Out for blood: causal inference in clinical transfusion research

Valk, S.J.

6

Chapter 6

Clinical transfusion-outcomes research: A practical guide

Authors: Sarah J Valk[1,2], Camila Caram-Deelder[2], Rolf HHGroenwold[2], Johanna G van der Bom[2]

1. Jon J van Rood Center for Clinical Transfusion Research, Sanquin/LUMC, Leiden, The Netherlands,

2. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands,

ABSTRACT

Clinical transfusion research on the effectiveness and safety of blood products differs greatly from other medical research fields. There are three main intricacies which are specific to research of transfusion products: 1) patients frequently require more than one unit, 2) each unit originates from a different donor, and 3) the likelihood of receiving a unit with certain characteristics depends on a variety of external circumstances, that are commonly not within the control of the investigator.

This commentary addresses methodological challenges when investigating effectiveness and safety of blood products from observational data. As an example, we focus on the association between donor sex, pregnancy history of female donors and transfusion recipient mortality. We describe the current best methodological practices and illustrate statistical analyses using an example dataset, which allows other researchers to implement these practices in their own research.

Introduction

Clinical transfusion research aims to provide insight into the benefits and harms of transfusions. Evidence from observational studies is the predominant source of information about effects of donor and product characteristics of blood transfusions on patient outcomes, as it can be used to study various factors using large datasets. In contrast, evidence from randomized controlled trials often has a more limited, specific scope, e.g. a single threshold for comparing restrictive to liberal transfusion practices[1], or comparing fresh vs. older red blood cell transfusions based on a predefined cutoff[2, 3]. Moreover, in randomized trials follow-up is often limited in duration and sample size is relatively small to avoid unnecessary burden for participants and reduce cost. However, the randomized trial design is less suited to investigate research questions about blood product characteristics. Provided observational studies are designed and executed rigorously, such that potential bias is sufficiently mitigated, evidence from observational studies can complement the evidence based on randomized trials.[4, 5]

In this commentary, we shine the spotlight on methodological aspects related to confounding and selection bias of longitudinal observational data in clinical transfusion research. The goal of this commentary is to inform readers and researchers of such studies, to provide practical guidance and to encourage discussion about these important topics. Specifically, we *(1)* discuss intricacies of observational data of blood product characteristics, *(2)* present an overview of methods used in studies of blood product characteristics, *(3)* discuss these methods, including considerations for designing and analyzing clinical transfusion studies of donor and product characteristics, and *(4)* provide a tutorial for the use of marginal structural models in investigating transfusion exposures.

1. The challenges pertaining to blood product characteristics research

There are several specific challenges that contribute to the difficulty of studying efficacy and safety in the clinical transfusion setting. First, because every transfusion is linked to a specific donor, there is a wide variation in the pool of available blood products. Depending on the research question, particular products might be very common or very rare, potentially leading to limited statistical power. Second, patients are frequently exposed to multiple transfusions. Although restrictive transfusion practices have become more common, on average patients in the Netherlands receive two transfusions per transfusion episode, with more transfusions given depending on the indication.[6] Third, external factors (e.g. calendar time, patient blood group and geographic region) influence the probability of receiving a unit with any of these different characteristics. Last, the

existence of a possible bidirectional relationship between donor characteristics and patient outcomes is a recent insight that warrants increased scrutiny.[7]

Before addressing the different statistical analyses that could be applied to accommodate these challenges, we first introduce the epidemiological concepts that are important in research addressing the clinical outcomes of patients treated with blood transfusion. From the standpoint of modern causal inference, the concepts of confounding and selection bias are barriers in the estimation of a causal effect, that can be overcome by identifying the minimally sufficient adjustment set of covariates from a directed acyclic graph (DAG).[5] A causal DAG identifies which variables to adjust for, and which not, to be able to estimate a causal effect of the exposure of interest on the outcome. Drawing the DAG can be challenging, as transfusion exposure investigations are complex studies, involving longitudinal data, often including time-varying confounding and censoring of follow-up. In contrast to single timepoint interventions, or 'point treatments', transfusions are given over time and therefore conventional approaches to adjust for confounding might not be appropriate. The causal contrast of interest in these studies is commonly defined as initiating and adhering to the initial exposure assignment, that is, the characteristics of the first received transfusion. The exposure of interest is then compared to a chosen reference category. However, in longitudinal studies, intercurrent events need to be taken into account. As patients are exposed to multiple transfusions over time, they often do not solely receive the same exposure category throughout their follow-up. The question arises, what should be done with the follow-up from these 'cross-over' patients?

The answer was generally thought to be: to adjust for the time-varying cumulative number of transfusions by censoring the follow-up time of patients when they no longer adhere to their earlier exposure category. Because the number of transfusions is associated with the exposure (a particular product characteristic), and the outcome (mortality), the causal effect of exposure to the product characteristic of interest is estimated by adjusting for the cumulative number of transfusions received over time. Follow-up should be included using time-varying approaches, because patients who receive multiple transfusions are less likely to adhere to their initial transfusion exposure, and selecting only patients who adhered to their `assigned' exposure will therefore lead to bias.[5] Thus, rather than standard adjustment for covariates at baseline, control for confounding when time-varying confounding is present requires adjustment for time-varying covariates during follow-up of individual patients and censoring of follow-up at the time of non-adherence to the initial transfusion exposure category. However, depending on assumptions about the reasons for non-adherence to the initial

transfusion exposure category, more advanced statistical modelling techniques may be required. This is because, when non-adherence is both 1. affected by prior exposure and 2. informative of the outcome, traditional methods can fail, and consequently yield biased results. This phenomenon is known as treatment-confounder feedback, which is discussed in more detail in the next section.

2. Treatment-confounder feedback in studies of transfusion exposures

When time-varying confounders are affected by prior treatment, traditional methods (e.g. stratification, matching, outcome regression) are generally not suitable for confounding adjustment, as these may adjust away part of the effect of the exposure, yet also introduce a spurious association between exposure and outcome.[5] In studying any exposures that are tied to the subsequent probability of receiving additional transfusions, i.e. exposures associated with consistent product hemoglobin increment differences, this hence becomes a problem that can no longer be solved easily. We refer to this as treatment-confounder feedback by product hemoglobin content. This concept, previously described by Zhao et al.[7], is illustrated in Figure 1.

In Figure 1, panel A shows the partial DAG for the investigation of donor characteristics and mortality. The number of transfusions received over time (L) is associated with the probability of receiving female donor-only units (A) and the underlying disease severity (U) and is therefore part of the minimally sufficient adjustment set. Adjustment for L is required to estimate the effect of A on mortality (Y); this can be done using traditional methods (not shown) or g-methods (shown). With exposure to female donor-only units, however, comes a decreased hemoglobin 'dose' and therefore an increased need for additional transfusions (panel B). This can be illustrated by creating separate timepoints for treatment A and confounder L, thereby providing the complete DAG for this research question (panel C). This DAG shows that adjustment for L using traditional methods is not appropriate when the combined effect of A_t and A_{t+1} is of interest, as L is now located in the causal path of A_t on Y, in addition to being a confounder for the effect of A_{t+1} on Y. Alternative methods, such as g-methods (which include inverse probability of treatment weighting of marginal structural models, the parametric g-formula, and g-estimation of structural nested models[5]), are required here.

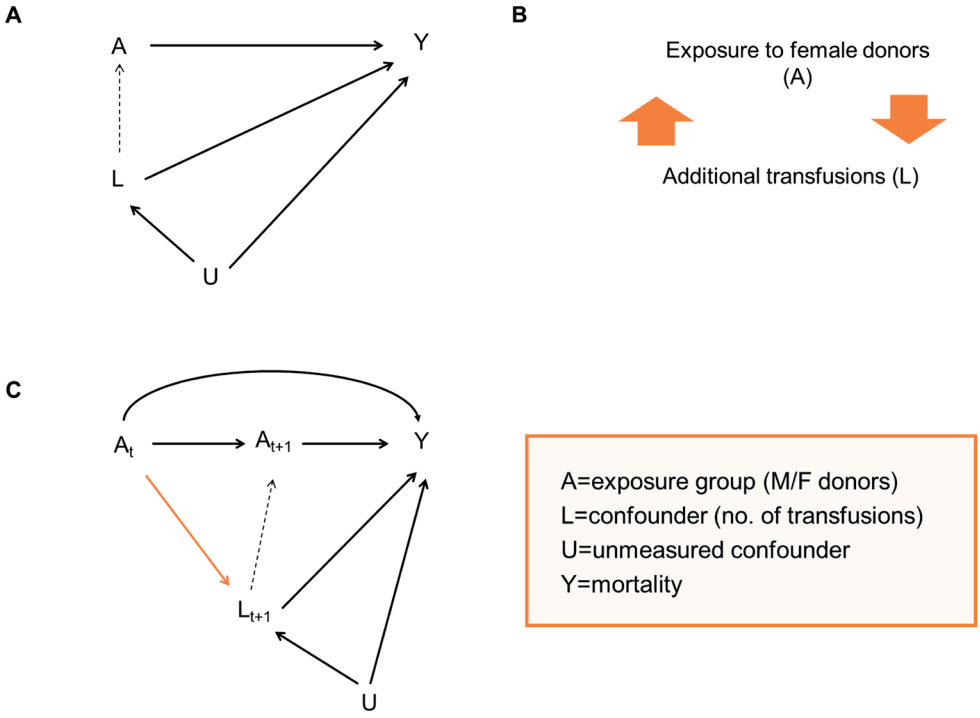


Figure 1. Different graphs to illustrate when advanced statistical modelling using g-methods is required.

A) Partial directed acyclic graph (DAG) of the effect of exposure to female donors (A) on mortality (Y) in transfusion recipients, confounded by unmeasured confounders (U, e.g. disease severity) through the cumulative number of transfusions (L). Dashed arrow represents the use of g-methods for the estimation of a causal effect of A on Y in the absence of treatment-confounder feedback, by removing the dependence of A on L.

B) Perceived bidirectionality if time is not taken into account, resulting in a cyclic graph, when assessing the effect of A on Y.

C) Complete DAG for the effect of exposure to female donor units including the treatment-confounder feedback over two timepoints (t, t+1) by lower hemoglobin concentration of units from female donors. Orange arrow represents the treatment-confounder feedback. Dashed arrow represents analysis using g-methods, removing the dependence of A_{t+1} on L, making estimation of the causal effect of A on Y possible in the presence of treatment-confounder feedback.

Specific situations where extra attention is expected to be warranted are the previously mentioned studies on donor sex, and pregnancy history of the donor. Also, storage duration of blood products can lead to smaller hemoglobin increments, and irradiation of red blood cell products would similarly require caution if chosen as exposure. Note that this is a non-exhaustive list, and researchers are encouraged to think carefully if their research question necessitates the use of alternative methods which can be used to estimate treatment effects in the presence of treatment-confounder feedback.

3. Appropriateness of methods applied in clinical transfusion research of product characteristics

Several statistical analysis methods have been applied in the field of transfusion product characteristics research (*Table 1*).

Table 1. Overview of methods used to study blood product characteristics as exposure

Methodology	Description of application in clinical transfusion research	Can handle treatment-confounder feedback	References
Traditional methods – restriction approach	Selection based on exposure classification at end of follow-up Stratification, matching, outcome regression (including propensity score regression adjustment and matching)	No	Middelburg, Alshalani[8, 9]
Traditional methods – time-varying approach	Exposure and confounder information modelled as time-varying variables Cox proportional hazards model with time-varying treatment and confounders	No	Caram-Deelder, Edgren[10-12]
G-methods – inverse-probability of censoring weighting	Time-varying exposure and confounder information used for reweighing population to mitigate bias due to treatment-confounder feedback Inverse probability-weighted marginal structural models	Yes	Zhao[7], MATER study (Chapter 5

Restriction approaches were employed, assessing the risk of exposure for groups of patients that were exposed to a single exposure type, without time-varying components.[8] This method is at risk of introducing selection bias, as the patients who only received one type of exposure throughout the follow-up period are inherently different from those who receive more transfusions, and are removed from the analysis because they did not adhere to their initial exposure at the start of follow-up. A more in-depth discussion of selection bias in cohort studies can be found elsewhere.[13] Specific for the clinical transfusion field, an example would be the selection of male-donor only and female-donor only exposure in a 'unisex' recipient cohort (i.e. selective inclusion of patients who received transfusions of single-sex donor origin Selection based on classification at the end of follow-up is not appropriate in the presence of treatment-confounder feedback, as this can lead to biased estimates of risk for transfusion characteristics.

Time-varying exposure and confounding adjustment has also been applied, with the number of units received with a specific (donor or product) characteristic included in the model as a continuous variable.[10-12] A potential pitfall in applying this method is the inclusion of continuous variables without properly taking into

account nonlinearity.[12, 14] Stratified Cox proportional hazards regression models with time-dependent exposures have recently been applied in this field.[11] The time-varying approach is not appropriate if there is treatment-confounder feedback, as it can lead to biased effect estimates.

Other possible analysis strategies include inverse probability of censoring weighting, to account for patients in certain exposure categories being more likely to receive additional transfusions and no longer being compliant to the initial blood product exposure, and therefore having to be censored.[7, 15]

Alternative approaches for defining exposure exist: binary exposure (coded as 0 or 1; female donor exposure vs. no female donor exposure) and full data utilization exposure (unexposed male vs. exposed male and female).[16] As these are either by design included in (or variations of) the methods presented here, we will not discuss them further. Lastly, methods reclassifying the exposure into a ratio have been proposed, but their complexity and computational intensity for survival data make them fall outside the scope of this commentary.[17]

4. Example dataset with applied methods illustrating that some approaches can lead to biased results

We applied the above-described methods to an example dataset to allow for a comparison of their performance in a semi-controlled setting. For this dataset, the study population consisted of male patients included in an earlier publication[10]. These male patients received transfusions in one of six included hospitals between 2005 and 2015. The complete exposure information was sourced from the Dutch municipality registration (see Chapter 5) to overcome the limitation of the original publication where 44% of the units donated by female donors had missing information about the pregnancy history. In Table S1, patient and blood product characteristics are described for this example dataset. Associations described in Table 2 apply to the patient population from the original, earlier publication and data were not altered or manipulated. This, opposed to the dataset for which the results are described in Table S2, which underwent an anonymization procedure that removed the empirical data, for the purpose of a publicly accessible tutorial.

Table 2. Results for different methods applied to the example dataset

Analysis	No. of Deaths	No. of Recipients*	HR (95% CI)
Restriction method			
Male (reference)	1,916	6,430	1 (reference)
Ever-pregnant female	207	770	1.22 (1.05-1.42)
Time-varying exposure and confounding adjustment method			
Male (reference)	1,916	10,901	1 (reference)
Ever-pregnant female	207	1,494	1.21 (1.04-1.41)
Inverse probability of censoring weighting method			
Male (reference)	1,916	10,901	1 (reference)
Ever-pregnant female	207	1,494	1.01 (0.85-1.20)

*Population included all male transfusion recipients that were identified in both datasets[10] with approx. 10% of patients not identified in the new dataset because of changes to the hospital administration records. HR hazard ratio; CI confidence interval

In Table 2, the risk for exposure to ever-pregnant donor-only units compared to the reference group of male-only unit exposure is presented for the three methods described in Section 2 applied to an example dataset. The inverse probability of treatment- and censoring-weighted analysis, estimating the average treatment effect of exposure to donors with a positive pregnancy history on mortality, is unbiased by treatment-confounder feedback present in the data (hazard ratio 1.01, 95% confidence interval 0.85-1.20). In contrast, the application of the time-varying adjustment method and restriction method give an estimate that is further away from 1, which is likely because of treatment-confounder feedback by hemoglobin increment differences between the two compared blood product exposures. In conclusion, statistical choices have considerable influence on the estimated hazard ratio in the investigation of blood product characteristics.

5. Tutorial for the application of marginal structural models as a way to estimate causal associations in the presence of treatment-confounder feedback

The use of inverse-probability weighted marginal structural models is not widespread in the field of clinical transfusion research, because their importance for studying transfusion exposures has not been recognized until recently. By providing an open-access example dataset with donor and patient characteristics, as well as concise R code, we hope to engage the scientific community, and encourage researchers to be more aware of the specific problems that arise when studying donor and product characteristics that relate to product hemoglobin content.

We provide a structured tutorial to perform the inverse probability of censoring weighting method described in Section 3 on a provided dataset (*Supplemental materials*, page 4). The dataset used in Section 4 is made available, after having applied an anonymization procedure to avoid sharing of personal patient data, and can be obtained by contacting the author (hyperlink to be added upon publication). The results for the inverse probability of censoring weighted analysis applied to the anonymized dataset can be found in Table S2. Because the original structure in the dataset was lost and treatment-confounder feedback is not present, all methods perform similar and are unbiased.

6. Conclusions

The importance of thorough epidemiological study design in clinical transfusion research cannot be overstated. In this commentary, recent insights about hemoglobin increments and their impact on blood product characteristics research were extensively discussed, and an overview including an appraisal of these methods was provided. As an example, we made use of a large observational dataset of transfusion and patient data. We applied several methods used in the past and present, from which inverse probability of censoring weighting should be considered in the presence of treatment-confounder feedback because this method can adequately account for time-varying confounding in the presence of such feedback. We also provide a detailed tutorial to guide those pursuing similar research.

Evidently, clinical transfusion outcomes research using observational data can be complex. Specifically for blood product characteristics research, these challenges include the adjustment for time-varying confounders, the censoring of follow-up time when mixed exposure occurs, and treatment-confounder feedback by product hemoglobin content. The appropriate statistical methodology can be difficult to identify, and especially when complex research questions are of interest, target trial emulation can provide useful insights. Target trial emulation, where a hypothetical randomized controlled trial is imagined and replicated with observational research, can be a useful tool to avoid both basic mistakes, and more complex analytical pitfalls.[18] With a target trial emulation approach, the hypothetical randomized trial is imagined which the observational study aims to mimic, including the specification of the start of follow-up, the exposure definition, the approach for how intercurrent events are handled, and so on. Of note, assumptions and decisions about the analysis are best specified up front, to avoid the problems associated with 'researcher degrees of freedom'.[19] When the aforementioned challenges are appropriately handled, it is possible to draw causal conclusions from observational transfusion data.

We emphasize that, while there are certainly limitations to several study designs used in the past, there is always a tradeoff between bias and precision where in some cases, a simpler method might be preferable. This can include the choice of changing the exposure of interest to single timepoint exposures, as opposed to sustained exposure over time. Researchers can and should give sufficient attention to the strengths and limitations of their chosen approach, and sensitivity analyses can be employed to test the impact of assumptions on the robustness of the estimate.

To conclude, we addressed the appropriateness of specific statistical methods in the presence of treatment-confounder feedback in the clinical transfusion research field and have provided guidance for future research. The suitability of any method depends on assumptions about the underlying causal relations in the data, and careful consideration about this is needed to ensure interpretations are valid.

Data availability statement

The original data used in this article and an earlier publication is available for inspection upon request. An anonymized dataset which can be used to run the provided syntax on will be made available in a repository upon publication. Anonymization was performed by random permutation.[20] Note: the original data structure is not completely retained following anonymization, but more advanced anonymization methods that can retain the original data structure have not yet been developed for survival analysis.[21]

Supplementary materials

The Supplementary materials contain the tutorial with syntax for use in R (*Supplementary materials*). Additional tables with results for the provided, anonymized dataset available from the repository are reported in Table S2.

REFERENCES

1. Carson JL, Stanworth SJ, Dennis JA, et al. Transfusion thresholds for guiding red blood cell transfusion. *Cochrane Database of Systematic Reviews*. 2021;(12) doi:10.1002/14651858.CD002042.pub5
2. Shah A, Brunskill SJ, Desborough MJ, Doree C, Trivella M, Stanworth SJ. Transfusion of red blood cells stored for shorter versus longer duration for all conditions. *Cochrane Database Syst Rev*. Dec 22 2018;12:Cd010801. doi:10.1002/14651858.CD010801.pub3
3. McQuilten ZK, French CJ, Nichol A, Higgins A, Cooper DJ. Effect of age of red cells for transfusion on patient outcomes: a systematic review and meta-analysis. *Transfusion Medicine Reviews*. 2018/04/01/ 2018;32(2):77-88. doi:https://doi.org/10.1016/j.tmrv.2018.02.002
4. Hernán MA. The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data. *American Journal of Public Health*. 2018;108(5):616-619. doi:10.2105/ajph.2018.304337
5. Hernán MA RJ. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2020:chap Chapter 20.
6. van Hoeven LR, Koopman MM, Koffijberg H, Roes KC, Janssen MP. Historical time trends in red blood cell usage in the Netherlands. *International journal of clinical transfusion medicine*. 2016;4:67-77.
7. Zhao J, Sjölander A, Edgren G. Mortality Among Patients Undergoing Blood Transfusion in Relation to Donor Sex and Parity: A Natural Experiment. *JAMA Internal Medicine*. 2022;182(7):747-756. doi:10.1001/jamainternmed.2022.2115
8. Middelburg RA, Briet E, van der Bom JG. Mortality after transfusions, relation to donor sex. *Vox Sang*. Oct 2011;101(3):221-9. doi:10.1111/j.1423-0410.2011.01487.x
9. Alshalani A, Uhel F, Cremer OL, et al. Donor-recipient sex is associated with transfusion-related outcomes in critically ill patients. *Blood Adv*. Jun 14 2022;6(11):3260-3267. doi:10.1182/bloodadvances.2021006402
10. Caram-Deelder C, Kreuger AL, Evers D, et al. Association of Blood Transfusion From Female Donors With and Without a History of Pregnancy With Mortality Among Male and Female Transfusion Recipients. *Jama*. Oct 17 2017;318(15):1471-1478. doi:10.1001/jama.2017.14825
11. Edgren G, Murphy EL, Brambilla DJ, et al. Association of Blood Donor Sex and Prior Pregnancy With Mortality Among Red Blood Cell Transfusion Recipients. *Jama*. Jun 11 2019;321(22):2183-2192. doi:10.1001/jama.2019.7084
12. Edgren G, Ullum H, Rostgaard K, et al. Association of Donor Age and Sex With Survival of Patients Receiving Transfusions. *JAMA Intern Med*. Jun 1 2017;177(6):854-860. doi:10.1001/jamainternmed.2017.0890
13. Hern, xe, n MA, et al. A Structural Approach to Selection Bias. *Epidemiology*. 2004;15(5):615-625.
14. Chasse M, Tinmouth A, English SW, et al. Association of Blood Donor Age and Sex With Recipient Survival After Red Blood Cell Transfusion. *JAMA Intern Med*. Sep 1 2016;176(9):1307-14. doi:10.1001/jamainternmed.2016.3324
15. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*. Sep 2000;56(3):779-88. doi:10.1111/j.0006-341x.2000.00779.x

16. Heddle NM, Cook RJ, Liu Y, et al. The association between blood donor sex and age and transfusion recipient mortality: an exploratory analysis. *Transfusion*. Feb 2019;59(2):482-491. doi:10.1111/trf.15011
17. Bruun-Rasmussen P, Andersen PK, Banasik K, Brunak S, Johansson PI. Estimating the effect of donor sex on red blood cell transfused patient mortality: A retrospective cohort study using a targeted learning and emulated trials-based approach. *eClinicalMedicine*. 2022;51doi:10.1016/j.eclinm.2022.101628
18. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol*. 2016;183(8):758-764. doi:10.1093/aje/kwv254
19. Simmons JP, Nelson LD, Simonsohn U. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*. 2011;22(11):1359-1366.
20. Ängquist L. Stata tip 92: Manual implementation of permutations and bootstraps. *The Stata Journal*. 2010;10(4):686-688.
21. Kroes SKS, van Leeuwen M, Groenwold RHH, Janssen MP. Generating synthetic mixed discrete-continuous health records with mixed sum-product networks. *Journal of the American Medical Informatics Association*. 2022;30(1):16-25. doi:10.1093/jamia/ocac184

Supplement: Clinical transfusion-outcomes research: A practical guide

Contents:

Supplemental methods

Inverse probability of censoring weighting method (IPW)

The dataset was organized as longitudinal survival data (with *t_begin* representing start of follow-up and *t_end* representing the end of follow-up for each patient row), for use in the *ipw* and *survey* package in R.[1] Initial follow-up is ordered as daily intervals for the first 28 days, followed by 4-week intervals ("blocks"). Weighted Cox proportional hazards models were fitted to correct for censoring and confounding.[1] Analyses were performed in R (version 3.6.3) and R Studio (version 2022.02.0+443) software.

The following variables were included in the multinomial logistic regression to estimate the baseline inverse probability of treatment weights: year of first transfusion exposure (*Transfusion_Year_first,* continuous), patient blood group (*Patient_ABORh*, categorical), hospital (*Hospital*, categorical). The outcome variable for the logistic regression was the categorical variable *Arm* (taking 0 if exposure was to the reference of male donors, 1 if exposure was to ever-pregnant female donors, and 9 if exposure was to other/mixed products).

The cumulative number of transfusions was included as the only covariate in the model for the generation of inverse probability of censoring weights (*Arm_Total_cum*), as a time-varying continuous variable. The outcome for this model was the censoring variable (*Censored*). Because patients could contribute multiple transfusion episodes, robust standard errors were used for the computation of

the confidence limits.[2] Only patients exposed to the reference arm (male, *Arm* taking the value 0) donors or the exposure arm (ever-pregnant female, *Arm* taking the value 1) were included in the estimation of censoring weights. Censoring weights were generated for the dataset weighted by the inverse probability of treatment weights generated earlier. Weights were plotted within strata of follow-up time to determine the distribution of the weights with *ipwplot*.

The resulting weights were multiplied to create the final weights. Truncation, or trimming, of the weights in case of extreme weights (e.g. >10) is optional. The spread of the weights was assessed by calculating the 0.5^{th} and 99.5^{th} percentiles of the weights.

If patients were censored or died in a block, they were interval-censored. The actual end of follow-up, the variable *t_end_new,* was then used to replace the block time *t_end* for use in the Cox proportional hazards model.

The weighted Cox proportional hazards model was specified with the exposure (*Arm*), the outcome (*Death*), the time variables (*t_begin*, *t_end)* and the final weights. Only uncensored lines (*Censored* = 0) were included in the model.

A detailed R code including all steps described above is available at the end of the Supplemental materials.

Time-varying exposure and confounding adjustment method

Cox proportional hazards models were fitted, adjusted for: cumulative number of transfusions (restricted cubic spline with three knots); hospital (categorical); blood group (categorical); calendar year (categorical); age of the donor (cumulative number of transfusions from donors aged >50 years, continuous); interaction term for cumulative number of transfusions and hospital.(see Chapter 5) Exposure is included as a binary, categorical variable.

This method is expected to be biased if treatment-confounder feedback is present due to limitations of traditional regression analysis. Analyses were performed in Stata, version 16 (StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC).

Restriction method

Similar to method described above, with the distinction that only patients who received transfusions from the same exposure category as the first, are included and Cox PH regression is performed without a time-varying component.

This method conditions on information from the future follow-up of the patients, and is also expected to lead to bias. Analyses were performed in Stata, version 16 (StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC).

Tutorial for use of IPW for transfusion-outcomes research in R

The below provided syntax can be used to perform an inverse probability of treatment- and censoring-weighted analysis(see Chapter 5) for blood product exposures related to hemoglobin increment raising capacity of the product, on a provided, anonymized dataset. Note that this dataset does not retain all original features of the real dataset, and the treatment-confounder feedback structure was lost due to the anonymization process. The provided anonymized dataset is a representative example of a dataset generated with random permutation of the variables *Arm* (exposure, assigned randomly from original distribution), *Hospital* (category for the hospital where the patient received the transfusion, assigned randomly in one of four categories from original distribution of six hospitals), *Patient_ABORh* (category of the blood group ABO and Rhesus type, assigned randomly from original distribution) and *Transfusion_Year_first* (year of the first transfusion of the patient, i.e. year of patient's start follow up, assigned randomly from original distribution). All other variables were kept identical to the original dataset.

Tutorial syntax in R:

The tutorial is organized as follows:

Step 0. Specify working directory and prepare files

Step 1. Inverse probability of treatment weights (IPTW) estimation with multinomial logistic regression

Step 2. Inverse probability of censoring weights (IPCW) estimation with weighted Cox regression

Step 3. Multiplication of weights (IPTW*IPCW) to create final weights

Step 4. IPW-corrected Cox model

Step 0

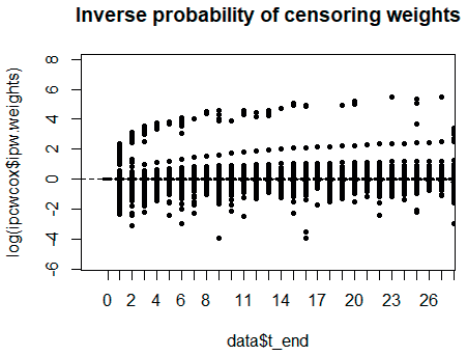
```
#Tutorial Clinical transfusion-outcomes research: A practical guide #required: file
= "Datafile-clinicaltransfusion.Rdata" (available upon request)
##############################################
#Tutorial
#Male patients only
#Comparison: Male (0) vs Ever-pregnant female (1)
#Variables in the dataset are:
```

```
#PIN: unique patient identifier.
#Arm: 0: control, patients whose first transfusion was donated by a male donor; 1: exposure, patients whose first transfusion was donated by a female donor who had been pregnant; 9: patients whose first transfusion was donated by blood donated by any other than exposure and control, i.e. female without history of pregnancy or sex of the donor unknown, and/or mixed exposure on day 1).
#Transfusion_Year_first:  year of the first transfusion of the patient, i.e. year of patient's start follow up).
#Patient_ABORh: patient blood group, category.
#Hospital: hospital name, category.
#Censored: censoring indicator (0 if patient received all transfusions from the same Arm group, 1 if patient no longer adhered to initial group assignment).
#Arm_Total_cum: cumulative number of transfusions, continuous.
#t_begin and t_end: time variables, each line refers to a single time period (t_begin refers to the start of the follow up, as required by the ipw package; all t_begin lines are rescheduled having -1 as reference; the first 28 days of follow up are included as one line per day and after day 28 the lines refer to blocks of 28 days).
#t_end_new: time variable, adjusted from block size (only blocks of 28 days are allowed) to real end of follow-up (individual days are allowed, e.g. if the patient has died at day 30, t_end_new would be '30', while t_end would be '56').
#Death: indicator for event at time t_end.
################################################
#step 0. specify working directory and prepare files
#install packages
#install.packages("ipw")
#install.packages("survival")
#install.packages("survey")
#install.packages("dplyr")
#load packages
library(ipw)
library(survival)
library(survey)
library(dplyr)
#set working directory: complete the path with the local where the datafile "Datafile-clinicaltransfusion.Rdata") is located
setwd("C:\\dir")
#clear workspace
rm(list=ls())
#load files
load(file= "Datafile-clinicaltransfusion.Rdata")
```

Step 1-4

```
################################################
#step 1. inverse probability of treatment weights (IPTW) estimation with multinomial
logistic regression
confounder_weight <- ipwpoint(exposure = Arm, family = "multinomial", numerator =
~1, denominator =~Transfusion_Year_first + Patient_ABORh + Hospital, data = data)
#OUTPUT
# weights: 6 (2 variable)
## initial value 925365.525208
## iter 10 value 345710.424406
## iter 10 value 345710.424400
## iter 10 value 345710.424397
## final value 345710.424397
## converged
## # weights: 42 (26 variable)
## initial value 925365.525208
## iter 10 value 420905.812229
## iter 20 value 388504.584486
## iter 30 value 361899.264984
## iter 40 value 349123.738592
## iter 50 value 345305.718723
## iter 60 value 345284.712165
## iter 70 value 345284.435350
## final value 345284.422337
## converged
data$iptwlogweights <- confounder_weight$ipw.weights
summary(data$iptwlogweights)
#OUTPUT
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.6927 0.9952 1.0009 1.0000 1.0066 1.3027
#selection of subset of exposed (ever-pregnant, F1: coded as 1) and reference (male,
M: coded as 0); excluding Unknown, F0 and mixed (coded as 9) after estimation of
iptwlogweights
data<-subset(data, Arm!=9)
################################################
#step 2. inverse probability of censoring weights (IPCW) estimation with weighted
Cox regression
#IPTCW is estimated in the population weighted by IPTW
ipcwcox <- ipwtm(
```

```
 exposure = Censored,
 family = "survival",
 numerator = ~ 1,
 denominator = ~ Arm_Total_cum ,
 id = PIN,
 tstart = t_begin,
 timevar = t_end,
 type = "first",
 data = data,
 weight = data$iptwlogweights)
data$ipcwcoxweights <- ipcwcox$ipw.weights
summary(data$ipcwcoxweights)
#OUTPUT
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.0039 0.8961 0.9788 1.0004 1.0004 2384.5971
#plot IPCW weights
ipwplot(weights = ipcwcox$ipw.weights, timevar = data$t_end,
 binwidth = 1, main = "Inverse probability of censoring weights" , xlim = c(0, 28))
#OUTPUT
```




```
#interpretation:
#weights are depicted for the first 28 days; the distribution of the weights is
balanced with the exception of some large weights. Weights are selected for only the
uncensored lines in step 3., leading to less extreme weights.
#preparation of data for IPW-corrected model
#selection of non-censored observations only to limit the model to follow-up time
eligible for analysis (Arm=0 or Arm=1)
data2<-subset(data, Censored!=1)
################################################
#step 3. multiplication of weights (IPTW*IPCW) to create final weights
```

```
data2$weights <- (data2$ipcwcoxweights*data2$iptwlogweights)
summary(data2$weights)
#OUTPUT
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.6753 0.8999 0.9664 0.9792 0.9995 60.2151
#store ranges of weights for assessment of extreme weights and weights distribution
min <- min(data2$weights)
max <- max(data2$weights)
pct005 <- quantile(data2$weights, c(.005))
pct995 <- quantile(data2$weights, c(.995))
#store extreme weights
extreme <- subset(data2, weights>10)
#truncate weights (optional: large weights lead to instability of the IPW estimator; truncation can reduce variance, but increase bias)
#data2["weights"][data2["weights"] >10] <- 10
#change t_end (to no longer be the 'block t_end', but the 'real t_end' from patient final follow-up date)
data2$t_end <- data2$t_end_new
################################################
#step 4. IPW-corrected Cox model
surveydesign1<-svydesign(id = ~ PIN, strata = ~ Arm, weights = ~ data2$weights, data = data2)
summary(svycoxph(Surv(t_begin, t_end, Death) ~ as.factor(Arm), design = survey-design1))
#OUTPUT
## Stratified 1 - level Cluster Sampling design (with replacement)
## With (12395) clusters.
## svydesign(id = ~PIN, strata = ~Arm, weights = ~data2$weights,
## data = data2)
## Call:
## svycoxph(formula = Surv(t_begin, t_end, Death) ~ as.factor(Arm),
## design = surveydesign1)
##
## n= 830334, number of events= 2297
##
## coef exp(coef) se(coef) robust se z Pr(>|z|)
## as.factor(Arm)1 0.01564 1.01576 0.06244 0.07283 0.215 0.83
##
## exp(coef) exp(-coef) lower .95 upper .95
```

```
## as.factor(Arm)1 1.016 0.9845 0.8806 1.172
##
## Concordance= 0.503 (se = 0.004 )
## Likelihood ratio test= NA on 1 df, p=NA
## Wald test = 0.05 on 1 df, p=0.8
## Score (logrank) test = NA on 1 df, p=NA
##
## (Note: the likelihood ratio and score tests assume independence of
## observations within a cluster, the Wald and robust score tests do not).
msm <- svycoxph(Surv(t_begin, t_end, Death) ~ as.factor(Arm), design = survey-
design1)
a <- exp(coef(msm))
b <- exp(confint(msm))
#store counts for Deaths/Recipients, by exposure (0/1)
n_distinct(data$PIN)
## [1] 12395
data00 <- subset(data2, Arm==0)
c <- n_distinct(data00$PIN) #Recipients 0, total
data01 <- subset(data2, Arm==0 & Death==1)
d <- n_distinct(data01$PIN) #Recipients 0, died
data10 <- subset(data2, Arm==1)
e <- n_distinct(data10$PIN) #Recipients 1, total
data11 <- subset(data2, Arm==1 & Death==1)
f <- n_distinct(data11$PIN) #Recipients 1, died
#create output table
Tutorialclinicaltransfusion <- data.frame(expcoef = a,
 confint = b,
 total0 = c,
 deaths0 = d,
 total1 = e,
 deaths1 = f,
 min = min,
 max = max,
 pct005 = pct005,
 pct995 = pct995,
 name = "Tutorialclinicaltransfusion")
#view output
View(Tutorialclinicaltransfusion)
```

Output

Supplemental results

The characteristics of the anonymized dataset are presented in Table S1.

Table S1. Patient and product characteristics for the anonymized dataset

Characteristics	Complete population	No-mixture subset*	Restriction subset†
Number of patients	N=18,206	N=13,361	N=7,659
Number of deaths, (%)	7,092 (39%)	2,234 (17%)	2,234 (29%)
Follow-up, median (IQR), days‡	1,819 (389-2,744)	341 (7-2,253)	2,051 (679-2,977)
Person-time, sum in years	87,382	42,999	41,107
Age of patients, median (IQR), years	65 (49-75)	65 (44-75)	64 (27-74)
0 to 17	2,754 (15%)	2,589 (19%)	1,796 (23%)
18 to 50	1,947 (11%)	1,287 (10%)	660 (9%)
51 to 70	6,825 (37%)	4,737 (35%)	2,568 (34%)
≥71	6,680 (37%)	4,748 (36%)	2,635 (34%)
Transfusions of red blood cell units per patient, median (IQR)	3 (2-6)	2 (1-2)	2 (1-2)
Units of red blood cells transfused, Number (%)§	103,016	25,600	14,172
male donor	65,239 (63%)	22,454 (88%)	12,617 (89%)
female donor, ever-pregnant	22,931 (22%)	1,939 (8%)	982 (7%)
female donor, never-pregnant	14,474 (14%)	1,207 (5%)	573 (4%)

* Consists of all the follow-up time during which patients either received all their red blood cell transfusions exclusively from one exposure category: female donors with a history of pregnancy (ever-pregnant donors), never-pregnant female donors, or male donors. The IPW analysis and Time-varying analysis use this definition. Follow-up time was censored at the time this inclusion criterion was violated.
† Consists of patients who received only one type of exposure (ever-pregnant, never-pregnant or male donor only) during the period in which they were followed up. Complete follow-up from these patients was included in the Restriction analysis.
‡ Median follow-up time is defined as the longest time any patient is in one of the comparisons. Exposure categories are: ever-pregnant donors and male donors.
§ Includes 372 (0.4%) transfusions with unknown donor sex and pregnancy history in the Complete population.

Below, the results for the anonymized dataset are presented (Table S2).

Table S2. Results for the different methods applied to the anonymized dataset

Analysis	No. of Deaths	No. of Recipients	HR (95% CI)
Restriction method			
Male (reference)	1,860	6,316	1 (reference)
Ever-pregnant female	263	884	1.00 (0.88-1.14)
Time-varying exposure and confounding adjustment method			
Male (reference)	1,860	10,901	1 (reference)
Ever-pregnant female	263	1,494	1.01 (0.89-1.15)
Inverse probability of censoring weighting method			
Male (reference)	1,860	10,901	1 (reference)
Ever-pregnant female	263	1,494	1.02 (0.88-1.17)

Here, due to the random permutation of the different variables, the original structure of the data was not maintained. Thus, the treatment-confounder feedback necessitating the use of the here described Inverse probability of censoring weighting method is not present, and all methods perform similarly. This, as opposed to the performance of these methods on the original data, can be seen in Table 2 of the main article.

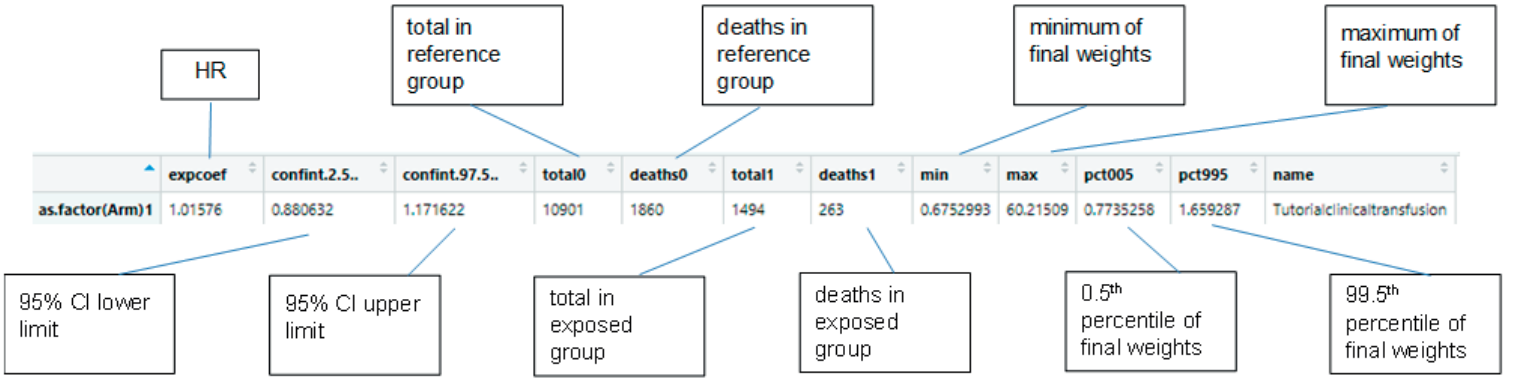


Figure S1. Output of tutorial syntax in R