Published in final edited form as:

Epidemiology. 2016 September ; 27(5): 624–632. doi:10.1097/EDE.0000000000000515.

Unmeasured Confounding in Observational Studies with Multiple Treatment Arms: Comparing Mortality of Severe Trauma Patients by Trauma Center Level

Junxin Shja,b, Bo Lu^d, Krista K. Wheeler^{a,b}, and Huiyun Xiang^{a,b,c}

^aCenter for Pediatric Trauma Research, The Research Institute at Nationwide Children's Hospital

^bCenter for Injury Research and Policy, The Research Institute at Nationwide Children's Hospital

^cThe Ohio State University College of Medicine

^dThe Ohio State University College of Public Health, Division of Biostatistics

Abstract

Comparing emergency department mortality across different levels of trauma care (non-trauma centers, Level I and Level II centers) is important in evaluating regionalized care. Patient population characteristics differ across different levels of trauma care and it is essential to adjust for baseline covariates to make valid comparisons. Propensity score matching has been established as a more robust method to infer causal relationship in observational studies than conventional regression adjustment. We designed and implemented a three group matching methodology. First, we conducted optimal pair matching between the treatment group (non-trauma centers) and the first control group (Level I trauma centers); second, we conducted optimal pair matching between the non-trauma centers and the second control group (Level II trauma centers); the final step was to link the two sets of matched pairs by the common treatment subjects to form matched triplets. We then implemented a sensitivity analysis with three treatment arms, Lu's imputation based method, to assess the impact due to potential unmeasured confounding. The results showed that if the most severe adult trauma patients treated in non-trauma centers were to be treated in Level I or II trauma centers, the odds of emergency department death would be reduced dramatically (OR=0.61 and OR=0.49). There was no difference between being treated in Level I or II trauma centers. The sensitivity analysis revealed that the positive association between being treated at Level I or II trauma centers and the reduced odds of mortality would remain present even in the presence of strong unmeasured confounding.

Keywords

Propensity score matching; matching on multiple groups; sensitivity analysis; trauma; ED mortality

CORRESPONDING AUTHOR: Huiyun Xiang, MD, MPH, PhD, Center for Pediatric Trauma Research, The Research Institute at Nationwide Children's Hospital, 700 Children's Drive, Columbus, OH 43205; Huiyun.Xiang@NationwideChildrens.org; Phone: 614-355-5893; Fax: 614-355-5897.

Conflicts of interest

The authors have no conflicts of interest to report.

INTRODUCTION

Trauma centers in the U.S. are categorized according to the resources and expertise required to provide various levels of care for trauma patients.^{1,2} An important aspect of trauma system evaluation is to compare the mortalities across different levels of trauma care. Previously, research found that Level I trauma centers had lower in-hospital mortality among trauma patients when compared with non-trauma centers.² However, the emergency department mortality of trauma patients at a national level has only recently been studied.³ The emergency department mortality in different levels of trauma centers is important information for optimal utilization of medical resources. Our previous study showed that if major trauma patients treated in emergency departments of non-trauma centers or Level III trauma centers were triaged to Level I or II trauma centers, the capacity of trauma centers would have to be increased dramatically.⁴ The current practice of transferring a trauma patient typically is from a non-trauma center to a Level I trauma center. To be treated in a Level I trauma center, patients may require longer transportation and greater utilization of medical resources. If severe trauma patients could be treated in non-trauma centers and have similar emergency department survival probabilities, it would be unnecessary to triage all severe trauma patients directly to trauma centers; and if Level II trauma center patients have similar survival rates as Level I patients, Level II trauma centers could serve as an appropriate alternative to Level I centers. Fewer transfers could help save lives by shortening the time between the injury event and treatment, and it might also be less costly. Our key questions are: (1) if major trauma patients treated in the emergency department of non-trauma centers were to be treated in either Level I or II trauma center emergency departments, would they have lower emergency department mortalities and (2) and would patients treated in non-trauma centers have similar emergency department survival probabilities if first treated in the emergency departments of Level II instead of Level I trauma centers? To answer the questions, we need to compare the emergency department mortalities among the three groups. In our study, the treatment effect is defined as the emergency department mortality difference (in terms of odds ratios) as the result of different trauma care provided by non-trauma centers and Level I or II trauma centers.

Since the patients in these three groups of emergency departments have different baseline characteristics, adjustment for potential confounding is needed. Regression analysis is common practice in covariance adjustment, but Rosenbaum and Rubin's propensity score matching technique provides us with another choice.⁵ Rubin showed that models based on [propensity score] matched samples are less sensitive to model misspecification than regression adjustment. They also showed that covariance adjustment of matched samples consistently reduced bias.^{6,7}

Because we wanted to compare non-trauma center patients to both Level I and II trauma center patients, our treatment group was adult severe trauma patients treated in the emergency departments of non-trauma centers. We naturally had two control groups, similar patients treated in Level I trauma centers and those treated in Level II centers. The benefit of introducing a second control group was studied by Campbell and Rosenbaum.^{8,9} They reported that the second control group, if carefully selected, may render less plausible

concerns about biases from certain unobserved covariates, thereby strengthening the evidence that the treatment caused its ostensible effects.⁸

Pair matching between one treatment group and one control group is the most common matching design. When comparing three groups, however, new techniques are needed. Although several researchers have explored how to deal with multiple treatment arms,^{10–12} previous methods either had limitations,¹³ or were designed to solve different problems.¹² We aimed to develop and implement a 1:1:1 triplet matching strategy, a new approach in trauma outcome research. Propensity score matching, like regression, can only adjust for observed covariates. To investigate the impact of unmeasured confounding in this study, we developed a sensitivity analysis to examine the robustness of our conclusions.

METHODS

Optimal pair matching

When there are two treatment arms, optimal pair matching is often preferred in practice. Optimal pair matching is the process of developing matched sets $(A_1, \dots, A_s; B_1, \dots, B_s)$ with size of (α, β) in such a way that the total distance among all matched pairs is minimized. Formally, optimal matching minimizes the total distance defined as $\Delta = \sum_{s=1}^S \omega(|A_s|, |B_s|) \delta(A_s, B_s)$, where $\omega(|A_s|, |B_s|)$ is a weight function and δ is the difference between treated and control in terms of their observed covariates, such as their difference on propensity scores or Mahalanobis metrics.^{14, 17} In our study, we used Mahalanobis metrics as our distance measure since there were multiple treatment arms. We used Hansen's *Optmatch* R package to conduct optimal pair matching.¹⁵ This package used Kuhn's methods¹⁶ and implemented Bertsekas's algorithms.¹⁷

The procedures of 1:1:1 triplet matching and outcome analysis

Optimal triplet matching presents a computational problem that cannot be readily solved.¹⁸ To the best of our knowledge, there is no existing software package that can perform this task. We have proposed a partial-optimal triplet matching strategy, which includes three steps: First, optimal 1:1 pair matching between the treatment group (non-trauma centers) and the first control group (Level I trauma centers); second, optimal 1:1 pair matching between the treatment group and the second control group (Level II trauma centers); third, linking the two sets of matched pairs by the subjects in the common treatment group to form matched triplets. Through the matching, we kept the profile of patient mix of the treatment group. This matching idea is straightforward; however, it does not guarantee the triplet matching in terms of minimizing the three-way distance among all possible triplets. The extent of similarities among the groups may be different. The two control groups were more likely to be similar to the common treatment group, while the total difference between the two control groups was not optimally minimized.

Before and after matching, we calculated the standardized differences for each baseline covariate, and standardized differences greater than 0.25 were considered evidence of imbalance.¹⁹ We followed a recommendation from Rosenbaum and put those unbalanced covariates into the outcome regression model to eliminate residual imbalance.²⁰

The outcome analysis, comparing emergency department mortality across the three matched groups, was performed by fitting conditional logistic regression models, using SAS 9.3.²¹ In the model, emergency department mortality served as the outcome variable, the nominal trauma level indicator (non-trauma center, Level I trauma center, Level II trauma center) was the independent variable, and each triplet was a stratum.

Sensitivity analysis

Rosenbaum proposed use of conventional sensitivity analysis for observational studies with two treatment arms, that is, evaluating the potential change of study conclusions due to a hypothetical unmeasured confounder.¹⁴ We extended the idea of Lu et al.'s imputation-based sensitivity analysis for survival data to cope with triplet matching analysis.²² It involved creating a binary unobserved covariate, U , based on a hypothetical model, $\text{Logit}[P(U=1)] = \beta_0 + \beta_1 T + \beta_2 Y$, and including U as a covariate in the outcome analysis, then examining how the result changes when increasing the magnitude of U . In this way, a confounding relationship was set up. U was related to the treatment T by $\text{OR}_1^* (\text{OR}_1^* = \exp(\beta_1))$ conditional on outcome Y , and related to Y by $\text{OR}_2^* (\text{OR}_2^* = \exp(\beta_2))$ conditional on T . For easier interpretation, the marginal odds ratio of (U, T) , say OR_1 , and (U, Y) , say OR_2 , were calculated based on the imputed data to capture the unconditional association. If OR_1 and OR_2 are not equal to one, then U is a confounder. Since we were primarily concerned with the impact of potential unmeasured confounding on the observed positive relationship between T and Y , without loss of generality, we explored only the situations with OR_1 and OR_2 larger than one. In this case, the larger OR_1 and OR_2 get, the greater the confounding effect. Therefore, the combination of OR_1 and OR_2 identifies potential deviation from the "strong ignorability" treatment assumption. Our simulation showed that the sensitivity analyses were also affected by the prevalence of U among those alive in control group (referred as baseline prevalence), which was related to β_0 . Conceivably, if the baseline prevalence of U is high, even a large odds ratio due to treatment or outcome won't increase the proportion of U by much, and we will not observe much difference in U between groups. Therefore higher baseline prevalence of U produces results less sensitive to the associations between (U, T) and (U, Y) . We imputed U with OR_1^* and OR_2^* taking various values from 1 to 15 (by changing β_1 and β_2), and the baseline proportion of U being 0.2, 0.3, 0.4, and 0.5 (by changing β_0). We used stochastic regression imputation to impute U based on the logistic model. To smooth the final sensitivity analysis curve, we imputed U ten times for each pair of OR_1^* and OR_2^* and combined the results following Rubin's rule for multiple imputation.

The result from our matched data showed the treatment group had higher emergency department mortality compared with the control groups. To simplify the sensitivity analysis, we assumed the unobserved binary covariate U was more likely to be present in the treatment group ($\text{OR}_1 > 1$) and positively related with emergency department mortality ($\text{OR}_2 > 1$). The rationale behind this setup was that we expected the unobserved U to explain some of the positive relationship between treatment and outcome (higher emergency department mortality in non-trauma centers). Since we did not see much difference in emergency department mortalities between patients treated in Level I and II trauma centers in the matched samples, we presented a simplified version of our results by generating U for

the two control groups similarly (by setting OR_1^* the same in the pair of non-trauma center/ Level I and non-trauma center/ Level II trauma centers). We also examined the effect of generating U differently for the two control groups, which required one more parameter. The results did not differ substantially from the simplified one, but the contour plots were much harder to visualize due to additional parameters.

We assumed four scenarios of baseline of U in control groups (Level I and II trauma centers) by changing β_0 :

A: $P(U)=0.2$; B: $P(U)=0.3$; C: $P(U)=0.4$; D: $P(U)=0.5$

The specific models used to generate U for the groups were:

In non-trauma center : $\text{Logit}[P(U=I)] = \beta_0 + \beta_1 + \beta_2 Y$

In Level I and Level II trauma center: $\text{Logit}[P(U=I)] = \beta_0 + \beta_2 Y$

By putting U along with the treatment indicator (three trauma care levels) in the outcome regression model, we evaluated how large of an OR_1 and an OR_2 were needed to overturn the qualitative conclusion (from $OR>1$ to $OR=1$). If large OR_1 and OR_2 are needed to change the qualitative conclusion, a result showing a positive relationship between being treated at Level I and II trauma centers and the reduced odds of emergency department mortality would be considered robust; otherwise, if a small OR_1 or OR_2 could alter the conclusion, the results would be sensitive to hidden bias.

Data source and study variables

We utilized five years of data from the Nationwide Emergency Department Sample (NEDS, 2006–2010).²³ The NEDS is a national dataset that has been designed by the Agency for Healthcare Research Quality (AHRQ) to enable analyses of emergency department utilization patterns. The NEDS is a publicly available de-identified dataset. This project does not fit the definition of human subjects' research under Title 45 Code Federal Regulations (CFR) part 46 or 21 CFR part 50. For this reason, it is considered exempt from IRB review by our institution. We used the National Trauma Data Standard Patient Inclusion Criteria to define trauma.²⁴ Children and older adults may respond to the treatment differently, so we limited our analysis to adults aged 18–64.

- Died in emergency department.

This was the outcome variable. "Emergency department visits in which the patient died in the department" was included in "disposition from emergency department." We grouped other emergency department dispositions as "did not die in emergency department."

- Trauma center designation.

Assigned by the American College of Surgeons' Committee on Trauma (ACS/COT), trauma center Level I and II have comprehensive resources and are able to care for the most severely injured. Level I trauma centers also provide leadership in education and research. Level III

trauma center hospitals are not included in this analysis. Other hospitals are called non-trauma centers.²⁵

- Injury Severity Score.

Injury severity score was calculated using a Stata program (ICDPIC) (ranging from 1 to 75 with 75 being the most severe).²⁶ The injury severity score correlates linearly with mortality, morbidity, length of hospital stay and other measures of severity.²⁷ The triage of patients is not based directly on injury severity score, but it can serve as a proxy of injury severity before treatment. Because the majority of the emergency department deaths were in the group of most severe patients (injury severity score ≥ 25), and for the purpose of illustration, our analyses were limited to trauma patients with scores ≥ 25 .

- Other covariates

included age, sex, comorbidity of chronic conditions, multiple injuries, median household income by zip code, expected primary payer, and urban-rural designation for patient's county of residence. The detailed definitions could be found on the website of Healthcare Cost and Utilization Project (HCUP)- NEDS.²⁸

RESULTS

Matching results

To maintain the profile of the patients treated in the treatment group (non-trauma centers), the treatment group needs to be the smallest to ensure there was one control patient for each treated subject. In our study, the number of patients treated in non-trauma centers was greater than in Level II trauma centers. To solve this problem, we randomly split the patients treated in non-trauma centers into two groups, H1 and H2. Each of them was matched with the two control groups separately (we also used H1 and H2 to indicate the matched subsamples). This setup also provided us the opportunity to examine whether the two subsamples produce the same conclusion.

The balance for each of the baseline covariates got better after matching (the absolute values of standardized differences were smaller after matching). Two variables, chronic conditions and multiple injuries, were not well balanced between non-trauma centers and Level II trauma centers (absolute value of standardized difference > 0.25). We tried to balance these two unbalanced covariates with a fine balance method and by setting up “sinks” to forcefully delete some patients with chronic conditions or multiple injuries from the Level II trauma center group.²⁹ Due to the large discrepancies in proportions of chronic conditions and multiple injuries between the two groups, we could not obtain a balanced level of standardized difference, even if all the unmatched patients in the Level II trauma center group were those with chronic conditions or multiple injuries. For this reason, we chose to adjust for residual confounding using a regression method. We completed the outcome analyses in two ways, with and without the unbalanced covariates, and we found only small differences. In the matched sample, the proportions of patients with chronic conditions or with multiple injuries were both higher in Level I and II trauma centers than in non-trauma centers. So, if the two variables were positively related with emergency department

mortality, as would be expected, and if the data were balanced, the difference in emergency department mortality between non-trauma centers and Level I or II trauma centers would be larger than what we reported here (Table 1).

Outcome analysis.

Our results showed that trauma centers (Levels I and II) had much lower odds of emergency department mortality than non-trauma centers. Although Level II trauma centers had lower emergency department mortality compared to Level I trauma centers, the difference was not statistically meaningful (the 95% confidence interval of the odds ratio contained 1) (Table 2).

Sensitivity analysis

The sensitivity analyses focused on H1, because in the subsample H1 the relationship between treatment and outcome was weaker (indicated by OR closer to 1) than in subsample H2. Figure 1 is the contour plot depicting the borderline (the combination of OR_1 and OR_2) when the association of being treated at Level I or II trauma centers and the reduced odds of mortality changed qualitatively (from $OR > 1$ to $OR = 1$). When the baseline prevalence of U decreased, the sensitivity increased (the curves shifted to lower-left). If the baseline prevalence was set at 0.2, OR_1 and OR_2 both needed to be slightly greater than 6 to overturn the study conclusion of reduced odds of mortality at the Level I or II trauma centers. This translated to a unmeasured confounder, U , that needed to be more than six times as prevalent (in terms of odds) in the non-trauma center as in the Level I trauma center and also more than six times as prevalent (in terms of odds) among those who died in the emergency department than among those that did not die there. The confounding effect of U at $OR_1=6$ and $OR_2=6$ should be deemed very strong for comparative trauma outcome research. We also conducted the sensitivity analysis for the comparison between non-trauma centers and Level II trauma centers for the four baseline scenarios, but these are not seen in Figure 1. The other five curves do not appear in this figure, because even at $OR_1^*=15$ and $OR_2^*=15$, the difference between non-trauma centers and trauma centers was not altered to null. The curves for non-trauma/ Level II trauma centers were on the upper-right of corresponding curves for non-trauma/ Level I trauma centers. This meant that the difference between non-trauma and Level II trauma centers was even more difficult to overturn. We also conducted sensitivity analyses with U distributed differently in Level I and Level II trauma centers. For example, we let OR_1^* in the pair of non-trauma and Level II trauma centers be 0.5 and 2.0 times of the OR_1^* in the pair of non-trauma and Level I trauma centers. None of the curves were lower than the curve of non-trauma/Level I trauma centers at the baseline 0.2. For simplicity of presentation, we did not show the results from these analyses.

DISCUSSION

We designed and implemented a three-step technique to conduct 1:1:1 triplet matching, and we extended Lu's imputation-based sensitivity.²² To our knowledge, we are the first to use this technique to conduct triplet matching and to perform sensitivity analysis after matching in trauma outcome research. Although comparison of trauma emergency department mortality between three trauma center levels was the example chosen to illustrate our

methods, we believe the methodology could be used more broadly among epidemiology studies. Our results indicated that if the most severe adult trauma patients treated in non-trauma centers were to be treated in Level I or Level II trauma centers, the odds of emergency department death would be reduced dramatically, and there were no statistically meaningful differences between Level I or Level II trauma centers. This finding suggested that Level II trauma centers are a good alternative to Level I centers, if severe trauma patients need to be triaged or transferred to a trauma center.

Propensity score matching is a common method of adjustment in observational studies.³⁰ Many matching techniques have been developed on the basis of two groups. Imbens proposed a weighting method using the generalized propensity score to account for multiple levels of treatment.¹⁰ Rubin showed that Imbens' method may be more sensitive to whether the propensity score has been accurately estimated.¹³ Lu and his colleagues explored the theory and methodologies of applying optimal non-bipartite matching to conduct a study with incomplete block design.^{12,31} Their design can be illustrated by Fig 2A (optimal three-way pair matching), while our triplet matching is illustrated by Fig 2B.

With non-bipartite matching, one subset of the treatment group is matched with the first control group and another subset of the treatment group is matched with the second control group. These two subsets of treatment group are disjoint by design; that is, they may have dissimilar distributions of the matching variables. This design is not ideal for our study, because our goal in matching was to form three similar matched groups with respect to the distribution of the non-trauma center patient population. Rassen and colleagues proposed another three-way matching method.¹¹ They used an algorithm for caliper-based nearest neighbor greedy matching using generalized propensity score. Although this method also creates matching triplets, it is still a greedy matching method. Simulation studies have shown that the pairwise matching procedure, proposed in this study, tended to produce matched triplets with a smaller total distance more often than the nearest neighbor algorithm. Hence, it likely leads to better balance in the results.³²

Like other experimental designs in observational studies, matching can only adjust for observed covariates.⁵ Unobserved factors that affect assignment to treatment cannot be accounted for in the matching procedure. The basic idea of Rosenbaum's sensitivity analysis is to examine, in Rosenbaum's words, "how much hidden bias can be presented before the qualitative conclusion begins to change".³³ In our study, we modified Lu's imputation based sensitivity analysis.²² There are three major differences between Lu's method and the method implemented here: (1) Lu's method used with survival data was essentially for comparing two treatment arms, whereas in our analysis we implemented the method for three arms; (2) We considered the baseline proportion change of U , and found that the sensitivity results were closely related with different baseline distributions of U ; (3) We used $OR=1$ as the cut point, instead of $P\text{ value}=0.05$. When $OR=1$ is the cut point, the definition of qualitative change is from $OR>1$ (reduced odds of emergency department mortality in Level I or Level II centers) to $OR \leq 1$ (higher odds of emergency department mortality than in non-trauma centers or no effect). Using $OR=1$ as the cut point is common in epidemiology, and it is more difficult to overturn study conclusions when compared with a cut point based on a $P\text{ value}=0.05$.

In the sensitivity analyses, we tried to find a real world example of an unobserved variable of U . One candidate we considered was the percentage of patients transferred from other hospitals. A higher percentage of patients transferred in is believed to be associated with higher mortality,³⁴ and it occurs more frequently in trauma centers than non-trauma centers, especially in Level I trauma centers, but this distribution is opposite to our simulated U (we assumed U to be positively related with emergency department mortality and more likely present in non-trauma centers). So if we had mimicked the percentage of transfer patients as the U , our conclusions would be even more difficult to overturn (more robust).

Although our matching method worked well for our data and design, there are several issues and limitations that researchers should keep in mind. First, our method used optimal pair matching between the treatment group and the two control groups; this method does not promise the overall distance of the matched triplet to be minimal among all possible ways of forming triplets.

Second, the maximum number of possible triplets equals the sample size of the smallest group (usually the treatment group), so that when there are large differences in the sample sizes of the groups, more subjects are excluded from the analysis. On the other hand, the greater the difference in sample sizes, the better chance of achieving balance in the matched sample. In contrast, if the sizes of the groups are similar and there are big differences in the distributions of matching covariates among groups, it is then difficult to balance all covariates. Thus, there is a tradeoff between using more subjects and finding the optimally balanced subset for comparison.

Third, in the ideal situation, all covariates should be balanced after matching. The causal effect can then be estimated by contrasting patients with similar propensity scores but from different treatment groups. In our analysis, two covariates could not be balanced without discarding a substantial portion of the patients. Therefore, we chose to employ further regression adjustment to control for the residual confounding after matching. But such an approach comes with some limitations: (i) it is essentially an extrapolation/interpolation on these non-overlapping regions of the covariates between the treatment and the two controls, (ii) this approach removes the “blinding” aspect of matching, and you cannot view the treatment assignment within matched sets as “randomized”, and (iii) it relies on a parametric modeling assumption between the outcome and the imbalanced covariates.³⁵

Fourth, since the primary outcome of interest is mortality, which is binary, we used logistic regression as the outcome model. Due to the non-collapsibility of the odds ratio, one needs to be careful while interpreting the treatment effect. As pointed out by Greenland,³⁶ we estimated the individual level causal effect in terms of the odds ratio (not the population average causal effect). We also implicitly assumed that there was a constant individual level causal effect in terms of odds ratio. This assumption also excluded the possibility of interaction between the causal effect and any covariate or propensity score strata.

Acknowledgements:

Thank Jin Peng for preparing the dataset. Thank Sarah Johnson in helping editing the original manuscript.

Funding Source

This work was supported by a research grant (PI: Dr. Huiyun Xiang, Grant No.:1R03-HS022277) from the Agency for Healthcare Research and Quality. The views and conclusions are solely the responsibility of the authors and do not represent official views of the funding agencies.

REFERENCES

1. Durham R, Pracht E, Orban B, et al. Evaluation of a mature trauma system. *Ann Surg* 2006;243(6): 775–83; discussion 783–5. [PubMed: 16772781]
2. MacKenzie EJ, Rivara FP, Jurkovich GJ, et al. A national evaluation of the effect of trauma-center care on mortality. *N Engl J Med* 2006;354(4):366–78. [PubMed: 16436768]
3. Vickers BP, Shi J, Lu B, et al. Comparative study of ED mortality risk of US trauma patients treated at level I and level II vs nontrauma centers. *Am J Emerg Med* 2015.
4. Xiang H, Wheeler KK, Groner JI, et al. Undertriage of major trauma patients in the US emergency departments. *Am J Emerg Med* 2014;32(9):997–1004. [PubMed: 24993680]
5. Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 1983;70(1):41–55.
6. Rubin DB. Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies. *J Am Stat Assoc* 1979;74(366):318–328.
7. Rubin DB. Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies. *Matched Sampling for Causal Effects*, 2006;142–159.
8. Campbell D. Artifact and control. In: Rosenthal R, Rosnow R, eds. *Artifact in behavioral research*. San Diego, CA: Academic Press, 1969.
9. Rosenbaum PR. The role of a second control group in an observational study. *Stat Sci* 1987;2(3): 292–306.
10. Imbens GW. The role of the propensity score in estimating dose-response functions. *Biometrika* 2000;87(3):706–710.
11. Rassen JA, Shelat AA, Franklin JM, et al. Matching by propensity score in cohort studies with three treatment groups. *Epidemiology* 2013;24(3):401–9. [PubMed: 23532053]
12. Lu B, Rosenbaum PR. Optimal pair matching with two control groups. *J Comput Graph Stat* 2004;13(2):422–434.
13. Rubin DB. On principles for modeling propensity scores in medical research. *Pharmacoepidem Dr S* 2004;13(12):855–857.
14. Rosenbaum P *Observational Studies*. New York, NY: Springer, 2002.
15. Hansen BB. Optmatch: Flexible, optimal matching for observational studies. *R News* 2007;7:18–24.
16. Kuhn HW. The Hungarian method for assignment problem. *Naval Res Logist Quart* 1955;2:83–97.
17. Bertsekas DP. A new algorithm for the assignment problem. *Math Program* 1981;21:152–171.
18. Pierskalla WP. The Tri-Substitution Method for the Three-Dimensional Assignment Problem. *Canadian Operational Research Society Journal* 1967;5(2).
19. Stuart EA. Matching Methods for Causal Inference: A Review and a Look Forward. *Stat Sci* 2010;25(1):1–21. [PubMed: 20871802]
20. Rosenbaum PR. Covariance adjustment in randomized experiments and observational studies. *Stat Sci* 2002;17(3):286–304.
21. SAS/STAT. Version 9.3 ed. Cary, NC: SAS Institute, 2014.
22. Lu B, Qian Z, Cunningham A, et al. Premarital cohabitation and marital disruption: selectivity or causality. *Sociol Methods Res* 2012;(In press).
23. Agency for Healthcare Research and Quality. Medical Expenditure Panel Survey (MEPS). http://meps.ahrq.gov/mepsweb/about_meps/survey_back.jsp Accessed February 18, 2014.
24. National Trauma Data Bank. National Trauma Data Standard Data Dictionary. http://www.ntdsdictionary.org/dataElements/documents/NTDS2012_xsd_Final_102011.pdf Accessed August 16, 2012.

25. Committee on Trauma American College of Surgeons. Resources for the Optimal Care of the Injured Patient 2006. Chicago, IL, 2006.
26. Clark DE, Osler TM, Hahn DR. ICDPIC: Stata module to provide methods for translating International Classification of Diseases (Ninth Revision) diagnosis codes into standard injury categories and/or scores. <http://ideas.repec.org/c/boc/bocode/s457028.html> Accessed October 10, 2013.
27. Brohi K Injury Severity Score: Overview and Desktop Calculator. <http://www.trauma.org/index.php/main/article/383/> Accessed October 10, 2013.
28. Distributor HC. NEDS Description of Data Elements. <http://www.hcup-us.ahrq.gov/db/nation/neds/nedsdde.jsp> Accessed 5/10/2015, 2015.
29. Rosenbaum PR. Design of Observational Studies. Springer Series in Statistics New York: Springer, 2010.
30. Rosenbaum PR. Optimal matching in observational studies. J Am Statist Assoc 1989;84:1024–1032.
31. Lu B, Greevy R, Xu X, et al. Optimal Nonbipartite Matching and Its Statistical Applications. Am Stat 2011;65(1):21–30. [PubMed: 23175567]
32. Lu B. Cross-time Matching in Observational Studies; Mid-Atlantic Causal Inference Conference; 2009.
33. Rosenbaum PR, Rubin DB. Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome. Journal of the Royal Statistical Society Series B-Methodological 1983;45(2):212–218.
34. Newgard CD, McConnell KJ, Hedges JR, et al. The benefit of higher level of care transfer of injured patients from nontertiary hospital emergency departments. J Trauma 2007;63(5):965–71. [PubMed: 17993937]
35. Rubin DB, Waterman RP. Estimating the causal effects of marketing interventions using propensity score methodology. Stat Sci 2006;21(2):206–222.
36. Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. Am J Epidemiol 1987;125(5):761–8. [PubMed: 3551588]

Fig 1.

Contour plot of sensitivity analyses, subsample H1, controlling the baseline proportion of U in Level I trauma centers (TC I) at A: $P(U)=0.2$, B: $P(U)=0.3$, C: $P(U)=0.4$, D: $P(U)=0.5$, with U similarly distributed in Level I and Level II trauma centers (TC II). OR1 is the indicator of association between U and treatment, and OR2 is the indicator of association between U and outcome (emergency department death). The line indicating the contour (based on combinations of OR1 and OR2) where the causal effect of treatment on outcome begins to change qualitatively (odds ratio changed from greater than one to less than one). The combinations in the area to the right and above the border line shows the region where the difference in emergency department mortality between non-trauma centers (NTC) and Level I or Level II trauma centers is changed qualitatively. The total number of curves should be $4 \times 2 = 8$. Five curves missing from the figure are those the difference between non-trauma centers and Level I or Level II trauma centers cannot be changed qualitatively even at $OR1^*=15$ and $OR2^*=15$ (conditional ORs). OR indicates odds ratio. TC I indicates Level I trauma centers; TC II indicates Level II trauma centers; NTC indicates non-trauma centers. “A” indicates baseline distribution of $U=1$ in Level I trauma center equal to 0.2; “B” indicates baseline distribution of $U=1$ in Level I trauma center equal to 0.3; “C” indicates baseline distribution of $U=1$ in Level I trauma center equal to 0.5. “01” indicates comparison between non-trauma centers and Level I trauma centers; “02” indicates comparison between non-trauma centers and Level II trauma centers.

A. Optimal three-way pair matching

B. Triplet matching (1:1:1)

Fig 2.A-B.

Optimal three-way pair matching and triplet matching

Table 1A.

Balance checking before and after matching, first subsample, H1

	Original sample				Matched sample			
	Mean/Percent		Standardized difference		Mean/Percent		Standardized difference	
	Non-trauma centers (N=2655)	Level I trauma centers (N=13383)	Level II trauma centers (N=3158)	Non- vs. level I trauma centers 0.0968	Non-trauma centers (N=2655)	Level I trauma centers (N=2655)	Level II trauma centers (N=2655)	Non- vs. level I trauma centers 0.0160
Age (mean, years)								
Mean	38.5	37.2	37.8	0.0968	38.5	38.3	38.5	0.0160
(standard deviation)	(13.7)	(13.6)	(13.9)		(13.7)	(13.5)	(13.8)	
ISS (mean)								
Mean	34.2	32.5	32.6	0.1233	34.2	33.6	33.0	0.0417
(standard deviation)	(16.6)	(10.9)	(11.9)		(16.6)	(15.1)	(12.7)	
Female	26.9	25.4	24.6	0.0353	26.9	24.5	25.7	0.0548
With chronic condition	51.9	71.3	71.3	-0.4086	51.9	56.6	67.9	-0.0988
Multiple injury	66.8	92.3	89.1	-0.6652	66.8	73.7	87.0	-0.1800
Patient location								
Large central metropolitan area	15.4	24.5	13.9	-0.2276	15.4	17.7	15.2	-0.0559
Large fringe metropolitan area	20.4	22.2	24.6	-0.0440	20.4	19.8	21.9	0.0147
Medium metropolitan area	21.4	20.9	28.8	0.0118	21.4	21.1	25.5	0.0066
Small metropolitan area	11.0	7.9	9.1	0.1063	11.0	10.7	10.2	0.0089
Metropolitan area	17.5	13.6	12.0	0.1075	17.5	17.2	14.2	0.0083
Not metropolitan or micropolitan area	12.5	9.1	10.8	0.1106	12.5	12.4	11.9	0.0035
Median household income by patient zip code								
Q1 (0–25%)	29.7	35.4	26.2	-0.1229	29.7	32.0	27.5	-0.0492
Q2 (25%–50%)	27.8	26.3	32.4	0.0322	27.8	27.5	32.1	0.0068
Q3 (50%–75%)	21.9	20.6	21.4	0.0325	21.9	21.1	21.5	0.0193
Q4 (75%–100%)	16.8	13.4	17.3	0.0942	16.8	16.5	16.1	0.0095

	Original sample				Matched sample			
	Mean/Percent		Standardized difference		Mean/Percent		Standardized difference	
	Non-trauma centers (N=2655)	Level I trauma centers (N=13383)	Non- vs. level I trauma centers	Non- vs. level II trauma centers	Non-trauma centers (N=2655)	Level I trauma centers (N=2655)	Non- vs. level I trauma centers	Non- vs. level II trauma centers
Primary expected payer								
Medicare	5.5	4.7	0.0383	0.1025	5.5	5.5	0.0000	0.0714
Medicaid	12.2	16.7	-0.1286	-0.0686	12.2	12.2	-0.0009	-0.0265
Private insurance	43.4	47.6	-0.0850	-0.2379	43.4	43.3	0.0014	-0.1837
Self-pay	26.1	18.9	0.1726	0.2404	26.1	25.8	0.0062	0.1761
No charge	1.2	1.7	-0.0367	0.1153	1.2	1.2	0.0000	0.1095
Other	10.8	9.2	0.0544	0.0252	10.8	10.7	0.0050	-0.0013

^a considered as unbalanced according to the criterion of 0.25.

Note: ISS = Injury Severity Score

Table 1B.

Balance checking before and after matching, second subsample, H2

	Original sample				Matched sample			
	Mean/Percent		Standardized difference		Mean/Percent		Standardized difference	
	Non-trauma centers (N=2659)	Level I trauma centers (N=13383)	Level II trauma centers (N=3158)	Non- vs. level I trauma centers	Non- vs. level II trauma centers	Non-trauma centers (N=2659)	Level I trauma centers (N=2659)	Level II trauma centers (N=2659)
Age (mean, years)				0.0795	0.0374			
Mean	38.3	37.2	37.8			38.3	38.2	38.5
(standard deviation)	(13.7)	(13.6)	(13.9)			(13.7)	(13.6)	(13.9)
ISS (mean)				0.1184	0.1065			
Mean	34.1	32.5	32.6			34.1	33.5	33.0
(standard deviation)	(16.6)	(10.9)	(11.9)			(16.6)	(14.9)	(12.7)
Female	26.4	25.4	24.6	0.0224	0.0404	26.4	23.9	25.5
With chronic condition	49.5	71.3	71.3	-0.4573	-0.4569	49.5	55.6	67.4
Multiple injury	68.1	92.3	89.1	-0.6373	-0.5297	68.1	74.0	87.1
Patient location								
Large central metropolitan area	16.3	24.5	13.9	-0.2046	0.0665	16.3	18.4	15.3
Large fringe metropolitan area	18.5	22.2	24.6	-0.0912	-0.1470	18.5	18.4	21.3
Medium metropolitan area	20.1	20.9	28.8	-0.0188	-0.2025	20.1	19.8	26.0
Small metropolitan area	12.5	7.9	9.1	0.1545	0.1117	12.5	12.5	10.5
Metropolitan area	18.1	13.6	12.0	0.1226	0.1710	18.1	17.4	14.0
Not metropolitan or metropolitan area	12.8	9.1	10.8	0.1189	0.0636	12.8	12.6	12.0
Median household income								
Q1 (0-25%)	29.4	35.4	26.2	-0.1289	0.0712	29.4	33.1	27.1
Q2 (25%-50%)	29.6	26.3	32.4	0.0729	-0.0597	29.6	28.9	33.0
Q3 (50%-75%)	22.2	20.6	21.4	0.0400	0.0199	22.2	21.3	22.1
Q4 (75%-100%)	14.4	13.4	17.3	0.0280	-0.0800	14.4	14.2	15.0

τ considered as unbalanced according to the criterion of 0.25.

Emergency department mortality comparisons before and after matching for the most severe adult trauma patients (ISS 25, age:18–64), by trauma center level, NEDS 2006–2010

Table 2.

	Original sample			Matched sample		Logistic regression	
	N	No. died in emergency department	Mortality in emergency department (%)	N	No. died in emergency department	Mortality in emergency department (%)	OR ^a 95% CI
First subsample, H1							
Non-trauma centers	2,655	362	13.6	2,655	362	13.6	0.61 (0.50,0.74)
Level I trauma centers	13,383	503	3.8	2,655	207	7.88	0.49 (0.38,0.63)
Level II trauma centers	3,158	134	4.2	2,655	122	4.6	0.81 (0.63,1.04)
Second subsample, H2							
Non-trauma centers	2,659	398	15.0	2,659	398	15.0	0.50 (0.41,0.60)
Level I trauma centers	13,383	503	3.8	2,659	197	7.4	0.42 (0.33,0.53)
Level II trauma centers	3,158	134	4.2	2,659	123	4.6	0.84 (0.66,1.09)

^a After controlling for chronic condition and multiple injury, the two variables were not balanced very well.

^b From a second model, same data, but with TC I as reference group.