*Title Page

Type of manuscript:

Original contribution

Manuscript title:

Designing an Observational Study to be Less Sensitive to Unmeasured Biases: Effect of the 2010 Chilean Earthquake on Posttraumatic Stress

Author names and affiliations:

José R. Zubizarreta: Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA

Magdalena Cerdá: Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY

Paul R. Rosenbaum: Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA

Corresponding address:

José R. Zubizarreta
Department of Statistics
The Wharton School
University of Pennsylvania
3730 Walnut Street, 431-3 JMHH
Philadelphia, PA 19104-6340
Phone: 215-573-6124
Fax: 215-898-1280
Email: josezubi@wharton.upenn.edu

Running head:

Effect of the 2010 Chilean Earthquake on Posttraumatic Stress

Conflicts of interest and source of funding:

The authors have no conflicts of interest to declare. This work was partly supported by a grant from the Measurement, Methodology and Statistics Program of the US National Science Foundation.

Acknowledgements:

The authors thank Chile's Ministry of Planning and Cooperation for access to their data, Carolina Casas Cordero for explanations about the data, and Sandro Galea for comments on a draft manuscript.

*Main Text

Designing an Observational Study to be Less Sensitive to Unmeasured Biases:

Effect of the 2010 Chilean Earthquake on Posttraumatic Stress

Revision: 29 June 2012

Submitted to Epidemiology

Abbreviations:

CASEN: National Socioeconomic Characterization Survey

EPT: Post Earthquake Survey

PGA: Peak ground acceleration

USGS: United States Geological Survey

95% CI: 95 percent confidence interval

KS-distance: Kolmogorov-Smirnov distance

ABSTRACT

In 2010, a magnitude 8.8 earthquake hit Chile, devastating parts of the country. Having just completed its national socioeconomic survey (CASEN), Chile interviewed a subsample of respondents, creating unusual longitudinal data about the same individuals before and after a major disaster. The follow-up evaluated posttraumatic stress symptoms (PTSS) using Davidson's Trauma Scale. We use these data with two goals in mind. Most studies of PTSS following disasters rely on recall to characterize the state of affairs prior to the disaster. In contrast, we use the CASEN to study effects of the earthquake on PTSS with prospective data on pre-exposure conditions, free of recall bias. Second, we illustrate recent developments in statistical methodology for the design and analysis of observational studies. In particular, we use new and recent methods for multivariate matching to control 46 covariates that describe demographics, housing quality, wealth, health and health insurance prior to the earthquake. Also, we use the statistical theory of design sensitivity to select a study design whose findings are expected to be insensitive to small or moderate biases from failure to control some unmeasured covariate. We find that posttraumatic stress symptoms were dramatically but unevenly elevated among residents of strongly shaken areas of Chile when compared to similar individuals in largely untouched parts of the country. In 96% of exposed-control pairs exhibiting substantial PTSS, the exposed individual experienced elevated symptoms (95% CI:[0.91,1.00]).

MeSH headings: Causality, Confounding Variables, Disasters, Earthquakes, Matched-Pair Analysis, Posttraumatic Stress Disorders.

At 3:34 on Saturday morning the 27th of February 2010, an earthquake of magnitude 8.8 struck in the Pacific 65 miles west of Concepción, Chile, earth's 6th most severe earthquake since 1900 (1). The quake and tsunami caused more than $30 billion in damages, damaged or destroyed 370,000 houses, 4,013 schools and 79 hospitals (1). More than 500 people were crushed, drowned or perished in fires (1).

Posttraumatic stress symptoms (PTSS) are often reported following earthquakes (2,3). The literature suggests effects of trauma are not uniform, varying from one person to another, so that two people undergoing overtly similar traumas exhibit different PTSS, possibly because of differing resilience to psychological stress (4,5). Also, two individuals may be equally physically shaken but experience different traumatic events, such as physical injury or loss of a loved one, and may be differently affected by displacement, loss of income and destruction of property (6,7). Because most people who experience injury or loss of income do not develop PTSD, heterogeneous experiences explain only part of heterogeneous symptoms. A limitation of many studies is that, because an earthquake or similar disaster is not anticipated, investigators rely on retrospective recall of key variables, such as intensity of exposure and of conditions prior to exposure. A person in distress following an earthquake may recall exposure to the earthquake as more severe than does someone not in distress (8,9), or recall conditions before the earthquake differently. If distress distorts memory, it may consequently distort associations between current status and recalled exposure adjusting for recalled pre-exposure status. In contrast, the current investigation uses prospectively recorded conditions prior to the earthquake, and objective geologic measures of ground shaking.

Natural disasters strike without purpose or deliberate target, but are less equitable than randomized experiments. Lower-income residential areas may be more often found along fault lines, low-income homes may be constructed of poorer materials that are less resistant to earthquake damage, and poor health or limited financial resources may limit ability to respond and recover after the earthquake has struck.

Chile's earthquake provides a unique view because shortly before the earthquake, Chile's Ministry of Planning and Cooperation completed its national socioeconomic survey or CASEN, and shortly after the earthquake they re-interviewed a subsample. It is unusual to have detailed data on the same individuals for a large sample both before and after a major disaster.

In addition to conducting a study of effects on PTSS, we illustrate recent developments in statistics concerning the planning of observational studies to reduce sensitivity to unmeasured biases. In a particular study, a sensitivity analysis quantifies the magnitude of unmeasured bias that would need to be present to materially alter the conclusions of the study (10-15). Once quantified, studies are seen to vary substantially in their sensitivity to unmeasured biases, something that is not revealed by the magnitude of a P-value computed assuming there is no bias. Not unlike statistical power, design sensitivity anticipates the outcome of a sensitivity analysis when the data are generated by a particular design and analyzed by particular methods. Once quantified, study designs and methods of analysis are seen to vary substantially in their tendency to yield results sensitive to unmeasured biases, something that is not revealed by power calculations performed assuming there is no bias (16-19).

The specific tactics used here to reduce sensitivity to unmeasured biases include the following.

- Prospective natural experiment with extensive covariates. One seeks a situation in which the treatment is inflicted on one person and withheld from another in a somewhat haphazard fashion, without deliberate purpose, and in which many relevant pretreatment covariates are measured prospectively. In such a situation, one does not anticipate large unmeasured biases, in part because large biases of all kinds are not anticipated and in part because many potential sources of bias are measured, not unmeasured. As noted above, among studies of disasters, the Chilean earthquake is unusual in satisfying the second condition.
- Extreme exposures, not a continuum with many minor exposures. One seeks to compare a treated and control group that experienced very different exposures to the treatment (16-18). For this reason, we compare severely shaken and largely untouched regions of Chile.
- Methods tailored to anticipated patterns of response. If one anticipates a particular pattern of treatment effects, one uses methods that can detect this specific pattern. Here, we anticipate severe PTSS symptoms in some, but far from all, individuals severely shaken by the earthquake. Using methods designed to detect this pattern will reduce sensitivity to bias if the anticipated pattern does, in fact, occur (17-21). If one represented severe versus negligible exposure by a term in a model, such as indicator variable or a continuous measure of ground shaking in a regression, the model would estimate a typical effect, whose

magnitude would greatly understate the effects of the trauma on some individuals, and by understating the true magnitude of effect, it would overstate the sensitivity to unmeasured biases.

Investigating sensitivity to unmeasured covariates is facilitated by the use of simple, transparent methods to control for measured covariates, for instance, multivariate matching (17,22). Multivariate matching compares matched treated and control groups that look comparable prior to treatment with respect to many measured covariates. Whether groups look comparable in measured covariates is a fact in the data that can be settled unambiguously before entertaining the inevitably more contentious and uncertain issues involving unmeasured covariates. When these tasks are merged, when adjustments for observed covariates are merged with sensitivity analysis for unobserved covariates, it can be difficult to determine the extent to which either task has been completed successfully (17, Chapter 6).

Our analysis uses recent techniques for multivariate matching for observed covariates found in the `mipmatch` package (23) for `R`. See the methods section for description of the matching techniques.

DATA AND METHODS

The EPT data

A new version of the CASEN was collected in November and December 2009, before the earthquake of 27 February 2010, and a representative subsample of 22,456 of the original 71,460 households was re-interviewed in the Post Earthquake Survey or EPT between May and June of 2010.

Estimation of the intensity of the earthquake

Peak ground acceleration (PGA) is a physical measure of how strongly the earth shakes in different geographic areas. Using the PGA values provided by the United States Geological Survey (USGS; 1), we estimated the PGA in each of the counties where the EPT was collected (see online appendix).

Figure 1 depicts the force of the earthquake in terms of PGA, with a black asterisk at the epicenter and circles at counties where EPT was collected. The PGA is shown by the intensity of the color grey.

Study design

In clinical trials, Peto et al. (24, page 590) write: “A positive result is more likely and a null result is more informative if the main comparison is of only 2 treatments these being as different as possible.” In observational studies, this design, with two very different treatments, is expected to yield results that are least sensitive to unmeasured biases. In other words, under simple models for dose-response, the design with two very different treatments has the largest design sensitivity, whereas inclusion of marginal exposures in a dose response analysis tends to make conclusions more sensitive to unmeasured biases (16; 17, §17.2; 18, §6.4). For these reasons, we matched 2,520 survey respondents who experienced little or no shaking as measured by a PGA smaller than 0.014 (the control group) to 2,520 individuals who experienced strong shaking as measured by a PGA greater than or equal to 0.275 (the exposed group); see, again, Figure 1.

The survey included 11,485 individuals with PGA's in the relevant range from whom psychological and health outcomes were requested. We excluded 181 individuals who did not provide either psychological or health outcomes, 1191 individuals who reported a younger age when re-interviewed, 56 individuals who reported a different sex, leaving 7305 exposed and 2752 controls. In the indigenous ethnic group, there were more controls than exposed individuals (see Table 1), so exposed were matched to controls, yielding 210 pairs, while for others there were more exposed than controls, so controls were matched to exposed yielding 2310 pairs, making 2520 pairs in total, or 5040 individuals, half of whom were exposed.

Posttraumatic Stress

The EPT collected the self-rated Davidson Trauma Scale (25). Each of 17 PTS symptoms from DSM-IV is rated twice on a five point scale, once for frequency (1="not at all", 5="every day"), once for severity (1="not at all distressing", 5="extremely distressing"), yielding a 2-to-10 score for each symptom, or a 34 to 170 total score.

Construction of the matched samples

We used new, recent, and standard matching techniques. The new techniques (23) use mixed integer programming in IBM's CPLEX, made available in `R` as `mipmatch`. See the Discussion for alternative software.

For the covariates in Table 1, we estimated a propensity score (26) for the conditional probability of avoiding severe exposure to the earthquake given the covariates. A robust covariate distance was computed between each exposed individual and each potential control with a caliper on the propensity score (17, §8). Matching was exact for sex, indigenous ethnic group, and five-year age categories. Self-rated health and housing quality were finely balanced (23,27-30). Fine balance makes groups have identical distributions without pairing identical individuals. The entire distribution of income, quantile by quantile, was balanced (by constraining their Kolmogorov-Smirnov KS-distance, 23).

Analysis of the outcomes

In an observational study, after adjustment for observed covariates, association between an outcome and exposure to a treatment may reflect either an effect of the treatment or a bias from some unmeasured covariate. We conducted a sensitivity analysis to assess the magnitude of unmeasured bias that would need to be present to alter the conclusions of a naïve analysis that presumes adjustment for observed covariates suffices to remove all bias (10-13). Conventionally, a study is said to be sensitive to bias of a particular magnitude if a bias of that magnitude could cause the 95% confidence interval to include effects that change the sign of the effect.

The sensitivity analysis used methods tailored to an effect in which trauma strongly affects PTSS for some exposed individuals and has little or no effect on many others (20,21). In particular, based on results in (17, §16; 18, §6.6; 19), we used

Stephenson’s (31) generalization of Wilcoxon’s signed rank statistic with his m=8. We first describe the motivation for this statistic, then describe some of its mechanics.

Salsburg (32) was concerned that some drugs appear to work for some patients but not for others. There are biological reasons to anticipate such effects: humans are biologically heterogeneous in ways only partially understood, so a drug that works for one person may do nothing for another. Indeed, there are psychiatric reasons to anticipate such effects: humans are also cognitively and emotionally heterogeneous. As a statistician, Salsburg was especially concerned that available statistical methods were optimized to detect small effects that affect everyone in the same small way, not large effects confined to a small subpopulation, and available methods often missed effects that were dramatically evident, say in boxplots. Conover and Salsburg (1988) built a test optimized to detect large effects confined to small subpopulations; it is quite different from conventional tests that respond to the average treatment effect. Their optimized test does not yield an easily interpreted confidence interval; however, a similar test was proposed by Stephenson (31), who was motivated by different considerations, and Stephenson’s test does yield confidence intervals and sensitivity analyses (21). Instead of looking at exposed-minus-control pair differences one at a time, Stephenson’s test looks at m=8 pair differences, focusing on the one pair with the largest absolute difference in outcomes. In this way, the Stephenson’s-Conover-Salsburg test shifts attention to large-but-less-common effects while remaining robust to outliers. For numerical evaluation of the specific value m=8, see Rosenbaum (19, Table 3).

With 2520 pairs, Stephenson's statistic looks at all 4 x 1022 subsets of m=8 exposed-control pairs. In each subset Stephenson finds the one pair of eight pairs with the largest absolute difference in PTSS. Unlike a typical pair, in this one pair there is a larger than typical absolute difference in PTSS. In (21), this one pair is called the "peak response." In the set of 8 pairs, a 1 is scored if the peak response involves an exposed individual with elevated PTSS, and a 0 is scored if it is the control with elevated PTSS. If the earthquake had no effect, we expect half 1's, half 0's. As seen in the results, in Chile it was not 50% 1's but rather 96% 1's: when there was a big difference in symptoms, it was almost always the exposed individual who exhibited elevated symptoms. Stephenson's statistic is Wilcoxon's statistic for m=2 and is the sign statistic for m=1. Software for this is available in the on-line supplement to (19).

At the price of certain amount of notation, Stephenson's statistic may be expressed as similar to an attributable risk (21). PTSS will occur among some controls, perhaps resulting from domestic or criminal violence unrelated to the earthquake. Speaking informally, half of the peak responses should occur in the exposed group by chance, and 96% is 46% higher than chance, so 92% = .46/(1-.5) of the excess peak responses not due to chance are estimated to be attributable to effects of the earthquake; see (21) for a precise statement with a confidence interval.

RESULTS

Covariate balance

Covariate balance is displayed in Table 1. Table 1 contains means for 46 covariates measured prior to the earthquake, before and after matching. One hopes to see similar means after matching. Before matching, the group exposed to the earthquake had fewer members of the indigenous ethnic group, fewer years of education, lower employment, lower income, and paid less in rent; however, these differences are nearly absent in the matched sample.

Table 2 shows three exactly matched variables, sex, age group, and indigenous ethnic group, and two finely balanced variables, self-rated health and housing quality. This means men were matched to men, women to women, but housing quality was balanced without being paired (30). Figure 2 shows the income distribution differed before but not after matching. The balance for 46 measured covariates is substantially greater than expected by complete randomization of 5040 individuals to two groups of equal size; see online appendix. Randomization balances unmeasured covariates whereas matching cannot do this.

Effect on PTSS

Table 3 illustrates heterogeneous effects for the one question about difficulty falling asleep (33), ignoring who is paired with whom. Notable in Table 3 is a pattern seen for all symptoms: many more individuals exposed to the earthquake than controls reported extensive difficulty sleeping, and yet more than half of the exposed individuals (1382/2520 = 55%) reported no difficulty.

Figure 3 looks at all 17 symptoms, depicting the exposed-minus-control pair difference in Davidson Trauma Scores. Figure 3 has a density estimate or smoothed histogram (from `density` in `R`). If the effect of the earthquake were constant, the same for every exposed individual, then the density would be symmetric about that constant, and that constant could be estimated as the coefficient of an indicator or dummy variable in a regression; however, that is not what we see in Figure 3. The density is sharply skewed right, with large positive differences occurring much more often than large negative differences, and yet the mode or peak is close to zero. Many pairs exhibit no indication of PTSS, while some pairs exhibit large positive differences in PTSS, so it would be a mistake to estimate the effect of the earthquake on PTSS as the coefficient of an indicator, or as a mean, median or other typical difference. For instance, the point estimate and 95% confidence interval for the typical difference based on inverting the Wilcoxon test are 20.0 and [18.5, 21.0]; however, these and similar measures assume a constant effect, so they are very misleading. It appears that some people are severely affected and others little affected: a typical effect of 20 distorts many negligible effects and some dramatic effects. If the effect were a constant effect of 20, then Figure 3 would be symmetric about 20, which it clearly is not.

Recall from the Methods Section that Stephenson's (31) statistic looks at all subsets of m=8 pairs, and estimates the probability that the largest absolute difference in m=8 pairs has higher PTSS for the exposed individual rather than the control. A probability of ½ suggests no effect of the earthquake. In Figure 3, the estimate of the probability for m=8 is not ½, but rather 96% (with one-sided 95% confidence interval from (21) of [0.91, 1.00]), that is, in 8 pairs, the largest absolute difference is a positive difference 96% of the time, with a larger trauma score for the exposed individual than the control. In a matched pair, if there is a substantial difference in PTSS, it is almost always the exposed subject who has greater symptoms.

Moving away from the naïve assumption that adjustments for observed covariates remove all bias, we examine the sensitivity of this finding to biases of specific magnitudes (14; 15, §4; 21). Bias is measured by a parameter Γ that describes a pair of one exposed and one control subject matched on the 46 observed covariates: it says their relative risk of exposure to the earthquake may be as high as Γ or as low as $1/\Gamma$ because they differ in terms of unobserved covariates (19, expression (1)). In a clinical trial, random assignment of treatments ensures $\Gamma=1$, that is, each subject in a pair has the same ½ chance of receiving the treatment, with relative risk of treatment of (½)/(½)=1. If $\Gamma=2$, one subject might have a probability as high as 2/3 of treatment, the other as low as 1-2/3=1/3, so the relative risk of exposure to the treatment is at most (2/3)/(1/3)=2. So $\Gamma=2$ is a moderately large departure from a randomized experiment. We ask: How much bias, measured by Γ, would need to be present for the naïve 95% confidence interval, [0.91, 1.00], in the previous paragraph, to be moved to include ½ and smaller effects, which would imply that even the sign of the effect is in doubt. For this 95% interval to include ½ due to biased exposure to the earthquake, an exposed and a control subject matched on the 46 observed covariates would need to differ in their relative risk of exposure to the earthquake by a factor of more than $\Gamma=14$ because they differ in terms of an unobserved covariate not controlled by the matching, and this covariate would need to be a near perfect predictor of posttraumatic stress. (More precisely, the two-sided 95% confidence interval just reaches [½, 1] for $\Gamma=14.6$.) If the unobserved covariate were less than a perfect predictor of posttraumatic stress, then it would need to increase the relative risk of exposure by more than a factor of 30 and increase the relative risk of a positive difference in PTSS by more than a factor of 30 to explain the observed association (i.e., Γ=14 amplifies to Λ=30 and Δ=30 in the notation of (34)). For comparison, Hammond's (35) study of heavy smoking as a cause of lung cancer becomes sensitive at Γ=6, and Hammond's study is one of the least sensitive studies in epidemiology (15, table 4.1).

If Wilcoxon's statistic (m=2) is used, the difference in PTSS is judged more sensitive to unmeasured bias, becoming sensitive at Γ=6.0, rather than Γ=14 for m=8. For m=20, also evaluated in (19), the difference in PTSS becomes sensitive at Γ=22.4. If one looks at the largest absolute difference in PTSS in all sets of m=20 pairs, the difference is positive 98% of the time reflecting greater symptoms for the exposed individual in the pair, as distinct from 96% of the time for m=8 as seen above. This pattern is expected for a treatment that harms some people and spares many others.

The online appendix presents parallel analyses for: (i) subgroups defined by gender, age, and ethnic group, (ii) for pairs reporting no physical health problems, (iii) for standard subscores of the Davidson Trauma Scale, and (iv) for a subscore that avoids reference to specific trauma. These analyses vary slightly, with somewhat greater PTSS for women than men, and for older individuals, but the difference in PTSS has the same form as Figure 3 in all subgroups and subscores and is highly insensitive to unmeasured biases.

DISCUSSION

The Chilean earthquake offered a unique opportunity to assess the effect of trauma on PTSS because of longitudinal data free of recall bias, with detailed information about the same people before and after the earthquake. Guided by the statistical theory of design sensitivity (16-18), the comparison was designed to be as insensitive as possible to biases from unmeasured covariates, and in the end the results were highly insensitive. Specifically, severely shaken regions of Chile were compared to untouched regions, and exposure to the earthquake had a strong and heterogeneous effect on PTSS. In particular, in 96% of pairs exhibiting substantial PTSS, the exposed rather than the unexposed individual exhibited elevated symptoms (CI: [91%-100%]). Because heterogeneous effects were anticipated, statistical techniques with power to detect such effects were used.

Combining standard (17,22) and new (23,30) techniques, the matching balanced 46 observed covariates measured prior to the earthquake including socioeconomic status, housing quality, health and health insurance.

Figure 3 clearly indicates the heterogeneous pattern of PTSS effects in pairs similar in terms of 46 covariates. Adjustment for covariates using a model would not yield a simple depiction of heterogeneous effects among similar individuals. Indeed, some model-based analyses never entertain the possibility of heterogeneous treatment effects, instead estimating a typical effect as the coefficient of an exposure indicator in a model. If only a subset of exposed individuals respond to exposure, an estimate of a typical effect may miss a stable and substantial atypical effect (20,21).

Advice about study design from design sensitivity derives from mathematical proof and simulation (11-13); however, its findings are often intuitive. Ever since the first sensitivity analysis (10), it is known that larger treatment effects are less sensitive than smaller effects; however, the meaning of this once one moves beyond the 2x2 table is vague. If a treatment strongly affects some people and leaves others unaffected, then is that a large effect? The mean or median effect may be small, but the effect on affected individuals may be large and unambiguous; see Figure 3. Statistics built to detect heterogeneous effects (20,21) judge this a large effect, hence an effect insensitive to small unmeasured biases (17, §16; 18, §6.6). When two (or more) patterns of effect are plausible, an appropriate analysis may look for two patterns correcting for multiplicity (36). Similarly, if more intense treatment produces larger effects, the largest and therefore least sensitive effects will be found in a comparison of intense exposure and no exposure (16; 17, §17.3; 18, §6.4). We have discussed design sensitivity as it is relevant to the study of the earthquake in Chile, but other aspects of design sensitivity are not discussed here but may be relevant to other studies (17,18).

An investigator wishing to use the statistical techniques described here may find the following remarks helpful. The matching was done using the new `R` package `mipmatch` (23). Once installed, `mipmatch` is fairly straightforward to use, but installation is not as effortless as for most `R` packages because IBM's CPLEX must also be installed. IBM makes CPLEX available for free to academics. The `mipmatch` package does some unique things, such as finely balance many variables at once, force balance on the mean of a continuous variable, or balance a continuous variable such as income at every quantile; however, somewhat less general packages, such as Hansen's `optmatch` (37) and Yang's `finebalance` (29), produce excellent matched comparisons, and install effortlessly. Hansen's `optmatch` is illustrated in (17, §13). Software in `R` and a practice session for the sensitivity analysis are in the online appendix to (19).

One limitation is that PTS is based on self-report. There is no evidence internal to the survey about the relationship between self-reported symptoms and what a psychiatrist would determine. It is conceivable that a natural disaster reduces stigma associated with reporting PTS and this might account for some of the increase in reported symptoms.

Posttraumatic stress symptoms were dramatically but unevenly elevated among residents of strongly shaken areas of Chile when compared to similar individuals in largely untouched parts of the country. Contrary to some prior expectations (38), exposure to disasters is less equitable than random. Residents of highly-affected areas had less education, lower income, lower employment, and lived in cheaper housing; however, these measured differences were removed by matching. Possibly the exposure was inequitable in other unmeasured ways as well, but the sensitivity analysis shows these unmeasured biases would need to be very large to alter the main conclusions.

REFERENCES

1. United States Geological Survey. (http://earthquake.usgs.gov/earthquakes/world/10_largest_world.php; http://earthquake.usgs.gov/earthquakes/recenteqsww/Quakes/us2010tfan.php#summary). (Accessed June 8, 2012).
2. Neria Y, Nandi A, Galea S. Post-traumatic stress disorder following disasters: a systematic review. *Psychological Medicine*. 2008; 38:467-480.
3. Goenjian AK, Steinberg AM, Najarian LM, et al. Prospective Study of Posttraumatic Stress, Anxiety, and Depressive Reactions After Earthquake and Political Violence. *Am J Psychiatry*. 2000; 157;911-895.
4. Rutter M. Resilience in the face of adversity: Protective factors and resistance to psychiatric disorders. *British Journal of Psychiatry*. 1985; 147:598-611.
5. Bonanno GA. Loss, trauma and human resilience: Have we underestimated the human capicity to thrive after extremely aversive events? *Am Psychol*. 2004; 59:20-28.
6. McLaughlin KA, Fairbank JA, Gruber MJ, et al. Serious emotional disturbance among youths exposed to Hurricane Katrina 2 years postdisaster. *J Am Acad Child Adolesc Psychiatry.* 2009; 48:1069-78.
7. Galea S, Tracy M, Norris F, et al. Financial and social circumstances and the incidence and course of PTSD in Mississippi during the first two years after Hurricane Katrina. *J Trauma Stress.* 2008; 21:357-68.

8. Roemer L, Litz BT, Orsillo SM, et al. Increases in retrospective accounts of war-zone exposure over time: the role of PTSD symptom severity. *J Traum Stress*. 1998;11(3):597-605.
9. David AC, Akerib V, Gaston L, Brunet A. Consistency of retrospective reports of peritraumatic responses and their relation to PTSD diagnostic status. *J Traum Stress*. 2010;23(5)599-605.
10. Cornfield J, Haenszel W, Hammond EC, et al. Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst*. 1959; 22:173-203.
11. Marcus SM. Using omitted variable bias to assess uncertainty in the estimation of an AIDS education treatment effect. *J Educ Stats*. 1997; 22:193-201.
12. Rosenbaum PR, Rubin DB. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J Roy Stat Soc B Met*. 1983; 45:212-218.
13. Yanagawa T. Case-control studies: assessing the effect of a confounding factor. *Biometrika*. 1984; 71:191-194.
14. Rosenbaum PR. Sensitivity analysis for certain permutation tests in matched observational studies. *Biometrika*. 1987; 74:13-26.
15. Rosenbaum PR. *Observational Studies*. 2nd ed. New York, NY: Springer; 2002.
16. Rosenbaum PR. Design sensitivity in observational studies. *Biometrika*. 2004; 91:153-164
17. Rosenbaum PR. *Design of Observational Studies*. New York, NY: Springer; 2010.
18. Rosenbaum PR. What aspects of the design of an observational study affect its sensitivity to bias from covariates that were not observed? In: Dorans N and

Sinharay S, eds. *Looking Back: Proceedings of a Conference in Honor of Paul W. Holland*. New York, NY: Springer; 2011:87-114.

19. Rosenbaum PR. A new u-statistic with superior design sensitivity in observational studies. *Biometrics*. 2011; 67:1017-1027. R-software/instructions at http://www-stat.wharton.upenn.edu/~rosenbap/rsession.txt
20. Conover WJ, Salsburg DS. Locally most powerful tests for detecting treatment effects when only a subset of patients can be expected to 'respond' to treatment. *Biometrics*. 1988; 44:189-196.
21. Rosenbaum PR. Confidence intervals for uncommon but dramatic responses to treatment. *Biometrics*. 2007; 63:1164-1171.
22. Stuart EA. Matching Methods for Causal Inference: A Review and a Look Forward. *Statist Sci*. 2010; 25:1-21.
23. Zubizarreta JR. Using mixed integer programming for matching in an observational study of kidney failure after surgery. *J Am Statist Assoc*. 2012; accepted. `R` software `mipmatch` at http://www-stat.wharton.upenn.edu/~josezubi/
24. Peto, R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. *Brit J Cancer*. 1976; 34:585-612.
25. Davidson JRT, Book SW, Colket JT, et al. Assessment of a new self-rating scale for post-traumatic stress disorder. *Psychol Med*. 1997; 27:153-160.
26. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70:41-55.
27. Rosenbaum PR. Optimal matching in observational studies. *J Am Statist Assoc*. 1989; 84:1024-1032.

28. Rosenbaum PR, Ross RN, Silber JH. Minimum distance matched sampling with fine balance in an observational study of treatment for ovarian cancer. *J Am Statist Assoc*. 2007; 102:75-83.
29. Yang D, Small DS, Silber JH, Rosenbaum PR. Optimal matching with minimal deviation from fine balance in a study of obesity and surgical outcomes. *Biometrics*. 2012 to appear. `R` package `finebalance`.
30. Zubizarreta JR, Reinke C, Kelz RR, et al. Matching for several sparse nominal variables in a case-control study of readmission following surgery. *Am Statist*. 2011; 65:229-238.
31. Stephenson WR. A general class of one-sample nonparametric test statistics based on subsamples. *J Am Statist Assoc*. 1981; 76:960-966.
32. Salsburg DS. Alternative hypotheses for the effects of drugs in small-scale clinical studies. *Biometrics*. 1986;42:671-674.
33. Lavie P. Sleep disturbances in the wake of traumatic events. *N Engl J Med.* 2001; 345:1825-1832.
34. Rosenbaum PR, Silber JH. Amplification of sensitivity analysis in observational studies. *J Am Statist Assoc*. 2009; 104:1398-1405.
35. Hammond EC. Smoking in relation to mortality and morbidity. *J Natl Cancer Inst*. 1964;32:1161-1188.
36. Rosenbaum PR. Testing one hypothesis twice in observational studies. *Biometrika*. 2012, to appear.
37. Hansen BB. Optmatch. *R News*. 2007;7:18-24. `R` package `optmatch`.

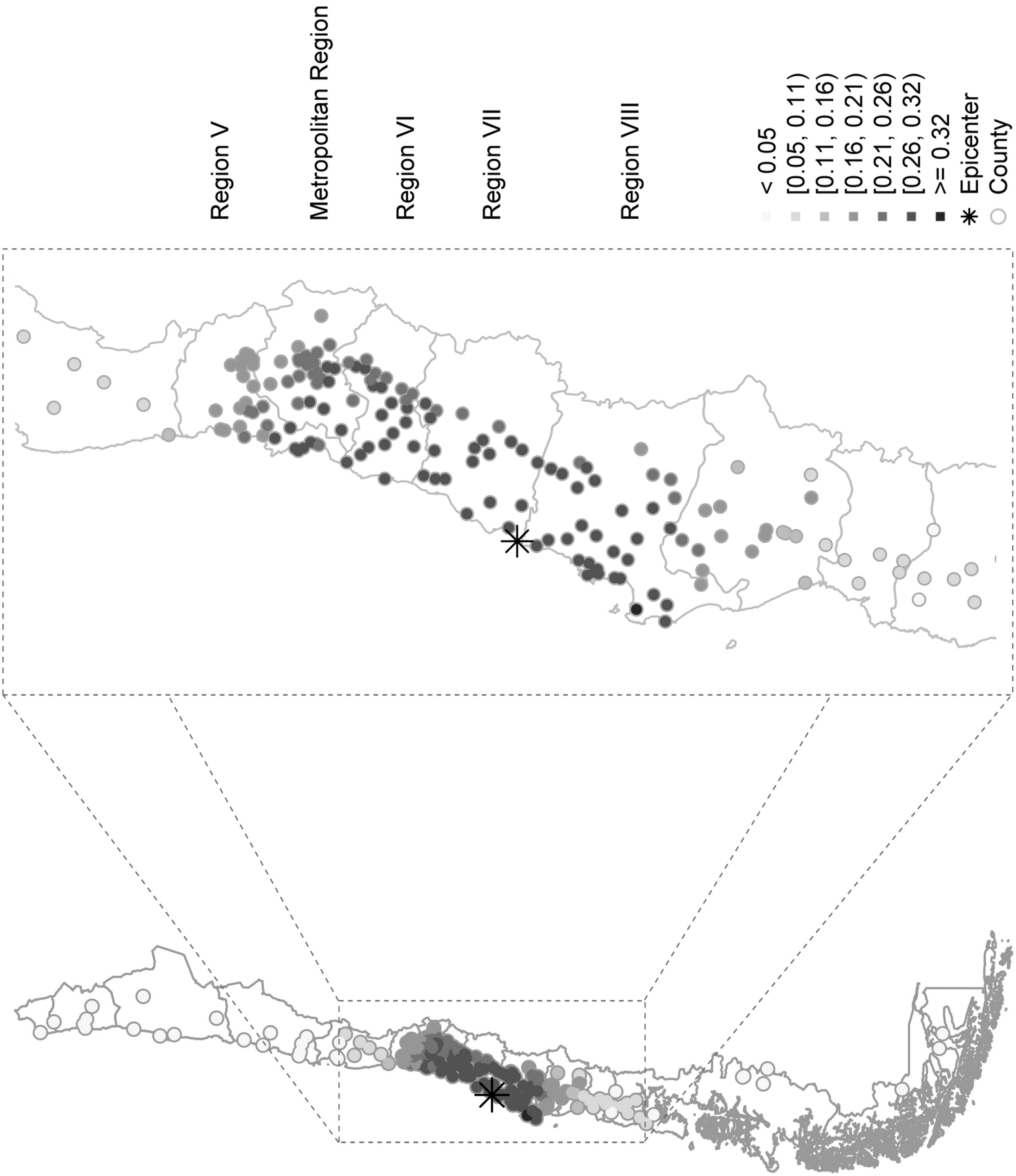
38. Bromet EJ, Havenaar JM. Mental Health Consequences of Disasters. In: Sartorius N, Gaebel W, López-Ibor J, et al., eds. *Psychiatry in Society*. Chichester, UK: John Wiley & Sons, 2002.

Figure 1: Map of Chile with the estimated peak ground accelerations expressed in g for all the counties in the study. As explained in the legend, a circle denotes a county and the intensity of the color grey shows its estimated shaking intensity. Our study paired respondents in non-affected areas (PGA < 0.014) with respondents in very affected areas (PGA > 0.275).

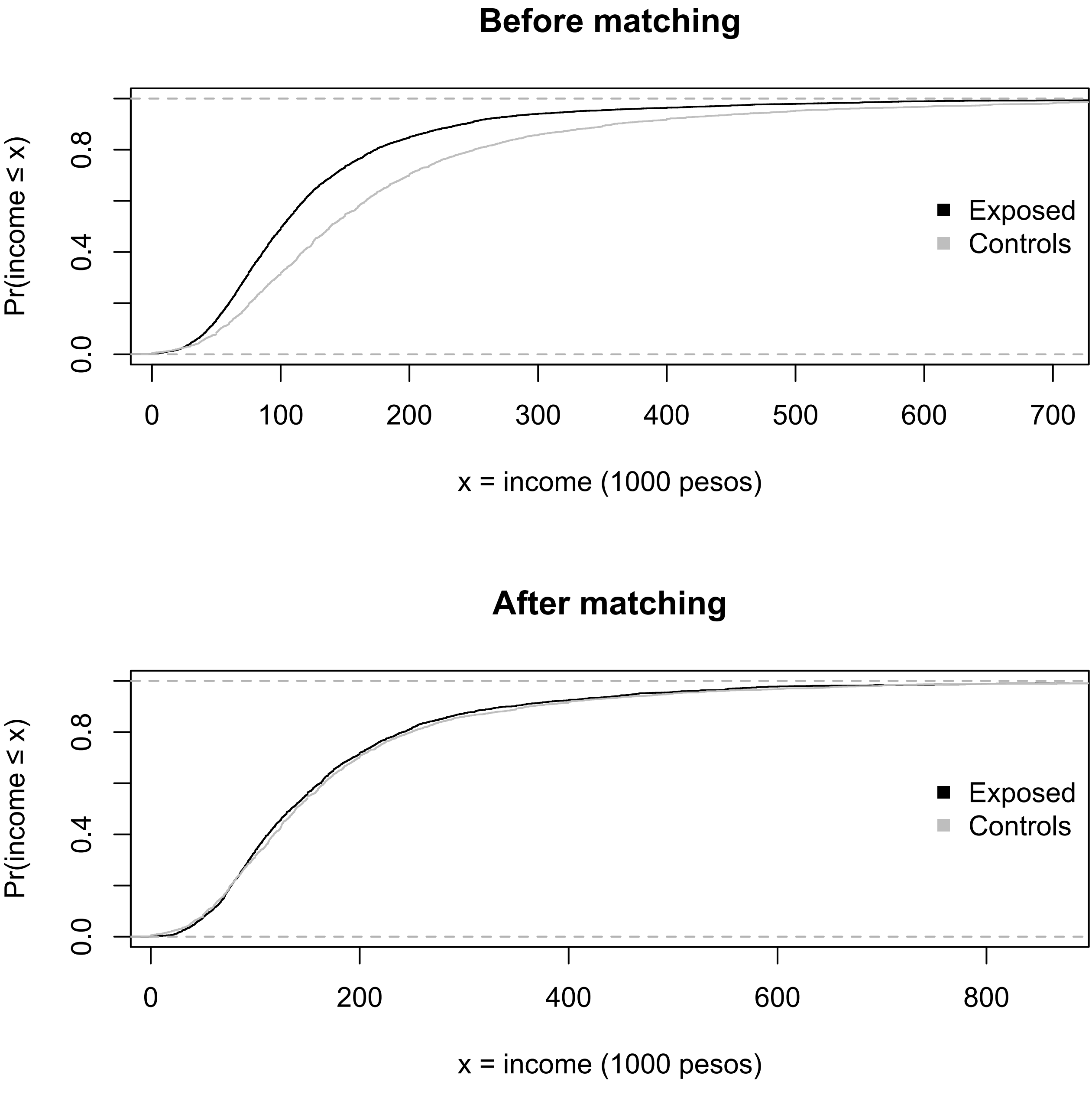
Figure 2: Cumulative distribution of household total per capita income for exposed and control groups before and after matching. The cumulative distribution is the proportion of individual with income less than or equal to x. Before matching, exposed subjects had lower income. After matching for KS-distance, the entire distributions are almost the same. The KS-distance is the largest vertical distance between the two curves.

Figure 3: Exposed-minus-control matched pair differences in posttraumatic stress (PTS) scores for 2520 matched pairs. The upper figure is a nonparametric density estimate (using default settings in R) and the lower figure is a boxplot. To aid the eye, the vertical line is at zero difference. The graphs indicate that many pair differences were close to zero, consistent with no difference in symptoms, but differences that were far from zero were mostly positive: when PTS symptoms differed, it was the exposed individual who typically experienced severe symptoms.

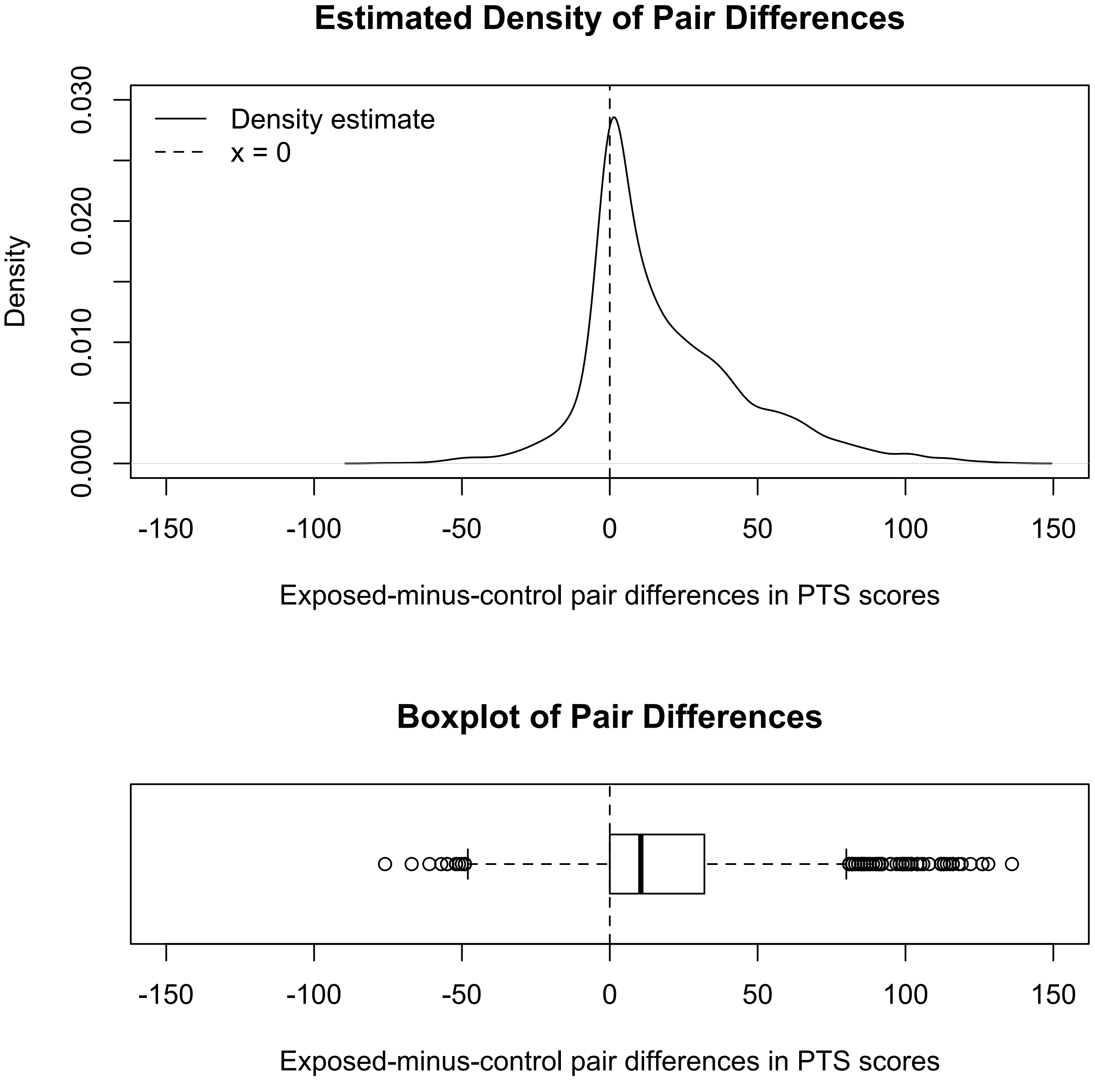
Figure


Region V
Metropolitan Region
Region VI
Region VII
Region VIII
< 0.05
[0.05, 0.11)
[0.11, 0.16)
[0.16, 0.21)
[0.21, 0.26)
[0.26, 0.32)
>= 0.32
Epicenter
County

Figure


Before matching
Pr(income ≤ x)
0.0
0.4
0.8
0
100
200
300
400
500
600
700
x = income (1000 pesos)
Exposed
Controls
After matching
Pr(income ≤ x)
0.0
0.4
0.8
0
200
400
600
800
x = income (1000 pesos)
Exposed
Controls

Figure


Estimated Density of Pair Differences
Density estimate
x = 0
Density
0.000
0.010
0.020
0.030
-150
-100
-50
0
50
100
150
Exposed-minus-control pair differences in PTS scores
Boxplot of Pair Differences
-150
-100
-50
0
50
100
150
Exposed-minus-control pair differences in PTS scores

Table

Table 1: Balance on 46 pre-earthquake covariates, before and after matching. All the means are proportions unless noted. Standardized differences express the difference in means as a fraction of a pooled standard deviation before matching. There are substantial imbalances for several covariates before matching, and only negligible imbalances after matching.

	Before Matching			After Matching		
	Mean Exposed n=7305	Mean Controls n=2752	Std Diff	Mean Exposed n=2520	Mean Controls n=2520	Std Diff
Demographic covariates						
Age (mean, years)	49.05	47.48	0.09	47.75	47.76	0.00
Gender	0.69	0.67	0.05	0.67	0.67	0.00
Indigenous ethnic group	0.03	0.16	-0.46	0.08	0.08	0.00
Household size (persons)	3.64	3.72	-0.05	3.72	3.70	0.01
Married or cohabitating	0.66	0.65	0.03	0.67	0.66	0.03
Divorced or widow	0.15	0.16	-0.03	0.16	0.16	0.00
Single	0.18	0.18	-0.01	0.17	0.18	-0.03
Socioeconomic covariates						
Years of education (years)	8.35	9.24	-0.21	9.34	9.20	0.03
Employed	0.39	0.50	-0.22	0.50	0.49	0.01
Unemployed	0.04	0.04	0.02	0.03	0.04	-0.04
Inactive	0.56	0.46	0.21	0.47	0.47	0.00
Individual work income (1000 pesos)	74.47	120.87	-0.26	119.87	121.22	-0.01
Household own per capita income (1000 pesos)	106.71	148.58	-0.25	147.77	149.37	-0.01
Household total per capita income (1000 pesos)	134.62	187.29	-0.30	183.70	188.11	-0.03
Poor	0.19	0.12	0.19	0.10	0.12	-0.04
Housing prior to the earthquake						
Own housing or paying to own it	0.72	0.71	0.04	0.70	0.71	-0.01
Rented housing	0.09	0.13	-0.14	0.14	0.14	0.02
Ceded housing	0.18	0.15	0.08	0.14	0.15	-0.01
Irregular use of housing	0.01	0.01	-0.01	0.01	0.01	0.02
Housing rent: 0-25000 pesos	0.12	0.10	0.06	0.10	0.10	-0.01
Housing rent: 25001-50000 pesos	0.38	0.14	0.56	0.16	0.15	0.02
Housing rent: 50001-75000 pesos	0.17	0.12	0.16	0.12	0.12	0.00
Housing rent: 75001- pesos	0.33	0.64	-0.66	0.62	0.63	-0.01
Acceptable housing structure	0.67	0.67	0.00	0.69	0.69	0.00
Reparable housing structure	0.31	0.30	0.03	0.29	0.29	0.00
Irreparable housing structure	0.01	0.03	-0.08	0.02	0.02	0.00
No overcrowding	0.89	0.87	0.06	0.88	0.88	0.00
Medium overcrowding	0.10	0.11	-0.04	0.11	0.11	0.01
Critical overcrowding	0.10	0.11	-0.04	0.11	0.11	0.01
Health prior to the earthquake						
Health problem (last month)	0.18	0.18	0.00	0.16	0.18	-0.05
Hospitalized (last year)	0.06	0.06	-0.01	0.06	0.07	-0.03
Has a psychiatric problem	0.00	0.00	0.05	0.00	0.00	-0.02

Poor self-rated health	0.08	0.05	0.14	0.05	0.05	0.00
Fair self-rated health	0.52	0.59	-0.15	0.59	0.59	0.00
Good self-rated health	0.40	0.36	0.08	0.36	0.36	0.00
Public health insurance (FONASA)	0.91	0.82	0.27	0.83	0.82	0.01
Private health insurance (ISAPRE)	0.04	0.08	-0.16	0.08	0.08	0.01
Other health insurance	0.02	0.04	-0.13	0.04	0.04	-0.02
No health insurance	0.02	0.04	-0.14	0.04	0.04	-0.01
Does not know health insurance	0.01	0.01	-0.04	0.02	0.02	0.00
Self-sufficient or low disability	0.12	0.12	0.02	0.11	0.12	-0.04
Moderate or severe disability	0.01	0.01	0.01	0.01	0.01	-0.03
No disability	0.87	0.87	-0.02	0.89	0.87	0.04
No data about disability	0.00	0.00	0.00	0.00	0.00	0.00
Other						
Rural zone	0.30	0.22	0.18	0.18	0.20	-0.05
Propensity score	0.23	0.40	-0.94	0.35	0.36	-0.05

Table 2: Distributions of sex, age, self-rated health and housing quality prior to the earthquake. Tabled values are counts. Sex and age were matched exactly, while health and housing quality were finely balanced.

	After Matching	
Sex	Exposed	Controls
Male	831	831
Female	1689	1689
Total	2520	2520
	After Matching	
Age Group	Exposed	Controls
15-24	195	195
25-34	412	412
35-44	561	561
45-54	474	474
55-64	406	406
65-	472	472
Total	2520	2520
	After Matching	
Ethnic Group	Exposed	Controls
Indigenous	210	210
Non-indigenous	2310	2310
Total	2520	2520
	After Matching	
Self-Rated Health	Exposed	Controls
Poor	122	122
Good	1487	1487
Fair	911	911
Total	2520	2520
	After Matching	
Housing Quality	Exposed	Controls
Acceptable	1738	1738
Unacceptable	739	739
Beyond repair	43	43
Total	2520	2520

Table 3: For 2520 matched exposed and control individuals, the table counts the responses to the question: Do you have difficulty falling asleep and remaining asleep? The score is the sum of a frequency response and a severity response, both scored 1 to 5, so the possible scores are 2 to 10, with 10 indicating both the highest frequency and the highest severity. Many more exposed individuals than controls reported difficulty sleeping, and yet more than half of the exposed individuals (1382/2520=55%) reported no difficulties.

Sleep Score	2	3	4	5	6	7	8	9	10
Exposed	1382	10	398	127	209	88	118	67	121
Control	2331	3	73	33	26	16	19	7	12