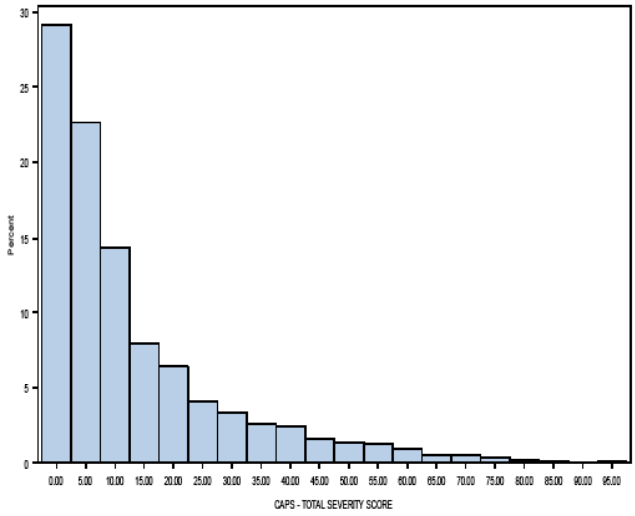

Figure 1: CAPS total severity score (n=2960).

Table 1: Results from Quantile and Linear Regression (Response Variable: Total CAPS Score)

	Quantile Regression				Linear Regression
	p=.50	p=.75	p=.90	p=.95	
	β	β	β	β	β
Perceived threat to one's life while working at the site	3.94	11.01	15.35	19.21	7.21
Education level	-0.56	-1.40	-1.83	-2.68	-1.20

Note. p=.50, p=.75, p=.90 and p=.95 are the 50th, 75th, 90th and 95th quantiles of CAPS distribution. Reference categories are as follows: Perceived threat to one's life while working at the site = no (versus yes); Education level: high school diploma or less (versus at least some college). Analyses are adjusted for occupational exposure (*e.g.* duration of work, areas of operation), personal exposure (*e.g.* injury of death of a family member, friend, or colleague), lifetime trauma and psychiatric history, age, gender, race/ethnicity and marital status.

linear regression estimate by showing that post-traumatic stress symptomatology for workers with a high school diploma or less is not only more severe but also more positively skewed, as compared to workers with at least some college education.

DISCUSSION AND CONCLUSION

The purpose of this report is to illustrate the utility of quantile regression in disaster research in contrast to commonly used linear regression models. We show the advantages of this methodology for assessing the differential effects of select risk factors for post-traumatic stress symptoms among the World Trade Center non-rescue disaster workers across quantiles of the symptom distribution. This approach has the potential to provide a richer understanding of the complex relationships between the extent of disaster exposure, individual characteristics and adverse mental health outcomes. The findings of the study are in line with previous WTC research that documented the link between perceived danger associated with disaster work, low education and elevated post-traumatic stress symptomatology [22, 23]. However, the use of quantile regression demonstrates the robust and differential association between these variables throughout the entire distribution of the post-traumatic stress symptomatology. For instance, we show that the effect of disaster exposure and education level is more strongly associated with post-traumatic stress symptoms in the upper tail of the distribution, after adjusting for covariates. While the analysis of the same data using a linear regression model found both factors were associated with the response variable, this analysis fails to disclose an important trend present in the data and generally underestimates the magnitude of the effects.

It should be noted that while quantile regression provides an interesting direction of novel disaster research, it is not without limitations [26, 27]. Limited information is available regarding minimal sample sizes overall and at quantile locations of interest. This methodology may yield elevated Type I error due to multiple hypotheses being tested. Lastly, computing time required for bootstrapping procedures may be substantial in some applications. However, some of these limitations are common in most of regression models and can be remedied by using well-established correction methods. Methodological research in quantile regression is ongoing and new developments are regularly implemented into mainstream statistical software [28-29].

In conclusion, this paper does not aim to provide a comprehensive study of risk factors for post-traumatic stress in WTC non-rescue workers. Instead, we illustrate the benefits of utilizing quantile regression in disaster research in order to increase the use of this methodology in the field, particularly when analyzing population-based data. Given that disasters, such as terrorist attacks, have been shown to be associated with persistent adverse mental health consequences, including PTSD, innovative research is needed to further elucidate the etiology, risk factors, and longitudinal course of these outcomes. Future avenues of research also include the examination of other outcomes associated with disaster related mental health outcomes, such as post-disaster occupational and social functioning as well as quality of life. Furthermore, longitudinal quantile regression can be utilized to model heterogeneous mental health outcomes over time. In our view quantile regression methodology has the potential to enrich disaster research by tackling research questions that were previously unanswered.

REFERENCES

[1] Norris FH. Disasters in urban context. J Urban Health 2002; 79(3): 308-314. https://doi.org/10.1093/jurban/79.3.308

[2] Norris FH, Tracy M, Galea S. Looking for resilience: understanding the longitudinal trajectories of responses to stress. Soc Sci Med 2009; 68: 2190-2198. https://doi.org/10.1016/j.socscimed.2009.03.043

[3] American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5. Washington, D.C: American Psychiatric Association 2013.

[4] Kessler RC. Posttraumatic stress disorder: the burden to the individual and to society. J Clin Psychiatry 2000; 61(Suppl 5): 4-12; discussion 13-14.

[5] Cukor J, Wyka KE, Jayasinghe N, Difede J. The nature and course of subthreshold PTSD. J Anxiety Disord 2010; 24(8): 918-923. https://doi.org/10.1016/j.janxdis.2010.06.017

[6] Norris FH. Disaster research methods: past progress and future directions. J Trauma Stress 2006; 19(2): 173-184. https://doi.org/10.1002/jts.20109

[7] Koenker R, Bassett G. Regression quantiles. Econometrica 1978; 46 (1): 33-50. https://doi.org/10.2307/1913643

[8] Koenker R, Hallock K. Quantile Regression: An Introduction. J Econom Perspect 2001; 15: 143-156. https://doi.org/10.1257/jep.15.4.143

[9] Hao L, Naiman D. Quantile Regression, Sage Publications, Thousand Oaks, 2007. https://doi.org/10.4135/9781412985550

[10] Yu K, Lu Z, Stander J. Quantile regression: applications and current research areas. The Statistician 2003; 52(3): 331-350. https://doi.org/10.1111/1467-9884.00363

[11] Fitzenberger B, Wilke RA. Quantile Regression Methods. Emerging Trends in the Social and Behavioral Sciences, John Wiley & Sons 2015; p. 1-18.

[12] Koenker R, Portnoy S, Ng PT, Zeileis A, Grosjean P, Ripley BD. R package Quantreg 2017. https://CRAN.R-project.org/package=quantreg.

[13] Chen C. An Introduction to Quantile Regression and the QUANTREG Procedure. SUGI 30 Proceedings; p. 213-230.

[14] Mu G, Cai Y, Han Y. Quantile Regression in Pharmaceutical Marketing Research, SAS Global Forum 2013; p. 1-4.

[15] Austin PC, Schull MJ. Quantile Regression: A Statistical Tool for Out-of-hospital Research, Acad Emerg Med 2003; 10(7): 789-797. https://doi.org/10.1111/j.1553-2712.2003.tb00075.x

[16] Costa-Font J, Fabbri D, Gil J. Decomposing Body mass Index gaps between Mediterranean countries: A Counterfactual Quantile Regression Analysis. The University of York HEDG Working paper 08/02, 2008.

[17] Gebregziabher M, Lynch CP, Mueller M, Gilbert GE, Echols C, Zhao Y, *et al*. Using quantile regression to investigate racial disparities in medication non-adherence. BMC Med Res Methodol 2011; 11(88): 1-11. https://doi.org/10.1186/1471-2288-11-88

[18] Fenwik EK, Xie J, Ratcliffe J, Pesudovs K, Finger RP, Wong TY, *et al*. The Impact of Diabetic Retinopathy and Diabetic Macular Edema on Health-Related Quality of Life in Type 1 and Type 2 Diabetes. Investigat Ophthalmol Vis Sci 2012; 53(2): 677-684. https://doi.org/10.1167/iovs.11-8992

[19] Payne AJ, Knight, JA, Abarin T. Effect of Environmental Factors on Obesity: A Quantile Regression Approach. J Biometrics and Biostatistics 2016; 7(2): 292-299. https://doi.org/10.4172/2155-6180.1000293

[20] Feizi A, Aliyari R, Roohafza H. Association of perceived stress with stressful life events, lifestyle and sociodemographic factors: a large scale community-based study using logistic quantile regression. Computational and Mathematical Methods in Medicine 2012; 2012: 1-12. https://doi.org/10.1155/2012/151865

[21] Cohen O, Bolotin A, Lahad M, Goldberg A, Aharonson-Daniel L. Increasing sensitivity of results by using quantile regression analysis for exploring community resilience. Ecological Indicators 2016; 66: 497-502. https://doi.org/10.1016/j.ecolind.2016.02.012

[22] Difede J, RobertsJ, Jayasinghe N, Leck P. Evaluation and treatment of emergency services personnel following the World Trade Center attack. In Neria Y, Gross R, Marshall R, Susser E.(Eds.). September 11, 2001: Treatment, Research and Public Mental Health in the Wake of a Terrorist Attack. New York: Cambridge University Press 2006; p.333-354.

[23] Cukor J, Wyka KE, Jayasinghe N, Weathers F, Giosan C, Leck P, *et al*. Prevalence and predictors of posttraumatic stress disorder in a population of 2,960 utility workers deployed to the World Trade Center following the attacks of September 11, 2001. Depress and Anxiety 2011; 28(3): 210-217. https://doi.org/10.1002/da.20776

[24] Blake DD, Weathers F, Nagy LM, Kaloupek DG, Klauminzer G, Charney DS, *et al*. A Clinician rating scale for assessing current and lifetime PTSD: the CAPS-1. Behavior Therapist 1990; 13: 187-188.

[25] SAS Institute Inc. The QUANTREG Procedure. Available at: http://support.sas.com/documentation/cdl/en/statugquantreg/61825/PDF/default/statugquantreg.pdf. Accessed March 15, 2016.

[26] Austin PC, Tu JV, Daly PA, Alter DA. The use of quantile regression in health care research: a case study examining gender differences in the timeliness of thrombolytic therapy. Statistics in Medicine 2005; 24: 791-816. https://doi.org/10.1002/sim.1851

[27] Petscher Y, Logan JAR. Quantile regression in the study of developmental sciences. Child Dev 2014; 85: 861-881. https://doi.org/10.1111/cdev.12190

[28] Koenker R. Quantile Regression. Cambridge UK: Cambridge University Press. 2005; p. 9-14, 32-33, 173-221, 293-294. https://doi.org/10.1017/CBO9780511754098

[29] Koenker R. Quantile Regression in R: A Vignette. Available at: http://www.econ.uiuc.edu/~roger/research/ rq/vig.pdf. Accessed March 15, 2016.

Received on 11-07-2017 Accepted on 07-08-2017 Published on 30-08-2017