

Utilizing Propensity Score Analyses in Prehospital Blood Product Transfusion Studies: Lessons Learned and Moving Toward Best Practice

Thomas J. Greene, PhD*; Stacia M. DeSantis, PhD*; Erin E. Fox, PhD†‡; Charles E. Wade, PhD†‡; John B. Holcomb, MD†‡; Michael D. Swartz, PhD*†

ABSTRACT Recently, observational studies analyzing prehospital blood product transfusions (PHT) for trauma have become more widespread in both military and civilian communities. Due to these studies' non-random treatment assignment, propensity score (PS) methodologies are often used to determine an intervention's effectiveness. However, there are no guidelines on how to appropriately conduct PS analyses in prehospital studies. Such analyses are complicated when treatments are given in emergent settings as the ability to administer treatment early, often before hospital admission, can interfere with assumptions of PS modeling. This study conducts a systematic review of literature from military and civilian populations to assess current practice of PS methodology in PHT analyses. The decision-making process from the multicenter Prehospital Resuscitation on Helicopter Study (PROHS) is discussed and used as a motivating example. Results show that researchers often omit or incorrectly assess variable balance between treatment groups and include inappropriate variables in the propensity model. When used correctly, PS methodology is an effective statistical technique to show that aggressive en route resuscitation strategies, including PHT, can reduce mortality in individuals with severe trauma. This review provides guidelines for best practices in study design and analyses that will advance trauma care.

INTRODUCTION

Although randomized controlled trials (RCTs) remain the gold standard, RCTs may be either impractical or impossible to implement, especially in emergency care fields such as trauma. Specific barriers in trauma include administration of treatment that is tailored to an individual's specific injury, inexperience with performing Exception from Informed Consent studies, cost, and logistical issues involved in ensuring timely randomization assignment and treatment administration.^{1,2} When RCTs are not ethically or practically possible, observational studies with non-random treatment assignment are often implemented. Although observational studies do not result in the same level of evidence as an RCT, they can provide useful and accurate results when properly implemented.

In properly designed and executed randomized studies, treatment assignment and patient characteristics are guaranteed to

be independent as variables such as age, gender, and disease severity do not influence the choice of treatment.^{3,4} However, this is not the case in observational studies, where treatment assignment is chosen at the physician's discretion and therefore influenced by the patient's baseline characteristics. Thus, patient variables may not be similar (or balanced) between treatment groups. For example, the Prehospital Resuscitation on Helicopter Study (PROHS) was an observational study investigating the effectiveness of prehospital blood product transfusion (PHT) where patients with a traumatic injury who received PHT were more injured, on average, than those who did not receive PHT.⁵ This issue is both expected and unavoidable. When used correctly, propensity scoring methodology can be an effective statistical technique to balance treatment groups and accurately estimate a causal treatment effect.³

Hemorrhage is a leading cause of preventable death among trauma patients in both civilian and military populations.⁶⁻⁸ To improve outcomes after traumatic injury, numerous recent studies utilizing propensity scoring methodology have investigated whether or not early administration of blood products affects mortality.⁹⁻¹³ However, difficulties with the propensity score (PS) analysis arose throughout PROHS. This led to a search of available literature for guidance on the nuances of conducting a causal analysis when the treatment is administered before hospital admission. Although one paper detailed critiques and guidelines for proper PS use in cardiovascular surgery, there was only one brief review that detailed results of PHT studies.^{14,15} Thus, there is not adequate referential material in this area of research. The objective of this article is to review the current PHT literature, assess strengths and

*Department of Biostatistics, School of Public Health, University of Texas Health Science Center at Houston, 1200 Pressler St, Houston, TX 77030.

†Center for Translational Injury Research, University of Texas Health Science Center at Houston, 6410 Fannin St, Houston, TX 77030.

‡Department of Surgery, McGovern Medical School, University of Texas Health Science Center at Houston, 6431 Fannin St, Houston, TX 77030.

The opinions or conclusions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of any sponsor. This manuscript has been reviewed by the PROHS Publication Committee for scientific content and consistency of data interpretation with previous PROHS publications.

doi: 10.1093/milmed/usx137

© Association of Military Surgeons of the United States 2018. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

limitations of current analytic practice, and present clear guidelines for how to plan and implement PS analyses in studies focusing on prehospital blood product administration.

Propensity Score Basics

This manuscript provides guidance for the broad overall process of PS analyses and recommends other literature for more detailed reviews of PS implementation.^{16–18} At the most basic level, the PS is a single value defined as a patient's probability of receiving treatment given their observed variables.³ On average, patients with similar propensity scores also have similar observed variables.¹⁸ Therefore, the PS is a natural metric used to group subsets of patients who are similar. Conducting a PS analysis is a three-step procedure that is described below:

1. First, a model is fitted with treatment as the outcome variable.
 - a. All baseline variables that influence treatment selection, but none that are influenced by treatment selection, should be included as independent variables in this model.
2. Results from the previous model are then used to calculate each subject's conditional probability of receiving treatment given their observed baseline clinical variables, that is, the PS.
3. Once each patient's PS is obtained, a final outcome model is built to estimate the treatment effect on outcome (e.g., mortality) that adjusts for the PS by either stratifying, matching, weighting, or including the PS in the regression model.

It is important to perform diagnostic checks of the model after the propensity model is fitted. First, the distribution of the variables used to obtain the PS should be compared across treatments to ensure that application of the propensity scoring technique produces a balanced sample. The similarity of distributions should be assessed using qualitative methods such as comparison of box plots or histograms for continuous variables, or cross-tabulations for discrete variables.^{17,19} Although it may seem intuitive to assess balance using formal statistical tests, experts warn against this “balance test fallacy.”^{20–22} This fallacy exists since statistical tests use a sample to make inference about the overall population. When conducting PS analyses, it is of no interest to determine if the overall population is balanced. Instead, it is only necessary that the observed sample is balanced. *p*-Values also have the undesirable property of being influenced by sample size. Therefore, to quantitatively determine if balance has been achieved, experts suggest comparing standardized differences, which are not affected by sample size.^{20,23,24}

As the PS can be thought of as a single number representing the summary of observed baseline clinical variables, another way to roughly assess balance is by comparing distributions of the PS between treated and untreated patients.^{3,21,25} By

comparing these distributions, an area of common overlap of propensity scores can be observed. This area represents individuals who have similar propensity scores and therefore have, on average, similar baseline characteristics.²¹ If the distributions do not overlap, that is when there exist treated subjects with propensity scores outside of the range of control subjects, no valid estimation of the treatment effect can be made.²⁵ Overlap is directly tied to the assumption of positivity, which requires all subjects to have a probability of receiving treatment that is greater than zero.^{3,18} Therefore, it is essential that any study utilizing propensity scoring methodology assesses the conditions of balance and overlap to ensure accurate results. Figure 1A displays an example of treated and control distributions, which have good overlap, whereas Figure 1B shows poor overlap, as displayed by the medium gray region.

METHODS

Study Identification

In order to make recommendations for future research, it is first necessary to assess current practice. We conducted a systematic review in order to identify recent clinical studies that used propensity scoring techniques to assess the effect of PHT on mortality. To identify pertinent articles, we performed online searches using the National Center for Biotechnology Information's PubMed database and EMBASE for articles containing MeSH headings of “emergency medical services” (the recommended MeSH term to identify prehospital treatment) and “blood transfusion” and text: “propensity score,” “transfusion,” “hemorrhage,” “resuscitation,” or “red blood cell.” This search identified 16 potential papers for initial investigation.

Assessment Criteria

In order to determine how closely current practice agrees with PS theory, details regarding the propensity model building process in each study were thoroughly examined and assessed for validity in the areas of:

1. selection of the control population
2. propensity model building
3. application of selected PS method
4. covariate balance assessment

RESULTS

A flowchart detailing the selection process and primary reason for article exclusion can be seen in Figure 2. Five retrospective cohort studies, along with PROHS, met all selection criteria and were chosen for review.^{5,9–13} Estimates of the effect of PHT on the selected mortality endpoints for each article are shown in Figure 3 and an evidence table can be seen in Table I.

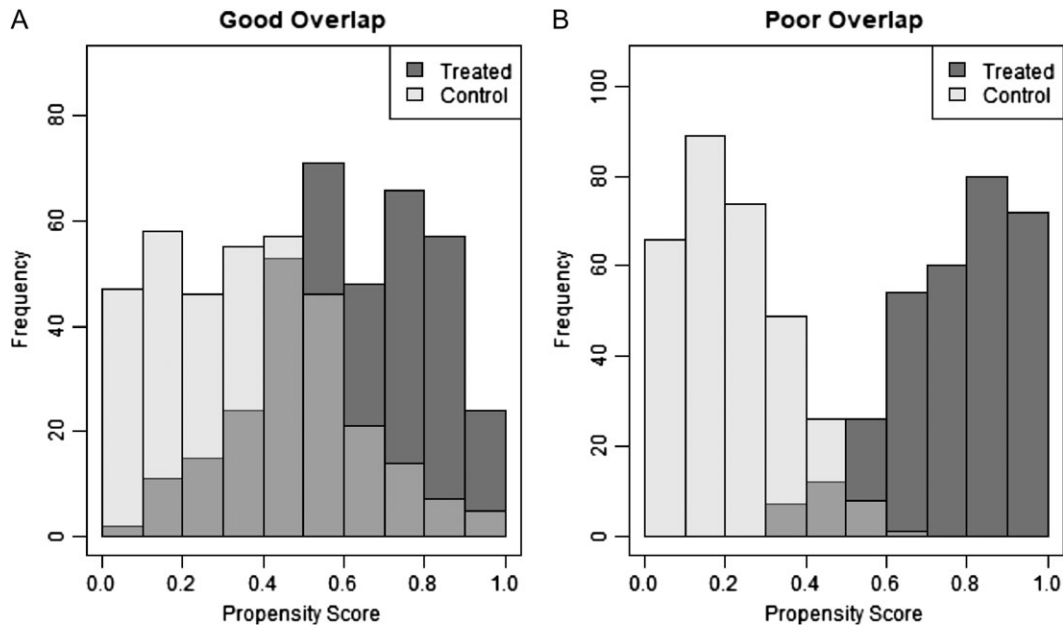


FIGURE 1. The medium gray area represents the area of common support (overlap). An example of good overlap is displayed in panel (A) and poor overlap is displayed in panel (B).

Review of the Relevant Literature

In their 2013 paper, Brown et al used propensity scoring to assess outcomes for prehospital crystalloid volume in patients with and without prehospital hypotension suffering from a severe blunt traumatic injury.⁹ Although administration of crystalloid is not identical to PHT, this study utilized propensity scores and was important in paving the way for future research on PHT.¹⁰ Data were collected between 2003 and 2010 from the Inflammation and Host Response to Injury Large Scale Collaborative Program, a large prospective multicenter observational study of blunt injured adults with hemorrhagic shock who lived long enough for admission to an intensive care unit.²⁶ Researchers used multiple logistic regression to build a PS model for the binary treatment variable indicating high amounts of prehospital crystalloid received (greater than 500cc). This model calculated a patient's probability of receiving a high amount of crystalloid, adjusting for variables influencing treatment assignment listed in the summary of evidence Table I. The authors fit a Cox proportional hazards model controlling for the selected independent variables as well as the PS indicating probability of receiving high amounts of crystalloid to assess the effect of PHT on mortality.¹⁰ Separate Cox proportional hazards models were constructed for patients with and without prehospital hypotension, defined as systolic blood pressure (SBP) < 90 mmHg. For subjects without prehospital hypotension, high prehospital crystalloid administration was not significantly associated with 24-hr mortality (hazard ratio [HR₂₄] = 3.68; 95% CI [0.78, 17.24], $p = 0.10$) but was associated with a significant increase in risk of 30-d in-hospital mortality (HR₃₀ = 2.45; 95% CI [1.25, 4.83], $p = 0.01$). However, among subjects with prehospital hypotension, high prehospital crystalloid volume was not

significantly associated with 24-hr or 30-d mortality (HR₂₄ = 1.40; 95% CI [0.33, 6.03], HR₃₀ = 0.81).

O'Reilly et al conducted a study considering casualties admitted to the field hospital at Camp Bastion, Afghanistan, from May 2006 to March 2011.¹² Administration of PHT, consisting of packed red blood cells (RBCs) and fresh frozen plasma, to military casualties in Afghanistan became possible in July 2008. The authors conducted a matched PS analysis to attempt to quantify the effect of PHT on mortality in combat casualties. First, a multiple logistic regression model was used to calculate a PS, interpreted as the probability of a subject receiving PHT given their observed variables, for each subject who was treated after July 1, 2008. The binary outcome variable was an indicator of whether or not the subject received PHT and the model controlled for the variables listed in Table I. Subjects treated with PHT after July 1, 2008, were then matched to one or more subjects with common observed variable profiles treated before July 1, 2008. A matching algorithm created matched pairs where one subject was treated with PHT and the other would have likely been treated with PHT, had it been available. Results from matched McNemar tests indicated that the 30-d mortality rate in PHT recipients was significantly less than the mortality rate in non-recipients (8.2% vs 19.6%, $p = 0.013$). In order to ensure comparability between studies, a relative risk (RR) point estimate of mortality for PHT recipients compared with non-recipients was calculated (RR 8.2/19.6 = 0.42; 95% CI [0.19, 0.92]).

In 2015, Brown et al published the first civilian study that utilized propensity scoring to indicate that PHT is associated with reduced mortality and coagulopathy in severely injured patients with blunt trauma.¹⁰ The 2015 study also used subjects

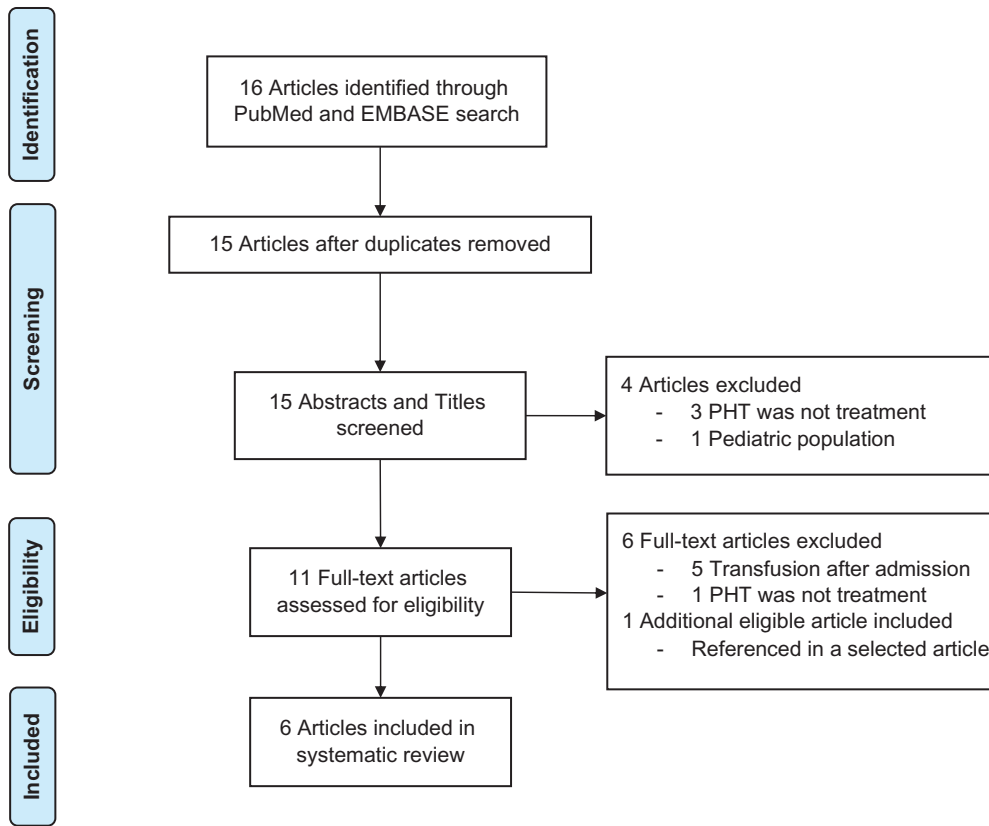


FIGURE 2. PRISMA flowchart of search strategy and study selection. *Note.* PHT, prehospital blood product transfusion.

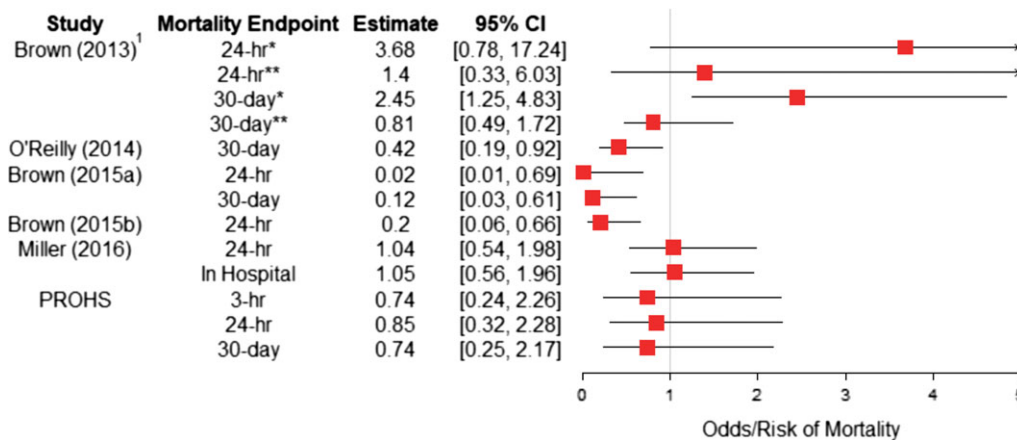


FIGURE 3. Forest plot of point estimates and variability of the effect of PHT on mortality endpoints for the six selected studies. ¹Patients treated with prehospital crystalloid not prehospital blood product transfusion. *Subset of patients without prehospital hypotension. **Subset of patients with prehospital hypotension.

from the Inflammation and Host Response to Injury Large Scale Collaborative Program but focused on assessing the effect of PHT, rather than crystalloid administration.²⁶ Since only 50 patients received PHT compared with 1,365 subjects who did not receive PHT, a 3:1 matched PS analysis was conducted. Similar to the 2013 study, researchers constructed a multiple logistic regression model controlling for important prehospital variables to calculate each subject's probability of receiving

PHT. A total of 113 subjects were matched, 35 received PHT and 78 did not. In the matched cohort of patients, a conditional logistic regression model for matched pairs indicated a 98% reduction in 24-hr mortality (odds ratio [OR] = 0.02; 95% CI [0.01, 0.69]). Similarly, the Cox proportional hazards model indicated an 88% reduction in risk for 30-d mortality (hazard ratio = 0.12; 95% CI [0.03, 0.61]). These promising results led the authors to make two primary suggestions: further investigation

TABLE I. Table of Evidence for Selected Studies

Study	Population	Treatment	Control Population	PS Model	PS Variables	PS Application Method	Mortality Endpoint (s)	Outcome Model and Estimate	Outcome Variables
Brown (2013)	Civilian	Crystalloid volume	Subjects from Glue Grant database with blunt trauma treated with < 500cc of prehospital crystalloid (2003–2010)	Logistic regression	Time from injury to hospital, PH SBP, units of PH blood products, ISS, initial base deficit	Including PS as independent variable in outcome model	24 h, 30 d	Cox proportional hazards	Probability of receiving > 500cc of crystalloid (propensity score), age, gender, PH packed RBC volume, PH heart rate, PH GCS, total PH time, ISS, admission: base deficit, Hgb level, and INR, ED body temperature, ED hypotension, vasopressor use, laparotomy or thoracotomy within 48 h, trauma treatment center, 24 h volume of packed RBCs, FFP, platelet, and crystalloid
O'Reilly (2014)	Military	PHT	Military casualties treated in Afghanistan (May 2006 to July 1, 2008)	Logistic regression	Sex, age, nationality, mechanism of injury, three most severe AIS codes	Matching	30 d	McNemar test	NA
Brown (2015a)	Civilian	PHT	Subjects not treated with PHT from Glue Grant database (2003–2010)	Logistic regression	Sex, age, PH time, PH SBP, PH heart rate, PH GCS, PH crystalloid volume, ISS, trauma center, admission values of INR, hemoglobin level, base deficit	Matching	24 h, 30 d	Logistic regression (24hr), Cox proportional hazards (30 d)	Age, gender, year of enrollment, transfer status, PH time, PH SBP, PH crystalloid volume, admission GCS, admission INR, initial base deficit, ISS, ED hypothermia, vasopressor use, urgent laparotomy or thoracotomy, 24-h volume of packed RBCs, FFP, platelets, and crystalloid

Brown (2015b)	Civilian	PHT	HEMS patients transported to UPMC who were not treated with PHT (2007–2012)	Logistic regression	Age, transfer status, PH SBP, PH heart rate, RBC volume before HEMS arrival, crystalloid before arrival and during HEMS transport, MOI, HEMS transport distance	Matching	24 h	Conditional logistic regression	Sex, race, ISS, admission values for SBP, heart rate, GCS, and INR, alcohol intoxication, ICU admission, emergent abdominal thoracic or vascular operation, ventilation, trauma mortality prediction model (TMPM) predicted mortality
Miller (2016)	Civilian	PHT	HEMS patients transported to VUMC who did not receive PHT (2007–2013)	Logistic regression	Age, MOI, scene pulse, scene SBP, scene GCS, travel duration, ISS, total blood products received within 24 h	Matching	24 h, in-hospital	Conditional logistic regression	Age, ISS, HCT, ED pulse, ED SBP, 24-h blood in hospital, travel duration, sex, race, MOI, ED GCS
PROHS (2017)	Civilian	PHT	HEMS patients transported on HEMS without blood available (January to November 2015)	GBM	Age, gender, race, ISS, PH SBP PH DBP, PH pulse, highest risk indicator, MOI, PH LSI, time from air team call time to arrival at ED, PH bleeding site identified, site volume	Matching	3 h, 24 h, 30 d	Conditional logistic regression	Age, race, gender, SBP, PH LSI, ISS, > 1 high-risk criteria, identification of bleeding source, pulse, time from air team call time to arrival at ED

PH, prehospital; ISS, injury severity score; RBC, red blood cell; GCS, Glasgow coma scale; INR, internal normalized ratio; FFP, fresh frozen plasma; AIS, abbreviated injury scale, SBP, systolic blood pressure; DBP, diastolic blood pressure; ED, emergency department; MOI, mechanism of injury (blunt/penetrating); HEMS, Helicopter Emergency Medical Services; LSI, lifesaving intervention; HCT, hematocrit.

of PHT in resuscitation strategies and addressing logistical challenges that prevent the RBC products from being widely available in the prehospital setting.¹⁰

Brown et al, 2015b investigated the relationship between PHT and early survival¹¹ among patients transported by helicopter to the University of Pittsburgh Medical Center Presbyterian Hospital from 2007 to 2012. As the treatment assignment was not random, patients receiving PHT were typically more severely injured and had a higher risk of mortality.¹¹ The authors used PS matching to pair patients treated with PHT with similar control patients. Of the 8,616 air medical trauma patients identified, 255 received PHT and 240 of these were PS matched to 480 control patients. Conditional logistic regression, adjusting for selected variables, was used in the outcome analysis to assess the effect of PHT among patients. Results indicated that PHT was associated with a very large increase in 24-hr survival (OR = 4.92; 95% CI [1.51, 16.04]). These values were converted to risk of mortality in Figure 3 by inverting OR and CI limits (OR = 0.20; 95% CI [0.06, 0.66]). The authors also reported decreased odds of shock (OR = 0.28; 95% CI [0.09, 0.85], not shown in Fig. 3), a secondary outcome, in subjects who received PHT compared with those who did not. These findings agree with other studies and support the notion that PHT is an effective treatment for patients with severe hemorrhage.

The study conducted by Miller et al in 2016 included all adult trauma patients from 2007 to 2013 who were transported by air from the scene to Vanderbilt University Medical Center.¹³ A multiple logistic regression model, controlling for the selected variables listed in Table I, was used to calculate the probability of each patient receiving PHT. After matching 195 recipients of PHT to 195 control patients, most, but not all, of the variables showed evidence of being balanced. A conditional logistic regression model was used to evaluate the matched data. Results did not show evidence of an effect of PHT on 24-hr in-hospital mortality (OR = 1.04; 95% CI [0.54, 1.98]) or overall in-hospital mortality (OR = 1.05; 95% CI [0.56, 1.96]). As these results seem counterintuitive to current early resuscitation strategy with blood products, the authors suggest that more prospective studies are needed to assess the potential risks, benefits, and costs of widespread use of PHT during transport.

PROHS enrolled 1,058 patients across nine US Level I trauma centers from January to November 2015.⁵ Five sites in PROHS had the ability to administer PHT while four did not. The objective of the study was to find comparable patients who did and did not receive PHT and implement PS matching to determine the effect of PHT on 3-hr, 24-hr, and 30-d mortality. A generalized boosted model was used to calculate each patient's PS. However, systematic differences in patient characteristics between sites with PHT capabilities and those without led to uncontrollable variable imbalance between treatment groups and severe lack of overlap. This lack of overlap led to only 109 of 1,058 eligible patients being matched. After matching, conditional logistic regression was used to calculate odds

of mortality. There was no significant difference observed between patients treated with PHT and those who were not at 3-hr (OR = 0.74; 95% CI [0.24, 2.26]), 24-hr (OR = 0.85; 95% CI [0.32, 2.28]), or 30-d (OR = 0.74; 95% CI [0.25, 2.17]) mortality. The PROHS results suggest that large randomized studies will be required to definitively assess the effectiveness of PHT.⁵

Assessment of Selected Articles

Figure 4 presents a flowchart describing the steps for implementing propensity scoring analysis according to the assessment criteria outlined in the introduction of this manuscript. Table II presents a summary of the performance of each study for the defined assessment criteria; a check indicates that the criteria were satisfied and an exclamation point indicates that the criteria were not fully met. The criteria are briefly discussed for each article below.

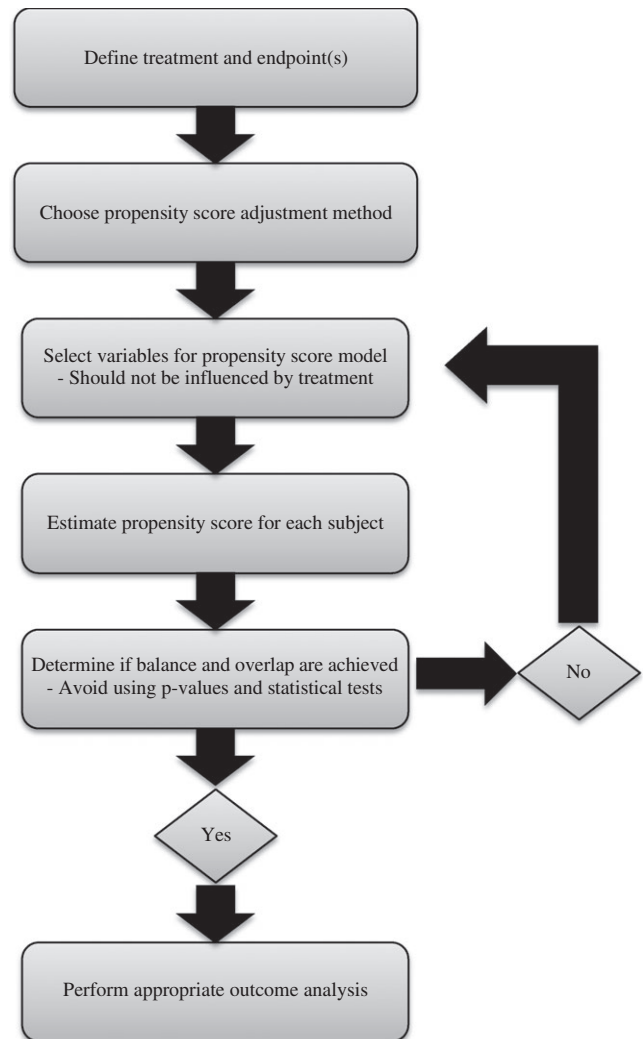


FIGURE 4. Detailed suggested steps for implementing a propensity score analysis with prehospital intervention.

TABLE II. Assessment of Study Quality

Study	Selection of Control Population	PS Model Building	Application of PS Model	Balance Assessment	Overall Quality
Brown <i>et al</i> ⁹	✓	✓	✓	✓	✓
O'Reilly <i>et al</i> ¹²	!	✓	✓	!	✓
Brown <i>et al</i> ¹⁰	✓	✓	✓	!	✓
Brown <i>et al</i> ¹¹	✓	✓	✓	✓	✓
Miller <i>et al</i> ¹³	✓	!	✓	!	✓
PROHS ⁵	✓	✓	✓	✓	✓

PS, propensity score.

Note. A check mark indicates that the criteria was sufficiently met. An exclamation point indicates the criteria was not sufficiently met.

As the Brown 2013 study utilized a large amount of data from the multicenter glue grant, there was a large population from a similar time period from which treated and control patients could be selected.²⁶ Furthermore, both treated and control patients were enrolled throughout the entirety of the study, preventing selection bias due to differences in temporality. The authors thoroughly explained the propensity model building process and specifically listed which variables were included in the propensity model. Variables were considered for inclusion into the logistic regression PS model based on whether they were available to prehospital care providers and whether they could be reasonably assumed to affect treatment assignment.⁹ The authors justify the inclusion of injury severity score (ISS) and initial base deficit as variables in the PS model by noting that, while they would not be available to prehospital providers, they are a good surrogate for observed injury severity that a prehospital provider would subjectively evaluate and use to guide resuscitation.⁹ The PS was included as an independent variable in the Cox proportional hazards outcome model. The authors presented a brief assessment of balance by comparing quartiles of propensity scores between treated and untreated patients. Although not providing the detail that a plot of overlapping densities provides, it does indicate that there is a significant overlap between the treatment groups. Overall, this study is of high quality because it satisfies pertinent requirements of PS theory.

In the O'Reilly article, patients treated with PHT between July 2008 and March 2011 were matched to similar patients resuscitated between May 2006 and July 2008 before PHT was available. The authors note that numerous changes in practice, as well as decreased prehospital times, due to increased deployment of forces occurring between 2006 and 2011, could have influenced mortality.¹² Ideally, one would control for this temporal effect by including the year of injury as a variable in both the PS and the outcome models; however, this is not possible since year cannot be disentangled from treatment assignment. Due to study design and selection of controls, it cannot be determined if effects observed are due to treatment or evolving changes in practice, a limitation that the authors fully recognize. The authors did not avoid the "balance test fallacy" and assessed covariate balance within the matched subset using statistical tests and *p*-values. Although this article has some unavoidable limitations due to availability of the control sample, the authors present clear and informative results.

The control population used for the Brown *et al* (2015a) manuscript also came from the glue grant database, an adequate source for obtaining an untreated control sample.²⁶ The authors carefully selected numerous prehospital measurements to include as independent covariates in the propensity model. However, two of these variables were admission hemoglobin level, and admission internal normalized ratio, which are measured after, and may be influenced by treatment assignment.²⁷ No justification was provided for inclusion of these two variables. It is not an appropriate practice to use covariates that may be influenced by treatment when building a PS model. PS models that are incorrectly specified by including post-treatment covariates have been shown to lead to improper PS estimation and biased results.^{28,29} Histograms of propensity scores for each of the treatment groups within the matched sample are shown to demonstrate overlap. A table displaying the summary statistics of important covariates for each treatment group within the matched sample presents evidence of balance; however, *p*-values are presented, which is not recommended.

The control patients used in the single-center Brown (2015b) study were drawn from the same time and population as the treated patients, protecting against possible selection bias. All independent variables included in the PS model were selected *a priori* and were either measured before PHT or were not influenced by the administration of PHT. Furthermore, the paper presents the most detailed balance assessment of all studies considered for review, including standardized difference to determine if relevant variables were adequately balanced after matching. This method has been recommended by experts and is preferable to relying on *p*-values from statistical tests.^{23,25,30} This study most closely adheres to guidelines established by PS theory.

In the 2016 study conducted by Miller *et al*, both the treatment and the control samples were drawn from Vanderbilt University Medical Center (VUMC) over the same time period.¹³ Numerous air medical transport services bring patients to VUMC, and some have the ability to administer PHT while others do not. Variables selected for inclusion into the propensity model were those used by medical personnel to determine whether to administer PHT. However, the authors also included the amount of in-hospital blood administered within 24 hr of admission in the PS model. As this is a post-treatment variable likely to be influenced by treatment itself, it should not be

included in the propensity model. The authors created a 1:1 matched sample and performed conditional logistic regression to assess the outcomes. They assessed balance within the matched set, but did so using statistical tests and *p*-values.

PROHS and Guidelines for Future Studies

Reviewing the selected articles informed the design and analysis plan for PROHS. All nine participating sites agreed not to change their prehospital resuscitation protocol for the duration of the study. As it was not possible to randomize treatment in PROHS, it was believed that having centers without the ability to administer PHT would enroll an untreated sample in which some patients had a high propensity for receiving PHT during transport. Unfortunately, key variables predicting treatment (including ISS, prehospital Glasgow coma scale [GCS], and prehospital SBP) differed drastically between sites with and without blood available. This level of heterogeneity between sites ultimately proved impossible to control for. As some sites were not able to administer PHT, we included site volume as a continuous variable in the PS model rather than a site indicator and utilized generalized estimating equations, with a random effect for site, in the outcome model.⁵ Balance assessment in PROHS avoided statistical tests and included density plots, illustrating the study's issue with lack of overlap.

DISCUSSION

The objective of the current study is to review observational PHT studies and determine how closely their statistical methods met the assumptions that PS theory requires. Two of the five retrospective studies reviewed, as well as PROHS, met all of the PS assessment criteria.^{5,9,10} Even though the other articles did not explicitly meet all of the established criteria, their contribution should not be diminished, as they set the stage for randomized studies.

The most common issue was not adequately addressing balancing techniques and using *p*-values to assess balance. We recommend that overlap and balance diagnostics always be included in the publication (without *p*-values) and furthermore should be a condition for publication when reviewing submitted articles utilizing PS. However, it is also possible that statistical details were removed in the revision and editing, possibly due to journal requirements. Another common issue was utilizing post-treatment variables in predicting the probability of treatment (the PS model). Although this is not an appropriate practice, in general, collecting a variable's value after treatment is administered does not necessarily mean that it was influenced by treatment. Permanent patient characteristics such as age, or gender, along with injury-related variables such as mechanism of injury, or injury severity score may be recorded later and included in the propensity model as they are not affected by treatment assignment.

One limitation of the current study is that so few articles exist that meet the inclusion criteria of the systematic review. It should also be noted that three of the six articles had the

same lead author and are drawn from the same database. These factors limit the amount of variation observed in the propensity scoring methods implemented. However, the small existing body of research using propensity scoring to assess effectiveness of PHT also enabled this review to capture all of the relevant studies. As PHT is a treatment that appears to be growing in popularity, it is helpful to provide uniform guidance for designing these studies. Future research in this area could include developing guidelines for interim monitoring techniques to assess balance throughout the course of a pre-planned prospective observational study and provide remedial action if unexpected issues arise.

Finally, we recognize the difficulty in planning and implementing novel studies where there is limited specific reference material available. Our critiques of the selected studies are not meant to diminish their authors' work or suggest that their results are invalid. Instead, we hope that this manuscript encourages best practice in future studies and provides guidance for researchers desiring to enhance trauma care in both civilian and military populations.

FUNDING

The Prehospital Resuscitation on Helicopter Study (PROHS) was sponsored by the U.S. National Heart, Lung, and Blood Institute (U01HL077863) and the U.S. Department of Defense. Additional author funding provided by: National Institute of General Medical Sciences T32 GM074902.

PRESENTATIONS

Presented as a poster at the 2016 Military Health System Research Symposium (abstract number: MHSRS-16-0320).

ACKNOWLEDGMENTS

Prehospital Resuscitation on Helicopter Study (PROHS) Study Group

Clinical Coordinating Center: John B. Holcomb, MD; Charles E. Wade, PhD; Erin E. Fox, PhD; Jeanette M. Podbielski, RN; Jeffrey S. Tomasek, MD; and Deborah J. del Junco, PhD. Data Coordinating Center: Michael D. Swartz, PhD; Stacia M. DeSantis, PhD; Savitri N. Appana, MS; Thomas J. Greene, PhD; Misung Yi, MS; Michael O. Gonzalez, MS; and Sarah Baraniuk, PhD. Resuscitation Outcomes Consortium at the University of Washington: Gerald van Belle, PhD; and Brian G. Leroux, PhD.

PROHS Clinical Sites (listed in the order of number of highest risk patients enrolled)

University of Texas Health Science Center at Houston: Carrie L. Howard, MA, MBA; and Amanda Haymaker. Shock, Trauma and Anesthesiology Research – Organized Research Center (STAR-ORC), R Adams Cowley Shock Trauma Center, University of Maryland Medical Center: Deborah M. Stein, MD, MPH; Thomas M. Scalea, MD; Benjamin Ayd; Pratik Das; and Anthony V. Herrera, MS. University of Washington: Eileen M. Bulger, MD; Bryce R. H. Robinson, MD; Patricia Klotz, RN; and Anika Minhas, BS. University of Alabama at Birmingham: Jeffrey D. Kerby, MD, PhD; Sherry M. Melton, MD, MSHA; Carolyn R. Williams, RN, MSHI; and Shannon W. Stephens, EMTP. University of Cincinnati: Michael Goodman, MD; Jay A. Johannigman, MD; Jason McMullan, MD; Richard D. Branson, MSc, RRT; Dina Gomaa, BS, RRT; and Christopher Barczak, BS, MT(ASCP). Oregon Health and Science University: Martin A. Schreiber, MD; Samantha J. Underwood, MS; and Cheri Watson, BS. Mayo Clinic: Martin D. Zielinski, MD; James R. Stubbs, MD; and Amy Headlee. University of Arizona: Terence O'Keefe, MBChB, MSPH; Peter Rhee, MD;

Laurel L. Rokowski, RN, BSN, MKT; John Santoro Jr, AA; Andrea Seach, BS; David Bradford, BS; Michelle Fealk, BA; and Fortesa Latifi, BS. University of Southern California; Kenji Inaba, MD; Henry Kim, MD; Carl Chudnofsky, MD; and Monica D. Wong, MS.

REFERENCES

1. Baraniuk S, Tilley BC, Del Junco DJ, et al: Pragmatic randomized optimal platelet and plasma ratios (PROPPR) trial: design, rationale and implementation. *Injury* 2014; 45(9): 1287–95.
2. Holcomb JB, Tilley BC, Baraniuk S, et al: Transfusion of plasma, platelets, and red blood cells in a 1: 1: 1 vs a 1: 1: 2 ratio and mortality in patients with severe trauma: the PROPPR randomized clinical trial. *JAMA* 2015; 313(5): 471–82.
3. Rosenbaum PR, Rubin DB: The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70(1): 41–55.
4. Rubin DB: Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 1974; 66(5): 688–701.
5. Holcomb JB, Swartz MD, DeSantis SM, et al: Multicenter observational prehospital resuscitation on helicopter study (PROHS). *J Trauma Acute Care Surg* 2017; 83(1): S83–91.
6. Kauvar DS, Lefering R, Wade CE: Impact of hemorrhage on trauma outcome: an overview of epidemiology, clinical presentations, and therapeutic considerations. *J Trauma Acute Care Surg* 2006; 60(6): S3–11.
7. Teixeira PG, Inaba K, Hadjizacharia P, et al: Preventable or potentially preventable mortality at a mature trauma center. *J Trauma Acute Care Surg* 2007; 63(6): 1338–47.
8. Kelly JF, Ritenour AE, McLaughlin DF, et al: Injury severity and causes of death from Operation Iraqi Freedom and Operation Enduring Freedom: 2003–2004 versus 2006. *J Trauma Acute Care Surg* 2008; 64(2): S21–7.
9. Brown JB, Cohen MJ, Minei JP, et al: Goal directed resuscitation in the prehospital setting: a propensity adjusted analysis. *J Trauma Acute Care Surg* 2013; 74(5): 1207–14.
10. Brown JB, Cohen MJ, Minei JP, et al: Pretrauma center red blood cell transfusion is associated with reduced mortality and coagulopathy in severely injured patients with blunt trauma. *Annals Surg* 2015; 261(5): 997–1005.
11. Brown JB, Sperry JL, Fombona A, Billiar TR, Peitzman AB, Guyette FX: Pre-trauma center red blood cell transfusion is associated with improved early outcomes in air medical trauma patients. *J Am Coll Surg* 2015; 220(5): 797–808.
12. O'Reilly DJ, Morrison JJ, Jansen JO, Apodaca AN, Rasmussen TE, Midwinter MJ: Prehospital blood transfusion in the en route management of severe combat trauma: a matched cohort study. *J Trauma Acute Care Surg* 2014; 77(3): S114–20.
13. Miller BT, Du L, Krzyzaniak MJ, Gunter OL, Nunez TC: Blood transfusion: in the air tonight? *J Trauma Acute Care Surg* 2016; 81(1): 15–20.
14. MacDonald RD: Articles that may change your practice: prehospital blood products. *Air Med J* 2015; 34(6): 317–9.
15. Austin PC: Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: a systematic review and suggestions for improvement. *J Thoracic Cardiovasc Surg* 2007; 134(5): 1128–35.
16. Yanovitzky I, Zanutto E, Hornik R: Estimating causal effects of public health education campaigns using propensity score methodology. *Eval Program Plann* 2005; 28(2): 209–20.
17. Austin PC: An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011; 46(3): 399–424.
18. Guo S, Fraser MW: Propensity Score Analysis: Statistical Methods and Applications, pp 134–135. Edited by Knight V, Guarino K, McDuffee Y, Palermi S, Dickens G Thousand Oaks, CA, Sage Publications, 2014.
19. Austin PC: The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. *Med Decis Making* 2009; 29(6): 661–77.
20. Imai K, King G, Stuart EA: Misunderstandings between experimentalists and observationalists about causal inference. *J R Stat Soc [Ser A]* 2008; 171(2): 481–502.
21. Ho DE, Imai K, King G, Stuart EA: Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Polit Anal* 2007; 15(3): 199–236.
22. Harder VS, Stuart EA, Anthony JC: Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychol Methods* 2010; 15(3): 234–49.
23. Austin PC, Grootendorst P, Anderson GM: A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Stat Med* 2007; 26(4): 734–53.
24. Flury BK, Riedwyl H: Standard distance in univariate and multivariate analysis. *Am Stat* 1986; 40(3): 249–51.
25. Rubin DB: Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Serv Outcomes Res Methodol* 2001; 2(3): 169–88.
26. Inflammation and the Host Response to Injury Large Scale Collaborative Program. The Inflammation and Host Response to Injury Investigators: Glue Grant Clinical Protocols. National Institute of General Medical Sciences. Available at <http://www.gluegrant.org/clinical-protocols.htm>; accessed March 10, 2017.
27. Berkow L: Factors affecting hemoglobin measurement. *J Clin Monit Comput* 2013; 27(5): 499–508.
28. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T: Variable selection for propensity score models. *Am J Epidemiol* 2006; 163(12): 1149–56.
29. Drake C: Effects of misspecification of the propensity score on estimators of treatment effect. *Biometrics* 1993; 49(4): 1231–6.
30. McCaffrey DF, Ridgeway G, Morral AR: Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychol Methods* 2004; 9(4): 403–25.