Inverse Probability of Treatment Weighting (Propensity Score) using the Military Health System Data Repository and National Death Index

Joshua D. Mitchell¹, Brian F. Gage², Nicole Fergestrom³, Eric Novak¹, Todd C. Villines⁴

¹Cardiovascular Division, Washington University School of Medicine, St. Louis, MO, USA

²General Medical Sciences, Washington University School of Medicine, St. Louis, MO, USA

³Center for Advancing Population Science, Medical College of Wisconsin, Milwaukee, WI, USA

⁴Cardiology Service, Department of Medicine, Walter Reed National Military Medical Center, Bethesda, MD, USA

Abstract

When randomized controlled trials are not feasible, retrospective studies using big data provide an efficient and cost-effective alternative, though they are at risk for treatment selection bias. Treatment selection bias occurs in a non-randomized study when treatment selection is based on pre-treatment characteristics that are also associated with the outcome. These pre-treatment characteristics, or confounders, can influence evaluation of a treatment's effect on the outcome. Propensity scores minimize this bias by balancing the known confounders between treatment groups. There are a few approaches to performing propensity score analyses, including stratifying by the propensity score, propensity matching, and inverse probability of treatment weighting (IPTW). Described here is the use of IPTW to balance baseline comorbidities in a cohort of patients within the US Military Health System Data Repository (MDR). The MDR is a relatively optimal data source, as it provides a contained cohort in which nearly complete information on inpatient and outpatient services is available for eligible beneficiaries. Outlined below is the use of the MDR supplemented with information from the national death index to provide robust mortality data. Also provided are suggestions for using administrative data. Finally, the protocol shares an SAS code for using IPTW to balance known confounders and plot the cumulative incidence function for the outcome of interest.

SUMMARY:

When randomized controlled trials are not feasible, a comprehensive health care data source like the Military Health System Data Repository provides an attractive alternative for retrospective

Corresponding Author: Joshua D. Mitchell (jdmitchell@wustl.edu), Tel: 314-494-3124.

Disclaimer: Additionally, the views expressed in this article are those of the author only and should not be construed to represent in any way those of the United States Government, the United States Department of Defense (DoD), or the United States Department of the Army. The identification of specific products or scientific instrumentation is considered an integral part of the scientific endeavor and does not constitute endorsement or implied endorsement on the part of the author, DoD, or any component agency.

DISCLOSURES:

The authors have nothing to disclose.

analyses. Incorporating mortality data from the national death index and balancing differences between groups using propensity weighting helps reduce biases inherent in retrospective designs.

Keywords

inverse probability of treatment weighting; propensity score; national death index; Military Health System Data Repository; bias; treatment selection; confounding; big data

INTRODUCTION:

Randomized, placebo-controlled trials are the strongest study design to quantify efficacy of treatment, but they are not always feasible due to cost and time requirements or a lack of equipoise between treatment groups¹. In these instances, a retrospective cohort design using large-scale administrative data (“big data”) often provides an efficient and cost-effective alternative, though the lack of randomization introduces treatment selection bias². Treatment selection bias occurs in non-randomized studies when the treatment decision is dependent on pre-treatment characteristics that are associated with the outcome of interest. These characteristics are known as confounding factors.

Because propensity scores minimize this bias by balancing the known confounders between treatment groups, they have become increasingly popular³. Propensity scores have been used to compare surgical approaches⁴ and medical regimens⁵. Recently, we have used a propensity analysis of data from the United States Military Health System Data Repository (MDR) to assess the effect of statins in primary prevention of cardiovascular outcomes based on the presence and severity of coronary artery calcium⁶.

The MDR, utilized less frequently than the Medicare and VA data sets for research purposes, contains comprehensive administrative and medical claims information from inpatient and outpatient services provided for active duty military, retirees, and other Department of Defense (DoD) healthcare beneficiaries and their dependents. The database includes services provided worldwide at US military treatment facilities or at civilian facilities billed to the DoD. The database includes complete pharmacy data since October 1, 2001. Laboratory data is available from 2009 but is only limited to military treatment facilities. Within the MDR, cohorts have been defined with methods including use of diagnoses codes (e.g., diabetes mellitus⁷) or procedure codes (e.g., arthroscopic surgery⁸). Alternatively, an externally defined cohort of eligible beneficiaries, such as a registry, can be matched to the MDR to obtain baseline and follow-up data⁹. Unlike Medicare, the MDR includes patients of all ages. It is also less biased towards males than the VA database since it includes dependents. Access to the MDR is limited, however. Generally, only investigators that are members of the Military Health System can request access, analogous to requirements for use of the VA database. Non-government researchers seeking access to Military Health Systems data must do so through a data sharing agreement under the supervision of a government sponsor.

When using any administrative data set, it is important to bear in mind the limitations as well as strengths of administrative coding. The sensitivity and specificity of the code can vary based on the related diagnosis, whether it is a primary or secondary diagnosis, or whether it

is an inpatient or outpatient file. Inpatient codes for acute myocardial infarction are generally accurately reported with positive predictive values over 90%¹⁰, but tobacco use is often undercoded¹¹. Such undercoding may or may not have a meaningful effect on a study's results¹². Additionally, several codes for a given condition may exist with varying levels of correlation to the disease in question¹³. An investigative team should perform a comprehensive literature search and review of the International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM) and/or ICD-10-CM coding manuals to ensure that the appropriate codes are included in the study.

Several methods can be employed to improve the sensitivity and accuracy of the diagnostic codes to define comorbid conditions. An appropriate "look-back" period should be included to establish baseline comorbidities. The look-back period includes the inpatient and outpatient services provided prior to study entry. A period of one year may be optimal¹⁴. Additionally, requiring two separate claims instead of a single claim can increase specificity, while supplementing coding data with pharmaceutical data can improve sensitivity¹⁵. Select manual chart audits on a portion of the data can be used to verify accuracy of the coding strategy.

Once comorbidities have been defined and assessed for the cohort in question, a propensity score may be used to balance differences in covariates between treatment groups. The propensity score is derived from the probability that a patient is assigned to a treatment based on known covariates. Accounting for this propensity treatment reduces the effect that the covariates have on treatment assignment and helps generate a truer estimate of the treatment effect on the outcome. While propensity scores do not necessarily provide superior results to multivariate models, they do allow for assessment of whether the treated and untreated groups are comparable after applying the propensity score³. Study investigators can analyze the absolute standardized differences in covariates before and after propensity matching or inverse probability of treatment weighting (IPTW) to ensure known confounders have been balanced between groups. Importantly, unknown confounders may not be balanced, and one should be aware of the potential for residual confounding.

When executed properly, though, propensity scores are a powerful tool that can predict and replicate results of randomized controlled trials¹⁶. Of the available propensity-score techniques, matching and IPTW are generally preferred¹⁷. Within IPTW, patients are weighted by their propensity or probability for treatment. Stabilizing weights are generally recommended over raw weights, while trimming of the weights can also be considered^{18–21}.

Once study groups are balanced, they may be followed until the outcome of interest. Studies utilizing administrative data may be interested in outcomes such as readmission rates and time-to-event analyses. In studies interested in mortality, the Military Health System Data Repository includes a field for vital status that can be further augmented using the national death index (NDI)^{22,23}. The NDI is a centralized database of death record information from state offices that is managed by the Center for Disease Control. Investigators can request basic vital status and/or specific cause of death based on the death certificate.

The following protocol details the process of conducting an administrative database study using the MDR augmented with mortality information from the NDI. It details the use of IPTW to balance baseline differences between two treatment groups including SAS code and example output.

PROTOCOL:

The following protocol follows the guidelines of our institutional human ethics committees.

1. Defining the cohort

1.1. Determine and clearly define the inclusion and exclusion criteria of the planned cohort using either 1) a registry or 2) data points that can be extracted from the MDR such as administrative codes for diagnoses or procedures (i.e., all patients with more than two outpatient diagnoses or one inpatient diagnosis of atrial fibrillation).

1.1.1. If using a registry, include two or more patient identifiers for accurate matching with the Military Health System Data Repository such as medical record number (listed in different data sets as patuniq and edipn), full name, date of birth, and/or sponsor's social security number.

NOTE: As with all studies utilizing personal health information, safeguards are required and must be adhered to. Proper encryption and data management must be employed during the collection process, and information should be de-identified as soon as possible.

NOTE: When referencing the sponsor's social security number (sponssn), all patients are listed with regard to their relationship to the military member (or sponsor), including an identifier for the sponsor, spouse, and children. Be aware that the relationship code and sponsor's social security number may change over time in the data set when patients become adults and get married or divorced. Thus, multiple patient identifiers help to ensure accuracy.

1.1.2. If defining the cohorts through administrative coding, perform a comprehensive literature search to identify prior studies that have potentially validated the codes of interest. Review ICD-9-CM²⁴ and/or ICD-10-CM²⁵ manuals to clarify code definitions and neighboring codes to ensure the appropriate range of codes is being used. Additionally, review the cross-reference tables included in the manuals for consideration of additional codes for inclusion/exclusion. Prior validation studies contain reports of positive predictive value, sensitivity and specificity for various administrative coding strategies. These aid in optimization of cohort selection as well as outcome identification.

1.2. Determine if there are restrictions (e.g., based on age) on the desired cohort or other exclusion criteria to include in the data request.

1.3. Define the study period to include time prior to index date for collection of baseline covariates (generally 12 months in administrative data research) as well as study end date.

2. Defining covariates and outcomes

- 2.1. Define administrative codes for confounding conditions through literature searches and use of the ICD-9-CM²⁴ and/or ICD-10-CM²⁵ manuals as done in step 1.1.2 above.
- 2.2. Determine other necessary covariates including demographics, medication, and laboratory data.
- 2.3. Review available data fields in the MDR Data Dictionary here: <https://health.mil/Military-Health-Topics/Technology/Support-Areas/MDR-M2-ICD-Functional-References-and-Specification-Documents>.

3. Submitting a request for the MDR

- 3.1. Obtain approval of the Institutional Review Board.
- 3.2. Complete a data sharing agreement application that can be found here: <https://health.mil/Military-Health-Topics/Privacy-and-Civil-Liberties/Submit-a-Data-Sharing-Application?type=All#RefFeed>. As part of the application, specify data fields and files being requested on the DRT Military Health System Data Repository (MDR) Extractions worksheet (linked from application form). Specify whether the team is requesting a data analyst supply the raw data or if the team will access the MDR directly. Further specify whether the request is for a one-time data pull or if regular pulls are requested daily, monthly or yearly.

NOTE: To obtain MDR data by any method, there must be a sponsor that is a government employee (active duty military or GS), who is usually a member of the investigator team.

- 3.3. If accessing the MDR directly, complete the “MDR Authorization Request Form” and “MDR CS 2875 Form” that can be found here: <https://health.mil/Military-Health-Topics/Technology/Support-Areas/MDR-M2-ICD-Functional-References-and-Specification-Documents>.

4. Accessing the MDR and extracting relevant data

- 4.1. If accessing the MDR directly, follow instructions for accessing and using the MDR including software requirements and example SAS programs that are available in the “MDR User’s Guide” and “MDR Functional Guide” found here: <https://health.mil/Military-Health-Topics/Technology/Support-Areas/MDR-M2-ICD-Functional-References-and-Specification-Documents>.

NOTE: Files are saved in SAS format and accessed through a unix shell generally using putty.exe as well as an ftp program. Knowledge of SAS is required.

- 4.2. For a helpful overview of the MDR setup, review the DOD Guide for DOD Researchers on using MHS Data <https://health.mil/Reference-Center/Publications/2012/10/10/Guide-for-DoD-Researchers-on-Using-MHS-Data>.

4.3. As done in step 2.3, review the MDR Data Dictionary for detailed information on all available data files <<https://health.mil/Military-Health-Topics/Technology/Support-Areas/MDR-M2-ICD-Functional-References-and-Specification-Documents>>.

NOTE: Not all data files include all patient identifiers for matching/merging. The data dictionary helps list identifiers that are available for each data file. The DOD ID number, also referred to as “patuniq” or “edipn”, is needed to extract pharmacy information, for instance. Having all appropriate patient identifiers in the data mining step is therefore important to ensure the ability to match all patient information across multiple years and multiple data sets. It is important to reiterate that inherent in research that involves PHI, strict adherence to data safeguarding procedures is required after acquiring necessary approval, and PHI should be destroyed after it is no longer needed.

4.4. Obtain necessary patient identifiers for the cohort by accessing the vm6 beneficiary data (Sep 2002–present) or pben file (Sep 2000–Sep 2002).

4.4.1. Use the macro below or a similar program to match vm6 data to the cohort file. In this case, the code can be used as written to find patient medical record numbers (MRNs) for a given patient social that is already in a cohort file. Use different variable names in the vm6 data draw and cohort files for patient names and birthdates to help check for errors later. To safeguard PHI, store the data with patient identifiers on the service node in the space provided as part of the data request (see MDR User’s Guide).

NOTE: MRNs are referred to as the DOD ID number, PATUNIQ or EDIPN in the MDR).

```
LIBNAME my '/hsm1/army-other/username/folder';

%macro patuniq (fy2,fm2); *extracts each monthly file and appends to
my.cohort_patuniq;
filename invm3 pipe "uncompress -c /mdr/pub/deers/detail/vm6ben/fy&fy2./
fm&fm2..txt.Z";
DATA vm6; *imports name and birthdt as well for later error checks;
infile invm3 lrecl=700 missover;
input
@9 SPONSSN $char9.
@21 DDS_CD $char2.
@24 BENSSN $char9.
@34 birthdt yymmdd8.
@201 last_name $char26.
@227 first_name $char20.
@495 PATUNIQ $char10.;
if patuniq ne "";
format birthdt yymmdd10.;
RUN;
PROC SQL;
```

```

CREATE TABLE patuniq&fy2.&fm2. AS
SELECT *
FROM cohort as a, vm6 as b
WHERE a.benssn=b.benssn
;
QUIT;

*adds each file to permanent file in library;
proc append data = patuniq&fy2.&fm2. base=my.cohort_patuniq force;
run;

%mend patuniq;

*cohort file with limited variable set for matching to MDR data;
data cohort (keep = benssn fmp_ssn dob lastname firstname);
set my.complete_cohort;
run;

*calls macro for fiscal year 2015 (October 2014-September 2015);
%dodid(fy2=15, fm2=02) /*November 2014*/
%dodid(fy2=15, fm2=01) /*October 2014*/

```

4.4.2. As database entries are never completely free of error, perform error checks after each major step in addition to checking the program log and output for any potential concerns. Use the data step below to review potential mismatches with the code above (patient files are matched based on the patient/beneficiary social). When comparing names from the cohort file (lastname, firstname) with the vm6 file (last_name, first_name), only match on the first three letters to reduce false errors that arise with differences in spelling/spacing between files.

```

Data checkname;
set my.cohort_patuniq;
    if UPCASE(SUBSTR(lastname,1,3)) NE UPCASE(SUBSTR(last_name,1,3)) and dob
ne birthdt then output; *if last name and date of birth don't match;
    else if UPCASE(SUBSTR(firstname,1,3)) NE UPCASE(SUBSTR(first_name,1,3))
then output; *if first names don't match;
run;

```

4.4.3. Review error data file ("checkname"). Ignore errors caused by punctuation (O'Reilly vs. OReilly). Check other errors of concern with manual review of the health record or consider discarding relevant patient/patient information if significant errors exist and if verification is not possible.

4.5. Extract the remaining needed data from the MDR.

4.5.1. If needed, obtain race and sex from vm6ben files (pben files prior to September 2002), merge with the cohort file, and check for errors as done above:

```
filename vmben pipe "uncompress -c /mdr/pub/deers/detail/vm6ben/fy##/
fm##.txt.Z";
```

```
DATA vmbenssn;
infile vmben lrecl=800 missover;
input @24 patssn $char9. @;
if (patssn ne "") then do;
input
    @43 bensex $char1.
    @44 race $char1.
end;
else delete;
RUN;
```

4.5.2. Obtain death data from the death master file, merge with the cohort file, and check for errors as done above:

```
DATA death (keep = sponssn patssn dob dds dthdate source edipn);
SET "/mdr/pub/death/master/mpidth"; *may need to add .sas7bdat;
if patssn ne "";
format dob dthdate mmddyy10.;
RUN;
```

4.5.3. Obtain additional data files needed for analysis (see MDR Functional User's Guide for data location and additional helpful SAS macros and code).

NOTE: Data is stored in separate files depending on whether it was directly provided by military health care system or delivered elsewhere and billed to the military health care system. Example files are shown below.

CAPER – direct care, outpatient files from fy 2004–present

SADR – direct care, outpatient files from 1998–2005

SIDR – direct care, inpatient hospitalizations (direct care) from 1989–present

TEDI – billed care, institutional claim files fy 2001–present

HCSRI – billed care, institutional claims fy 1994–2005

TEDNI – billed care, non-institutional claims fy 2001–present

HCSRNI – billed care, non-institutional claims fy 1994–2005

PDTS – pharmacy file with individual prescriptions fy 2002–present

5. Merging data and constructing summative files

5.1. Whether data is obtained from a data analyst or obtained directly from the MDR as done in section 4 above, data files will need summated and merged together to form the analysis file. Throughout the process, utilize methods that improve data accuracy, including error checks and review of logs and output as also previously discussed.

5.1.1. When merging data, use at least two patient identifiers when possible to ensure a strong match (such as medical record number and date of birth), since errors can exist in any field. After the data merge, review the data to ensure expected results. Run the code to ensure that the first three letters of the name match in addition to another identifier or two is useful to verify proper matches (see step 4.5.1).

NOTE: The last name may not match if the patient was married during the time period in question. Minor variations may also exist in name fields due to apostrophes or spacing as well as typos.

5.1.2. Pay particular attention to matches at terminal steps in the process such as defining patients who had outcomes.

5.2. Extract baseline comorbidities using ICD-9-CM or ICD-10-CM codes from the period before the index date, the date the patient is considered as entering the study. Generally, use 12 months prior to index date to define comorbidities.

5.2.1. Ensure patients had eligibility for the military healthcare system during the baseline period (can be verified monthly in the vm6ben file).

5.2.2. Search baseline diagnosis codes in outpatient and/or inpatient files to establish baseline comorbidities during the baseline 12-month period prior to index date. Use ICD-9-CM or ICD-10-CM codes established in section 1. If using Elixhauser comorbidities, use available software from HCUP, making sure to modify the names of diagnosis variables and files as needed. (<https://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp#download>)

5.3. Search inpatient and/or inpatient files after the index date for outcomes of interest defined by ICD-9-CM or ICD-10-CM codes, such as hospitalization for myocardial infarction as primary diagnosis (search for 410.x1 in SIDR).

5.4. Set a study end date for all patients as a cutoff for follow-up for patients who have not demonstrated the outcome of interest. Determine which patients need to be censored prior to study end.

5.4.1. Search vm6ben file to ensure eligibility for healthcare through the study end date.; otherwise, censor the patient at the time of loss of eligibility.

5.4.2. If it is important to limit the study to active users of the healthcare system independent of eligibility, such as active users of the pharmacy, then determine the last health care contact (such as last medication fill) within the data files and censor the patients at that date.

NOTE: Be careful using telephone encounters, as they can be present in the health record after a death has occurred or if the beneficiary has exited the healthcare system in another manner.

6. Match to the national death index (NDI)

6.1. Once the full cohort is identified, send the information to the national death index for matching if mortality is an endpoint.

6.1.1. First, include the intent to match to the NDI in the requests for MDR data and IRB approval. Ensure approval and complete all data encryption steps completed before sending protected health information (PHI) to the NDI for matching.

6.2. The “National Death Index (NDI) Application Form” and directions for requesting death data from the National Death Index can be found here: <https://www.cdc.gov/nchs/ndi/index.htm>.

6.3. Send the data on a password-protected CD by overnight mail to the NDI. Results will be sent back approximately 2 weeks later in the same manner.

6.4. After receiving the NDI results, review partial matches for potential inclusion/exclusion.

6.4.1. “Chapter 4 - Assessing NDI Output” provides a helpful overview of reviewing results and can be found on the same webpage: <https://www.cdc.gov/nchs/ndi/index.htm>. Matches on social security number generally provide the strongest match.

6.4.2. When needed, cross-check deaths in the Social Security Death Index and/or Veterans Affairs Beneficiary Identification Records Locator Subsystem (BIRLS) to improve accuracy. Be aware that service members who die overseas will likely not show up on an NDI search but are often recognized in the MDR vital status file or in the VA BIRLS.

6.5. Merge the death file with main cohort file after completing the review.

7. De-identifying data

7.1. Once all necessary information is acquired, de-identify the data files to help protect PHI. Generate a random patient identifier for each patient using “ranuni” (see MDR Functional User’s Guide). Remove patient social, medical record number, date of birth (after computing age), etc., from data files. If needed (and approved), store a key that links the random patient identifier to the PHI securely on the SCE node.

8. Computing the propensity score^{18,19,26}

8.1. Use logistic regression to model the probability of treatment (proc logistic in SAS).

8.1.1. Specify the data file (“dat” in the example).

8.1.2. Use class statement to specify categorical variables. Use “ref = first” to specify the lowest value (such as 0) as the reference value.

8.1.3. In the model statement, specify the treatment variable as the dependent variable (Rx) and set the value for the “event” as the value for receiving treatment (1 in this case).

8.1.4. Include any possible predictors of receiving treatment as covariates in the model, especially if they could be predictors of the outcome (such as death). Consider if interactions between terms may impact treatment. Include them in the model by using an “*” (such as male*ckd) or use the syntax shown below placing “|” between covariates and “@2” at the end to specify all 2×2 interactions, as appropriate for the specific model.

8.1.5. Use the output statement to specify that the predicted probability of treatment (prob) will be defined by “ps” and output to the file “ps_data.”

```
Proc logistic data=dat;
Class male ckd liver diabetes copd chf cad cvd pad / ref = first;
Model Rx(event='1') = male | ckd | liver | diabetes | copd | chf |
cad | cvd | pad | age @2;

Output out=ps_data prob=ps;
Run;
```

NOTE: Variables in model: male: male sex (binary), ckd: chronic kidney disease (binary), liver: chronic liver disease (binary), diabetes (binary), copd: chronic obstructive pulmonary disease (binary), chf: heart failure (binary), cad: coronary artery disease (binary), cvd: cerebrovascular disease (binary), pad: peripheral arterial disease (binary), age (continuous).

8.2. Calculate weights from the predicted probability (propensity score). If the patient received treatment (Rx = 1), then the propensity score weight is $1/(\text{propensity score})$. If the patient did not receive treatment, then the propensity score weight is $1/(1 - \text{propensity score})$.

```
Data ps_data;
Set ps_data;
If Rx=1 then ps_weight=1/ps;
Else ps_weight =1/(1-ps);
Run;
```

8.3. Stabilize the propensity score by dividing it by the mean weight. In the code below, Proc means outputs the mean weight into the variable “mn_wt” in the data file “m.” The data set below that retains the mn_wt from data file “m” then computes the stabilized propensity scores (st_ps_weight) for each observation.

```
Proc means data=ps_data;
Var ps_weight;
Output out=m mean=mn_wt;
```

```

Run;

Data ps_data;
  If _n_=1 then set m;
  Retain mn_wt;
  Set ps_data;
  st_ps_wt = ps_weight/mn_wt;
Run;

```

8.4. Verify balancing after applying the inverse probability of treatment weighting.

8.4.1. The stddiff macro simplifies computing standardized mean differences for covariates before and after weighting in SAS. The code for the macro can be found here: <http://www.lerner.ccf.org/qhs/software/lib/stdiff.sas>.

8.4.2. Calculate the standardized mean difference before weighting. As with all macros, the macro code should be run in SAS prior to calling it. An example call statement is below with the covariates of interest.

```

%stdiff(inds=ps_data,
  Groupvar=Rx,
  Charvars=male ckd liver diabetes copd chf cad cvd pad,
  Numvars=age,
  StdFmt=8.4,
  Outds=result);

```

Inds - input data set, groupvar - variable that defines the study groups, charvars – categorical variables, numvars – continuous variables, stdfmt – format of standardized difference, outds – output data set.

8.4.3. Call the stddiff macro again to calculate the standardized mean difference after weighting. “Wtvar” specifies the variable containing the standardized propensity score and is added to the macro call statement. If the standardized differences are all less than or equal to 0.1, then the balancing is considered successful.

```

%stdiff(inds=ps_data,
  Groupvar=Rx,
  Charvars= male ckd liver diabetes copd chf cad cvd pad,
  Numvars=age,
  Wtvar=st_ps_wt,
  StdFmt=8.4,
  Outds=result);

```

8.5. The ASD before and after weighting can be reported in tabular or graph format. For directions for utilizing a SAS macro to generate a plot, please see the supplementary materials.

8.6. The IPTW-adjusted data can now be used in a univariate analysis after ensuring balancing of measured confounders.

9. Creating the outcome model and generating a plot of cumulative incidence function

9.1. There are a few ways that the resultant time-to-event analysis can be plotted, including using proc lifetest to generate a survival plot. Use the weight statement to indicate the standardized propensity weight.

9.2. To generate a cif plot using a propensity weight, use proc phreg.

9.2.1. In proc phreg, reference a covariate file to specify covariate values to be used when generating the plot. In this case, the covariate file only contains the single variable Rx, which can be 1 or 0.

```
data treat;
    Rx = 1; output;
    Rx = 0; output;
run;
```

9.2.2. Toggle ods graphics on. Use additional statements as needed to specify output files for the graph or file type (jpeg, etc.; see <https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_odsgraph_sect014.htm>).

```
ods graphics on;
```

9.2.3. In the proc phreg syntax, use the weight statement to specify the standardized propensity score variable. Specify values for baseline covariates using the baseline statement in order to be able to plot the cumulative incidence function. Specify the strata to use for the plot using “rowid” (in this case RX 1 vs. 0). The number in parentheses following the outcome variable (“event”) specifies the value(s) of the variable that should be censored which should include the censoring date and any competing events. In this case, 0 is censored and 1 is a true event.

```
proc phreg data=ps_data plots(overlay)=cif;
    class Rx / order=internal ref=first;
    model time*event(0) = Rx /eventcode = 1;
    weight st_ps_wt;
    baseline covariates=treat / rowid=Rx;
    Title "Treatment vs. No Treatment";
```

```
run;  
  
ods graphics off;
```

REPRESENTATIVE RESULTS:

Upon completion of IPTW, tables or plots of the absolute standardized differences can be generated using the `stddiff` macro code or the `asdplot` macro code, respectively. Figure 1 shows an example of appropriate balancing in a large cohort of 10,000 participants using the `asdplot` macro. After application of the propensity score, the absolute standardized differences were reduced significantly. The cutoff used for the absolute standardized difference is somewhat arbitrary, though 0.1 is often used and denotes negligible difference between the two groups. In a small cohort, proper balancing is more difficult to achieve. Figure 2 shows the unsuccessful results of attempting to balance covariates in a cohort of 100 participants.

Once the standardized propensity score is generated, the study team can proceed with outcome analysis. Survival analysis is often employed due to the need to censor participants with uneven follow-up information, and Figure 3 depicts an example of the use of `proc phreg` with standardized propensity score weights to generate a cumulative incidence function (CIF) plot. The CIF plot depicts the increasing number of events over time. In this case, the untreated, or control, group (No Rx) has a larger number of events and is comparatively worse than the treated group (Rx).

DISCUSSION:

Retrospective analyses using large administrative datasets provide an efficient and cost-effective alternative when randomized controlled trials are not feasible. The appropriate data set will depend on the population and variables of interest, but the MDR is an attractive option that does not have the age restrictions seen with Medicare data. With any data set, it is important to be intimately familiar with its layout and data dictionary. Care should be taken along the way to ensure that complete data are captured, and data are accurately matched and merged.

Codes for diagnoses should be defined using existing literature and a thorough understanding of the ICD-9-CM and ICD-10-CM coding system to maximize the value of the assigned diagnoses. Existing sets of comorbidity codes, including the Elixhauser²⁷ or refined Charlson comorbidity index^{28,29}, can be used to define comorbid conditions that may influence the outcome of interest. Likewise, validated coding algorithms in administrative data and should be leveraged. Validation should remain an area of active research, as there is continued learning on the optimal use of ICD-9-CM and ICD-10-CM coding algorithms to maximize accurate classification of a wide-range of diseases.

Propensity scores can be used to address the bias inherent in any retrospective analysis. Effective propensity score weighting or matching should reduce the absolute standardized difference (ASD) below the desired threshold, generally set at 0.1. Appropriate balancing helps ensure comparability of the treatment groups with regard to known confounders, and

appropriately employed propensity score techniques have been used to successfully replicate randomized trial results. Once properly balanced, the treatment groups can be compared with univariate time-to-event or other analysis.

Even with appropriate balancing, there is potential for residual confounding³, so the investigative team should limit the effect of unmeasured confounders. Additionally, if the effects of the covariates on treatment selection are strong, bias may still remain³⁰. In small cohorts, the propensity scores are unlikely to fully reduce the ASD below 0.1 for all variables and regression adjustment can be employed to help remove residual imbalance³¹. Regression adjustment can also be used in subgroup analysis when appropriate balance is no longer assured.

When done correctly, research with administrative data provides timely answers to important clinical questions in the absence of randomized clinical trials. While it is impossible to remove all bias from observational studies, bias can be limited by using propensity scores and remaining meticulous analyses.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS:

Research reported in this publication was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1 TR002345. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES:

1. Concato J, Shah N, Horwitz RI Randomized, controlled trials, observational studies, and the hierarchy of research designs. *New England Journal of Medicine*. 342 (25), 1887–1892 (2000). [PubMed: 10861325]
2. Austin PC, Platt RW Survivor treatment bias, treatment selection bias, and propensity scores in observational research. *Journal of Clinical Epidemiology*. 63 (2), 136–138 (2010). [PubMed: 20122500]
3. Sturmer T, Wyss R, Glynn RJ, Brookhart MA Propensity scores for confounder adjustment when assessing the effects of medical interventions using nonexperimental study designs. *Journal of Internal Medicine*. 275 (6), 570–580 (2014). [PubMed: 24520806]
4. Schermerhorn ML et al. Long-Term Outcomes of Abdominal Aortic Aneurysm in the Medicare Population. *New England Journal of Medicine*. 373 (4), 328–338 (2015). [PubMed: 26200979]
5. Williams RJ II et al. A Propensity-Matched Analysis Between Standard Versus Tapered Oral Vancomycin Courses for the Management of Recurrent *Clostridium difficile* Infection. *Open Forum Infectious Diseases*. 4 (4), (2017).
6. Mitchell JD et al. Impact of Statins on Cardiovascular Outcomes Following Coronary Artery Calcium Scoring. *Journal of the American College of Cardiology*. 72 (25), 3233–3242 (2018). [PubMed: 30409567]
7. Rush T, McGeary M, Sicignano N, Buryk MA A plateau in new onset type 1 diabetes: Incidence of pediatric diabetes in the United States Military Health System. *Pediatric Diabetes*. 19 (5), 917–922 (2018). [PubMed: 29446519]
8. Rhon DI, Greenlee TA, Marchant BG, Sissel CD, Cook CE Comorbidities in the first 2 years after arthroscopic hip surgery: substantial increases in mental health disorders, chronic pain, substance

abuse and cardiometabolic conditions. *British Journal of Sports Medicine*. 10.1136/bjsports-2018-099294 (2018).

9. Mitchell J, Paisley R, Moon P, Novak E, Villines T. Coronary Artery Calcium Score and Long-term Risk of Death, Myocardial Infarction and Stroke: The Walter Reed Cohort Study. *Journal of the American College of Cardiology: Cardiovascular Imaging*. (2017).
10. McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA. Validity of myocardial infarction diagnoses in administrative databases: a systematic review. *PLoS ONE*. 9 (3), e92286–e92286 (2014). [PubMed: 24682186]
11. Huo J, Yang M, Tina Shih Y-C. Sensitivity of Claims-Based Algorithms to Ascertain Smoking Status More Than Doubled with Meaningful Use. *Value in Health*. 10.1016/j.jval.2017.09.002 (2017).
12. Nayan M. et al. The value of complementing administrative data with abstracted information on smoking and obesity: A study in kidney cancer. *Canadian Urological Association Journal*. 11 (6), 167–171 (2017). [PubMed: 28652873]
13. Birman-Deych E. et al. Accuracy of ICD-9-CM codes for identifying cardiovascular and stroke risk factors. *Medical Care*. 43 (5), 480–485 (2005). [PubMed: 15838413]
14. Preen DB, Holman CD, Spilsbury K, Semmens JB, Brameld KJ. Length of comorbidity lookback period affected regression model performance of administrative health data. *Journal of Clinical Epidemiology*. 59 (9), 940–946 (2006). [PubMed: 16895817]
15. Rector TS et al. Specificity and sensitivity of claims-based algorithms for identifying members of Medicare+Choice health plans that have chronic medical conditions. *Health Services Research*. 39 (6 Pt 1), 1839–1857 (2004). [PubMed: 15533190]
16. Hernán MA et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology (Cambridge, Mass.)*. 19 (6), 766–779 (2008).
17. Austin PC. The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. *Medical Decision Making*. 29 (6), 661–677 (2009). [PubMed: 19684288]
18. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 11 (5), 550–560 (2000). [PubMed: 10955408]
19. Robins J. Marginal structural models. 1997 Proceedings of the American Statistical Association, section on Bayesian statistical science 1–10 (1998).
20. Thoemmes F, Ong AD. A Primer on Inverse Probability of Treatment Weighting and Marginal Structural Models. *Emerging Adulthood*. 4 (1), 40–59 (2016).
21. Xu S. et al. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. *Value in Health: the Journal of the International Society for Pharmacoeconomics and Outcomes Research*. 13 (2), 273–277 (2010). [PubMed: 19912596]
22. Cowper DC, Kubal JD, Maynard C, Hynes DM. A primer and comparative review of major US mortality databases. *Annals of Epidemiology*. 12 (7), 462–468 (2002). [PubMed: 12377423]
23. Skopp NA et al. Evaluation of a methodology to validate National Death Index retrieval results among a cohort of U.S. service members. *Annals of epidemiology*. 27 (6), 397–400 (2017). [PubMed: 28641759]
24. Buck CJ. 2015 ICD-9-CM for Hospitals, Volumes 1, 2, & 3, Professional Edition. (Elsevier Saunders, 2015).
25. Buck CJ. 2018 ICD-10-CM for Hospitals, Professional Edition. (Elsevier Saunders, 2018).
26. Guo S. & Fraser WM. Propensity Score Analysis: Statistical Methods and Applications, Second Edition (Sage Publications, 2015).
27. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Medical Care*. 36 (1), 8–27 (1998). [PubMed: 9431328]
28. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases*. 40 (5), 373–383 (1987). [PubMed: 3558716]

29. Deyo RA, Cherkin DC, Ciol MA Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *Journal of Clinical Epidemiology*. 45 (6), 613–619 (1992). [PubMed: 1607900]
30. Austin PC, Stuart EA The performance of inverse probability of treatment weighting and full matching on the propensity score in the presence of model misspecification when estimating the effect of treatment on survival outcomes. *Statistical Methods in Medical Research*. 26 (4), 1654–1670 (2017). [PubMed: 25934643]
31. Austin PC Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*. 28 (25), 3083–3107 (2009). [PubMed: 19757444]

Figure 1: Example of successful balancing.

In a large cohort ($n = 10,000$), IPTW achieved balancing of the covariates with all absolute standardized differences reducing to less than 0.1.

Figure 2: Example of unsuccessful balancing.

In a small cohort ($n = 100$), IPTW was unable to achieve balancing of the covariates with many absolute standardized differences remaining greater than 0.1.

Figure 3: Example of cumulative incidence function plot comparing treatment groups. Over time, the cumulative incidence of mortality increases in both groups, though it is higher in the untreated group (No Rx). Thus, in this example, the treated group has improved survival.