Transfusion data:
from collection to reflection

Loan R. van Hoeven

Transfusion data: from collection to reflection.
PhD thesis. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands and Sanquin Research, Sanquin Blood Supply, Amsterdam, the Netherlands

ISBN: 978-90-393-6793-3
Author: Loan R. van Hoeven

The studies in this thesis were performed as part of the project 'Data analysis for ongoing optimization of the transfusion chain' (PPOC-11-042) which was funded by Sanquin Blood Supply Foundation.

Transfusion data:
from collection to reflection

Transfusiedata:
van verzameling tot interpretatie
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

donderdag 29 juni 2017
des middags te 4.15 uur

door
Loan R. van Hoeven
geboren op 24 april 1987
te Amersfoort

Promotor:	Prof.dr. K.C.B. Roes
Copromotoren:	Dr.ir. M.P. Janssen
	Dr.ir. H. Koffijberg

Contents

Chapter 1	General introduction	7
Chapter 2	Design of a national blood transfusion data warehouse from donor to recipient	13
Chapter 3	Validation of multisource electronic health record data: an application to blood transfusion data	31
Chapter 4	Why was this transfusion given? Identifying clinical indications for blood transfusion in health care data	51
Chapter 5	Aiming for a representative sample: Simulating random versus non-probabilistic strategies for hospital selection	69
Chapter 6	Historical time trends in red cell component usage in the Netherlands	85
Chapter 7	Prediction of the anti-RhD donor population size for managerial decision making	101
Chapter 8	General discussion	115
	Appendices	125
	Summary	133
	Samenvatting	137
	Dankwoord	141
	Curriculum vitae	145
	List of publications	147

Chapter 1

General introduction

From blood donor to transfusion recipient

Need for transfusion chain data

Undoubtedly blood transfusion has become an essential medical intervention, saving the lives of an estimated 20.000 patients per year in the Netherlands [1]. Blood use in the Netherlands is steadily decreasing, with the exception of blood plasma used for medicines such as the rhesus-D prophylaxis. Blood use in the Netherlands is also relatively low compared to other European countries [2]. Nonetheless, there may still be transfusions given unnecessarily. Ideally, the use of blood should be as efficient as possible without compromising patients' health, minimizing potential adverse effects in patients, the burden placed on blood donors, and costs. The fact that there is a substantial difference in the amount of blood transfused, between countries, hospitals, and even within hospitals, suggests that blood use is not optimal yet.

Besides safely reducing blood use, better alignment of donor and product characteristics with patient characteristics is desirable as well. The optimal blood product, the optimal amount and the optimal moment to transfuse might be different for various patient subgroups, potentially requiring different characteristics of blood products and/or donors. This also introduces logistical concerns related to the optimal matching strategy between donor and patient, and between collection and demand for blood. Examples of questions to be answered in this context are: 'What blood use can we expect in the future for the most demanding patient groups, given how their blood use has developed in the past?' and 'How can we estimate the number of donors to be recruited, given the ageing of the population and current shortage of specific donors?'. Analyzing transfusion chain data can help answering such questions.

Research in 'the blood transfusion chain' covers a broad spectrum of topics: from donor to patient, from study design to validation, and from analysis to policy support. A process at the start of the chain is connected to all that follows, and might thus affect later outcomes. Therefore, in order to comprehensively study transfusion practice, we need data on the complete blood transfusion chain.

Development of the Dutch Transfusion Data warehouse

As no comprehensive data warehouse existed in the Netherlands, we established the Dutch Transfusion Data warehouse (DTD). The data warehouse can be seen as an observational cohort which is regularly updated with recent data from hospitals and the national Dutch blood bank. The DTD contains information on donors (e.g., age, blood group, antibodies), blood products (e.g., type of product, expiration date, storage time), and transfusion recipients (e.g., transfusions administered, patient characteristics, diagnosis, surgical procedures, laboratory parameters). Linking data from multiple sources and over multiple years is the only way to continuously monitor blood use, identify best practices (by benchmarking hospitals), and investigate risk factors throughout the complete transfusion chain. A unique asset of the DTD is that it offers the opportunity to investigate donor-product-patient associations (for

example the potentials effects of storage duration and donor gender on patient health outcomes). Before we have valid and useable data however, various aspects need to be considered when collecting, processing, analyzing and interpreting data in the donor-recipient continuum.

Aspects of transfusion data

Over 555.000 blood products are transfused every year to patients in Dutch hospitals. Of these 77% are red blood cells, 12% plasma and 11% platelets. These products are obtained from the blood of over 300.000 voluntary donors. These numbers indicate that we are leaning towards 'big data', especially when data over multiple years are collected. One of the problems with these data is the abundance of items registered. Many diagnosis and procedure codes are collected per patient per hospital admission, whereas our main interest is in the primary diagnosis or procedure that necessitated the transfusion(s) given. Other aspects relevant for processing the data beside their size, is that they are observational, multisource, national and longitudinal. Observational data have the advantage that these can relatively easily be used, also when a randomized controlled trial is not possible (i.e., too costly or unethical to withhold patients from transfusion) [3]. However, inherently to observational data is the higher risk of errors occurring in each processing step they undergo:

Registration → Extraction → Interpretation

Clearly, data from many different sources are used: the blood bank, hospitals, the hemovigilance organization, and also different databases within hospitals (e.g., laboratory and administrative data). As each source might have different (registration) policies that might also change over time, it is important to standardize and harmonize the data. The data reflect current care processes, which change continuously over time: therapies are constantly evolving, new blood products and tests are developed, and new guidelines for transfusion triggers – stating the conditions for which a transfusion is required– are introduced. As the DTD aims to be a national data warehouse, the composition of hospitals included (which changes when new hospitals are added) must be considered carefully to ensure representativeness of the data for Dutch transfusion practice. It is crucial that the building of a data warehouse accounts for all these aspects, by finding appropriate, reproducible and preferably automated solutions.

Aim

We set out to answer various research questions, for example on trends in blood use in hospitals and on prediction of the future size and composition of the anti-RhD donor population. In the process of answering these questions however, we identified key methodological challenges, which became the primary focus of our research. As this would

benefit future research systematically, the methods used for dealing with transfusion data are shared in this thesis. The objectives are:

- to create the Dutch Transfusion Data warehouse (DTD)
- to validate the quality of the data collected in the DTD and develop a validation approach
- to develop an automated way to use and interpret data on transfusion indications
- to analyze trends in blood use and the types of recipients who receive blood, and
- to predict changes in the anti-RhD donor population in order to support recruitment decisions

Thesis outline

This thesis can be roughly divided into two themes: the preparation (Chapters 2-5) and the utilization (Chapters 6-7) of transfusion data.

In the first part of this thesis, data collection and processing are addressed**. Chapter 2** describes how the Dutch Transfusion Data warehouse was set up, the challenges involved and the utilization of the data warehouse. **Chapter 3** thoroughly examines the quality of the data, resulting in an approach that can be used for validating multisource electronic health record (EHR) data more generally. To facilitate the interpretation of the data, **Chapter 4** presents an automated algorithm that selects the most likely indication for transfusion. In **Chapter 5** a simulation study compares different strategies for deciding on which hospitals to include for a representative selection of Dutch hospitals.

In the second part of this thesis, the application of transfusion data is key with analyses of trends in blood use and blood donors. In **Chapter 6**, historical trends in the use of red blood cell products (and combinations with other products) are investigated, using transfusion recipient data from the previously performed Proton study [4]. In **Chapter 7**, retrospective donor data are modelled to predict how many new donors for anti-RhD immunizations are needed annually, given the donor drop-out rate and ageing of the current donor population. These last two chapters do not yet utilize the donor-recipient continuum as this type of analysis is more powerful when data from a larger number of hospitals are available. Still, these applications illustrate how transfusion data can add to the understanding of blood use and collection in practice, and have the potential to support decision making.

The common thread running through this thesis is transfusion data. All chapters in some way describe methods for data collection, interpretation or validation, finally using transfusion data to analyze and elucidate aspects of donor recruitment and clinical blood use in practice. **Chapter 8** reflects on the value of transfusion data, discusses the most important lessons learned with respect to acquiring a valid data warehouse and elaborates on the interpretation of transfusion indications. We conclude with recommendations for future research topics and questions that can be employed using the Dutch transfusion data

warehouse and to this end suggest various potential extensions that will enhance the answering of these questions.

References

1. Sanquin Blood Supply. Facts and figures. Accessed at February 2017 at: http://www.sanquin.nl/over-sanquin/pers/facts-and-figures/
2. van Hoeven LR, Janssen MP, Rautmann G. The collection, testing and use of blood and blood components in Europe, 2012 report. European Directorate for the Quality of Medicines & HealthCare of the Council of Europe (EDQM), 2015, Strasbourg, France
3. Dominick E, Migliaccio-Walle D, Molsen E (eds). Reliability and validity of data sources for outcomes research & disease and health management programs. 2013.
4. Borkent-Raven BA, Janssen MP, Van Der Poel CL, et al. The PROTON study: profiles of blood product transfusion recipients in the Netherlands. Vox Sanguinis 2010;99:54–64.

Chapter 2

Design of a national blood transfusion data warehouse from donor to recipient

BMJ Open 2016

L.R. van Hoeven
B.H. Hooftman
M.P. Janssen
M.C. de Bruijne
K.M. de Vooght
P.F. Kemper
M.M.W. Koopman

Abstract

Introduction

Blood transfusion has health related, economical and safety implications. In order to optimize the transfusion chain, comprehensive research data are needed. The Dutch Transfusion Data warehouse (DTD) project aims to establish a data warehouse where data from donors and transfusion recipients are linked. This paper describes the design of the data warehouse, challenges, and illustrative applications.

Methods

Quantitative data on blood donors (e.g., age, blood group, antibodies) and products (type of product, processing, storage time) are obtained from the national blood bank. These are linked to data on the transfusion recipients (e.g., transfusions administered, patient diagnosis, surgical procedures, laboratory parameters), which are extracted from hospital electronic health records.

Applications

Expected scientific contributions are illustrated for four applications: determine risk factors, predict blood use, benchmark blood use and optimize process efficiency. For each application, examples are given of research questions and analyses planned.

Conclusion

The DTD project aims to build a national, continuously updated transfusion data warehouse. These data have a broad range of applications, on the donor/production side, recipient studies on blood utilization and benchmarking, and donor-recipient studies, which ultimately can contribute to the efficiency and safety of blood transfusion.

Introduction

In 1874, a first review, or 'short resume', was published about the current evidence regarding blood transfusions, concluding that transfusion might be "an effective mean of saving life when all other means fail", yet this subject needed more investigation [1]. To date, it is widely accepted that blood transfusions can be lifesaving and can be used for treatment of various diseases. However, since blood transfusions may also have serious side effects [2], there is still much debate on optimal transfusion triggers [3]. There is growing but inconclusive evidence that a restrictive transfusion policy is more beneficial for patients than a more liberal policy [4,5] (exceptions might be patients with cardiac disease or oncological surgery [6]). The large variation that exists in the use of blood products between countries, between hospitals and even within hospitals [7,8,9,10] indicates that –at least in part of the patients– transfusion practice is not optimal yet and that there is uncertainty about the optimal transfusion policy. Importantly, transfusion policy concerns not only the timing and quantity of the transfusions, but also other characteristics of the blood product, the donor and the production process that might affect patient outcomes. In order to investigate the magnitude and nature of the observed differences as well as gain proficient understanding of efficiency and safety of the donor-product-recipient relationship, more data are needed.

Even though several individual hospitals and blood banks analyze data on donors and transfusion recipients [11,12], worldwide initiatives that permanently monitor transfusions on a large scale are sparse. The SCANDAT database from Sweden and Denmark, originally established in 2002, now covers all donor and transfusion data nationwide since 1968 (Sweden) and 1980 (Denmark). It includes 47 years follow-up data on health outcomes regarding hospital care, cancer and death [13,14]. The REDS-III program in the United States is currently constructing a similar blood donor and transfusion recipient database [15]. Finland established a recipient database originating in 2002, covering in the year 2007 70% of all blood units delivered for all potentially transfused patients [16]. Recently a Canadian donor-recipient study was initiated, containing data from hospitals in a specific region [17]. In the Netherlands, the PROTON database was created to identify PROfiles of TransfusiON recipients, with data on transfusion recipients in terms of age, sex, main diagnoses and operations, number of products per hospitalization [18].

These initiatives resulted in studies on the epidemiology of both donors and recipients, providing evidence on the effect of donation and of transfusion, as well as the link between donor and recipient. Examples of this are studies to investigate mortality risk in transfusion recipients [19], and length of hospital stay after receiving red blood cell units [20]. In the donor-recipient continuum, research topics include the risk of cancer in recipients who received a blood transfusion from donors with subclinical cancer [21,22], the effect of the match of donor and recipient sex on survival after plasma transfusion [23,24,25], safety of ABO-compatible non-identical plasma versus identical plasma [26], and the effect of storage duration on recipient survival [27,28]. Nowadays, there is a tendency to modify risk-adverse guidelines for donor selection into more liberal guidelines based on new evidence [29].

Although the evidence is yet scarce [30], there are successful examples, such as extending the upper age limit for donors without increasing the number of adverse events in patients [31,32]. Other results of transfusion data warehouse initiatives include the development of a model to predict the impact of demographic changes on the demand of red blood cell units [33]. Such a model may guide donor recruitment requirements. Moreover, benchmarking events have been organized, for example in Finland for different transfusion practices such as orthopedics, gynecology, hematology and heart surgery. Benchmarking discussions have led to adoption of best practices in several cases, reflected in the reduction of differences in blood use [34].

The Dutch PROTON database included hospital transfusion data starting in the year 1996 [17]. Unfortunately, data collection stopped after 2006. Also the database contained information on transfusion recipients, but not on the corresponding blood donors. In an effort to continue this database and expand its scope, the Dutch Transfusion Data warehouse (DTD) project started. In this project a data warehouse is developed that is intended for continuous storage, management and monitoring of transfusion data, linking donor to recipient. This means that the DTD facilitates research on blood utilization in hospitals, but it also offers the unique opportunity to study donor and product risk factors for recipient outcomes and examine efficiency over the complete transfusion chain. Thereby the creation of the DTD infrastructure will allow the comprehensive study of blood transfusion in the Netherlands. The four main applications of this data warehouse are to:

1. Determine risk factors
2. Predict future blood products needed
3. Benchmark blood use
4. Improve process efficiency

To illustrate how the DTD initiative will be used for these applications, we will propose four example studies. The successful completion of this cohort will contribute to the safety of transfusion practices, and provide insights that can improve efficiency in the complete blood transfusion chain.

Methods

Data collection and data set

The data warehouse can be seen as an observational research registry, in which routinely registered administrative data are collected continuously. Starting point was the previously conducted PROTON study [18], consisting of a single collection of blood transfusion data in the Netherlands from 1996-2006. This dataset is further extended further with additional recipient, donor and product data.

In the Netherlands the blood supply is organized at a national level by Sanquin which is the sole supplier, enabling a centralized extraction of data on donors and blood products. Sanquin provides data on donor demographics, blood groups and laboratory parameters, and blood product characteristics such as product type and expiration date (Table 2.1). In this

paper, the term blood bank refers to the national blood supplier. The participating hospitals provide data related to transfusion recipients from their electronic health records, including patient characteristics, hospitalizations, diagnoses, procedures, blood products received, blood groups, laboratory parameters and transfusion reactions (Table 2.2). In addition, each hospital is requested to provide aggregated information on the total number of patients per indication (including non-transfused patients), allowing computation of transfusion rates. Linkage of donor and transfusion recipient data is based on the uniquely identifying combination of donation identification code and the internationally used ISBT product code [35]. All Dutch hospitals (n= 91) are allowed to participate in the project, however in order to meet the research objectives, a minimum sample of 15 academic and general Dutch hospitals in total is aimed for. Data collection starts from 2010 and will include future transfusions as well. The current number of donors in our database is approximately 500,000, with 3,500,000 products issued by the blood bank covering the years 2010-2015 (this is a complete set for national coverage). These products are linked to recipient data from the participating hospitals. Based on inclusion of 15 hospitals, we now estimate that the number of recipients in our data warehouse for the years 2010-2015 (including academic, teaching and general hospitals) will be 150,000, with approximately 1,100,000 transfusions.

Future fusions of hospitals and shifts in type and complexity of care especially in academic hospitals will be monitored closely, as these factors directly affect blood use.

Table 2.1. Overview of donor and blood product data collected in the blood bank

Donor	Donor number	Donor identification number
	Date of birth	Date of birth
	Gender	Gender
	AB0 blood group	AB0 blood group
	RhD blood group	RhD blood group
	Kell blood group	Kell blood group
	Donor entry date	Date of registration at the blood bank
	Date of first donation	Date of first donation since 2007
	Weight	Donor weight
	Length	Donor length
	Number of donations	Total number of donations since 2007
	Number whole blood donations	Total number of whole blood donations since 2007
	Number plasma donations	Total number of plasma donations since 2007
	Other donations	Total number of other donations since 2007
	Stopping code	Stopping code
	Stopping reason	Reason to quit as a donor

Donation	Donation date	Date of donation
	DIN	Donation Identification Number (unique for each donation)
	Donation type	Type of donation (whole blood / plasmapheresis / erythrocytapheresis / plateletpheresis)
	Donation volume	Volume of the donation (in ml)
	Hemoglobin level	Hemoglobin level (in mmol/L)
	Platelet count	Number of platelets (x 10^9/l)
	Donation location	Blood center location
	Donation duration	Duration of donating
	apheresis machine	Code of the machine used for apheresis
	Blood pressure	Donor blood pressure
Product		
	Product code	Product code (specifying product type, location of the blood bank center, split product or not; according to ISBT 28 Standard Specification)
	Product modifiers	Optional attribute of a product (e.g., CMV neg./pos.)
	Pool DIN	Product Identification code; applies only to pooled products (thrombocytes)
	Expiration date	Date that the product expires
	Erythrocyte antibodies	Erythrocyte antibodies in donor blood
	Platelet phenotype	HPA phenotype
	Date of pooling	Date of pooling; applies only to pooled products (thrombocytes)
Transport	Transport date	Date of transport of the blood product
	Institute	Destination of the blood product
	Return date	Date of return of the blood product; only in case a blood product is sent back
	Return code	Return code specifying the reason for returning the blood product

Table 2.2. Overview of the data collected in the participating hospitals

Patient	Patient number	Encrypted patient identification number as used by the hospital
	Date of birth	Patient's date of birth
	Gender	Patient gender
Hospitalization	Hospital	Hospital name or code
	Hospitalization dates	Date and time of start and end of hospitalization during which a transfusion was given
Diagnosis	Diagnoses	All diagnoses**
	Hospital dismissal status	Patient status when dismissed from hospital (home / dead / institution)
Procedure	Procedures	All procedures and procedure dates**
Transfusion	Transfusion administration	Date and time of transfusion
	DIN	Donation identification number of the transfused unit
	Product code	Product code of the transfused unit (ISBT 128 Standard Specification)
Blood values	Hb	Patient hemoglobin level*
	Platelet count	Patient platelet count*
	Hct	Patient hematocrit level*
	PT	Patient prothrombin time*
	PTT	Patient partial thromboplastin time*
	Blood group	Patient AB0 blood group
	RhD	Patient RhD
	Irregular antibodies	Patient irregular antibodies
	Troponine	Patient troponine level
Transfusion reactions	Transfusion reaction type	Type of transfusion reaction
	Date	Date of transfusion reaction
	Severity	Severity of transfusion reaction
	Imputability	Likelihood that the transfusion reaction is caused by the transfusion

* All laboratory parameters measured during hospitalization, or in case of outpatient transfusion all laboratory parameters within the period 72 hours before and after transfusion. All laboratory measurements include time stamps.
** Diagnoses and procedures can be linked to a hospitalization post hoc, or to outpatient transfusions within a time interval around the transfusion. In the Netherlands, instead of one diagnosis date, a start and end date is registered of the 'diagnosis treatment combination' trajectory.

Data quality

Extracting and combining large amounts of data from electronic hospital and blood bank systems is challenging: often the data have to be split into different tables (e.g. by year, department or aggregation level), that afterwards have to be linked. In this process, errors can occur in the data, therefore validation of the data is very important. This starts with a uniform format and filters; we ask the participating centers to deliver the data in the same format for every update of the data.

In order to check and improve data quality, the data warehouse will be validated on the following aspects: completeness, uniqueness, time patterns, uniformity and plausibility. Also, external concordance of the number of blood products issued by the blood bank and the products transfused by the hospitals is assessed as a validity check. In the Netherlands, the blood bank registers donor and product data in one system. In contrast, some of the hospital data such as diagnoses and clinical procedures are registered in more heterogeneous ways across hospitals and sometimes even across departments within a single hospital. This means that more time is needed to validate and harmonize the hospital data. Moreover, as every registration system is subject to updates and changes, each time new data are sent to the data warehouse, the additional content will have to be validated. We intend to publish the outcomes of the validation check or at least make them available for other researchers who use the data warehouse.

Indication for transfusion

In order to facilitate the attribution of the main diagnosis (i.e. indication) for a transfusion, an automated algorithm will be developed for the DTD. This algorithm will determine the most likely indication for transfusion in the case of multiple diagnoses and/or procedures per transfusion event. The algorithm will be developed based on expert opinion regarding the prioritization of diagnoses, and will be externally validated by transfusion experts.

Security, ethical and privacy aspects

The data warehouse is hosted by the data management department of a university medical center, in a technical environment that meets ISO-9001:2008 quality requirements. DTD has been approved by a hospital medical ethical committee and meets the requirements of Dutch privacy law. Donors are asked for permission with a donor questionnaire before each donation. Patients are not actively asked for permission but they can opt out for use of their medical data for research purposes. Donor and patient data are transferred and stored in an de-identified format. The encryption is carried out by the contact person of the hospital, and the key to reverse encryption is stored exclusively in the hospital. In addition, non-traceability of blood donors and recipients is maximized by excluding privacy sensitive information such as name and postal code.

Organization structure

The DTD project team, consisting of experienced researchers in the areas of transfusion medicine, data modelling and health care research, is responsible for the management of the data warehouse. An advisory board, consisting of representatives of all involved disciplines, is established to handle all data requests. Main objective of the board is to guarantee that the interests of all participating parties are secured. Every data provider has one contact person who, for instance, arranges the formal permission for data exchange. Researchers planning a project can gain access to the data warehouse by completing a data request form. The advisory Board will determine whether the request is granted, thereby guaranteeing the interests of all parties involved.

Framework blood supply chain

The framework as presented in Figure 2.1 provides an overview of the different steps in the transfusion chain and can be used to systematically identify and highlight areas with room for improvement. The four main applications (see Introduction) are linked to these steps, showing which data are necessary for each application. The main contribution of the data warehouse is to allow insight in the association between blood donor characteristics and clinical outcomes (left broad arrow) and in the link between transfusion triggers and clinical outcomes (right broad arrow).

Applications

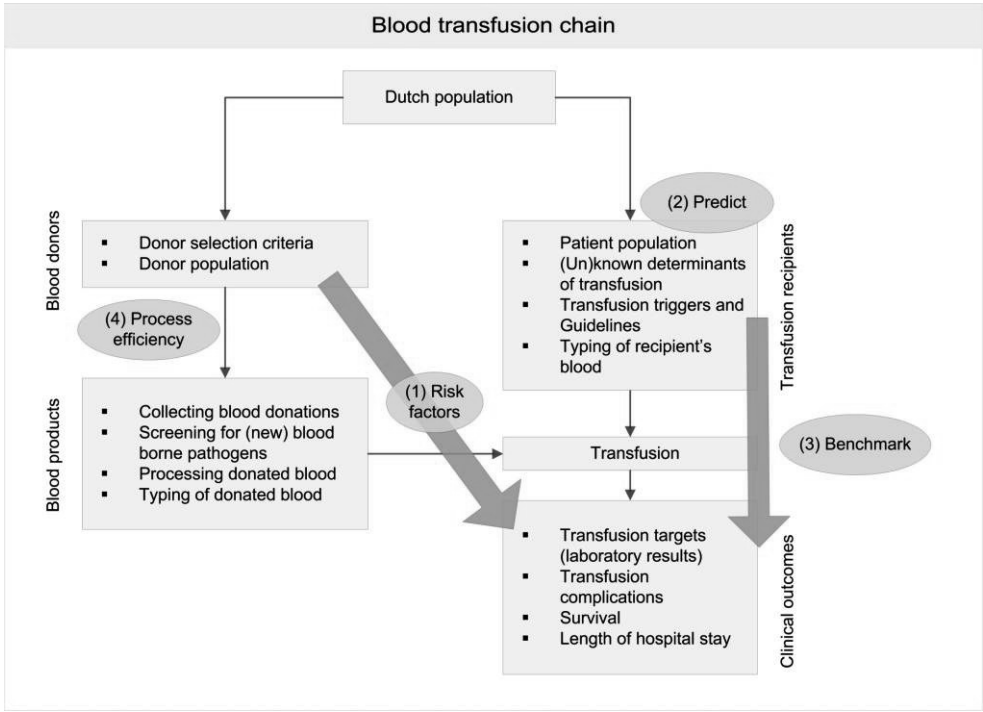
The DTD data warehouse will be available for a broad range of purposes. To illustrate expected scientific contributions, we describe exemplary studies that could be conducted with the DTD dataset.

Application 1: Risk factors

Example research question: What is the effect of donor characteristics and season on the risk of (febrile) non-hemolytic transfusion reactions (FNHTR) experienced by recipients?

Non-hemolytic transfusion reactions are relatively common, especially among hematology patients, with median reported rates for FNHTR of 4.6% for platelets and 0.33% for RBCs.[36] This type of transfusion reaction seems to occur in particular with platelet transfusions but also with erythrocytes. With our data we can determine the association of (febrile) non-hemolytic reactions with season and with certain donor characteristics (age, sex, blood group, donation frequency). Donation frequency for example is hypothesized to affect iron storage, and might also affect patient outcomes. The primary outcome is risk of non-hemolytic transfusion reactions. Secondary outcomes are: risk of infections, other transfusion reactions, survival and duration of hospitalization.

Figure 2.1. Framework of the blood transfusion chain. Each part of the chain can be linked to one of the four applications.



Application 2: Predict future blood products needed

Example research question: What is the expected use of blood (medical versus surgical) in the Netherlands for the upcoming years?

Long-term data from 2010 up to the present will be examined for trends in blood use per product type. This information can be used to generate prognoses on the number of blood products needed in the future. Increasingly refined and specific predictions can be made by distinguishing between surgical and medical use of RBCs, as well as academic, teaching and general hospitals. Observed trends in de past will be extrapolated using a regression model. Furthermore, corrections for growth and ageing of the general population can be incorporated in the predictions of the amount of blood products demanded.

Application 3: Benchmark blood use

Example research question: What is the variation in blood use between hospitals, corrected for important determinants of blood use?

Differences in blood use between hospitals might be caused by different uses of transfusion (Hb) triggers and targets. A benchmark study could compare these triggers between hospitals in specified patient groups, while correcting for other determinants of blood use available in

the data warehouse, such as: age, sex, comorbidity burden, recent myocardial infarction, emergency or elective presentation, medical or surgical admission, diagnosis, type of surgical procedure, hospital department, preoperative hematocrit and preoperative or admission hemoglobin [9,10,37]. A multilevel random effects model can be specified with the following levels: hospital type, hospital, and patient. This allows estimation of the variation in blood use between hospitals compared to the variation within hospitals, while controlling appropriately for differences in patient characteristics.

Application 4: Improve process efficiency

Example research question: Is more extensive blood group matching between donor and recipient possible given the current donor population and is it cost-effective?

More extensive matching of donor and recipient blood groups (especially for ethnic minorities) would reduce the formation of red cell anti-bodies and ultimately also the risk of transfusion reactions. Data on donors and patients (which reflect the availability and consumption of blood and blood types) is used to obtain insight in the logistical requirements and limitations, costs and (health-)effects of various preventive matching schemes. In the ongoing BloodMatch study [38], several scenarios for matching strategies will be evaluated. These scenarios vary in the extent of blood type matching between donor and recipient for specific patient groups, and its anticipated impact on transfusion complications, the size and composition of the red blood cell stocks in both the blood bank and hospitals, as well as the requirements for typing of the donor base in order to fulfill the demand for typed red blood cell units. The findings will allow balancing various aspects of the blood transfusion chain and therefore provide the means for a global optimization of matching strategies.

Discussion

Importance

The DTD project aims to build a national, up-to-date transfusion data warehouse, linking donor to recipient. By gaining more insight in donor and product related risk factors for recipient outcomes, blood transfusions can be more tailored (minimizing risks) and unnecessary transfusions avoided, further diminishing transfusion reactions in patients. Especially today, facing increasing societal pressure for transparency in quality of care, multiple parties may benefit from a continuous feedback structure. For the blood bank, DTD creates the possibility to enhance the safety of transfusion on the donor and product side, as well as stock management (optimize the availability and minimize wastage of blood products). For healthcare institutions, DTD enables insight in efficient and safe use of blood products. Moreover, by participating in the project, hospitals can take better control in the way they are held accountable for blood use by external parties such as insurers and regulators. For researchers from or in collaboration with participating institutions, DTD offers access to

essential data as well as a network within the clinical field. Finally, patients benefit from optimal and evidence-based quality of care in transfusion medicine.

Applications and future directions

The data warehouse will be available for different types of users, including the blood bank, hospital management, doctors and researchers. In hospitals, blood reduction policies can be directly linked to trends in blood use [39,40,41], and new transfusion guidelines and quality indicators can be evaluated. Moreover, the availability of laboratory data can shed light on the impact and relevance of clinical and laboratory parameters (like hemoglobin level) that are used as transfusion triggers and targets. An important step in the overall process is to report (benchmark) results back to the caregivers [39]. Whereas the level of detail in the indicators themselves is of less importance (especially in complex practices such as heart surgery and hematology), the discussion between clinical experts it will instigate that might provide novel insights, solutions to existing problems and evolvement of best practices.

The data warehouse also enables comparison of transfusion practices internationally. A great advantage is that with the presence of various patient and hospitalization characteristics, the outcome can be adjusted for factors like age, sex, diagnosis and surgery. The scope of variables collected in DTD is similar to the SCANDAT2 database, which also focusses on donors' health using donor hospital information and already has national coverage. In the future it will be possible to expand the data warehouse with additional variables, either permanently or temporary, such as recipient survival. Additional data on vital signs (pulse, temperature and blood pressure) and laboratory parameters can also be included post-hoc, depending on the specific research. Moreover, we aim to add data from patients who did not receive a blood transfusion at all, in order to calculate transfusion rates and to compare profiles of transfused recipients to non-transfused recipients.

Barriers and facilitators

The advantage of the project's wide organizational structure is that the collaboration of hospitals, clinicians and researchers is facilitating multisite and multidisciplinary research. Moreover, the process of data validation needs to be performed only once, and everyone can benefit from this. Challenges are found in the rapid development of and changes in registration systems, the project financing structure, participation of hospitals and changes in legislation with respect to data usage. Currently, electronic health data are primarily registered for clinical use and a systematic interpretation for research purposes is often lacking. A related problem is the large registration burden on hospital personnel and the current focus on billable 'health products', which largely determines what is registered and how. These aspects are external factors that are mostly out of control of the project, but do complicate regular data extraction and therefore pose potential threats to the future of the data warehouse. Projects to improve source registration have already been set up in the Netherlands supported by the Federation of University medical centers [42], the Dutch Association of general hospitals and specialized

institutions [43] and the center of expertise for standardization and eHealth Nictiz [44], and for example in the US by the Centers for Medicare & Medicaid Services, promoting meaningful use of certified electronic health record (EHR) technology [45]. If uniform registration will be successfully implemented in hospitals, standardized source data could be used for the data warehouse, allowing real-time data extraction. However the analysis of imperfect data requires other solutions. For example, when patients have received multiple transfusions, we must take into account the potential for confounding in the analysis. Several analysis methods can be used, including restriction to certain cases and statistical correction using standardization or maximum likelihood methods [46].

Conclusion

The Dutch Transfusion Data warehouse contributes to the optimization of Dutch transfusion practice by enabling researchers to identify donor risk factors that affect recipients, monitor and benchmark the use of blood products both at national and international levels, and evaluate the effect of changes in the supply chain. This will contribute to optimally tailored transfusions and fewer transfusion reactions. Joint support from the blood bank, hospitals and external parties are key success factors for a future-proof and clinically relevant blood transfusion data warehouse.

References

1. Madge HM. On transfusion of blood. British medical journal 1874;1(680):42.
2. Goodnough LT. Blood management: transfusion medicine comes of age. The Lancet 2013;381(9880):1791-1792.
3. Goodnough LT, Levy JH, Murphy MF. Concepts of blood transfusion in adults. The Lancet 2013;381(9880):1845-1854.
4. Carson JL, Carless PA, Hébert PC. Outcomes using lower vs higher hemoglobin thresholds for red blood cell transfusion. Jama 2013;309(1):83-84.
5. Carson JL, Carless PA, Hébert PC. Transfusion thresholds and other strategies for guiding allogeneic red blood cell transfusion. The Cochrane Library 2012.
6. De Almeida JP, Vincent JL, Galas FR, et al. Transfusion requirements in surgical oncology patients: a prospective, randomized controlled trial. Anesthesiology 2015;122(1).
7. Council of Europe: The collection, testing and use of blood and blood components in Europe. 2011 Report. Accessed September 2015 at: https://www.edqm.eu/medias/fichiers/the_collection_testing_and_use_of_blood_and_blood_components_in_europe_2011_report.pdf.
8. Gombotz H, Rehak PH, Shander A, et al. The second Austrian benchmark study for blood use in elective surgery: results and practice change. Transfusion 2014;54(10pt2):2646-2657.
9. Jin R, Zelinka ES, McDonald, et al. Effect of hospital culture on blood transfusion in cardiac procedures. The Annals of thoracic surgery 2013;95(4):1269-1274.
10. Ejaz A, Spolverato G, Kim Y, et al. Variation in triggers and use of perioperative blood transfusion in major gastrointestinal surgery. British Journal of Surgery 2014;101(11):1424-1433.
11. Biggin K, Warner P, Prescott R et al. A review of methods used in comprehensive, descriptive studies that relate red blood cell transfusion to clinical data. Transfusion 2010;50:711–718.
12. Edgren G, Hjalgrim H. Epidemiological considerations for the use of databases in transfusion research: a Scandinavian perspective. Current Opinion in Hematology 2010;17:596-601.
13. Edgren G, Hjalgrim H, Tran TN, et al. A population-based bi-national register for monitoring long-term outcome and possible disease concordance among blood donors and recipients. Vox Sanguinis 2006;91:316-23.
14. Edgren G, Rostgaard K, Vasan SK, et al. The new Scandinavian Donations and Transfusions database (SCANDAT2): a blood safety resource with added versatility. Transfusion 2015;55(7):1600-1606.
15. Kleinman S, Busch MP, Murphy EL, Shan H, Ness P, Glynn SA, and The National Heart, Lung, and Blood Institute Recipient Epidemiology and Donor Evaluation Study (REDS-III). The National Heart, Lung, and Blood Institute Recipient

Epidemiology and Donor Evaluation Study (REDS-III): a research program striving to improve blood donor and transfusion recipient outcomes. Transfusion 2014,54: 942–955.

16. Palo R, Ali-Melkkilä T, Hanhela R, et al. Development of permanent national register of blood component use utilizing electronic hospital information systems. Vox Sanguinis 2006;91: 140–147.
17. Chassé M, McIntyre L, Tinmouth A, et al. Effects of blood donor characteristics in transfusion recipients: protocol of a framework to study the blood donor–recipient continuum. BMJ open 2015,5(1):e007412.
18. Borkent-Raven BA, Janssen MP, Van Der Poel CL, et al. The PROTON study: profiles of blood product transfusion recipients in the Netherlands. Vox Sanguinis 2010;99:54–64.
19. Borkent-Raven BA, Janssen MP, van der Poel CL, et al. Survival after transfusion in the Netherlands. Vox Sanguinis 2011;100:196–203.
20. Palo R, Ahonen J, Salo H, et al. Transfusion of red blood cells: no impact on length of hospital stay in moderately anaemic parturients. Acta Anaesthesiologica Scandinavica 2007;51(5):565-569.
21. Edgren G, Hjalgrim H, Reilly M, et al. Risk of cancer after blood transfusion from donors with subclinical cancer: a retrospective cohort study. The Lancet 2007;369(9574):1724-1730.
22. Hjalgrim H, Edgren G, Rostgaard K, et al. Cancer incidence in blood transfusion recipients. Journal of the National Cancer Institute 2007;99(24):1864-1874.
23. Middelburg RA, Briët E, Van der Bom JG: Mortality after transfusions, relation to donor sex. Vox Sanguinis 2011, 101: 221–229. doi: 10.1111/j.1423-0410.2011.01487.x
24. Tynell E, Andersson TM, Norda R, et al. Should plasma from female donors be avoided? A population-based cohort study of plasma recipients in Sweden from 1990 through 2002. Transfusion 2010;50(6):1249-1256.
25. Norda R, Andersson TML, Edgren G, et al. The impact of plasma preparations and their storage time on short-term posttransfusion mortality: a population-based study using the Scandinavian Donation and Transfusion database. Journal of Trauma and Acute Care Surgery 2012;72(4):954-961.
26. Shanwell A, Andersson TL, Rostgaard K, et al. Post-transfusion mortality among recipients of ABO-compatible but non-identical plasma. Vox sanguinis 2009;96(4):316-323.
27. Edgren G, Kamper-Jørgensen M, Eloranta S, et al. Duration of red blood cell storage and survival of transfused patients (CME). Transfusion 2010;50(6):1185-1195.
28. Lacroix J, Hébert PC, Fergusson D, et al. Age of Transfused Blood in Critically ill Adults. N Engl J Med 2015;372(15):1410-1418.
29. Williamson LM, Devine DV. Challenges in the management of the blood supply. The Lancet 2013;381(9880):1866-1875.

30. Eder A. Evidence-based selection criteria to protect blood donors. J Clin Apher 2010;25:331-37.
31. Eder AF, Dy BA, Kennedy JM, et al. The American Red Cross donor hemovigilance program: complications of blood donation reported in 2006. Transfusion 2008;48:1809–19.
32. Fan W, Yi QL, Xi G, et al. The impact of increasing the upper age limit of donation on the eligible blood donor population in Canada. Transfus Med 2012;22:395–403.
33. Borkent-Raven BA, Janssen MP, Van Der Poel CL. Demographic changes and predicting blood supply and demand in the Netherlands. Transfusion 2010,50: 2455–2460.
34. Maki T. Optimizing blood usage through benchmarking. Transfusion 2007; 47:145S–148S.
35. ISBT. ST-001-ISBT-128-Standard-Technical-Specification-v4.1.0 Accessed October 2014 at: http://www.iccbba.org/tech-library/iccbba-documents/cellular-therapy-documents.
36. Geiger TL, Howard SC. Acetaminophen and diphenhydramine premedication for allergic and febrile nonhemolytic transfusion reactions: good prophylaxis or bad practice?. Transfusion medicine reviews 2007;21(1):1-12.
37. Roubinian NH, Murphy EL, Swain BE, et al. Predicting red blood cell transfusion in hospitalized patients: role of hemoglobin level, comorbidities, and illness severity. BMC health services research 2014;14(1):213.
38. BloodMatch study. Accessed at June 2016 at: http://www.sanquin.nl/en/research/departments/clinical-transfusion-research/bloodmatch
39. Apelseth TO, Molnar L, Arnold E, et al. Benchmarking: applications to transfusion medicine. Transfusion medicine reviews 2012;26(4):321-332.
40. Verlicchi F, Desalvo F, Zanotti G, et al. Red cell transfusion in orthopaedic surgery: a benchmark study performed combining data from different data sources. Blood Transfusion 2011;9(4):383.
41. Heddle NM, Liu Y, Barty R, et al. Factors affecting the frequency of red blood cell outdates: an approach to establish benchmarking targets. Transfusion 2009;49(2):219-226.
42. Dutch Federation of University medical centers (NFU). Source registration (Dutch title: Registratie aan de bron), 2013. Accessed September 2015 at: http://www.nfu.nl/img/pdf/13.3694_Brochure_Registratie_aan_de_bron_versie_4-7-13.pdf.
43. Dutch Association of general hospitals and specialized institutions (NVZ). Information policy: four priorities (Dutch title: Informatiebeleidsplan NVZ: vier speerpunten). Accessed September 2015 at: https://www.nvz-ziekenhuizen.nl/onderwerpen/informatiebeleid.

44. Nictiz. Advice for improving the exchange of information in health care (Dutch title: Advies ter verbetering van informatie-uitwisseling in de zorg), 2015. Accessed September 2015 at: http://www.nictiz.nl/module/360/1214/Advies_informatie-uitwisseling_Nictiz_naar_minister_Schippers.pdf.
45. CMS. EHR Incentive Programs. Accessed September 2015 at: https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/ehrincentiveprograms.
46. Middelburg R, Le Cessie S, Briët, et al. A solution to the problem of studying blood donor–related risk factors when patients have received multiple transfusions. Transfusion 2010;50(9):1959-1966.

Chapter 3

Validation of multisource electronic health record data: an application to blood transfusion data

Submitted

L.R. van Hoeven
M.C. de Bruijne
P.F. Kemper
M.M.W. Koopman
J.M.M. Rondeel
A. Leyte
H. Koffijberg
M.P. Janssen
K.C.B. Roes

Abstract

Background

Although data from electronic health records (EHR) are often used for research purposes, systematic validation of these data prior to their use is not standard practice. Existing validation frameworks discuss validity concepts without translating these into practical implementation steps or addressing the potential influence of linking multiple sources. Therefore we developed a practical approach for validating routinely collected data from multiple sources and to apply it to a blood transfusion data warehouse to evaluate the usability in practice.

Methods

The approach consists of identifying existing validation frameworks for EHR data or linked data, selecting validity concepts from these frameworks and establishing quantifiable validity outcomes for each concept. The approach distinguishes external validation concepts (e.g. concordance with external reports, previous literature and expert feedback) and internal consistency concepts which use expected associations within the dataset itself (e.g. completeness, uniformity and plausibility). In an example case, the selected concepts were applied to a transfusion dataset and specified in more detail.

Results

Application of the approach to a transfusion dataset resulted in a structured overview of data validity aspects. This allowed improvement of these aspects through further processing of the data and in some cases adjustment of the data extraction. For example, the proportion of transfused products that could not be linked to the corresponding issued products initially was 2.2% but could be improved by adjusting data extraction criteria to 0.17%.

Conclusion

This stepwise approach for validating linked multisource data provides a basis for evaluating data quality and enhancing interpretation. When the process of data validation is adopted more broadly, this contributes to increased transparency and greater reliability of research based on routinely collected electronic health records.

Background

Electronic health databases are vastly expanding in both the amount and scope of the data available. For health care researchers it seems very attractive to utilize these data maximally [1,2]. Unfortunately, the use of routinely collected electronic health record (EHR) data potentially leads to quality issues, resulting from the fact that the data were not registered for research purposes but rather for clinical management or financial administration. This affects the basic quality of the data for research purpose and the ability to correctly interpret these data. Therefore, it is important to validate the quality of the data before they can serve as a source for health care research aimed to change clinical practice.

Data validity has previously been described as whether values 'make sense' [3]; data are considered valid if the data represent what it claims to represent [4]. In this paper we use the term validation to indicate the process of assessing and improving data quality. Benefits of performing data validation are that it provides guidance on strategies to improve data quality, and, by providing an overview of data quality, enables a fair appreciation and interpretation of study results [5].

Ideally a uniform, systematic method should be used to assess, report and improve data quality. However, as noted previously [3]: There is currently little consistency or potential generalizability in the methods used to assess EHR data. [...] researchers should adopt validated, systematic methods of EHR data quality assessment. A review of 35 empirical studies that used electronic health care data showed that 66% of the studies evaluated data accuracy, 57% data completeness, and 23% data comparability.6 Even if quality measures were reported, the accuracy of variables were highly variable, ranging from 45% to almost 100% [6]. Also, information about chronic and severely ill patients was more likely to be documented as compared to healthier patients [7], which in itself may be a source of bias.

Data quality assessment is especially important in studies using data from multiple sources, in order to distinguish true variations in care from data quality problems [8]. More sources will provide either more cases (multicenter studies) or additional information. Whereas adding more cases can be problematic for data harmonization because different sources may use different coding systems, linking additional information (for example from external data sources) requires that patients (or other entities) can be identified and linked in all sources [9]. In each linkage step, non-linking records might result in a selection of the data that is incomplete and possibly biased. However, existing validation frameworks rarely address multiple sources linkage [10]. Only the RECORD statement, a reporting checklist for observational research using routinely-collected health data [11], and a guideline for the reporting of studies involving data linkage mention that the percentage of linked records should be provided [12].

Existing EHR data quality assessment frameworks all have different approaches, with partly overlapping dimensions or components of data validity. Fundamental dimensions that in some form occur in most frameworks are: completeness, correctness and currency [3].

Although all of these frameworks list important aspects that should be reported, it is rarely mentioned how these aspects should be verified or appraised.

In this paper we aim to further standardize the process of data validation. To this end we developed a practical approach for assessing various dimensions of EHR data quality, directed specifically at linked data from multiple providers. The approach is applied to the Dutch transfusion data warehouse [13] and explicitly shows how to assess the validity outcomes. Thereby we provide a detailed example of how this type of data can be validated systematically. We hope that this will increase awareness among researchers of the importance and benefits of structured data validation.

Methods

Validation approach

The validation approach starts with selecting validity concepts from previous literature and applying these to the data. First, existing frameworks were identified in the literature, from which then relevant concepts were selected, and finally, the concepts were operationalized in terms of the final application. Each of the different steps are depicted in Figure 3.1 and further described below.

Identification of validation frameworks in literature

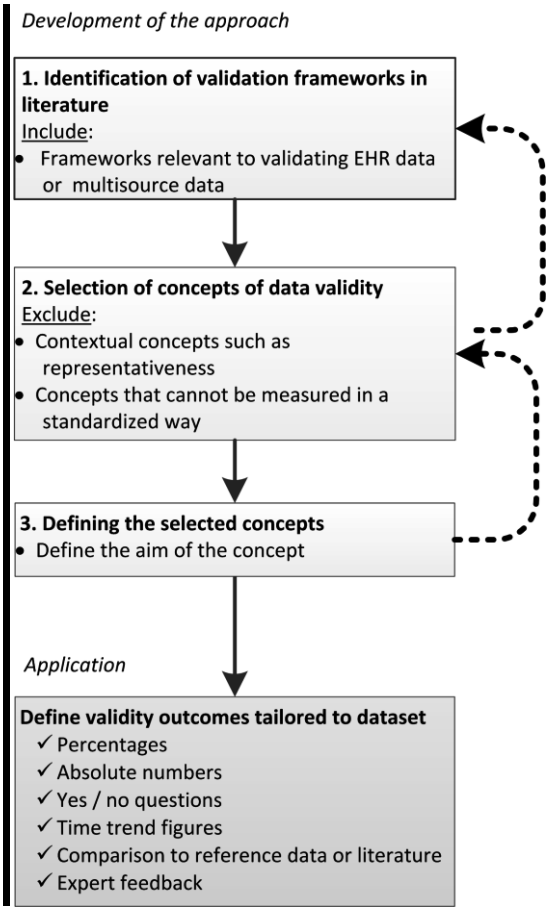
Previous frameworks on the validation of EHR data were identified in literature using the search terms 'data validation', 'data validity' or 'data quality', separately and combined with 'electronic health record', 'routine patient care record', 'routinely collected (health) data', 'routine administrative healthcare data', 'hospital registry data', 'joint registry data', 'linked data', 'administrative database', and via examining the references in those papers, until no new concepts seemed to emerge. From this literature, we selected those frameworks that might have relevance to EHR or to linked (multisource) data. In total, six data quality frameworks and two reporting guidelines were selected from literature [3,8,11,12,15,16,17,18,19].

Selection of data validity concepts

From the frameworks identified, data validity concepts were selected that were applicable to validating EHR data from multiple sources. Excluded were contextual concepts that differ for each research question such as currency, timeliness, representativeness, relevance, appropriate amount of data, and accessibility; these contextual concepts might eventually be addressed at a later stage. Also excluded were concepts that can only be assessed by manually reviewing the original medical records; instead only concepts that can be assessed in a more standardized way were included.

The following concepts were included in our approach (between brackets the number of frameworks that include these concepts in some form): External concordance (3), Linkage (3), Identity (7), Completeness (3), Uniformity (4), Time patterns (2), Plausibility (6) and Event

Figure 3.1. Development of the validation approach. First, validity concepts are identified, selected and defined. Second, concrete validity outcomes are established, tailored to the specific application or dataset.



attributes (1). The concept External concordance was split into four separate concepts: External concordance with (annual) reports from related organizations, External concordance with earlier findings in literature, External concordance with external clinical registries or databases, and External concordance with expert feedback. Also, because of the multisource character of our application, we added the concept Consistency of hospitals within the data warehouse. All selected validity concepts are depicted in a step-wise approach (Figure 3.2). The data warehouse is depicted in the center, with arrows leading to the concepts. Broadly, the concepts and outcomes can be categorized as either external or internal. External concordance (depicted in the upper part of Figure 3.2) is the agreement between aggregated numbers in the data warehouse and external sources. For example, the numbers in the data warehouse can be compared to (annual) reports from related organizations. Likewise, earlier findings in literature

can be used, or external clinical registries or databases might be available for comparison. Finally, the numbers and findings can be checked by presenting them to experts in the field. Internal consistency outcomes (depicted in the lower part of Figure 3.2) use expectations of what are considered valid values, often within one data source, or valid relationships between and within variables. Internal consistency concepts are: Linkage of entities occurring in multiple data tables within the data warehouse, Identity, Completeness, Uniformity, Times patterns, Plausibility, Event attributes and Consistency of results between hospitals within the data warehouse.

Defining the selected concepts

For each concept selected the aim was defined, i.e. what would be perfect validity in terms of this concept (Table 3.1 under the column header 'Aim'). In addition, an order is suggested in which to check the concepts that is efficient in identifying errors in the data (numbered steps 1-12 in Figure 3.2). In general, one would start with concepts that are relatively easy to check, and end with concepts that require further processing of the data. Applying this general logic, the first step is to start with External concordance of the raw numbers in the data warehouse as compared to external data from for instance an annual report (Step 1 in Figure 3.2). If crude numbers are incorrect a return visit to the data provider is necessary to check whether the correct data can be provided. In Step 2, it should be ensured that entities occurring in multiple data tables can be linked, and it must be decided whether records that cannot be linked will be excluded or not, before the data on the other concepts are validated. Next the application of the Identity (Step 3) and Completeness (Step 4) concepts is straightforward: The requested variables should be present, ideally have no missing values and single entities or events should be unique. If the dataset is incomplete or in case duplicates exist, this might bias the other validity outcomes. When data have no duplicates and is as complete as possible, the remaining Internal consistency concepts can be checked: Uniformity, Time patterns, Plausibility and Event attributes (Step 5-8). The Uniformity concept (Step 5) checks and ensures that measurements across time and departments all have the same units or duration. This is especially important for diagnoses and procedures; ideally hospitals should use similar coding systems, with the same level of detail, and use them in the same way also over time. Time patterns (Step 6) within one variable or linkage patterns between multiple variables might reveal the occurrence of registration or extraction errors through large gaps or unexplained changes that occur over time. The Plausibility concept (Step 7) examines the data on identifiable errors, using expectations of relationships between variables to check the accuracy of measurements, for instance the accuracy of date and time values. The Event attributes concept (Step 8) requires that for each event (e.g. a hospitalization or procedure) all relevant attributes are present (e.g. measurements). Finally, when validity outcomes have been

Figure 3.2. The validation approach. The approach distinguishes external validation concepts (upper part) and internal validation (lower part) concepts. The numbers indicate a suggested order in which to check to concepts in order to efficiently identify errors in the data.

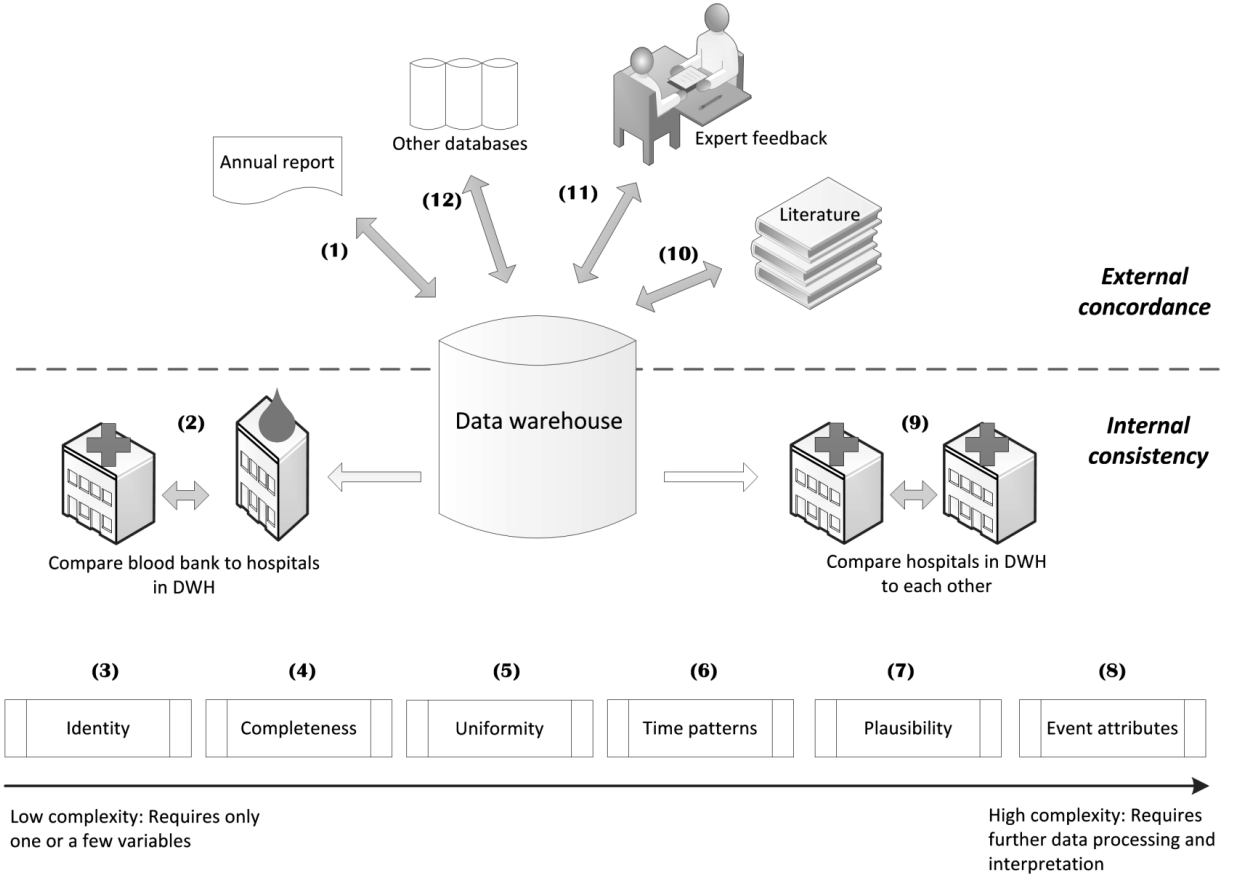


Table 3.1. Concepts and applied validation outcomes for n=2 hospitals from the DTD. DWH = data warehouse

Validation Approach			Application	
Order	Concept	Aim	Outcome	Average of two hospitals
External				
1	Concordance with report	Data are concordant with (annual) report	% agreement between number of products in annual blood bank report and DWH	98.7% (RBC 99.2%, PLT 97.6%, FFP 98.7%)
10	Concordance with literature	Data are concordant with previous findings in literature	Comparison of distribution of blood products by age and gender per product type in the Netherlands	Distributions were quite similar, only platelet use has shifted towards younger patients.
11	Concordance with experts	Data are concordant with expert opinions; findings can be explained in a clinical context	Plausibility of changes in Hb after blood transfusion	The experts concluded that the plausibility is acceptable; the 1% unexpected decreases might be explained by other factors.
12	Concordance with other databases	Findings are concordant with other databases	Comparison of findings with SCANDAT, a Scandinavian transfusion database	The SCANDAT database has similar external concordance, completeness and linkage rates.
Internal				
2	Linkage data sources within DWH	Entities occurring in multiple data tables can be linked	% transfusions linked to issued products by id of the end product	99.96% (no link for n=46 RBC, n=5 PLT, and n=1 FFP)

			% products issued linked to transfusion (indicates spilling rate)	97.65% (RBC 97.95%, PLT 99.25%, FFP 93.35%)
			% products that can be linked to donation(s); % products linked to donors	Initially 96.73%, after improving the donation numbers this increased to 99.99%; the link from product to donor was 99.98%
3	Identity	No duplicates	% duplicated transfusions (donation identification code + product type)	0.14% (initially this was 1%; it turned out that most duplicates were split products. Due to unavailable product codes in one hospital, the broader product type had to be used)
			% duplicated donations (donation identification code + product code)	0.005% (RBCs); 0% (FFP and PLT)
			% duplicated procedures codes	0% (all duplicates were removed, because it was expected that double registration would occur)
4	Completeness	No missing variables or values	% patient ID; date of birth; gender, procedure date; Hb and Thrombocyte counts; product code	100%; 99.99%; 99.99%; 100%; 99.8% and 97.5%; 50%
			% non-missing values for donor ID; date of birth; gender, Hb value, Expiration or Production date	100%; 99,995%; 100%; 98.8%; 100%

			% of transfusions that fall within at least one diagnosis start and end date	98%
5	Uniformity	Measures across time and data sources all have the same units and level of detail	% product codes that occur in the reference list of ISBT product codes	50% (for one hospital product code was not available)
			% Diagnosis codes that occur in the reference list (of national diagnosis codes and descriptions)	96.1%
			% of Hb measurements from hospitals and blood bank with the same level of precision	>99.6% uses 1 significant decimal in all sources
6	Time patterns	No unexplained changes over time	Compare number of donations, products and donors of subsequent (calendar) years	The observed decrease is in line with the known nationally decreasing trend.
			Examine number of transfusions per year per product type	The relatively high decrease for FFP use can be explained by the introduction of ROTEM, a hemostasis testing method.
			Examine linkage percentage of transfusions to products issued per year	In 2010 relatively many unlinked transfusions occurred. After blood bank data from the previous year 2009 was included, the linkage percentage increased to 99.8% or higher for all years.
7	Plausibility	Data are free of identifiable errors	% donation date < date of pooling	100%

% within limits for number of donations per donor per year (maximum is 3 (females) or 5 (males) for whole blood and 23 for plasma)	FFP 100%; WB 99.8% (0.2% exceeds the limit with in total 6 or 7 donations within a year)
% donor age > 18 and > 70 years (minimum and maximum age for donating)	100% (only 0.0006% was >70 and 0.0004% was <18 and these were mainly autologous donations)
% transfusion with increase (and decrease) in Hb level (Hb values +- 1 day around transfusion; difference >+- 8.8% is considered a clinical change)	54% increases; 6% decreases; 40% no change. Of those decreasing, 97% had a diagnosis indicating high bleeding risk
% patient age < 121 years	100%
Maximum number of transfusions per year	Max tr. per year 476 (mainly FFP) for diagnosis TTP.
% correct gender for Gynecology diagnoses	100%
% patients with transfusions/ surgery after date of death	0.0% (n=2 changed mortality status to NA)
% with admission date before discharge date) (zero-length rule)	100%

			% with non-negative difference between expiration and transfusion date	99.93%
8	Event attributes	All attributes relevant to an event description are present	% of pooled products that are linked to the correct number of unique donors (in this case 5 or 6 donors contribute to one pooled platelet product)	100%
			% of patients that are transferred to another hospital according to the 'discharge destination' variable	6%
			% transfusions linked to hospitalization (indicates outpatient transfusions)	99.16% (of which 23.64% day admissions, likely including transfusions given at the outpatient ward)
9	Consistency hospitals within DWH	No unexplained differences between hospitals	Comparison of (validity) outcomes of the hospitals	The two hospitals have very similar outcomes.

computed for various centers within the data warehouse, the observed differences between hospitals from within the data warehouse can be compared (Step 9). This will either support the validity of the data when the outcomes between centers are consistent, or might indicate errors in the data or findings; unexplained differences between centers might warrant further investigation.

After the Internal consistency concepts have been checked and -if necessary- improved, the final External concordance outcomes can be computed. These outcomes require some preliminary analyses to be done on the data, and then comparing the findings to previous literature (Step 10) and discussing the results with clinical experts (Step 11). Finally, the resulting validity outcomes can be placed in context by comparing them to similar databases, if available (Step 12).

Example case

Application

To apply the concepts to a specific dataset, tangible outcomes per concept need to be defined, preferably in quantifiable terms such as percentages or absolute numbers. Outcomes can also be yes/no questions, time trend figures, comparisons to reference data or literature, or expert feedback. As the outcomes are tailored to the specific application, this step of defining the exact outcomes has to be repeated for each unique dataset.

Table 3.2. Hospital characteristics (for the year 2014)

	Hospital A	Hospital B
Number of beds	1,100	471
Annual number of RBC transfusions	12,653	6,681
Presence of typical transfusion specialisms	Hematology, oncology, thoracic surgery, trauma center	Hematology, oncology, thoracic surgery, trauma center (heavy emphasis on major vascular / aneurysm surgery and obstetrics)

Data

Data on blood transfusion were used as an example case to illustrate the application of the validation approach. These data were collected in the context of the Dutch Transfusion Data warehouse (DTD) project [13], in which data from the national blood bank on blood donors and products are linked to patient data from two teaching hospitals for the period 2010-2014 (see Table 3.2 for hospital characteristics). Both hospitals use the national Blood transfusion Guideline [14]. Variables include the three most important blood products that are transfused: red blood cell (RBC), fresh frozen plasma (FFP) and platelet (PLT) products, as well as clinical information on transfusion recipients such as diagnoses, surgeries and laboratory

measurements. After collection, these data need to be validated in order to create a valid transfusion data warehouse that can be used for research purposes. As this dataset involves multiple centers and linked donor-recipient data, it is especially suitable to serve as an example case to illustrate the validation approach. In order to keep the Results table manageable, the average outcomes of the two teaching hospitals are shown.

Results

The validation approach was applied to the transfusion dataset, so that each concept was assessed by one or more outcomes (Table 3.1). A selection of outcomes that demonstrate how the validation process led to improvements of the data is discussed in more detail below. A more extensive discussion of all validity outcomes can be found in Appendix A.

We first checked the agreement between the number of blood products in the dataset and those reported in the annual blood bank report (Step 1: External concordance with report), which was 98.7%. The slight disagreement can be explained by potential differences in the way of counting composite and split blood products. Of the transfused products, initially 96.7% could be linked to the corresponding donation (Step 2: Linkage). We traced this difference back to a post-hoc modification in the coding of the product identification number at the blood bank, leading to different codes existing in the blood bank and the hospital system for the same product. When the coding was adjusted, the proportion linked products increased to 99.98%. Initially 1% of products were duplicated (Step 3: Identity). Investigation of product types revealed that most duplicates were split products. The products were given unique identifiers post-hoc, resulting in an improved duplication percentage of 0.14%. Most transfusions -98%- could be linked to one or more diagnoses (Step 4: Completeness). In most cases the diagnosis was even more than complete: the number of pending diagnoses ranged up to 15 diagnoses per transfusion. This means that it will be necessary to make a selection of those diagnoses in the future if we want to determine the main indication for a transfusion. Diagnoses were defined differently by the two hospitals and therefore had to be recoded using a uniform reference table (Step 5: Uniformity). The percentage of diagnosis codes that could be linked to the reference table was 96.1%. Investigation of time patterns (Step 6: Time patterns) revealed that for the year 2010, an exceptionally high percentage of transfusions could not be linked to products issued (2.2% versus 0.07% in other years). This percentage could be lowered to 0.17% by including blood bank data from the previous year 2009 (the unlinked products were mainly frozen plasma products that were issued in the year before the actual transfusion). Plausibility of registered dates and times (Step 7: Plausibility) was checked based on a priori expectations. For example hemoglobin (Hb) generally is expected to increase after transfusion. Indeed, in 54% of cases Hb did increase, 40% did not clinically change, however 6% decreased. This decreasing 6% might indicate incorrect date values, which could occur when the registered time of transfusion actually records the moment that a product or service (e.g. the blood product) was requested instead of administered. To check this, expert

feedback was asked regarding the plausibility of the observed Hb changes (Step 11: Concordance with expert feedback). Further investigation of the data showed that most recipients with a decrease in Hb had a diagnosis indicating high bleeding risk (87%), explaining the observed decrease. Taking this into account, the percentage of all transfusions with an unexplained decrease is lower than 1%, which according to the experts is acceptable. Event attributes (Step 8) include that each platelet product is attributable to five or six unique donations, which was also found in the dataset for 100% of platelet products.

These are all average outcomes, however a comparison of the two hospitals included shows that their validity outcomes were very similar (Step 9: Consistency of hospitals within data warehouse; results not shown), supporting the validity of the findings. Also, concordance with literature (Step 10) was checked by comparing the distribution of blood products over age and gender per product type with the previously reported PROTON study, of which the DTD is the successor. Distributions were very similar but platelet use has shifted towards older patients, especially men aged 60-80 years (Supplementary file 2). This can be explained in part by the ageing of the population and changes in policy in the past ten years; platelet use was increased in thorax surgery and hematological disorders, which both are more prevalent in men.

Lastly, the concordance of these findings with validity outcomes reported for other databases was investigated (Step 12: Concordance with other databases). The most extensive list of validation outcomes were reported by the SCANDAT study [19,20], therefore, these outcomes are shown next to the validity outcomes of the DTD (Table 3.3). SCANDAT and the DTD show similar results regarding the high external concordance of the data with external statistics and the fact that both studies identified missing data by investigating time patterns. Different is the proportion of hospitalized patients, which might be due to differences between the countries in the registration of patients (we found a consistently higher hospitalization rate for both of the DTD hospitals included). The estimated proportion of patients with incomplete information due to transference to another hospital was up to 6% for the DTD. This might actually be an underestimation, considering the finding that in SCANDAT 8.9% of recipients received a blood transfusion in two or more local registers, and because our 6% did not include patients who were hospitalized elsewhere prior to being hospitalized in our included hospitals.

Discussion

Being explicit about research methods is important for the reliability, reproducibility and credibility of research, and in our opinion this includes being explicit about data validity. We recommend that any study that involves electronic health record data should include an overview of the steps taken to ensure data validity. This applies in particular to more complex routinely registered data that are not designed for research purposes or are complex (for example data covering an extensive time period or linkage of several sources). Therefore, we

Table 3.3. Comparison of validity outcomes in the SCANDAT study and the current DTD results

Outcome	SCANDAT 1/2	DTD example
External concordance of database and official statistics on the number of transfusions	>97%	>98.7% for products and 99.96% for transfusions
% transfusions linked to the corresponding donor	95%	99.99%
% transfusions linked to hospitalization	88.7%	99.2% (of which 23.6% day admissions)
% duplicated donations and transfusions	4.9% (donations) and 9.1% (transfusions)	0% (donations) and 0.14% (transfusions)
% missing or invalid values for identification number or date values	Range between 0.1% to 3.6%	0%-0.01%
Time patterns for donations and transfusion counts	In one year approximately 160,000 transfusions were missing; it took two years for the number of donations and transfusions to stabilize after the start of a new registration system	In one year the link of transfusions to products could be made for 2.2%, however this could be improved by adding donation data from the previous year
% of recipients had records of receiving a blood transfusion in two or more local registers	8.9%	6% of patients are transferred to another hospital

documented a step-wise approach for systematically validating multiple source data from electronic health records. The approach integrates concepts from existing guidelines for EHR quality assessment with the specific challenges inherent to linked, multisource data. The proposed approach is practical as the validity concepts are directly related to what is needed for actually carrying out the validation. For some validation steps, data from only one center and of a single point in time might be sufficient, other steps require at least two centers or data from prolonged periods of time, or require the availability of external information such as previous literature or expert opinion.

The approach was applied to data from the Dutch Transfusion Data warehouse (DTD), resulting in an overview of validity outcomes and improvements of the quality of the data. In addition to improving the data, the validity outcomes, if made publicly available, increase transparency and contributes to the efficient use and reuse of existing sources of

information. A clear overview of data validity is informative for researchers who want to use an existing database and, vice versa, requests for using the data can be evaluated more easily.

What is 'good' quality data?

It can be argued that no universal cut-off values exist as to whether electronic health data are 'valid' or of 'high quality' [26], since the level of data quality required will also depend on the purpose of the study concerned. Still, objective measures for relevant quality concepts are necessary, and an adequate understanding of how to interpret validity outcomes is desirable.

Levels of validity as found in earlier studies might set a -more or less arbitrary but realistic- standard. Validation outcomes reported by previous transfusion data warehouse studies are sparse and vary greatly. Most often reported was the linkage rate of transfusions to donors, varying between 92%-99% [20,21,22,23,24] and, vice versa, estimates of wastage of blood products (i.e., issued but not transfused) of 1.3% and 7.7% [21,22]. The percentage missing values was also reported by some studies: clinical variables were missing for 13% (post-transfusion Hb), 14% (ASA code) [25], and 20% (specialty), with the degree of missingness varying between specialties from 2% to 47% [26] (a more extensive overview per transfusion database is provided in Supplementary file 3). Comparing the SCANDAT to the DTD outcomes, we found similar results regarding the completeness measures and the high external concordance of the data with external statistics. In this context, we think the data from the DTD shows sufficient validity. This is, however, time-bound; when new data (either hospital or external) are included in the future, these data must be validated as well.

Although data quality should ideally be checked continuously, such checks may be particularly relevant following: the first data extraction from a hospital or other data source, inconsistencies present in the data, the introduction of a new hospital information system, or reorganization or fusion of hospitals [1]. In order to keep track of changing registration systems in individual hospitals, a questionnaire might be submitted once a year on new developments in the hospital that could impact the registration or data extraction. Especially for the purpose of benchmarking hospitals (or other data sources), it is important that all selection and interpretation steps are performed in a similar way for each hospital. After all, the impact on the comparison is minimized when bias is similar for each hospital.

Remaining issues and future directions

The concept Accuracy or Correctness -which is particularly relevant for diagnosis and procedure codes- was not yet covered by the approach. As the data warehouse is too large to manually check accuracy of diagnoses for each individual patient, a sample of diagnoses and procedures will be validated by manual review of patients' local medical records.

It must be noted that it is a choice whether to aggregate the validation outcomes to the level of patients, variables, hospitals or even combining all sources in the database. Outcomes become less informative for higher aggregation levels, but are still useful for detecting large irregularities in the data. To simplify the interpretation of validity outcomes, we

encourage adding visualizations or summaries (especially when outcomes are similar), for example, Completeness could be summarized by giving the percentage of variables that is at least 95% complete [15].

It appears that the process of preparing data for analysis, including data harmonization and assessing data quality, is commonly taking place at the intersection of data management and research. In practice, part of the validation may be performed by the data manager. However, it is the researcher who ultimately is responsible for making and communicating any choices made in this process, and the implications for the validity and interpretability of study results.

Conclusion

The proposed approach provides a structure for validating multisource EHR data. By making the validation steps explicit and concrete, the applied example shows that the approach is feasible, enhances the interpretation of the data and improves data quality. Hopefully this will encourage researchers to consider and report data quality, in a way that goes beyond conceptual classifications and allows transparent assessment of the potential impact of data quality on research findings.

References

1. Cook JA, Collins GS. The rise of big clinical databases. Br J Surg. 2015;102:e93-101.
2. Kleinman S, Glynn SA. Database research in transfusion medicine: The power of large numbers. Transfusion. 2015;55(7):1591-1595.
3. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc. 2013;20.1:144-151.
4. Fischer C. Quality indicators for hospital care: reliability and validity. (Doctoral thesis, Erasmus Universiteit Rotterdam, Rotterdam, The Netherlands). Available from: http://repub.eur.nl/pub/79066/
5. Callen J. What is the impact of electronic health records on the quality of health data?. Health Information Management Journal. 2014;43(1):42.
6. Chan KS, Fowles JB, Weiner JP. Electronic health records and reliability and validity of quality measures: a review of the literature. Medical Care Research and Review. 2010.
7. Weiskopf NG, Rusanov A, Weng C. Sick patients have more data: the non-random completeness of electronic health records. In: AMIA. 2013.
8. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. Medical care. 2013;50.
9. Goldstein H, Harron K (2016). Bias in data linkage studies. In: Methodological Developments in Data Linkage. 2016; Ch.4.
10. Brown J, Kahn M, Toh S. Data quality assessment for comparative effectiveness research in distributed data networks. Medical care. 2013;51(8 0 3):S22.
11. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I; RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015;12(10), e1001885.
12. Bohensky MA, Jolley D, Sundararajan V, Evans S, Ibrahim J, Brand C. Development and validation of reporting guidelines for studies involving data linkage. Australian and New Zealand journal of public health. 2011;35(5):486-489.
13. van Hoeven LR, Hooftman BH, Janssen MP, de Bruijne MC, de Vooght KM, Kemper P, et al. Protocol for a national blood transfusion data warehouse from donor to recipient. BMJ open. 2016,6(8),e010962.
14. Quality Institute for health care (CBO). [Blood transfusion guideline] Richtlijn Bloedtransfusie. Alphen aan de Rijn: Van Zuiden Communications; 2004.
15. Wang R, Strong D. Beyond accuracy: what data quality means to data consumers. J Manag Inf Syst. 1996;12:5–34.

16. Maydanchik A. Data Quality Assessment. Bradley Beach, NJ: Technics Publications. 2007.
17. Van den Broeck J, Cunningham SA, Eeckels R, Herbst K. Data cleaning: detecting, diagnosing, and editing data abnormalities. PLoS med. 2005;2(10):966.
18. Davoudi S, Dooling JA, Glondys B, Jones TD, Kadlec L, Overgaard SM, et al. Data Quality Management Model (2015 Update). Journal of AHIMA. 86 2015;10.
19. Black N, Payne M. Directory of clinical databases: improving and promoting their use. Quality and Safety in Health Care. 2003;12(5):348-352.
20. Edgren G, Hjalgrim H, Tran TN, Rostgaard K, Shanwell A, Titlestad, et al. A population-based bi-national register for monitoring long-term outcome and possible disease concordance among blood donors and recipients. Vox Sanguinis. 2006;91:316-23.
21. Edgren G, Rostgaard K, Vasan SK, Wikman A, Norda R, Pedersen OB, et al. The new Scandinavian Donations and Transfusions database (SCANDAT2): a blood safety resource with added versatility. Transfusion. 2015;55(7):1600-1606.
22. Tinegate H, Chattree S, Iqbal A, Plews D, Whitehead J, Wallis JP. Ten-year pattern of red blood cell use in the North of England. Transfusion. 2013;53(3):483-489.
23. Allden RL, Sinha R, Roxby D J, Ireland S, Hakendorf P, Robinson KL. Red alert–a new perspective on patterns of blood use in the South Australian public sector. Australian Health Review. 2011;35(3):327-333.
24. Palo R, Ali-Melkkilä T, Hanhela R, Jäntti V, Krusius T, Leppänen E, et al. Development of permanent national register of blood component use utilizing electronic hospital information systems. Vox Sanguinis. 2006;91:140–147.
25. Palo R. Epidemiology of blood component use in Finland [dissertation]. Helskini: Helsinki University Hospital. 2013.
26. Barr PJ, Donnelly M, Morris K, Parker M, Cardwell C, Bailie KEM. The epidemiology of red cell transfusion. Vox Sanguinis. 2010;99(3):239-250.

Supplementary file references

27. Burtis CA, Ashwood ER, Bruns DE. Tietz textbook of clinical chemistry and molecular diagnostics. 2006 4th Edition. Elsevier Health Sciences, Ch.17:p.469.
28. Ricos C, Alvarez V, Cava F, Garcia-Lario JV, Hernandez A, Jimenez CV, et al. Desirable specifications for total error, imprecision, and bias, derived from intra- and inter-individual biologic variation. Updated in 2014. Available from: https://www.westgard.com/biodatabase1.htm
29. Borkent-Raven BA, Janssen MP, Van Der Poel CL, Schaasberg WP, Bonsel GJ, Van Hout. The PROTON study: profiles of blood product transfusion recipients in the Netherlands. Vox sanguinis 2010; 99(1):54-64.

Chapter 4

Why was this transfusion given? Identifying clinical indications for blood transfusion in health care data

Submitted

L.R. van Hoeven
A.L. Kreuger
K.C.B. Roes
P.F. Kemper
H. Koffijberg
F.J. Kranenburg
J.M.M. Rondeel
M.P. Janssen

Abstract

Introduction

To enhance the utility of transfusion data for research, ideally every transfusion should be linked to a primary clinical indication. In electronic patient records, many diagnostic and procedural codes are registered, but unfortunately it is usually not specified which one was the reason for transfusion. Therefore, a method is needed to determine the most likely indication for transfusion in an automated way.

Methods

An algorithm to identify the most likely transfusion indication was developed and evaluated against a gold standard based on expert review of 234 cases. In a second step, information on misclassification was used to fine-tune the initial algorithm. The adapted algorithm predicts, out of all data available, the most likely indication for transfusion, using information on medical specialism, surgical procedures, and diagnosis and procedure dates relative to the transfusion date.

Results

The adapted algorithm was able to predict 74.4% of indications in the sample correctly (extrapolated to the full dataset 75.5%). A kappa score, which corrects for the number of options to choose from, was found of 0.63. This indicates that the algorithm performs substantially better than chance-level.

Conclusion

It is possible to use an automated algorithm to predict the indication for transfusion in terms of procedures and/or diagnoses. Before implementation of the algorithm in other datasets, the obtained results should be externally validated in an independent hospital dataset.

Introduction

In blood transfusion research, it is important to know the clinical condition of the patient that motivated the physician to give the transfusion. Hereby the underlying disease or treatment is meant, not the immediate reason for transfusion such as low blood values. Typically, the reason for transfusion or transfusion indication, is not routinely registered with the request for blood [1], or only temporarily in the context of a study [2,3]. Alternatively, routinely registered diagnostic and procedural codes can be used to determine the transfusion indication retrospectively. Potentially the primary diagnosis code can be used, but this code is not always available and it has been shown that the primary diagnosis in many cases is not the indication for transfusion. For example in the EASTR study [4], transfused patients were assigned to indication groups based on surgical procedure (43% of all patients), ICD-10 primary diagnosis (36%) and ICD-10 secondary diagnosis (12%), with 9% of patients remaining unclassified. As each electronic patient record was reviewed by research nurses, this is a highly labor-intensive method for determining the transfusion indication.

The amount of data in transfusion research databases generally is very high, including many patient records as well as many diagnoses and procedures per record. Many patients have two or more diagnoses at the time of transfusion. As manual review is too time-consuming, an algorithm is needed that predicts, out of all diagnoses and procedures registered, the most likely indication for transfusion. No method currently exists for determining the main indication for transfusion in an automated way. Reported strategies for attributing transfusion events to patients' clinical information are: to classify a transfusion as either medical, surgical, or obstetric and gynecological; or to attribute a transfusion to the requesting hospital department, specialism or admitting service [1,2,5,6]. Although these broad classifications provide information on the global indication, they do not identify an indication as specific as a diagnosis or procedure. In addition, this information is not always reliable because patients may be transferred to other specialisms during hospitalization. Also, the accuracy of the used coding system has to be validated, as reportedly patients with a diagnosis code do not always have the condition it represents [7]. Therefore, indications for transfusion should be validated, but in practice often they are not [8].

The aim of this study is to establish an algorithm for the automated identification of transfusion indications and to evaluate the performance of this algorithm using expert evaluations. In addition, in order to validate the diagnostic and procedural codes, we checked transfusions with a single diagnosis as well as transfusions without any diagnostic information registered, to investigate whether the diagnosis was the most likely indication and whether information was missing in a systematic way.

Methods

In order to develop the initial version of the algorithm, expert opinions were asked and used to establish the rules of the algorithm. Subsequently, the algorithm was applied to hospital data, with the result that each transfusion was linked to its predicted most likely indication for transfusion. To evaluate the performance of the algorithm, the predicted indications were compared with the gold standard, which is the most likely indication for transfusion as determined independently by two clinical experts based on the medical record.

Data

We used data on all transfusions in a teaching hospital in the Netherlands over a 5-year period (Isala hospital; 2010-2014; n=86,043). In the Netherlands, diagnoses and procedures are currently coded using a national system adopted by all hospitals and are primarily for financial reimbursement [9]. Each diagnosis code consists of a specialism code and a more specific diagnosis code. For the current study, all diagnoses of a patient that were pending at the time of transfusion were selected as potential transfusion indications. Similarly, all surgical procedures falling within the hospitalization in which the transfusion was administered were selected, with an extra filter to include only procedures within a time interval of -7 and +1 days around the transfusion date. Procedures were linked to a specialism using the admitting specialism. Diagnoses that were registered at the same day under the same specialism as well as surgeries that were performed at the same day under the same specialism, were regarded as related and are clustered together into 'diagnosis clusters' and 'procedure clusters'.

Determination of the gold standard by reviewers

A sample of transfusion cases was manually reviewed by two reviewers (both medical doctors) to determine the true indication for transfusion, i.e. the gold standard. The indication always consists of a diagnosis, optionally complemented by a procedure. This was decided firstly because it reflects the way that this type of information is registered, and secondly because it makes sense that some transfusions are necessitated by a disease (i.e. diagnosis) and other transfusions are necessitated by a treatment (i.e. procedure) in the follow-up of a disease (i.e. diagnosis).

The reviewers were given a list of patient identification numbers and transfusion events (i.e. a transfusion date and the type and amount of products transfused). They were instructed to first look into the hospital electronic patient documentation (including electronic health records, correspondence and clinical outcomes), and determine the most likely indication for transfusion. After they had determined the indication, they looked at the answer categories provided (i.e. diagnoses and procedures), and selected the correct indication if available, and otherwise selected the option 'none of these'. A free text field was provided to fill in the correct indication. In order to enhance the reliability of the gold standard, the two

reviewers reviewed all cases independently of each other and afterwards discussed the cases on which they disagreed until consensus was reached on the correct indication.

Sample selection

The sample of transfusion cases to be presented to the reviewers was selected using a stratified random approach. The sample was stratified by the six most important specialisms according to experts with high blood use: Cardiopulmonary surgery, Gynecology, Gastroenterology, Internal medicine, Surgery, and Orthopedics. All remaining specialisms fell into the seventh category 'Other'. The reason for stratifying by specialism was that this ensures a sufficient number of observations per specialism and will provide information as to whether the algorithm works for the most prevalent indications. As we did not have information on the true specialism yet, we stratified by specialism as predicted by the initial version of the algorithm.

The sample size was based on the expected performance of the algorithm and the predefined, acceptable margin of error. For instance, when the algorithm is expected to correctly identify the transfusion indication in 90% of cases, a sample size of n=138 would be required (assuming power of 0.9 and alpha of 0.05); for an expected proportion of 80% this is n=246. We chose for a sample size of 234 cases, resulting in a margin of error of 3.8% for an expected proportion of 90% and a margin of error of 5.1% for 80%. The sample was divided over nine strata: the seven specialism strata, a 'data quality check' stratum of cases with only one diagnosis, and a stratum of transfusions that are not linked to any diagnosis or procedure at all (which is 3.1% of transfusions in the dataset). These last two strata were also used to validate the data: to check whether the single diagnosis was the correct indication, and why diagnoses were missing and whether there was a pattern (for example, the absence of a diagnosis might be specific for certain specialisms).

Development of the algorithm

The core of the algorithm is a set of decision rules. Using these rules the algorithm selects a diagnosis and if available a procedure as the most likely transfusion indication. These rules are based on the recommendations of transfusion experts (two medical doctors, and a doctor working in transfusion medicine). The experts were asked to make a list of the most likely specialisms to be responsible for transfusion. Similarly, experts were asked to make a prioritization based on the time between the diagnosis and procedure dates relative to the transfusion date. Based on these prioritizations, a decision tree algorithm was made and applied to the data. In a second step, the algorithm was adapted based on the performance of the initial algorithm.

Statistical analysis

The performance of the algorithm was evaluated by computing the percentage agreement between the algorithm and the gold standard. This was computed for the sample and also

extrapolated to the full dataset based on the sampling fraction per stratum (for the final algorithm). A secondary outcome was a quantification of the agreement corrected for agreement by chance. This score indicates whether and how much the algorithm performs better than random chance. The algorithm chooses between several diagnosis clusters but, due to the nature of the data, the number of diagnosis clusters varies per case. When for example the choice is between two diagnosis options, the chance-agreement is 50%, whereas a case with 6 diagnosis options has only a 16.7% chance-agreement. Likewise, a guessing probability equal to 1 divided by the number of outcome categories was assigned to each case. The underlying assumption was that a priori all outcome categories are equally likely. With this information, the total expected chance-level agreement can be calculated and compared with the observed proportion agreement. These two measures are used to calculate kappa (k), an agreement statistic similar to Cohen's kappa [10]. The formula for kappa is:

$k = (a_{obs} - a_{chance}) / (1 - a_{chance})$, where
a_{obs} = number of cases predicted correctly / total number of cases, and
$a_{chance} = \sum$ (1 / number of outcome categories) / total number of cases

The interpretation of kappa is as follows: kappa=0 indicates chance-agreement, kappa=1 indicates perfect agreement, kappa<0 indicates lower than chance-agreement, and kappa >0 indicates agreement better than random chance. The advantage is that kappa corrects for the level of difficulty of each case, by taking into account the number of outcome categories.

As a sensitivity analysis, the data were analyzed both including and excluding cases lacking a gold standard. Evaluating only those cases for which a gold standard could be found in the data registered might be fairest because the algorithm, which always selects one indication, can never match with a missing gold standard. Conversely, an analysis that includes all cases better demonstrates the usefulness of the algorithm for the complete dataset.

Results

Gold standard: Inter-rater consistency

The two reviewers initially agreed on 223 out of 234 cases (95%). After discussion of the cases not agreed upon, consensus was reached for all cases. For diagnoses, the gold standard was 'none of these' for n=15 cases, meaning that none of the diagnoses provided was found likely to be the indication for the transfusion. Reasons that the indication could not be found in the available diagnoses varied: anemia of unknown cause (n=6), registration lacking (n=6), anemia of the critical illness (n=2), or no indication for transfusion according to the reviewers (n=1). For procedures, the gold standard was 'none of these' for n=14 cases. Reasons that no procedure was selected were: supposedly incorrect procedure dates (n=6, these seemed to be mainly non-elective surgeries), the correct procedure fell outside the selected time period of -7 and +1 days around the date of transfusion (n=4), missing registration of the correct

procedure (n=3), and based on medical chart review it was not possible to select one procedure (n=1).

As the predicted specialism strata were not always the true strata according to the gold standard, the number of observations in each specialism stratum changed somewhat (Table 4.1A). Also during the determination of the true indication, the reviewers discovered that some of the diagnoses from which the algorithm had to choose, were overlapping or the same (n=24). This probably occurred because the same diagnoses were registered by different specialisms, resulting in different codes for the same diagnosis (of these cases, n=16 had one duplicated diagnosis, n=7 had two, and n=1 had three duplications). These duplicated diagnoses were recoded post hoc to make sure that equal alternatives would also be evaluated as such. This slightly changed the number of observations in each stratum, as n=13 clusters went from multiple to only one diagnosis cluster and therefore moved to the 'data quality check' cluster. For procedures, transfusion clusters with only one procedure cluster were combined into a 'data quality check' stratum (n=47). The n=21 remaining cases with multiple procedure clusters were grouped together, as a breakdown by specialisms would result in very low sample sizes per stratum.

Table 4.1A. Agreement between initial algorithm and gold standard for diagnoses as observed in the sample (n=234). The raw % correct in the sample is shown by specialism and in total, showing cases with only one diagnosis option ('data quality check'), cases without a gold standard, and cases without any diagnostic information as separate strata. Kappa provides a measure for chance-adjusted agreement for cases with at least two diagnosis options.

Stratum (sample size)	% correct	Kappa
Cardiopulmonary surgery (n=19)	94.7%	0.91
Gynecology (n=12)	75%	0.57
Gastroenterology (n=15)	86.7%	0.78
Internal medicine (n=61)	44.3%	0.15
Surgery (n=18)	66.7%	0.50
Orthopedics (n=16)	75.0%	0.58
Other (n=15)	20.0%	-0.25
Total specialisms (n=156)	60.2%	0.37
Data quality check (n=37)	100%	
Specialisms + data quality check (n=193)	67.9%	
No codes registered (n=26)	0%	
No gold standard (n=15)	0%	
Total (n=234, incl. cases without diagnoses)	56.0%	

Table 4.1B. Agreement between initial algorithm and gold standard for procedures as observed in the sample (n=234). The raw % correct in the sample is shown in total, and separately for cases with only one procedure option ('data quality check'), cases without a gold standard, and cases without a procedure registered in the time selection. Kappa provides a measure for chance-adjusted agreement for cases with at least two procedure options, excluding cases without gold standard.

Stratum (sample size)	% correct	Kappa
Total specialisms (n=17)	82.4%	0.71
Data quality check (one procedure) (n=47)	100%	
Specialisms + data quality check (n=64)	95.3%	
No gold standard (n=14)	0%	
Total (n=234, incl. cases without procedures)	92.7%	

Initial algorithm

Rules of the initial algorithm

The algorithm works like a decision tree (Figure 4.1 and 4.2). First, it selects diagnoses based on the prioritization of the specialisms of the available diagnoses (Table 4.2). If the patient underwent a procedure, the diagnosis matching the admission specialism (of the department the patient is hospitalized) highest in priority (according to 'Order procedures' in Table 4.2) will be selected. If no procedure was registered for this patient, the diagnosis with the specialism highest in priority (according to 'Order diagnoses' in Table 4.2) will be selected. Second, if after this first selection there is still more than one diagnosis option available, the algorithm selects the diagnosis that is closest in time to the transfusion (using the start date of the diagnosis). For procedures, the algorithm selects the procedure closest in time to the transfusion (prioritizing procedures one day before transfusion over one day after transfusion).

Performance of the initial algorithm: Diagnoses

For n=208 out of the 234 cases, one or more diagnoses were registered. The overall percentage raw agreement of the algorithm diagnoses with the gold standard was 56.0%. Excluding cases without a gold standard and those without diagnoses registered, resulted in a higher agreement rate of 67.9%. Agreement varied per specialism from 20.0% (for Other) to 94.7% (for Cardiopulmonary surgery) (Table 4.1A). Chance-adjusted agreement (excluding cases without a gold standard) was 0.37, varying per specialism from -0.25 (Other) to 0.91 (Cardiopulmonary surgery). This means that the algorithm performs better than chance level, both overall and for each specialism individually (except for the stratum Other).

Figure 4.1. Diagnosis selection: initial algorithm rules visualized by a decision tree

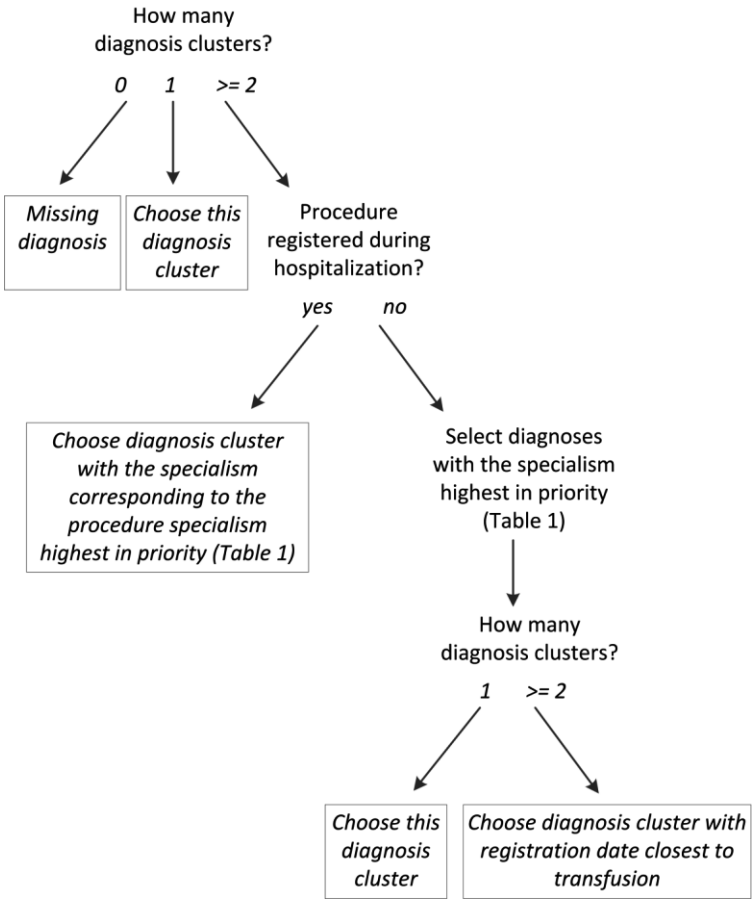


Figure 4.2. Procedure selection: initial algorithm rules visualized by a decision tree

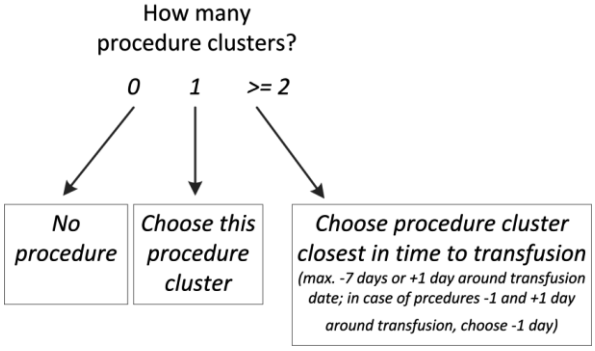


Table 4.2. Order of diagnosis and procedure specialisms used for attributing indication to transfusion in the initial algorithm and after adjustment in the adapted algorithm, from high to low priority.

Specialism	Order diagnoses	Update diagnoses	Order procedures
Cardiopulmonary Surgery	1	1	1
Gynecology	2	4	9
Gastroenterology	3	3	10
Internal medicine: Hematology	4	2	11
Surgery: Transplantation	5	5	2
Surgery: Vascular surgery	6	6	3
Surgery: Traumatology and first aid	7	7	4
Surgery: Oncology, lung and gastrointestinal surgery	8	8	5
Surgery: General surgery and pediatric surgery	9	9	6
Orthopedics	10	10	7
Urology	11	11	12
Anesthesiology	12	21	13
Neurosurgery	13	13	8
Throat Nose Ear	14	17	14
Plastic Surgery	15	12	15
Pediatrics	16	16	16
Consultative Psychiatry	17	18	17
Neurology	18	19	18
Cardiology	19	20	19
Internal medicine: Non-Hematology	20	14	20
Lung medicine	21	15	21
Ophthalmology	22	22	22
Clinical geriatrics	23	23	23
Radiotherapy	24	24	24
Dermatology	25	25	25
Rehabilitation medicine	26	26	26
Geriatric rehabilitation care	27	27	27
Rheumatology	28	28	28

Performance of the initial algorithm: Procedures

For n=78 out of the 234 cases, one or more procedures were registered. The overall percentage raw agreement of the algorithm with the gold standard was 92.7% (Table 4.1B).

Excluding cases without a gold standard, this was 95.3%. Chance-adjusted agreement, excluding cases without a procedure, was 0.71. This means that overall the algorithm performs substantially better than chance level.

Data validation

The transfusions that linked to only one diagnosis and/or procedure (the 'data quality check' stratum, n=26) corresponded to the indication according to the gold standard in 100%. The cases without any diagnoses registered (n=26) consisted mostly of neonates (n=25; 96.2%).

Adapted algorithm

Rules of the adapted algorithm

In the second step, the results of the initial algorithm, especially the misclassified cases, were used to adapt the algorithm (Figure 4.3). The following changes were made: Firstly, instead of selecting both a diagnosis and a procedure (if available) as the indication for transfusion, the adapted algorithm selects either a procedure or a diagnosis, prioritizing procedures over diagnoses. The reason for this is that in the gold standard a procedure, if present, was always selected as the indication. With this change in definition of the transfusion indication, the classification of cases into strata also changed somewhat (Table 4.3). Secondly, the prioritization of specialisms was adapted: diagnoses in Internal medicine-Hematology and Gastroenterology are prioritized over Gynecology, because the specialism Internal Medicine was misclassified relatively often (44.3% correct, Table 4.1A). In the adapted algorithm, Gastroenterology is only selected as indication if the patient underwent surgery under this specialism. In this way, a Hematology patient with a Gastroenterology diagnosis but no surgery who might not have had a bleeding, will be predicted to have had a Hematological indication for transfusion. In addition, both Internal medicine-Non-Hematology and Lung medicine are placed higher on the prioritization list (Table 4.2). Based on the results of the data validation, cases lacking any diagnostic information were classified by the algorithm as Neonatology. The R code of the algorithm is provided in Appendix B.

Performance of the adapted algorithm in the sample

The overall percentage raw agreement of the adapted algorithm diagnoses with the gold standard was 74.4%. Excluding cases without a gold standard and those without any diagnosis or procedure registered resulted in an agreement rate of 78.0%. Agreement varied per specialism from 38.9% (for Other) to 95.0% (for Cardiopulmonary surgery) (Table 4.3).

Performance of the adapted algorithm extrapolated to the full dataset

Weighting the agreement by the prevalence of predicted specialisms in the complete hospital dataset, the adapted algorithm was estimated to predict the transfusion indication correctly in 75.5% of transfusion clusters in the full dataset (results not shown).

Figure 4.3. Adapted algorithm rules visualized by a decision tree

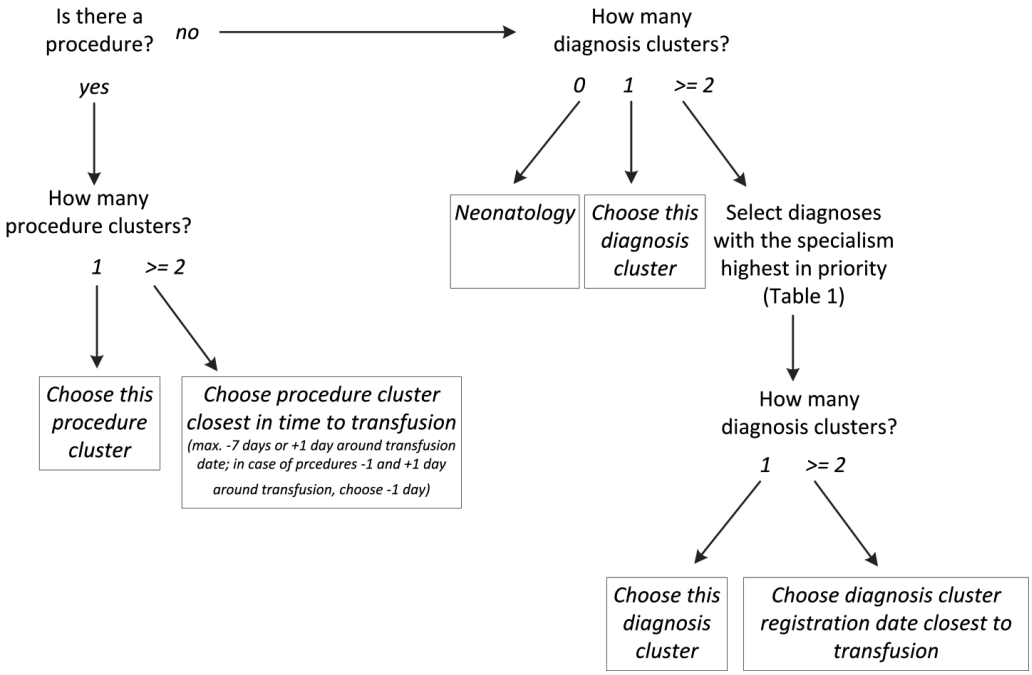


Table 4.3. Agreement between adapted algorithm and gold standard for the transfusion indication as observed in the sample (n=234). The raw % correct in the sample is shown by specialism and in total, showing cases with only one diagnosis option ('data quality check'), cases without a gold standard, and cases without any diagnostic information as separate strata. Kappa provides a measure for chance-adjusted agreement for cases with at least two options.

Stratum	% correct	Kappa
Cardiopulmonary surgery (n=20)	95.0%	0.93
Gynecology (n=17)	88.2%	0.81
Gastroenterology (n=16)	75.0%	0.59
Internal medicine (n=60)	73.3%	0.59
Surgery (n=22)	77.3%	0.66
Orthopedics (n=20)	85.0%	0.78
Other (n=18)	38.9%	0.07
Total specialisms (n=173)	75.7%	0.63
Data quality check (n=18)	100%	
Specialisms + data quality check (n=191)	78.0%	
No codes registered (n=26)	96.2%	
No gold standard (n=17)	0%	
Total (n=234, incl. cases without codes registered)	74.4%	

Discussion

We presented a systematic approach to develop and test a post-hoc algorithm for identifying the indication for transfusion, using expert opinion as starting point and a gold standard to validate and improve the algorithm. The final adapted algorithm was able to correctly identify the indication for transfusion in 74.4% of cases in the sample (75.5% when extrapolated to the full dataset). The algorithm can be utilized by implementing it in the Dutch Transfusion Data warehouse (DTD) [11]. Knowledge of transfusion indications not only facilitates the selection of specific patient groups for future studies and the studying of reasons for transfusion, but also allows benchmarking blood use in patient subpopulations.

Interpretation of results

As shown by the kappa value of 0.63, the algorithm performed above chance level and was able to successfully identify the majority of transfusion indications. Still, for approximately 25% of cases the algorithm's predictions did not agree with the gold standard. Part of the disagreement can simply be explained by the fact that the algorithm is not perfect; on a more detailed level (not available to the algorithm), clinical situations of patients might differ, leading to different indications for transfusion. Another part of the disagreement, however, is due to data quality issues. The gold standard could not be inferred from the data in some cases because of missing registration or suspected incorrect registration dates. For other cases however, even with perfect registration, it would be impossible to know the exact indication for transfusion. These cases often involve patients with multiple and complex morbidities (e.g. resulting in anemia of critical illness), making it impossible even for the treating physician to point out one particular disease or procedure that solely necessitated the transfusion. Finally, the cases without any diagnostic information registered (approximately 3% of all transfusions), showed a clear pattern: almost all were premature neonates. This is comparable to a European study in seven hospitals, where neonates also received 3% of all red blood cell units with a medical indication [3]. Upon request at the hospital, we found that the reason for missingness was that neonates are registered in a separate system. This is a useful outcome of the data check; now we know where to find the diagnostic information for this group if required for research.

Generalizability

Other data warehouses might use a similar approach to identify the indication for transfusion. Although the exact coding of diagnoses and procedures studied in this paper is specific for the Netherlands, the algorithm's prioritization rules are more generally applicable. The algorithm uses the broad category of specialisms and registration dates to select the indication, which is basis information that is generally available in hospitals. Moreover, transfusion indications are

often major, invasive diseases, which are expected to be registered quite consistently, allowing not much room for subjective interpretation of registration codes. In a next step, the transfusion indications resulting from the algorithm could be clustered post hoc into indication groups (for example [12]). This would increase generalizability and comparability of transfusion indications worldwide, if similar categories are used. We generalized the sample results to the full dataset by weighting the sample values according to the sampling fraction per stratum. As these strata were based on the predicted specialisms, which were not always the true specialisms, this might have induced some level of bias if certain specialisms would not end up in the sample in sufficient numbers. However, because within each predicted stratum a random sample of cases was drawn, the most important specialisms were included in the sample with at least n=16 cases.

Future recommendations

To ensure that the algorithm also works for a different case-mix of patients, the algorithm should be validated in one or more external hospital datasets. This validation should be performed by developing a gold standard in the external dataset against which the algorithm can be checked. If necessary, the algorithm may subsequently be adapted for a particular dataset. A way to improve the algorithm is to take into account more detailed patient information such as age, gender, previous treatments, and the number and type of blood products received. Also, the algorithm might be improved for certain patient groups by considering a broader time frame for procedures (we only included procedures within -7 and +1 days around transfusion and it is likely that some procedures are missed because of this selection). More specifically, the cases that were misclassified by the algorithm often concerned patients with chronic conditions like renal dialysis or malignancies. Therefore we prioritized internal medicine (Hematology) higher, so that when no surgery was present this indication was selected more often. Another solution to select these patients would be to take into account the frequency of and interval between transfusions; regular transfusions within a broader time interval point to a hematological transfusion indication. Finally, it would be interesting to investigate whether supervised machine learning techniques would be able to improve the algorithm. Note that such an approach would require more cases with a known gold standard in order to train the selection model.

In the long term, a structural solution for incomplete information on transfusion indications would be to nationally improve registration at the source. Projects that try to improve source registration have been set up in the Netherlands [13,14], for example implementing a diagnosis and procedure thesaurus that corresponds to the international standard of SNOMED CT, as well as in the US [15]. In Europe the EUROREC Institute (EuroRec), an independent not-for-profit organization, is promoting the use of high quality Electronic Health Record systems [16]. In time, projects like these will hopefully lead to enhanced data quality through better registration. Ideally, hospitals should register the diagnosis and/or procedure that motivated the transfusion for each blood product administered. Better registration is not only important for transfusion research, it is also in the

direct interest of the patient and care; especially for an in itself risky treatment as blood transfusion.

Conclusion

An expert-based algorithm is able to identify the indication for transfusion accurately for the majority of transfusions. The selected indications can be implemented in the Dutch Transfusion Data warehouse, where they can serve as a starting point for future studies. Before implementation of the algorithm in other datasets, the algorithm should be externally validated in one or more independent hospital datasets. The systematic approach can be used to apply, evaluate and improve the algorithm in other databases.

References

1. Biggin K, Warner P, Prescott R, et al. REVIEWS: A review of methods used in comprehensive, descriptive studies that relate red blood cell transfusion to clinical data. Transfusion 2010;50(3):711-718.
2. Tinegate H, Chattree S, Iqbal A, et al. Ten-year pattern of red blood cell use in the North of England. Transfusion 2013;53:483-489.
3. Bruun MT, Pendry K, Georgsen J, et al. Patient Blood Management in Europe: surveys on top indications for red blood cell use and Patient Blood Management organization and activities in seven European university hospitals. Vox Sanguinis 2016;111(4):391-398.
4. Wells AW, Mounter PJ, Chapman CE, et al. Where does blood go? Prospective observational study of red cell transfusion in north England. Bmj 2002;325(7368):803.
5. Frank SM, Resar L, Rothschild JA, et al. A novel method of data analysis for utilization of red blood cell transfusion. Transfusion 2013;53(12):3052-3059.
6. Shehata N, Forster A, Lawrence N, et al. Changing trends in blood transfusion: an analysis of 244,013 hospitalizations. Transfusion 2014;54(10pt2):2631-2639.
7. van Walraven C, Bennett C, Forster AJ. Administrative database research infrequently used validated diagnostic or procedural codes. Journal of clinical epidemiology 2011;64(10):1054-1059.
8. Kleinman S, Glynn SA. Database research in transfusion medicine: The power of large numbers. Transfusion 2015; 55:1591–1595.
9. Krabbe-Alkemade YJ, Groot TL, Lindeboom M. Competition in the Dutch hospital sector: an analysis of health care volume and cost. The European Journal of Health Economics, 2016:1-15.
10. Hallgren KA . Computing inter-rater reliability for observational data: an overview and tutorial. Tutorials in quantitative methods for psychology 2012;8(1): 23.
11. van Hoeven LR, Hooftman BH, Janssen MP, et al. Protocol for a national blood transfusion data warehouse from donor to recipient. BMJ open 2016, 6(8).
12. Llewelyn CA, Wells AW, Amin M, et al. The EASTR study: a new approach to determine the reasons for transfusion.Transfusion Medicine 2009;19(2):89-98.
13. Dutch Federation of University medical centers (NFU). Source registration [Registratie aan de bron], 2013. Accessed on January 2017 at: http://www.nfu.nl/img/pdf/13.3694_Brochure_Registratie_aan_de_bron_versie_4-7-13.pdf
14. Dutch Association of general hospitals and specialized institutions (NVZ). Information policiy: four priorities [Informatiebeleidsplan NVZ: vier speerpunten]. Accessed on January 2017 at: https://www.nvz-ziekenhuizen.nl/onderwerpen/informatiebeleid

15. CMS. EHR Incentive Programs. Accessed on January 2017 at: https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/ehrincentiveprograms
16. EuroRec Seal. Accessed on January 2017 at: http://www.eurorec.org/services/seal/index.cfm

Chapter 5

Aiming for a representative sample: Simulating random versus nonprobabilistic strategies for hospital selection

BMC Medical Research Methodology 2015

L.R. van Hoeven
M.P. Janssen
K.C.B. Roes
H. Koffijberg

Abstract

Introduction

A ubiquitous issue in research is that of selecting a representative sample from the study population. While random (probabilistic) sampling strategies are the gold standard, in practice, random sampling of participants is not always feasible nor necessarily the optimal choice. Before putting a lot of effort in recruitment or data collection, it may be worthwhile to carefully consider potential sampling strategies. In this paper, we evaluate several strategies for the case of estimating blood use in Dutch hospitals.

Methods

Available population-wide data on hospital blood use and number of hospital beds are used to simulate sampling strategies. Five pragmatic sampling strategies, both random and purposive, result in different samples of hospitals, that are each used to fit a model that predicts blood use. The subsequent prediction errors are used to indicate the quality of the sampling strategy.

Results

The strategy leading to the lowest prediction error in the case study was maximum variation sampling, followed by random, regional variation and two regions sampling, with sampling the largest hospitals resulting in worst performance. Out of all simulations, maximum variation sampling outperformed random sampling on the hospital level in 85% and the national level in 76%. Whereas a lower sample size increased preference for random strategies, increasing sample size did not change the ranking of the strategies and led to only slightly better predictions.

Conclusion

The optimal strategy for estimating blood use was maximum variation sampling. It is possible to evaluate probabilistic and non-probabilistic sampling strategies using simulations. The results enable researchers to make a more educated choice for an appropriate sampling strategy.

Introduction

When choosing a sample of participants, researchers often find themselves with a trade-off between the wish for randomization and pragmatic considerations. Random (probabilistic) sampling is the gold standard of sampling strategies because of its unbiasedness and the possibility to evaluate the reliability (precision) of the resulting estimates [1,2,3,4]. Random sampling is not, however, always feasible in practice due to constraints in time, resources and costs, and researchers in the medical field often use a 'convenience' or a purposive sample, i.e. choose participants that are easy to recruit or select participants based on preferences or expectations.

Fortunately, some studies suggests that such non-random strategies can lead to representative samples [5,6,7]. Also, the statement that a random sample is unbiased means that it will provide a representative estimate on average. The probability of randomly drawing an 'unrepresentative' sample is large if your population is small; the estimator is not robust, since data collection is done only once and not a thousand times. This can be illustrated by a study in the medical field that compared a randomized study design with a nonrandomized design. The nonrandomized design resulted in a more representative sample in 34% of cases [5]. In another study, comparison of several sampling strategies for surveillance of cases of injury and poisoning in accident and emergency departments showed that a well-planned systematic sampling strategy can generate data of equal quality to surveillance including all patients [7]. In a study estimating drug use characteristics, purposive samples were found to be sufficiently representative, as compared to probabilistic strategies, when these were drawn from a wide cross-section of participants and included a relatively large number of individuals [6]. Thus non-probabilistic strategies are sufficient at least in some cases.

If possible, strategies should be evaluated per study, in line with the 'fit for use' concept; see [8,9]). Preferably this evaluation should be done prior to the actual data collection so that this information can be used to choose the optimal sampling strategy. However, in the medical field to our knowledge no (simulation) studies exist on evaluating random versus preferential sampling strategies with respect to prediction accuracy before data collection; instead, methods exist for generalizing treatment effects in randomized trials from unrepresentative samples (see for example [10,5,11,12]), or studies that either compare only non-probabilistic [13,14] or only probabilistic strategies [15].

In the present study, in order to find the optimal strategy, we compare five stratified probabilistic and non-probabilistic sampling strategies that match real-life strategies used in practice. The case used is as follows: We want to study blood use in Dutch hospitals but, since the process of obtaining large amounts of data was found to be complicated and time-consuming, we can only collect data from 12 of the total 89 hospitals. The resulting database (containing data from the 12 selected hospitals) must include detailed information on patient diagnoses and clinical parameters that can be used to answer several research questions concerning blood use. For the simulation, we used a limited amount of data that was already available on each hospital before data collection. Five pragmatic sampling strategies were

simulated (stratified to hospital type as stratification has been proven beneficial [16]): 1). Largest hospitals sampling (resulting in a large database), 2) Maximum variation sampling (only hospitals on the most extreme ends of RBC use are sampled), 3) Random sampling, 4), Regional variation sampling (hospitals from each region are included) and 5) Two geographic regions sampling. Representativeness of the resulting samples is evaluated by performing model based inference and computing the prediction errors [17]. We assume that if a sample is representative in this restricted dataset, it will also be representative for other relevant population outcomes. The results will show whether or not, in the context of estimating blood use, the general consensus that non-probabilistic strategies are inferior to probabilistic strategies holds. More broadly, the case illustrates a method for evaluating different sample strategies. The results can be used to support an informed choice for a sampling strategy.

Methods

Five pragmatic sampling strategies (see below) are simulated using information that is already available prior to the actual data collection. The effect of sampling strategy is evaluated in terms of its prediction accuracy of the population estimates and margin of error.

Case and data

The target population consists of all non-specialized Dutch hospitals (n=89), comprising 8 academic centers, 28 teaching hospitals and 53 remaining general (smaller) hospitals. Specialized centers were excluded (n=3), because the majority of blood transfusions is already covered by the academic and peripheral hospitals. Limited data on all hospitals was already available and easily accessible. Firstly, the number of beds per hospital was extracted from annual hospital reports or the hospital website. Secondly, hospitals were classified by type, as described by Dutch Hospital Data [19]: a hospital is either an academic medical center, teaching or general hospital. Information on hospital blood use was available as well: Sanquin Blood Bank provided the number of issued blood products delivered to each hospital in the year 2013 for the three main product types: erythrocytes (RBC), fresh frozen plasma (FFP) and platelets (PLT) [18]. The number of issued blood products was used as a proxy for the number of transfused blood products. If information on blood use was not available for a hospital, that hospital was excluded from analysis (n=1). Classification of hospitals into organizational healthcare regions was done according to the Education and Research regions [20]; hospitals that were not classified by this structure (n=8) were manually assigned based on their location to the nearest region.

Sampling strategies

The simulation is confined to the following five strategies, which are all stratified to hospital type.1) Maximum variation sampling (MAXVAR) was used to sample hospitals that have the highest and lowest number of RBC transfusions, so that variation in total number of RBCs is

maximized. The theoretical advantage of maximum variation sampling is that extrapolating to extreme (impossible) values does not occur because the extremes are already in the sample. Selecting hospitals based on their RBC use is also supposedly sufficient for obtaining high variation in FFP and PLT use; the respective Spearman's rank correlations with RBC use are .88 ($p<.00001$) and .92 ($p<.00001$). 2) Sampling only the largest hospitals (LARG) has the obvious advantage that since larger hospitals have more patients, this yields the most data. 3) Random sampling (RAND) gives each hospital within a stratum an equal probability of being sampled. 4) Regional variation sampling (REGVAR) maximizes the number of randomly chosen organizational health care regions included for each stratum, based on the assumption that there is considerable variation between regions that must be reflected in the sample. 5) Sampling from two organizational health care regions (2REG). Including a large part of all hospitals from two regions, allows not only the benchmarking of hospitals, but also the benchmarking of (almost) complete regions. This form of sampling is simulated for all 21 combinations of two regions. If a region contains more hospitals than the preferred sample size per stratum, hospitals are selected randomly. In contrast, if a region contained fewer hospitals than the preferred sample size within a stratum, all hospitals within that stratum are included. Figure 5.1A illustrates which hospitals are sampled when using LARG and MAXVAR strategies. Figure 5.1B shows a possible result of sampling when the probabilistic RAND, REGVAR and 2REG strategies are used.

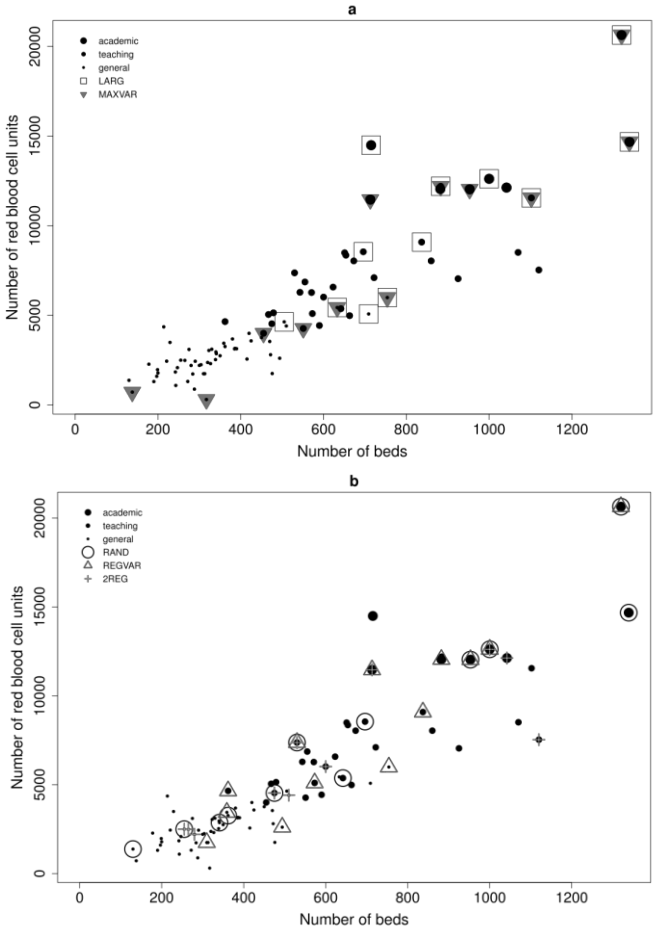
Sample size

Sample size is varied, starting with four hospitals per stratum (n total = 12). In each subsequent scenario the added value of including two more of each type of hospital (thus 6 hospitals per stratum) is examined, as well as the effect of including two hospitals fewer in each stratum (see Table 5.1 for all three sample size scenarios). The stratification by hospital type ensures a fixed sample size ratio of the strata of 1:1:1. The exception is the strategy of sampling two regions; here, since there is usually only one academic hospital per region, the number of included academic hospitals is fixed in all scenarios at two (and three when the region Noord-Holland is included).

Model-based inference

Model-based inference, as it may provide a viable alternative for design-based methods when design information is not available [21], seems most appropriate in case of non-probabilistic sampling. Even if a non-probabilistic sample in itself is not representative, the resulting model might very well be [22]. In the present case, model based inference was used to predict hospital blood use. In short, a data sample drawn according to one of the above strategies was used to fit a Poisson regression model that predicts blood use (i.e., the number of RBC, FFP and PLT

Figure 5.1. Illustration of sampling strategies. a) Hospitals that are selected when using the non-probabilistic strategies of sampling the largest hospitals (LARG) and maximum variation (MAXVAR); b) Possible selection of hospitals when the probabilistic RAND, REGVAR and 2REG strategies are conducted (since these methods involve a random element, the figure shows only one of many possible samples).



per hospital) as a function of hospital size (i.e., number of beds) and hospital type. With the obtained prediction models, RBC, FFP and PLT use was estimated for all Dutch hospitals and compared to the true population values, which are known for this case. Outcomes, expressed as a percentage of the population values, are the prediction error on hospital level (summed absolute errors at hospital level) and the national prediction error (absolute deviation of the national estimate from the population values). These two prediction errors types are of interest for different reasons. National level errors are important from the perspective of the national blood bank: since the blood bank produces blood products for the whole of the Netherlands, it is relevant to know how much blood is needed in total, for example on a yearly basis. However from the perspective of (clinical) studies, it might be considered more important to

have accurate predictions also within each hospital (i.e., on hospital level). Obviously, individual hospitals are also interested in their own expected blood use.

Table 5.1. Number of hospitals included per type for each sample size scenario

Scenario	N (academic)	N (teaching)	N (general)	N Total (%of all Dutch hospitals)
A	4	4	4	12 (13%)
B	6	6	6	18 (20%)
C	2	2	2	6 (7%)

Simulations

For each of the RAND, REGVAR and 2REG strategies, a random sampling process is simulated a thousand times. The median error percentages are reported, with the 95% centiles and the average error percentages. Since the strategy of sampling two regions encompasses 21 unique combinations of two regions, the median and average error percentages are taken over all sampled combinations. Striking differences between combinations of regions are described in the results section. All analyses are performed in R Version 3.0.0. R code that simulates the sampling strategies and creates an exemplary data set, is provided in Appendix 5.1.

Results

Prediction error on hospital level

Prediction errors for each sampling strategy are shown in Table 5.2, for the scenario of sampling 12 hospitals. For RBC, FFP and PLT, maximum variation sampling outperformed largest hospitals sampling in terms of hospital level error (Figure 5.2). When comparing the probabilistic strategies (RAND, REGVAR and 2REG) with the non-probabilistic strategies (LARG and MAXVAR), a mixed picture emerges. MAXVAR sampling resulted in 20% prediction error on hospital level for RBC, equal to the full population model. The random strategies (stratified random, regional variation and two regions) had a higher median error than MAXVAR for RBC (namely 22% for all three random strategies) and for FFP (varying from 34% (RAND), 35% (REGVAR) to 43% (2REG) versus 33% for MAXVAR respectively), but not for PLT (RAND and REGVAR 37%, 2REG 42%, versus 41% for MAXVAR). Of all the simulations, RAND resulted in a lower hospital level error than MAXVAR only 15% (RBC) and 29% (FFP) of the time. However for PLT, RAND resulted in a lower hospital level prediction error in 85% of the simulations (Table 5.3). Sampling only the

largest hospitals resulted in a median hospital level error of 44% for RBC, which is higher than the median error for the random strategies.

Prediction error on national level

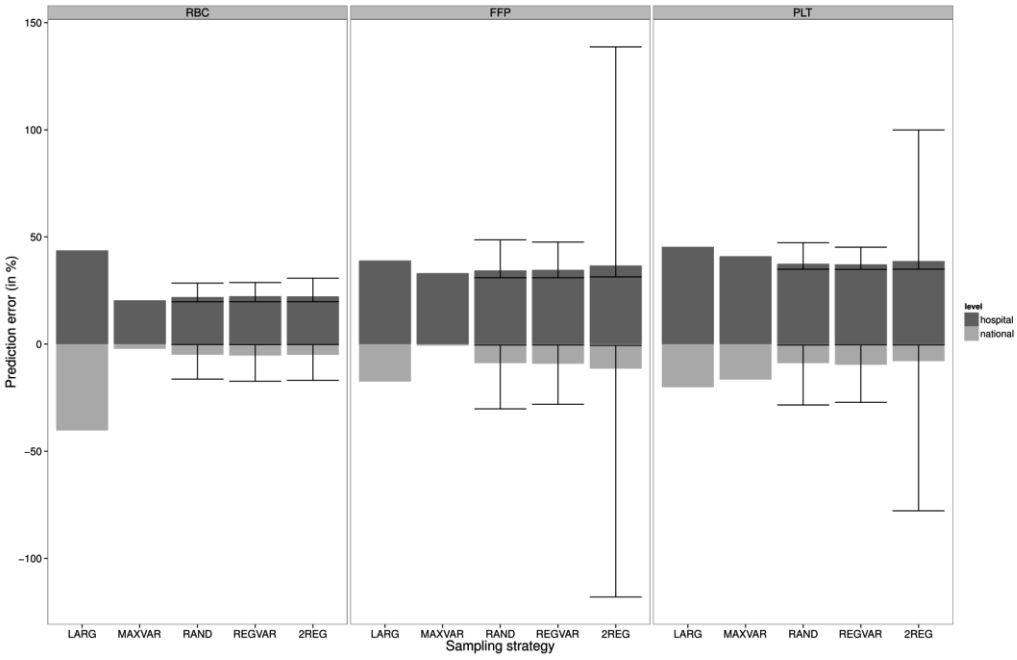
Comparing the national level prediction errors of the strategies resulted in a similar pattern, with MAXVAR outperforming the random strategies most of the time (Table 5.2 and Figure 5.2). In fact, random sampling resulted in a lower national level error than MAXVAR in 24% (RBC), 3.5% (FFP) and 81% (PLT) of the simulations (Table 5.3). The same pattern was found for REGVAR and 2REG.

Table 5.2. Comparison of prediction errors for the five sampling strategies, for n=12 (n=4 per hospital type)

Strategy	Prediction error on hospital level			Prediction error on national level		
	RBC	FFP	PLT	RBC	FFP	PLT
Population[a]	20%	32%	35%	444,674 (0%)	67,461 (0%)	55,572 (0%)
LARG	44%	39%	45%	40%	18%	20%
MAXVAR	20%	33%	41%	2%	1%	17%
RAND (median; mean, 95% centiles)	22%; 23% (20%-28%)	34%; 36% (31%-49%)	38%; 38% (35%-47%)	5%; 6% (0%-16%)	9%; 11% (0%-30%)	9%; 10% (0%-28%)
REGVAR (median; mean, 95% centiles)	22%; 23% (20%-29%)	35%; 36% (31%-48%)	37%; 38% (35%-45%)	5%; 6% (0%-17%)	9%; 11% (0%-28%)	10%; 11% (0%-27%)
2REG (median; mean, 95% centiles)	22%; 23% (20%-31%)	37%; 46% (31%-139%)	39%; 44% (35%-100%)	5%; 6% (0%-17%)	11%; 19% (1%-118%)	78%; 13% (0%-78%)

LARG = largest hospitals, MAXVAR = maximum variation in number of RBCs, RAND = random, REGVAR = regional variation, 2REG = two regions, RBC = red blood cell products, FFP = fresh frozen plasma products, PLT = platelet products. Output for RAND, REGVAR and 2REG is based on the average of 10 times 1000 simulations and accompanied by 95% centiles. Outcomes are the prediction error on hospital level (summed absolute errors at hospital level) and the national prediction error (absolute deviation of the national estimate from the population values), both expressed as a percentage from the population values. a Prediction errors for the models fitted with the complete population are shown as anchor point; naturally these are 0% at national level.

Figure 5.2. Prediction error on hospital and national level for n(academic)=4, n(teaching)=4 and n(general)=4. Median prediction errors of red blood cell (RBC), plasma (FFP) and platelet (PLT) use, for different sampling strategies. 95% centiles are provided for the strategies involving a random element. Number of simulations=1000. LARG = largest hospitals, MAXVAR = maximum variation in number of RBCs, RAND = random, REGVAR = regional variation, 2REG = two regions.



Effect of sample size

Adding two more academic, two more teaching and two more general hospitals to the sample (total n=18) reduced hospital and national level error as well as the 95% percentile error ranges by one or two absolute percent points (see Appendix 5.2). Reducing sample size to two hospitals per hospital type (total n=6) increased prediction errors considerably in some cases, especially for the LARG strategy. For LARG, hospital level error for RBC increased from 44% to 94% and national error from 40% to 92%. For MAXVAR, hospital level errors increased from 20% to 26% (RBC), 33% to 39% (FFP) and 41% to 42% (PLT); national level errors also increased by a few percent points. Similarly, the random strategies yielded moderately higher hospital and national errors and wider 95% centile ranges in the low sample size scenario (see Appendix 5.3).

Increasing the sample size did not affect the ranking of the sampling strategies by prediction error; the scenarios n=12 and n=18 both resulted in a preference for MAXVAR. However in the small sample size scenario of n=6, MAXVAR was outperformed by some of the random strategies. That is, when only six hospitals were sampled, hospital level prediction

Table 5.3. How often are random strategies (RAND, REGVAR and 2REG) better than purposive strategies (MAXVAR and LARG) in terms of hospital and national level prediction error for n=12 (n=4 per hospital type)?

	RBC	FFP	PLT	RBC	FFP	PLT
RAND versus:	MAXVAR			LARG		
Lower hospital level prediction error for RAND	15%	29%	85%	100%	80%	96%
Lower national level prediction error for RAND	24%	4%	81%	100%	81%	88%
REGVAR versus:	MAXVAR			LARG		
Lower hospital level prediction error for REGVAR	11%	32%	88%	100%	78%	98%
Lower national level prediction error for REGVAR	21%	5%	80%	100%	81%	89%
2REG versus:	MAXVAR			LARG		
Lower hospital level prediction error for 2REG	16%	17%	66%	100%	63%	84%
Lower national level prediction error for 2REG	24%	3%	83%	100%	71%	88%

of RBC, FFP and PLT was better for the RAND and/or REGVAR strategies, and national level prediction of PLT was better for the RAND, REGVAR and 2REG strategies.

Discussion

Currently, representativeness of a sample is often only checked after data collection has finished. Although such a post-hoc evaluation may provide some insight in the representativeness of the already selected sample, unfavourable outcomes can rarely be mitigated once the data collection process has ended. Therefore an evaluation of potential sampling strategies should ideally be performed prior to data collection. We evaluated five pragmatic sampling strategies for the case of estimating blood use in Dutch hospitals. The evaluation consists of simulating the sampling processes for five probabilistic and non-probabilistic strategies, using prior knowledge. Such a simulation study may help in deciding whether a random sampling design is necessary or whether variation should be aimed for in order to obtain a sample that is likely to be representative of population values. This type of evaluation is in theory applicable to a broad array of research fields, provided that there exists a relation between a predictor and outcome that can be modelled.

The case study illustrates that random sampling, which is considered the gold standard, is not necessarily the optimal sampling strategy. In fact, of the five strategies considered, the optimal strategy in our case was maximum variation sampling (MAXVAR). A sample selected using the MAXVAR strategy led to better predictions of red blood cell unit (RBC) use than a random sample in 85% (hospital level) and 76% (national level) of all simulations. In contrast, random sampling did perform much better than sampling only the largest hospitals. In general, the same pattern was found for both national and hospital level prediction errors, with national errors being lower since under- and overestimation of individual hospitals cancel each other out.

The preference for the non-probabilistic MAXVAR strategy over random sampling was not completely expected in the context of previous literature. In a previous study that simulated outcomes of a randomized design [5], non-probabilistic sampling was reported to be better in only 34% of simulations. Moreover, in a study on modelling species' distribution, non-probabilistic strategies were reportedly inferior to probabilistic strategies [23]. These contrasting findings could be due to the use of different measures for evaluating representativeness: In the present study a sample is considered representative if it gives us an unbiased estimate of the outcome studied, whereas in other studies, representativeness is defined in terms of whether participants in the sample have similar characteristics as those in a random sample. Moreover, these contrasting findings could be caused by the use of different data and models for inference, and the use in the present study of a convenience sample instead of systematic purposive samples. However, in line with earlier findings [5], differences between the median prediction errors for MAXVAR and the random strategies were quite small. This implies a trade-off between the certainty of a known prediction error with MAXVAR and the risk of potentially getting a higher (or a lower) error with one of the random strategies.

In the present study, MAXVAR seems the 'safest' option. However, MAXVAR was not consistently the preferred strategy. In accordance with findings from an ecological study [16], preference for either a non-probabilistic or a probabilistic strategy turned out to depend on which outcome was modelled. For example, in our study, MAXVAR sampling resulted in better predictions than two-regions sampling at national level, but not at hospital level. Whether prediction accuracy on hospital or on national level is given more importance depends on the aim and perspective of the researcher. From the perspective of the blood bank that produces blood products for the whole of the Netherlands, it is important to know how much blood is needed nationally on a yearly basis. However from the perspective of (clinical) studies and individual hospitals, it would be more important to have accurate predictions also within each hospital. In a second inconsistency in outcomes, MAXVAR outperformed random sampling for the outcomes RBC and plasma (FFP) use, but this was not the case for platelet (PLT) use. This result might be explained by differences in the underlying distributions: Whereas MAXVAR led to a selection of hospitals from the entire range of variation of RBC use, it did not for PLT use, because PLT use follows a different distribution (PLT use is relatively high in the largest hospitals and varies greatly between hospitals, partly due to its high

spilling rate and short shelf life). In case of a more or less linear relationship between predictor and outcome such as between number of beds and RBC use, MAXVAR will perform well (this result obviously also depends on the type of model used for making inferences). However for other relationships, such as between number of beds and PLT use, an amended MAXVAR strategy might be more suitable. This could for instance maximize the distances between subsequent hospitals (instead of sampling at the ends of the distribution). Choosing an appropriate variable at which to aim the maximum variation is important, since this choice can have substantial consequences for the estimates [16].

Increasing sample size from 12 to 18 hospitals did not alter the order of the strategies. In contrast, decreasing sample size to 6 hospitals increased preference for the random strategies over MAXVAR sampling. Apparently a very low sample size allows outlier (combinations of) hospitals to occur in the MAXVAR sample, leading to high prediction errors (in this case two academic hospitals which, in case these are sampled together, lead to a regression slope much too steep). In comparison, in a study on habitat suitability modelling [24], increasing or decreasing sample size did not change the order of strategies, however that study did not consider scenarios with a sample size as small as n=6. In that study, prediction accuracy increased with sample size, whereas in our study the difference in prediction accuracy between the n=12 and n=18 scenarios was quite small (around 1-2 absolute percentage points). However, if we had been able to increase sample size by a larger amount, we might also have found higher prediction accuracy. Presently, for estimating blood use it is not directly obvious that including six additional hospitals would be worth the additional effort. Instead, a more accurate prediction of blood use might have been obtained by extending the model with more detailed information (predictors), such as the presence of a cardiac center in a hospital and number of patients per admission diagnosis or type of surgery.

An important assumption that underlies our evaluation is that representativeness in terms of a known outcome can be used as a proxy for representativeness for other outcomes (which will be studied after the actual data collection). The reasoning behind this assumption is that if a sample is at least representative for the number of blood products, it is more likely to be representative for related outcomes as well. These related outcomes, such as the distribution blood products over diagnoses and surgeries, blood use in different patient subgroups, and transfusion triggers, are all expected to be related to the predictors hospital type and hospital size. Finally, we acknowledge that other considerations such as costs, feasibility, the need to include specific regions for benchmarking purposes, specific patient groups, or certain hospitals from which historical data are already available (which enables a trend analysis), might play a role in selecting potential participants. These conditions can also be included in the simulation. Last but certainly not trivial, the success of the data selection depends on the cooperation of the potential participants.

Conclusion

A simulation study as described above may offer guidance in choosing an appropriate sampling strategy and size before data collection is started. Following this guidance is straightforward and can be done with limited resources. Its only requirement is the a priori availability of a (limited) nationwide data set, which will often be available as long as the aggregation level is sufficiently high. In many situations, especially whenever data collection has large resource requirements, such a simulation will be worthwhile and should therefore be considered.

Appendix 5.1 (R code). R code that creates an exemplary data set and simulates the sampling strategies.
Appendix 5.2 (Table). Comparison of prediction errors for the five sampling strategies, for n=18 (n=6 per hospital type). Prediction errors for n=18 in a similar table as for n=12.
Appendix 5.3 (Table). Comparison of prediction errors for the five sampling strategies, for n=6 (n=2 per hospital type). Prediction errors for n=6 in a similar table as for n=12.
All Appendices available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4619525/

References

1. Levy PS, Lemeshow S. The population and the sample. In: Sampling of Populations: Methods and applications. 4th ed. John Wiley and Sons, New York, USA; 2008. p. 11-42.
2. Banerjee A, Chaudhury S. Statistics without tears: Populations and samples. Industrial Psychiatry Journal. 2010;19:60-5.
3. Kish L. New paradigms (models) for probability sampling. Survey methodology. 2002;28(1);31-34.
4. Tiwari N., Chilwal A. On Minimum Variance Optimal Controlled Sampling: A Simplified Approach. Journal of Statistical Theory and Practice. 2014;8(4):692-706.
5. Raaijmakers M, Koffijberg H, Posthumus J, van Hout B, van Engeland H, Matthys W. Assessing performance of a randomized versus a non-randomized study design. Contemp Clin Trials. 2008;29:293-303.
6. Topp L, Barker B, Degenhardt L. The external validity of results derived from ecstasy users recruited using purposive sampling strategies. Drug & Alcohol Dependence. 2004;73:33–40.
7. Morrison A, Stone DH. Injury surveillance in accident and emergency departments: to sample or not to sample? Inj Prev. 1998;4:50–3.
8. Wang RY, Strong DM. Beyond accuracy: What data quality means to data consumers. Journal of management information systems. 1996:5-33.
9. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. Med Care. 2012;50(Suppl):S21–9.
10. O'Muircheartaigh C, Hedges LV. Generalizing from unrepresentative experiments: a stratified propensity score approach. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2014;63(2):195–210.
11. Aronow PM, Middleton JA. A class of unbiased estimators of the average treatment effect in randomized experiments. Journal of Causal Inference. 2013;1:135–154.
12. Adhya S, Banerjee T, Chattopadhyay G. Inference on Polychotomous Responses in Finite Populations. Scandinavian Journal of Statistics. 2011;38(4):788-800.
13. Maghera A, Kahlke P, Lau A, Zeng Y, Hoskins C, Corbett T, Manca D, Lacaze-Masmonteil T, Hemmings D, Mandhane P. You are how you recruit: a cohort and randomized controlled trial of recruitment strategies. BMC medical research methodology. 2014;14(1):111.
14. Sherman KJ, Hawkes RJ, Ichikawa L, Cherkin DC, Deyo RA, Avins A L, Khalsa P S. Comparing recruitment strategies in a study of acupuncture for chronic back pain. BMC medical research methodology. 2009;9(1):69.
15. Ikeda N, Shibuya K, Hashimoto H. Improving population health measurement in national household surveys: a simulation study of the sample design of the

comprehensive survey of living conditions of the people on health and welfare in Japan. J Epidemiol. 2011;21:385–90.

16. Albert CH, Yoccoz NG, Edwards TC Jr., Graham CH, Zimmermann NE, Thuiller W. Sampling in ecology and evolution – bridging the gap between theory and practice. Ecography. 2010; 33:1028–1037.
17. Kruskal W and Mosteller F. Representative Sampling, I: Scientific Literature, Excluding Statistics. International Statistical Review / Revue Internationale de Statistique. 1980;47(2):111-127.
18. Sanquin Blood Bank. Number of issued blood products per hospital 2013. Accessed in 2014.
19. Dutch Hospital Data. Kengetallen Nederlandse ziekenhuizen 2012. Utrecht. https://www.nvz-ziekenhuizen.nl/_library/13152/Kengetallen%20Nederlandse%20Ziekenhuizen%202012.pdf. Accessed 4 November 2014.
20. Nederlandse Federatie van Universitair Medische Centra. OOR-zaak en gevolg. Opleidingen in de zorg. NFU-visiedocument 053059. 2005. http://www.nfu.nl/img/pdf/NFU_Oorzaak.pdf. Accessed 5 November 2014.
21. Buelens B, Boonstra HJ, Van den Brakel J, Daas P. Shifting paradigms in official statistics: from design-based to model-based to algorithmic inference. Discussion paper 201218, Statistics Netherlands, The Hague/Heerlen. 2012. https://www.cbs.nl/nl-nl/achtergrond/2012/38/shifting-paradigms-in-official-statistics-from-design-based-to-model-based-to-algorithmic-inference. Accessed 4 November 2014.
22. Rothman, Kenneth J., John EJ Gallacher, and Elizabeth E. Hatch. Why representativeness should be avoided. International journal of epidemiology. 2013;42(4):1012-1014.
23. Edwards Jr, TC, Cutler DR, Zimmermann NE, Geiser L, Moisen GG. Effects of sample survey design on the accuracy of classification tree models in species distribution models. ecological modelling. 2006;199(2):132-141.
24. Hirzel A, Guisan A. Which is the optimal sampling strategy for habitat suitability modeling? Ecological Modelling. 2002;157:329–339.

Chapter 6

Historical time trends in red cell component usage in the Netherlands

International Journal of Clinical Transfusion Medicine 2016

L.R. van Hoeven
M.M.W. Koopman
H. Koffijberg
K.C.B. Roes
M.P. Janssen

Abstract

Background

While the number of hospitalized patients in Dutch hospitals has increased since 1997, the demand for red blood cell units (RBCs) has simultaneously decreased. This implies a dramatic change in transfusion practice towards fewer blood transfusions on average per patient.

Objectives

In order to explain the RBC reduction, different patient groups (surgical, medical, obstetrical, specific age groups) were studied retrospectively in relation to RBC use. In addition, the use of combinations of RBCs, fresh frozen plasma and platelets during a transfusion episode was examined for trends over time.

Methods

Data from the PROTON database, containing information on all transfusions in 12 Dutch hospitals in the period 1996-2005, including corresponding patient data (age, diagnosis, treatment, hospitalizations) and blood unit data (type, amount, date) were analyzed.

Results

The proportion of RBCs used for surgical patients declined from 50% in 1996 to 40% in 2005, whereas medical use increased from 47% to 58% (the remaining 2-3% went to obstetrical patients). Changes were more marked in the higher age groups. Also, a trend was observed towards the use of only one or two RBC units during a transfusion episode rather than three or more. Amongst surgical patients who received blood, the use of combinations of blood units, as compared to RBCs only, increased from 32% to 39%.

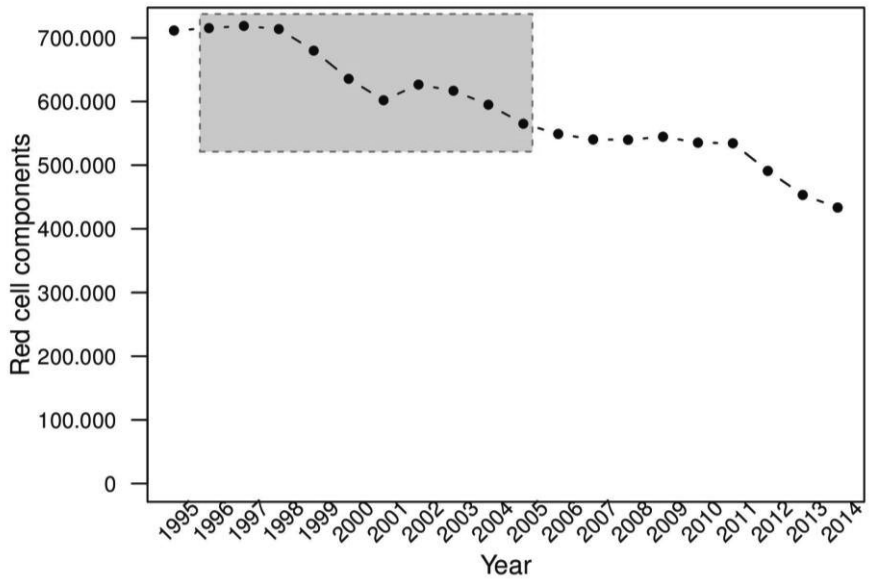
Conclusion

The results suggest a more restrictive transfusion policy for surgical patients as well as an increase in medical indications for transfusion. This fits well with the current focus towards more cost-effective transfusion policies.

Introduction

Over the past 18 years, the national Dutch Blood Supply (Sanquin Blood Supply) has been confronted with a steady decrease in the demand for red blood cell units (RBCs). During the years 1996-2005, the number of issued RBCs decreased by 20% which amounts to approximately 155.000 red cell units. This decrease is even more distinct in the most recent years (9% over 2012-2013) (Figure 6.1). At the same time, the number of hospitalized patients has increased since 1996 with 44% [1], implying a dramatic change in blood use in clinical practice. Currently, little is known about the exact reasons why and where these changes occur. Insight in the way in which transfusion practice changes is important to discover opportunities for restrictive transfusion policy on the one hand and predict future demands on the other hand. Therefore, in the present study, RBC use in specific patient groups (surgical, medical and obstetrical) was studied for trends, whether these trends could be confined to certain age categories, diagnoses or procedures, and whether they were consistent over time and hospital type (academic versus general). Furthermore we assessed whether the decrease in blood use was due to a reduction per operation. Finally, we identified trends in the practice of combining RBCs with fresh frozen plasma (FFP) or platelets (PLT) during a transfusion episode.

Figure 6.1. Total number of red cell components issued in the Netherlands over the past 20 years. Time under consideration in this study is 1996-2005 (marked area).



Methods

Data selection

A subset of data from the PROTON database was used, including information on blood transfusions in Dutch hospitals in the years 1996-2005 [2]. This subset was selected from PROTON based on completeness of information on diagnosis, procedure and other patient characteristics required for our analyses, resulting in a set of 4 academic, 7 general (of which 4 teaching) hospitals and one cancer hospital. This selected dataset encompasses information on 187.096 transfusion recipients who received 1.544.025 blood units (1.117.652 RBCs, 276.304 FFPs and 150.069 PLTs) during the years 1996-2005, comprising 19% (17% of RBCs, 26% of FFPs and 32% of PLTs) of total Dutch blood use.

Classification

The diagnoses and procedures were coded according to the International Classification of Diseases (ninth revision, clinical modification: ICD9-CM) and the Classification of Medical Specialistic Operations (CSMV) system. Based on these codes, transfusions were classified as either surgical, obstetrical of medical. Transfusions were classified as surgical when given within two weeks after surgical admission or were linked to a registered surgical procedure. Obstetrical patients were operationalized as having an obstetrical procedure and/or the main diagnosis 'complications of pregnancy, childbirth, and the puerperium'. The medical group includes patients who were neither surgical nor obstetrical. For surgical as well as obstetrical patients, a transfusion episode was defined as the time from admission to discharge. Since medical patients were not hospitalized in a large number of cases and trends were studied per calendar year, a transfusion episode for medical patients was defined as one calendar year [3]. It must be noted that part of the outpatient transfusions (i.e. without a hospitalization) could not be linked to patient diagnosis information and could therefore not be classified. As most outpatient transfusions would presumably be classified as medical, in the present study the proportion medical transfusions is likely underestimated. A detailed description of data quality, comprising completeness of transfusions per hospital and both accuracy and occurrence of missing values for diagnoses and procedures, is included in Appendix 6.1.

Statistical analysis

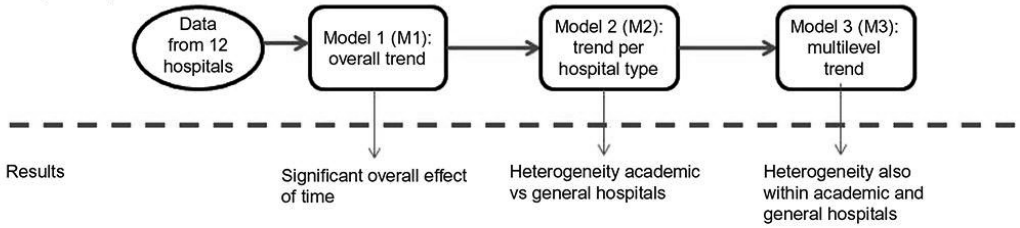
The overall trend in RBC use was modelled with time as predictor for number of RBCs, in three different ways which each provide different information about the trend (see upper part of Figure 6.2). Firstly, a linear regression model was fitted with the outcome total number of RBCs summed over all hospitals per year (M1). M1 provides an estimate of the overall effect of time. Secondly, a linear regression model was fitted with the outcome RBC use per hospital and including the interaction between time and hospital type (academic/general) (M2). M2 provides an estimate of the average trend per hospital. Thirdly, a multilevel model was fitted, with a fixed effect for time and random intercepts and slopes for each individual hospital (M3).

M3 accounts for a potential correlation between measurements within a hospital and estimates whether trends are consistent across hospitals.

In order to reveal changes in the quantity of RBCs used for medical, surgical and obstetrical patients, the numbers of RBCs per year were plotted against patient age. Changes per year were also identified for the most prevalent main diagnoses and procedures. To investigate changes in the number of transfusions given per transfusion episode, the relative occurrence of 1, 2, 3-4, 5-8, and 9 or more RBC transfusions per episode was plotted per year. Since the decrease in RBC use might also be related to a change in the practice of transfusing combinations of RBCs, FFPs and PLTs, the relative occurrence of the combinations during a transfusion episode was investigated over time (differences over time tested using the Chi-Square test of homogeneity). Moreover, the median number of RBC, FFP and PLT transfused when given in any combination was visualized over the years. All analyses were performed using R version 2.15.1 [4].

Figure 6.2. A) Overview of three ways of analyzing the trend over time of number of RBCs, resulting in three models: M1, M2 and M3, with an increasing level of detail and variation between hospitals taken into account in the model (upper part). B) The model coefficients are shown in the 'Results' part. **Significant at $p<0.01$; ***Significant at $p<0.001$.

A Analysis/setup



B

	M1	M2	M3
Intercept	117,796.3	18,496.4	18,496.5
Annual time trend (95% CI)	–1,137.9 (–1,809.5; –466.3)**	–420.7 (–689.0; –152.5)**	–420.7 (–830.2; –11.2)
Hospital type (95% CI, reference: academic)	Not applicable	–12,950.0 (–14,745.2; –11,154.9)***	–12,950.0 (–17,455.6; –8,444.4)
Interaction time × type (95% CI, reference: academic)	Not applicable	528.2 (191.9; 864.5)**	528.2 (14.9; 1,041.5)

Results

The distributions of age and main diagnosis of all hospitalized patients included in this study were, separately for academic and general hospitals, compared to those in the remaining Dutch

hospitals (n=82), (Appendix 6.2).[1] The similarity of the distributions suggests that the included hospitals are representative for Dutch hospitals in general with respect to age and diagnosis. The most notable difference was the percentage 65+ patients in academic hospitals, which is 26% for hospitals included and 23% for the remaining hospitals, whereas all other differences are limited to one or two percentage points.

Overall trend in RBC use

The overall trend in RBC use in the hospitals studied was downwards with a decrease of 1138 RBCs per year (95% CI -1810; -466, Figure 6.2: M1). However, the trend over time was different for academic and general hospitals: the interaction coefficient was +528 for general hospitals, indicating an average yearly decrease of 421 for academic hospitals but an average increase of 108 for general hospitals (Figure 6.2: M2).[2] The fact that the multilevel model (Figure 6.2: M3), although showing similar results as M2, is only marginally significant (as illustrated by the wide confidence intervals), indicates that the trend was not consistent across all hospitals.

Surgical, medical and obstetrical trends

In 1996, more than half of all RBC transfusions (58%) were administered to surgical patients as compared to 40% to medical patients. However, in 2005 this ratio had changed to 47:50% with medical transfusions becoming largest in number (see Figure 6.3a for absolute numbers). This shift is largely located in the general hospitals (surgical:medical 61%:35% in 1996 and 39%:57% in 2005), whereas in academic hospitals usage of both types decreased (Figures 3b and 3c). Obstetric transfusions represented 3% of all RBC transfusions and remained quite stable both in academic and general hospitals, with a small peak in 2000.

Distribution over age

For surgical RBC use, the decrease was present over almost the complete age range (except for the very young and very old); the absolute decrease was largest for the 65-85 year olds (Figure 6.4a). For medical transfusions (Figure 6.4b), there was a shift over time to the right, indicating an increased RBC use in the higher age groups (55-90 years). For obstetrical patients, transfusions also drifted slightly towards older obstetrical patients (Figure 6.4c).

Trends per diagnosis and procedure

RBC use decreased over time for cardiovascular diagnoses and procedures (-13% for medical and -36% for surgical use), as well as for surgery on blood vessels (-36%) and surgery on the

[1] The one cancer hospital is excluded in this comparison.
[2] The hospital type 'cancer hospital' is in the model but not shown in the results, since it was only one hospital. We repeated the analyses without the cancer hospital; this did not change the direction of the results.

Figure 6.3. Trend in total number of RBCs transfused to surgical (X), medical (O) and obstetrical (Δ) patients by year, in (A) all hospitals, (B) academic hospitals, and (C) general hospitals

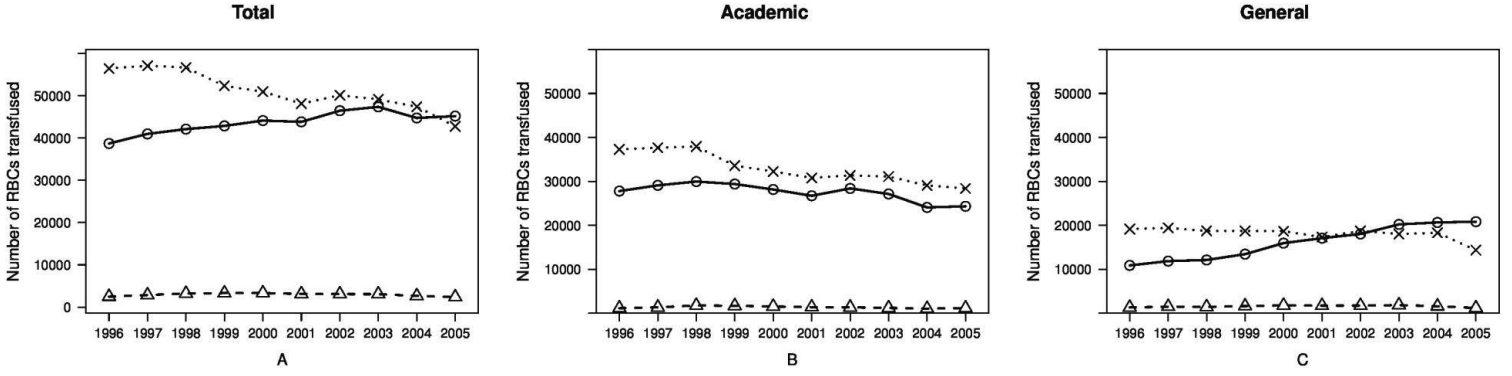


Figure 6.4. Trend in age distribution of RBC recipients for A) surgical, B) medical, C) obstetrical patients in the years 1996 (···), 2001 (– – –) and 2005 (——)

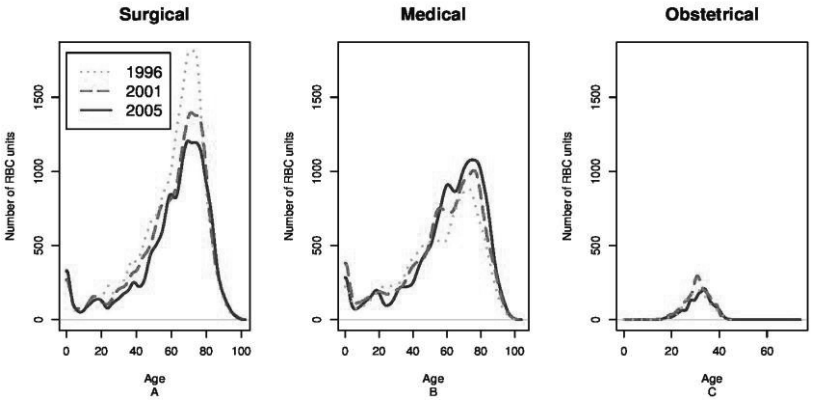
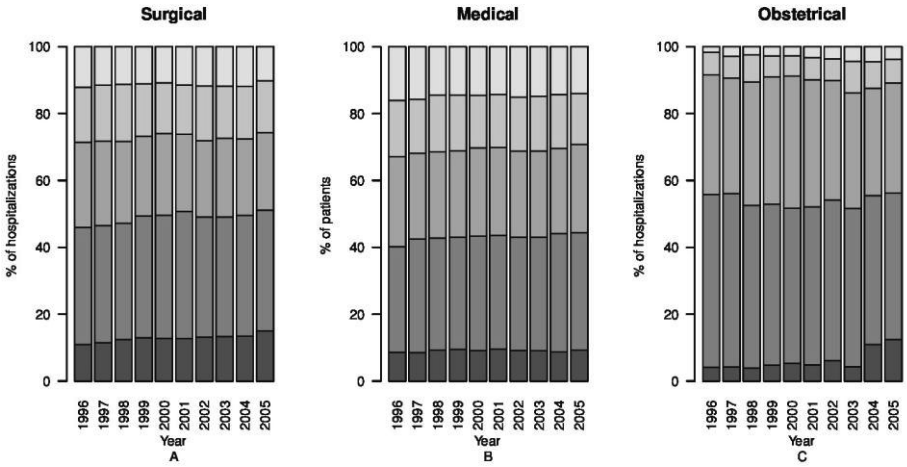


Figure 6.5. Trend in transfusion dose per transfusion episode for A) surgical, B) medical, C) obstetrical transfusions patients. Legend: 1 (■), 2 (■), 3-4 (■), 5-8 (■) and 9 (■) RBC units



musculoskeletal system and connective tissue (-27%). In contrast, an increase in RBC use was observed for diseases of blood and blood forming organs (+66%), neoplasms (+17%) and digestive system (+15%) (see Appendix 6.3). Neoplasms accounted for the largest medical use of RBCs each year, whereas the second to largest source of medical blood use changed from cardiovascular diseases (in 1996) to diseases of blood and blood forming organs (in 2005). For surgery-related RBC transfusions cardiovascular use decreased as well, whereas diseases of the digestive system became more prominent with an increase from 19% (in 1996) to 25% (in 2005) of all surgical RBCs used.

Dose of RBCs per transfusion episode

The most likely number of transfusions per episode was two (in 2005 36% for surgical, 35% for medical and 44% for obstetrical patients; Figures 5a, b and c). For both surgical and obstetrical patients, the use of one unit per transfusion episode increased over time (with 4 and 8 % points). For medical patients, over time the use of two units increased, whereas the use of more than three units decreased. In contrast, for obstetrical transfusion recipients the likelihood of receiving nine or more units increased from 2% (in 1996) to 4% (in 2005).

Occurrence and quantity of combinations of transfused units

During most transfusion episodes only RBC products were given. Over time however, combinations of RBC, FFP and PLT increased from 32% to 39% of all surgical hospitalizations (Appendix 6.4a). For medical patients, the proportion of transfusion episodes with only RBCs remained quite stable over time, whereas combinations of PLTs only and RBC&PLT increased (Appendix 6.4b). For obstetrical patients, no consistent trends were observed; notably in 2003 the proportion of hospitalizations with PLT only was especially high and the proportion of RBC only relatively low, but these both returned to lower occurrences in 2005 (Appendix 6.4c).

As shown in Appendix 6.5, patients who received a combination of different types of blood units, also received more units in total. For most combinations, the median number of blood units largely remained quite stable over time. For surgical patients however, the median number of PLTs in the combination RBC&PLT increased (Appendix 6.5a). Also for both surgical and medical patients, combinations of RBC&FFP&PLT had a lower median total number of blood products over time, primarily due to a decrease in number of RBCs (Appendix E6.5a and 6.5b). For obstetrical patients, the combination RBC&FFP&PLT showed an increase in median number of products for all types, but this trend was not consistent (Appendix 6.5c).

Discussion

In light of the steadily decreasing demand for RBCs in the Netherlands, we examined the changes in blood use for different patient groups historically from 1996 to 2005. Our selection of hospitals included was, when distinguishing between academic and general hospitals, comparable to the Dutch hospitals not included. Over the time period studied, average RBC use per hospital decreased with approximately 2.3% per year (equivalent to 421 red cell units). This trend however was not consistent across hospitals: Whereas in academic hospitals a large absolute decrease in RBC use was found, in general hospitals an average increase was observed (mostly due to an increase in the teaching hospitals). This might be an artefact of the particular selection of hospitals in our sample caused by fusion of hospitals, changing division of tasks, or possibly increased registration of transfusions, but it could also reflect a true increasing trend in general hospitals perhaps because of their older patient population and the centralization of complex care. Complex care (for example oncological treatments) is increasingly centralized in specialized hospitals. Consequently the initial treatment starts in the academic hospitals, after which the remaining care is transferred to a general hospital. Improved skills of clinicians due to specialization for example in complex surgery likely resulted in a decreased surgical RBC use in academic hospitals, whereas the increasing medical burden for general hospitals could explain the increase in medical RBCs. Time trends also differed between different areas of clinical usage. For surgical patients aged 65-85 years, the absolute number of RBCs transfused consistently declined during the ten-year time period, especially for surgery on the cardiovascular system, blood vessels, and musculoskeletal system and connective tissue. This surgical reduction can be attributed to developments in patient blood management strategies including cell saving techniques and non-invasive surgical procedures. Furthermore, the decline was likely to be stimulated in 2004 by the introduction in the Netherlands of new blood transfusion guidelines using hemoglobin level as a transfusion trigger ('4-5-6-flexinorm') [5]. Moreover, the use of RBCs only during a surgical hospitalization decreased relative to combinations of multiple product types. This might be explained by the increased PLT use and indicates that patients who would previously have received one RBC, are instead not transfused at all. In contrast, medical use of RBCs has been increasing, reflecting the ageing of the population. Since the year 2004, medical use accounts for the majority of RBC transfusions, mainly due to increased use for neoplasms, diseases on blood and blood forming organs, and the digestive system (of which the latter two also became more prominent over time relative to other diagnoses). Not surprisingly, the proportion of RBC used for obstetrical patients remained quite stable; postpartum bleedings are less suitable for patient management strategies as they are non-elective by nature. The slight shift towards higher age in obstetrics might be explained by the fact that the average age of the mother at delivery has increased from 30.32 years in 1996 to 31.06 in 2005 [6].

A common trend in the surgical, medical and obstetrical groups was that an increasing proportion of transfusion recipients received a smaller number of RBCs. We could not calculate the proportion of transfused patients (i.e., the 'transfusion rate') as only

information on transfusion recipients was available. However, we know that nationally the number RBCs per inhabitant is continuously decreasing: from 38 per 1,000 inhabitants in 2001, to 35 in 2005, and 27 in 2013 [7,8]. Moreover, the total number of hospitalized patients in the Netherlands increased, which, by increasing the denominator, leads to a lower transfusion rate. Accordingly, for most combinations, the median number of blood units per episode was stable or decreased over time. The transfusion rate would be useful to investigate in further studies, as it can also be used as a quality metric that is fed back to physicians, leading to less blood being transfused [9]. Moreover, collecting (aggregated) information on non-transfused patients adds to a more complete view of transfusion determinants.

Internationally, similar patterns in blood use have been reported: studies in the North of England, Northern Ireland and South Australia all identified medical patients as the main users of RBCs. Even though the exact numbers might not be completely comparable (because of differences between countries in case-mix of patients as well as differences in coding systems and classification of patients), common trends are observable. In roughly the same time period as the present study (1999-2009), the total RBC use in the North of England also dropped by 20%, due to a reduction in blood use per procedure [10]. In the same direction as the Dutch trends, North of England's RBC use in the years 2000-2009 changed from 41% (surgical):52%(medical) to 29%(surgical):64%(medical) (obstetrics and gynecology accounted for 6% of all RBC transfusions in both 2000 and 2009) [11,10]. These changes were observed only in patients aged between 50-80 years, while in the Netherlands also in the younger age groups a decrease in surgical and medical use was observed. In South Australia, medical transfusions in 2006 comprised the largest use of RBCs with 48% (Netherlands: 50% in 2005), with 46% to surgical patients and, 3.4% of RBC use was obstetrical, almost equal to the Netherlands [12]. Accordingly, South Australia found high RBC use for medical diagnoses such as hematology, medical oncology and gastroenterology. In Northern Ireland even 71% of the transfused were medical patients in 2010, leading to the conclusion that "with a likely plateau of efficiency having been reached in the surgical use of red cells, understanding the 'medical use' of red cells is increasingly imperative" [13]. In that study, surgical patients most likely received a single unit per transfusion episode (for comparison: in our present study this was two units), whereas medical patients, especially if being treated for cancer, were more likely to receive two-units, as in the present study. In Germany, the surgical (in this case also including intensive care and trauma patients) and medical use in a typical tertiary care hospital in 2008 was 58% (surgical) and 42% (medical), in line with the ratio in Dutch academic hospitals. In contrast to the Netherlands, RBC use increased between 2000 and 2009 by 10% [14]. Finland used 55% of RBCs for surgical patients (including obstetrical surgery) and 45% for medical patients in 2006 [15]. Differences in age distributions of patient populations could explain part of the differences between countries in RBC use, as both the surgical:medical ratio as well as the number of units transfused per episode is higher in the older age groups 10,140,15,16]. In addition, transfusion habits and hospital cultures lead to differences in use within both hospitals and countries (for example already 70% of the total variation in RBC use could be explained by the hospital effect in cardiac procedures [17]). Still these international

numbers, in conjunction with innovations in the areas of surgery, medication and alternative products, suggest that a medical:surgical ratio of 70:30 may be expected in the future. Insight in these trends and their impact on the demand for RBCs to be expected is important, as a reduced need for RBCs might negatively affect the availability and cost-effectiveness of blood products when the need for blood does not match investments of the blood bank in personnel and donors.

Conclusion

Over the period 1996-2005, RBC use in the Netherlands decreased, especially in the academic centers and for surgical patients. Transfusion recipients tended to receive fewer RBC units during a transfusion episode. For surgical recipients, transfusing combinations of blood products instead of only RBCs became more common over time. Similar to other countries, the data suggest a consistent trend towards conservative blood use for surgical patients, which is in line with the current focus on a restrictive, cost-effective transfusion policy. New data-warehouse initiatives should consider collecting detailed clinical information about transfusion recipients that could clarify blood use, such as hemoglobin status, comorbidities and use of medication, and, in order to calculate the transfusion rate, collect aggregated information on non-transfused patients.

Appendix 6.1. Description of data quality comprising completeness of transfusion data, and accuracy and missing values of diagnoses and procedures.
Appendix 6.2 (Table). Baseline characteristics of included (n=11) and remaining (n=82) academic and general hospitals in the Netherlands in 1996-2005.
Appendix 6.3 (Table). Change (in %) in number of transfused RBCs from 1996 to 2005 for various diagnoses and procedures .
Appendix 6.4 (Figure). Use of combinations per transfusion episode in a) surgical, b) medical, c) obstetrical patients by year.
Appendix 6.5 (Figure). Median number of blood products per transfusion episode, given a combination for a) surgical, b) medical and c) obstetrical patients.

All Appendices available online: https://www.dovepress.com/historical-time-trends-in-red-blood-cell-usage-in-the-netherlands-peer-reviewed-fulltext-article-IJCTM

References

1. Statistics Netherlands (CBS).
2. http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=71566NED&D1=0-1&D2=0&D3=0&D4=0&D5=l&D6=a&HD=140311-1454&HDR=G4,G2,T,G5&STB=G1,G3&CHARTTYPE=1, Accessed on 26 November 2014.
3. Borkent-Raven BA, Janssen MP, Van Der Poel CL, Schaasberg WP, Bonsel GJ, Van Hout BA. The PROTON study: Profiles of blood product transfusion recipients in the netherlands. Vox Sang. 2010;99(1):54-64. doi: 10.1111/j.1423-0410.2010.01312.x.
4. Biggin K, Warner P, Prescott R, McClelland B. REVIEWS: A review of methods used in comprehensive, descriptive studies that relate red blood cell transfusion to clinical data. Transfusion. 2010;50(3):711-718. doi: 10.1111/j.1537-2995.2009.02459.x.
5. R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.
6. Kwaliteitsinstituut voor de Gezondheidszorg CBO (CBO). Richtlijn Bloedtransfusie. Alphen aan de Rijn: Van Zuiden Communications; 2004.
7. Paone G, Brewer R, Likosky DS, et al; Membership of the Michigan Society of Thoracic and Cardiovascular Surgeons. Transfusion rate as a quality metric: is blood conservation a learnable skill? Ann Thorac Surg. 2013;96(4):1279-1286.
8. Trends and Observations on the Collection, Testing and Use of Blood and Blood Components in Europe - 2001-2011 Report; Directorate for the Quality of Medicines & HealthCare of the Council of Europe (EDQM), Strasbourg, CD-P-TS, 2014
9. The Collection, Testing and Use of Blood and Blood Components in Europe – 2013 Report; Directorate for the Quality of Medicines & HealthCare of the Council of Europe (EDQM), Strasbourg, CD-P-TS, 2016 in preparation.
10. Statistics Netherlands: Maternal age at birth over time; accessed 2 February 2016 at: http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=37744NED&D1=0&D2=0-35&D3=46-55&VW=T
11. Tinegate H., Chattree S., Iqbal A., Plews D., Whitehead J., Wallis JP and on behalf of the Northern Regional Transfusion Committee (2013), Ten-year pattern of red blood cell use in the North of England. Transfusion, 53: 483–489. doi: 10.1111/j.1537-2995.2012.03782.x
12. Wallis JP, Wells AW, Chapman CE. Changing indications for red cell transfusion from 2000 to 2004 in the north of England. Transfusion Medicine. 2006;16(6):411-417. doi: 10.1111/j.1365-3148.2006.00702.x.

13. Allden RLW, Sinha R, Roxby DJ, Ireland S, Hakendorf P, Robinson KL. Red alert: a new perspective on patterns of blood use in the south australian public sector. Aust Health Rev. 2011;35(3):327-333. http://dx.doi.org/10.1071/AH10957.
14. Barr PJ, Donnelly M, Morris K, Parker M, Cardwell C, Bailie KEM. The epidemiology of red cell transfusion. Vox Sang. 2010;99(3):239-250. doi: 10.1111/j.1423-0410.2010.01349.x.
15. Seifried E, Klueter H, Weidmann C, et al. How much blood is needed? Vox Sang. 2011;100:10–21.
16. Ali A, Auvinen M-K, Rautonen J. Blood donors and blood collection: the aging population poses a global challenge for blood services. Transfusion 2010;50:584-588.
17. Schramm W and Berger K (2007), Influencing blood usage in Germany: what is optimal use?. Transfusion, 47: 152S–154S. doi: 10.1111/j.1537-2995.2007.01374.x
18. Jin R, Zelinka ES, McDonald J, Byrnes T, Grunkemeier GL, & Brevig J (2013). Effect of hospital culture on blood transfusion in cardiac procedures. The Annals of thoracic surgery, 95(4), 1269-1274.
19. Paas RA, Veenhuizen, C.W. Onderzoek naar de betrouwbaarheid van de LMR. Rapportage voor de ziekenhuizen.Utrecht: Prismant, februari 2002.

Chapter 7

Prediction of the anti-RhD donor population size for managerial decision making

Vox Sanguinis 2016

L.R. van Hoeven
M.A. Berkowska
O.J.H.M. Verhagen
H. Koffijberg
C.E. van der Schoot
M.P. Janssen

Abstract

Background

RhesusD (RhD)-negative women pregnant with a RhD-positive child receive prophylactic injections to prevent hemolytic disease of the newborn. Because of the success of the prophylaxis, the number of naturally immunized women has decreased, thereby also decreasing the number of potential donors who provide the plasma from which the prophylaxis is made. As the current donor pool is ageing, the availability of the prophylaxis is threatened.

Objectives

Objectives are to investigate whether the anti-D population and the changes therein can be described by a relatively simple model, in order to determine the impact of ageing of the anti-D donors on the decline of the population and how many new donors should be recruited to meet future supply demand.

Methods

Data on Dutch anti-D donors in 1994-2013 were used to simulate the donor population size and age composition for various donor recruitment scenarios.

Results

With a continuous influx of 27 new donors per year and a donor stopping rate of 10% per year, the population size will stabilize at 195 donors, with 2.3% of donors stopping annually due to reaching the donor age limit. A formula is derived to estimate which donor recruitment and retention efforts are required to maintain a pre-specified donor pool.

Conclusion

A relatively simple model can already describe and predict the size of the anti-D donor population and the impact of ageing accurately.

Background

In The Netherlands pregnant women undergo rhesus-D (Rh-D) typing around week 12 of pregnancy, and the RhD-negative women are tested again at week 27 to type the foetus [1]. Each year approximately 27,000 women are found RhD negative, of which 60% are carrying an RhD-positive child [2]. To these women, anti-D injections are administered both at week 30 and within 48 hours after delivery to prevent the mother from producing antibodies against the RhD-antigen. These injections prevent hemolytic disease of the newborn when the mother becomes pregnant again with a RhD-positive child [3]. Before the introduction of the anti-D immunoprophylaxis, this disease was a major cause of perinatal death. Since 1969, women in the Netherlands (in some countries earlier) have been receiving the prophylaxis directly after delivery and since 1998 also antenatally in the 30th week of pregnancy [1]. This reduces the residual risk of immunization to 0,31% [4] (1,5 to 0,2% according to [5]).

In order to fulfill the need for the anti-D prophylaxis, several strategies (or combinations of strategies) can be considered. In the Netherlands the prophylactic anti-D immunoglobulins are partly imported and partly obtained from the plasma of immunized donors. Whether to be self-sufficient is a trade-off for policy makers between the feasibility of recruiting enough donors or immunizing the existing donor base, the desirability and cost of importing, and the financial resources available. In this paper we focus on securing an anti-D donor population. Sustaining an anti-D donor base is not straightforward as it requires recruitment of donors who are RhD negative, aged between 45 and 70 years, and immunized (i.e. have antibodies against the RhD antigen), either naturally by giving birth to a RhD positive child or intentionally by administering a small amount of D antigen. Apart from these practical constraints, the recruitment of anti-D donors is challenging because of the high donation frequency required and the burden of immunization. New donors are generally identified through targeted recruitment efforts. With a few exceptions, all current female anti-D donors are naturally immunized. Since the anti-D prophylaxis is successful in preventing the forming of antibodies, it decreases the number of naturally immunized women that would be eligible to be an anti-D donor. Women who got pregnant before the introduction of the prophylaxis in 1969 (when average age of pregnancy was 24 years [6]) will mostly have reached the age of 70 years between 2010 and 2020, and therefore drop out of the donor pool because of the age limit of blood donor eligibility (that is, 70 years in the Netherlands). Another reduction in the number of potential future donors is a result of the introduction of the antenatal prophylaxis in 1998: this further reduced risk of immunization, thereby halving the number of eligible donors that are immunized naturally (as the average age of first pregnancy was 29 years in 1998, women who were pregnant in 1998 reach the age of 45 as from 2015 [6]). Therefore, it is feared that the ageing of the current donor population might jeopardize the continuation of the anti-D program.

In the present study, historical donor records were analyzed in terms of donor age, influx of new donors and retention rate. Based on the results, we performed a simulation study in which the size and age distribution of the donor population is modelled. The aim was to

develop a relatively simple model to (1) examine the effect of ageing on the size of the anti-D donor population, and (2) determine how many new donors need to be recruited to stabilize the anti-D donor population size at a pre-specified level in order to meet supply demands.

Methods

Data

Administrative data on anti-D donors were provided by Sanquin the national blood bank for the time period between September 1994 and December 2013. This cohort includes data on demographic characteristics of the donors, boosts (injections of RhD positive cells to stimulate the production of RhD antibodies) and donations. Donor exclusion criteria for this study were age below 45 years as this is the minimum age for hyperimmunization [7], and low titer without hyperimmunization (these donors contribute only a small portion of all donations).

Donor classification

For the analysis, we classified donors as either repeat donor or potential donor. Repeat donors were defined as donors who either had more than four donations in total, or an average of more than 2 donations per year (in the years that they donated). Potential donors had both a low total number of donations (<=4) and a low yearly donation rate (<=2). In this study, the donor population was defined as all repeat donors (excluding potential donors) as they provide the vast majority of donations (i.e. 99%). New donors were defined as donors who did not donate between September 1994 (starting date of the cohort) and January 1996, in order to increase the likelihood that new donors were included from the beginning of their donor career and limit the effects of left-censoring. Only donors with their first registered donation in 1996 or later were considered as new donors and included in the analysis of the drop-out rate. In the final simulation of the Dutch anti-D donor population size, also repeat donors who donated in 1994 or 1995 (n=216), were included in the population at the start of the simulation.

Drop-out rates

It was checked whether age has an effect on the probability of becoming a potential donor (as opposed to a repeat donor) by fitting a linear regression model with age at first donation as a predictor of the proportion potential donors. The yearly stopping rate of repeat donors was estimated using Cox proportional hazard regression. Donor age was added to the model as a predictor to test whether the stopping rate was dependent of age. Taking into consideration the fact that that donors reaching the age limit of 70 years are forced to stop donating, donors who stop donating at an age over 69 years are considered to be censored. This age constraint is later incorporated in the model but does not affect the estimated drop-out rate. The model therefore explicitly differentiates between donors who are forced to stop because of age and

donors who drop-out for an external reason. Excluded from the analyses were donors who only donated in the last observation year 2013. The cumulative hazard and observation times resulting from the Cox PH model were used to fit a linear regression model to estimate the annual hazard rate.

Simulation

Combining these parameter values, a Monte Carlo simulation was performed to predict the size and age distribution of the donor population from 1996 onwards for the subsequent 50 years, as a function of the annual donor drop-out rate and the influx of newly recruited donors. The model contained the following parameters: the number of repeat donors at the beginning of a year, the number and age distribution of new repeat donors for each year, and the stopping rate for repeat donors. The outcome of the model was compared to the observed population values in order to verify the accuracy of the prediction. For the prediction of the future donor population size, assumptions concerning the new donor rate and its associated age distribution were required. Different scenarios with a varying number of new donors were simulated: In the first scenario it was assumed that 27 new donors are recruited each year (which was the average number of new donors per year over the study period). In the second scenario 15 new donors per year were assumed, which was the mean number of new donors recruited in the last 5 years. Finally, in the third scenario a decreasing number of new donors per year was assumed, starting with n=15 and declining by 5% per year (as was actually observed in recent years). For each scenario, it was predicted how many donors stop because of reaching the maximum age of 70 years, so called 'old age stoppers'. The simulation was repeated 1,000 times, which resulted in convergence of the population size and was therefore considered a sufficient number. The 95% confidence intervals around the estimates were computed as the 2.5% and 97.5% percentiles from all simulations. All analyses, including simulations, were performed using the statistical software program *R,* Version 3.2.0 [8].

Theoretical approximation

In addition, a formula was derived to theoretically approximate the stable population size. The formula was applied for several scenarios, varying the number of new donors (20, 40, 60 or 80 per year) and repeat donor stopping rates (5%, 10%, 15% or 20%). The reason for starting with the scenario of 20 new donors per year is that a lower number (as in most recent years) would not be enough for securing a stable donor base. The highest recruitment scenario of 80 new donors per year could be realistic with a recruitment program in place, as illustrated by the higher number of newly recruited donors in the past. It was assumed that the age distribution of new donors would be constant and similar to that of donors in the past. When both the simulation and the formula correctly estimate the size of the donor population their predicted stable population size should agree.

Results

Donor characteristics

In the selected time period from January 1996 up to December 2013, 823 donors were active resulting in 28,028 donations registered (Table 7.1). Fifty-seven (0.2%) donations were excluded from analyses due to missing donor identification numbers. Repeat donors comprise 80.4% of the donor population, accounting for 99.0% of all donations. Over 90% of all donors are female (90.6% of repeat donors and 96.0% of potential donors). The mean age at first donation was 52.8 for repeat donors and 53.4 for potential donors. Figure 7.1(a) clearly shows the shift in age of the donor population over time towards higher age. Repeat donors had a median of 27 donations in total per donor, with a mean length of donor career of 6.4 years. Potential donors delivered a median of one donation, with a mean length of donor career of 0.3 years.

Table 7.1. Anti-Rh(D) donor characteristics during 1996-2013

	Repeat donors	Potential donors
N (% of total number of active donors)	662 (80.4%)	161 (19.6%[3])
Total number of donations (% of all donations)	27,810 (99.0%)	272 (1.0%)
Women (%)	600 (90.6%)	161 (96.0%)
Mean (SD) and median (IQR) age at first registered donation	52.8 (6.5), 52.0 (46.4-58.0)	53.4 (7.8), 52.4 (46.9-58.7)
Mean duration donor career in years as captured in our database in years (SD)	6.4 (5.1)	0.3 (0.7)
Mean (SD) and median (IQR) number of donations per donor per year	6.0 (2.6); 27 (11-63)	1 (1-2); 1.6 (0.8)

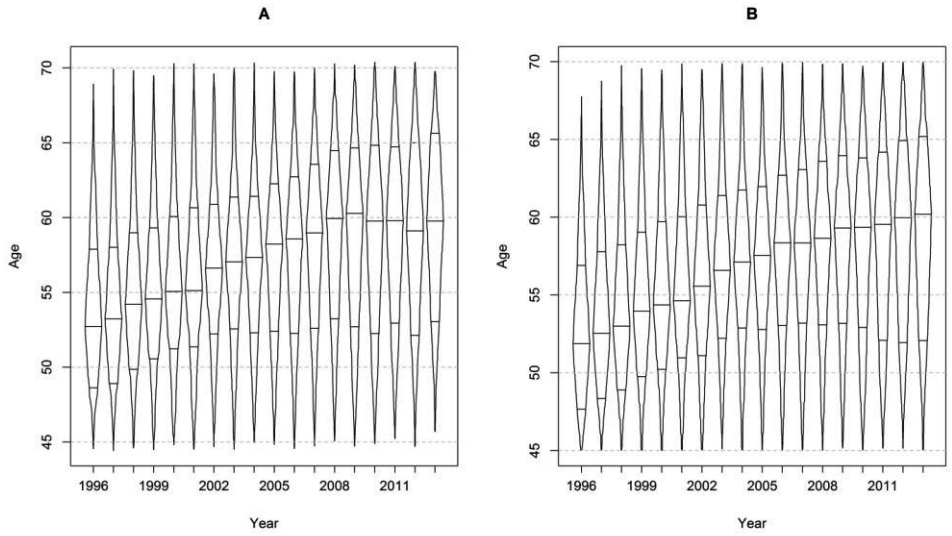
Drop-out rates

The probability for new donors to become a potential donor instead of a repeat donor was estimated at 31%, with no significant effect of age of -0.08 (p=0.8). This means that recruitment efforts should aim at a number of new donors per year that is the desired number of donors multiplied by a factor of 1.45 (1/(1-.31)), for a required number of repeat donors to

[3]Please note that the 19.6% of potential donors in Table 7.1 is unequal to the 31% mentioned under Drop-out rates. This is because Table 7.1 additionally includes repeat donors who could not be identified as new donors (their first registered donation was at the start of the study period), whereas the 31% is out of all new donors.

be included. For repeat donors, a constant drop-out rate of 10% per year was derived from the hazard rate. There was a clear tendency, but no statistically significant effect of age on drop-out risk (the hazard rate increased by factor 1.015 for a one-year increase in age; CI 1.00-1.03; p=0.1).

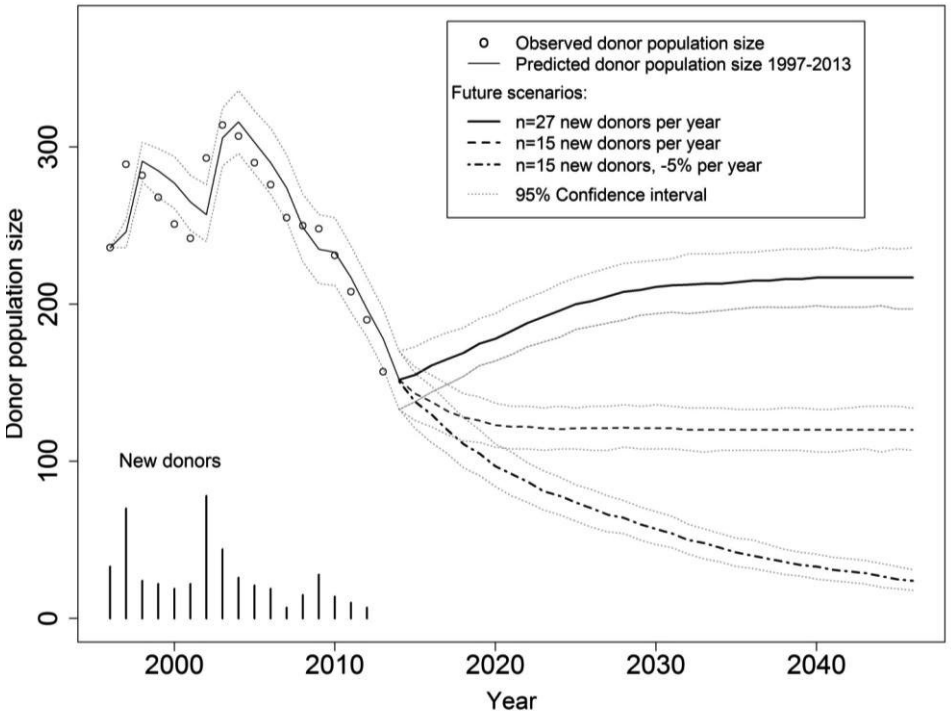
Figure 7.1. A) Observed repeat donor age distribution; B) Simulated donor age distribution per calendar year. In this box-percentile plot the proportion of individuals with a higher/lower age than indicated on the Y-axis is represented by the width of the bar relative to the median (which is the greatest width for each bar). For ages above the median the width of the bar represents the proportion of individuals with a higher age than indicated on the Y-axis, for ages below the median the width of the bar represents the proportion of individuals with a lower age than indicated on the Y-axis. The two other horizontal lines above and below the median line mark the 25th percentiles.



Simulation results

The simulation seems to predict donor age distribution (Figure 7.1b) and population size (Figure 7.2) quite accurately, with the observed population size mostly falling within the 95% confidence interval. In scenario 1 (annual number of new donors = 27) a stabilized population would consist of 195 donors (95% CI 168-222). Whereas the proportion of donors that stopped because of reaching age 70 was 0% in 2000 and initially increased, this proportion of old age stoppers is expected to converge to 2.3% of all donors in 2020 (Figure 7.3). In case of a steady number of new donors per year, the predicted future population size will stabilize (Figure 7.2, scenario 1 and 2), but as long as the number of new donors decreases each year, the size of the donor population will continue to decline as well.

Figure 7.2. Predicted (–) and observed (○) donor population size for different future donor recruitment scenarios: n=27 (—), n=15 (---) and n=15 decreasing with 5% each year (·-·-) (assuming a 10% stopping rate)



Theoretical approximation

An approximation of the donor population size, provided that there is a stable donor influx, is given by Formula 7.1. The input variables are the same as for the simulation: age distribution of new repeat donors, stopping rate and number of new donors. This formula results in similar estimates as compared to the simulation performed above, for example: In scenario 1 (n=27 and p=10%), the simulated long-term population size was 195 (95% CI 168-222), according to the formula it is 207; similarly the proportion of old age stoppers was 2.3% (95% CI 0.5%-4.5%) (simulation) and 2.8% (Formula 7.1).

Generally, as can be inferred from the first part of the formula ($N^* = nD$), when the number of new donors increases by a factor of 2, the population size also increases by a factor of 2. When the drop-out rate increases, the estimated population size decreases, but not proportionally: as the drop-out rate decreases, the increase in population size becomes relatively smaller. This is also evident from Table 7.2, which shows the effect of varying the new donor rate as well as the stopping rate of repeat donors (which in the above was fixed at 10%) on the estimated long-term donor population size.

Table 7.2. Theoretical approximation of long-term population size in the case of a stable donor recruitment scenario

Drop-out rate per year	New donors per year			
	20	40	60	80
5%	224	449	673	897
10%	158	315	473	631
20%	94	187	281	374

Formula 7.1. Theoretical approximation of donor population size N* in the case of a stable donor recruitment scenario

$$N^* = n\,D$$

where N* is the long term population size, n is the stable number of new repeat donors per year, and D is the mean number of years that a new donor will be actively donating. This donation time D will depend on the donor stopping rate (p), but also on the age of new donors, since their donating is restricted to age < 70. Here n(s) is the number of new donors by their age at first donation s.

$$D = \sum_{s=45}^{s=69} n(s)\,\frac{1}{p}\left(1-(1-p)^{(70-s)}\right)$$

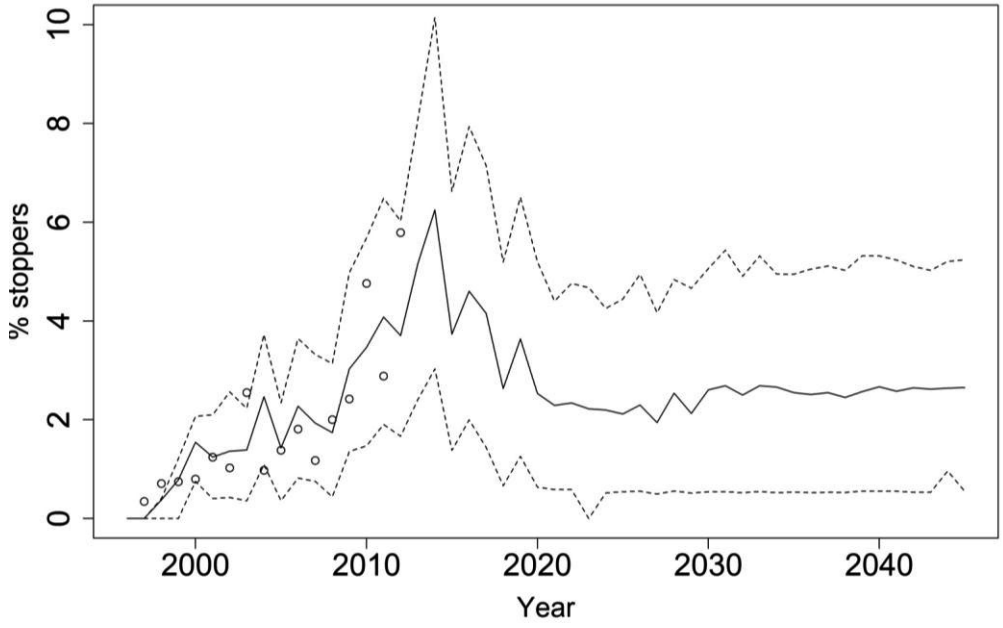
The older the newly starting donors, and the lower the stopping rate p, the larger the number of donors who stop due to reaching the age of 70.
Proportion stoppers each year of the total population due to reaching age 70:

$$\frac{\sum_{s=45}^{s=69} n(s)(1-p)^{(70-s)}}{N^*}$$

Discussion

With only three parameters, our model could quite accurately describe the course of the anti-D donor population from 1996 to 2013. Moreover, the model allows prediction of the future donor population size for various donor recruitment scenarios and drop-out rates. In general the repeat donor population (which is 69% of all new donors) is a steady group; the observed age-independent stopping rate of 10% means that, 10 years after donor inclusion, approximately 35% (i.e., 0.910) are still in the population. However, the fear of being unable to sustain a sufficiently large donor population, due to ageing of donors is not unfounded as the number of naturally immunized women decreased to an estimated 50 naturally immunized women per year (27,000*0.6*0.0031=50,2 women for whom anti-D injections are not

Figure 7.3. Percentage old age stoppers of all donors per year assuming a 10% stopping rate and annual number of new donors=27 including 95% confidence interval (---)



effective). Although some external factors such as immigration might increase the number of naturally immunized women, the effect on the total supply of plasma will be negligible. In the Netherlands, however, it is reassuring that for the current donor population, the proportion of donors stopping due to old age has already reached its peak in 2014. It is expected that from there on the proportion old age stoppers will decrease, given that a sufficient number of new donors will be recruited yearly. More specifically, simulations up to the year 2050 predict that, given the current dropout rate and donor population, from the year 2020 onward the proportion of old age stoppers will comprise approximately 2.7% of the total donor population. This is however provided a similar age distribution of new donors as in the past and a 10% drop-out rate. Yet in the future, the increasing need for boosting might lead to higher drop-out rates. It is apparent from our results that, for moderately high stopping rates, ageing would be less of a problem because donors drop out before reaching the maximum donor age. In contrast, a lower dropout rate would demand relatively higher recruitment efforts to maintain a stable donor population, as the proportion of old age stoppers would become higher. Insight into the reasons for donors to stop or to decline boosting will be helpful to sharpen assumptions on retention rates. Currently there is a lack of information on profiles of successful repeat anti-D donors.

How many donors are actually needed in the Netherlands is a decision for the policy makers. In order to meet the national demand for the prophylaxis, approximately 32,000 units

are needed corresponding to 3200 donations per year (one donation is sufficient for ten products) [9]. Assuming an average of six donations per (repeat) donor per year, 533 donors would be required to provide all units. During the study period, the need for the prophylaxis has decreased as a result of restricted targeting, first in 2008 by 40% when prophylaxis administration was restricted to mothers with children only, and in 2011 by 25% with the introduction of RhD determination of the foetus at week 27 [9]. It must be noted that this paper focusses on only one strategy for increasing the anti-D yield. Besides increasing the number of donors, other aspects of donation policy might be changed: for example in the last years, the maximum donation frequency has increased to once a week, a higher plasma volume is retrieved per donation, and restrictive exclusion criteria might be loosened.

The data and formula provided in this paper might be useful for policy makers to estimate the effect of recruitment efforts on the size of the donor population. For example, when 27 new donors are recruited per year, the simulated population size is 216 (95% CI 198-235). The uncertainty surrounding this estimate is attributable to the random drop-out of donors. If, due to chance, the drop-out is proportionally higher in younger donors, the estimated population size will turn out lower; vice versa if more of the elder donors drop-out, the population size will be higher. The simulation results agree with the formula, which estimated a stable population size of 213 donors. Implementation of the formulas described in this paper is available in R (Appendix 7.1). The user can specify the expected stopping rate, the expected number of new repeat donors recruited per year and their age distribution; the result is an estimate of the long-term population size and the proportion of donors that stop due to reaching the maximum donor age. The use of this formula is not restricted to anti-D donors. In principle the formula can be applied to all other donor types as well. As the problem of ageing has also emerged for whole blood donors in other countries [10,11], this type of calculation might be more widely applicable.

The model could be further refined by making a finer distinction between different types of donors and donation patterns [12], and predicting the number of donations instead of the number of donors. Still, the total number of donations can be estimated by multiplying the number of donors by the average number of donations per donor. Hence, with a relatively simple model it is possible to predict the composition of the Dutch anti-D donor population. This allows informing management on the efforts required in terms of donor recruitment and donor retention to maintain a sufficiently large donor population.

Appendix 7.1 (R code of Formulas). Theoretical approximation of donor population size and the proportion of donors stopping due to reaching age 70. Available online: http://onlinelibrary.wiley.com/store/10.1111/vox.12400/asset/supinfo/vox12400-sup-0001-SupInfo.txt?v=1&s=e2db583a606ab7877b5fa24bd30040d5a3debeac

References

1. RIVM: Bloedonderzoek zwangeren: Beleid bij bloedgroep Rhesus (D)-negatief. 2014; accessed 9 December 2015 at: http://www.rivm.nl/Onderwerpen/B/Bloedonderzoek_zwangeren/Voor_professionals/Aandoeningen/Bloedgroepantistoffen/Beleid_bij_bloedgroep_Rhesus_D_negatief
2. van der Ploeg CPB, Pal van der SM, Oomen P et al.: Procesmonitoring prenatale screening infectieziekten en erytrocytenimmunisatie 2007-2009. 2010 Leiden, TNO Kwaliteit van Leven.
3. Klein HG, Anstee DJ: Mollison's Blood Transfusion in Clinical Medicine 11th edition, Blackwell publishing, 2005; 114-23.
4. Koelewijn JM , de Haas M, Vrijkotte TG et al.: One single dose of 200 μg of antenatal RhIG halves the risk of anti-D immunization and hemolytic disease of the foetus and newborn in the next pregnancy. Transfusion 2008; 48:1721-29.
5. Crowther CA, Middleton P, McBain RD: Anti-D administration in pregnancy for preventing Rhesus alloimmunisation. Cochrane Database Syst Rev 2013; 2.
6. Statistics Netherlands: Maternal age at birth of first child over time; 2015; accessed 9 December 2015 at: http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=37744NED&D1=1&D2=a&D3=19,48&HDR=T&STB=G1,G2&VW=T
7. Sanquin Blood Supply: Richtlijn Immunisatie van donors en verkrijging van specifiek plasma. 2010.
8. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/
9. Sanquin blood supply: 11.000 anti-D injections less for future mothers; accessed 5 February 2016 at: http://2011.jaarverslagsanquin.nl/jaarverslag-2011/aDU1007_11000-anti-D-prikken-minder-voor-aanstaande-moeders.aspx
10. Zou S, Musavi F, Notari EP et al.: Changing age distribution of the blood donor population in the United States. Transfusion 2008; 48(2), 251-257.
11. Lattimore S, Wickenden C, Brailsford SR et al.: Blood donors in England and North Wales: demography and patterns of donation. Transfusion 2015; 55(1), 91-99.
12. Veldhuizen I, Folléa G, de KW et al.: Donor cycle and donor segmentation: new tools for improving blood donor management. Vox Sang 2013; 105(1), 28-371. Levy PS, Lemeshow S. The population and the sample. In: Sampling of Populations: Methods and applications. 4th ed. John Wiley and Sons, New York, USA; 2008. p. 11-42.

Chapter 8

General discussion

Many questions still exist regarding blood transfusion, most of which can be gathered under the following themes: 'what is the optimal composition of the blood donor pool (age, gender, blood groups)?'; 'what are optimal characteristics of blood products (type, storage duration, processing)?' and 'what is the optimal transfusion regime for a patient (which products, how many and when exactly)?'. By 'optimal' use of blood we predominantly mean 'leading to favorable patient outcomes', although other aspects such as feasibility and cost-effectiveness may also play a role when evaluating blood use. More long-term strategic questions include 'how many blood products are needed in the future to meet hospital demand?' and related, 'how many donors are needed in the future?'. In order to answer the questions above, data on the complete transfusion chain from donor to patient are required. Gathering these data, which have to be derived and linked from multiple sources, is not trivial and a challenge in itself. Within this thesis, we describe the development and validation of the Dutch Transfusion Data warehouse (DTD) and we analyzed sets of donor and patient data to answer some specific research questions.

In this Discussion we will reflect on the value of these data and transfusion data in general, the interpretation of indications for transfusion, and the most important lessons learned with respect to acquiring a valid data warehouse. We conclude with recommendations for future research and use of the DTD.

Value of transfusion data

Evidently, 'data' are not equivalent to 'useful information'. Therefore, one of our goals was to increase the usability of routinely registered transfusion data. The idea is that by using the DTD (Chapter 2), many processes and risks in the transfusion chain can be quantified, which may ultimately contribute to the development of clinical and operational guidelines. Even though convincing statistics alone might not be enough to change transfusion practice of doctors, valid information should form the foundation for the decision to transfuse a patient [1,2]. The prediction models as described in this thesis (Chapters 6 and 7), although based on relatively old data, have the potential to support managerial decision making. Currently, the production of blood products is a direct response to the request from hospitals. Similarly, the frequency with which donors are called to donate and the recruitment of new donors are based on the current demand. When the future demand for blood products is known as well, policies can timely be adjusted accordingly. For example, it might be investigated whether costly recruitment campaigns are necessary, or whether other solutions are more cost-effective, such as temporarily increasing donation frequency or donor calling efforts. This may solve shortages on the short term while also anticipating an expected future decrease in demand. This illustrates how analyses of time trends can be valuable, especially when performed in combination with cost-benefit analyses. The same principle holds for hospitals: feedback in terms of between- and within-hospital benchmark information on blood use, transfusion triggers and patient outcomes, should inform best clinical practice.

Clinical indications

One way to increase the usability of the data and the uniformity of their use, is to expand the DTD with an extra layer of information. If we are able to determine the indication for transfusion with an acceptable level of accuracy, for example using an automated selection algorithm (Chapter 4), this information can be incorporated in the data warehouse as an extra variable. First however, the algorithm should be externally validated and its performance in academic, teaching and peripheral hospitals should be verified. If necessary the algorithm could be adapted to improve the predictions. Ultimately the predicted indications are defined –and limited– by the way in which clinical indications are represented in hospital systems.

Representation versus reality

Clinical indications are forced into the format of the coding systems used by hospitals and this affects our way of thinking about and therefore analyzing them. The algorithm we developed selects the most likely indication out of all data available and hence is bound by the diagnosis and procedure codes in use. The algorithm selects either a procedure or a diagnosis code as the transfusion indication. In reality, this representation might be too simplistic and it may be worthwhile to investigate whether a more complex representation might better capture the right indication. Also, more variables could be included in the algorithm such as patient age, gender, previous treatments, and the number and type of blood products received. The added value of a more complex decision structure of the algorithm could be explored using machine learning techniques. As this would introduce a higher risk of overfitting, external validation is especially important to ensure generalizability and identify patient characteristics that affect the performance of the algorithm. Moreover, a relatively large sample would be needed that includes cases with a known 'true' indication; by either creating a gold standard in other hospital datasets or, if available, using other databases with a known gold standard to validate the indication selection.

Solutions at the source

A solution at an earlier stage would be to adapt the way clinical information is registered at the source. A possibility is to implement a standardized form that pops up whenever a transfusion is requested at the hospital, where doctors have to register the indication for the transfusion whenever ordering blood products. A barrier for implementing such a system is that registration takes extra time. Therefore the system must be as simple and intuitive as possible. A transfusion indication form might follow the following structure: first the doctor chooses whether the transfusion is necessitated by blood loss, and then whether it is "because of surgery", "because of illness" or "because of treatment of an illness". In addition, if a low hemoglobin value was the direct transfusion trigger, this should be indicated. To determine exactly which information is needed and in what format, close communication between researchers and clinicians should be leading.

Whether a pop-up system like this is more (cost-)effective than an algorithm that predicts the indication using routinely registered data is a subject for future research, as this will also depend on the potential to derive sufficiently accurate predictions. A balance must be found between the benefits of registering more and more precise data and the efforts required to turn this information into value.

Would such a registration be implemented in practice, this provides us with new information and potential uses. First, tracking and evaluating appropriate transfusion triggers is facilitated [3]. Second, it might even be possible to develop a reverse model: if we know what the transfusion indication was, we might be able to develop a model that uses electronic patient records to predict whether and when someone needs a transfusion to support clinical transfusion decision making. This would require the predictive information to be registered and processed before the transfusion is actually needed.

A valid data warehouse: Opportunities and pitfalls

At present the data included in the Dutch Transfusion Data warehouse seem sufficiently valid, also when compared to validity outcomes of similar international databases [5,6,7] (Chapter 3). This provides 'proof of principle' that developing a donor-recipient data warehouse is feasible and will be valuable for answering future research questions. We learned however that a thorough validation of the data is necessary and that this should take into account every step in data collection and processing. Referring back to the steps of data registration, extraction and interpretation, we will discuss several pitfalls and opportunities.

Registration

As the DTD is built from routinely registered data that originate from multiple sources, differences in data recording exist both between and within data providers, and the data registered might be inappropriate or insensitive [4]. We assessed and improved data quality aspects by examining –among other things– completeness, plausibility, uniformity, level of detail of information and time patterns (Chapter 3). Remaining issues include the unreliable registration of the time of transfusion and the time of measurement of blood values; such variables are more likely to reference the time that the blood product or measurement was requested rather than the time that it was administered. Recent developments such as the use of a 'cybertrack' system might make recordings more accurate in the long run. As the data are primarily intended for reimbursement of medical expenses, bias might occur due to selective registration. Examples we encountered include registration of only the most expensive diagnosis or procedure when there are multiple, procedures that are registered in duplicate under different specialisms, and procedures such as bypass surgery that in some cases are registered as a diagnosis and in other cases as a procedure. The practice of selective registration may lead to biased results, however if bias is similar in all hospitals the data can still be used for

benchmarking purposes. This underlines the importance of similar extraction and processing of the data in each hospital.

Extraction

The extraction and selection of data involves deciding which exact variables are collected, the time ranges that are included in the extraction and which filters are applied to the data. For example, when we want to be able to attribute a transfusion to a procedure, should we extract only procedures occurring within the same hospitalization as the transfusion, or all procedures within a week before or after transfusion (or for example 1 or 60 days instead)? This selection choice can lead –especially when dates are registered incorrectly– to procedures being missed in the data extraction. How many procedures are missed should ideally be checked for a sample of cases for which a wider selection is available, to determine the effect of different cut-off points. To overcome missing data due to incorrect registration of dates, the safest solution is to extract data with a one-day margin. In addition, subgroups of recipients who receive regular blood transfusions and treatment within a broad time interval such as hematological patients, should perhaps be extracted using customized selection methods.

In the final part of data extraction, patient hospital identification numbers are encrypted to ensure the patients' privacy. This makes it impossible to follow a unique patient who is transferred to another hospital. As this concerns a substantial proportion of patients, a more uniform way to encrypt privacy-sensitive data that enables linkage across centers might be worthwhile. Also, in some cases the data are –in light of certain research questions– rather outdated at the moment they are extracted. When we are able to further automate data extraction with a safe connection for remote transfer, this will facilitate timely use of the data. (Semi-)automated extraction will expectedly also facilitate the inclusion of more and more diverse centers, which is desirable for maximizing the representativeness of the data and for studying rare patient outcomes such as transfusion reactions. Currently successful extraction and participation is heavily dependent on the willingness of individual employees in the hospitals. Further standardization of extraction and participation by hospitals would be stimulated if participation in the DTD would be acknowledged as a quality indicator. Participation in the DTD is further promoted by the ongoing DTD consortium initiative, intended to initiate transfusion related research within the Dutch blood transfusion community.

Interpretation

For a correct interpretation of transfusion data, knowledge on the underlying processes is required such as how blood products are made and how this is registered. To identify a unique blood product and to link it from the blood bank to the hospital, information is needed on (the coding of) the type of product, the year it was produced, and whether it is split or not. Similarly, to interpret hospital data, multidisciplinary collaboration with medical doctors and clinical chemists is needed to interpret diagnosis codes and laboratory measurements and the

manner in which these data are registered in each hospital. Factors that might confound the interpretation of differences between centers are the presence of case-mix and other contextual differences, and chance variability [4]. Chance variability may become important if small subgroups of patients are considered, and will play a lesser role when sample size increases, which for the DTD will happen over time as more hospitals and years are included. Case-mix can be taken into account by correcting for –but first identifying– specific patient subgroups, for example using information on the transfusion indication (Chapter 4). This requires more complex clinical interpretation of the data (see paragraph 'Clinical indications') and appropriate analysis techniques such as multilevel models to account for differences between hospitals.

Scope and data validity

There is a balance between the wish for a wide applicability of the data warehouse and the need for a specific formulation of study goals. On the one hand we want the DTD to be useful for anyone who will potentially use it, demanding a broad dataset with an extensive range of variables and applications. On the other hand the danger in constructing a 'generic database' is that the targets are too wide, which complicates making decisions regarding the demarcation of variables. No straightforward solution exists, but to facilitate future expansions of the DTD and to increase its usability, it is imperative to systematically document the selection choices made and the outcomes of data validity checks. This is particularly important for a multi-purpose data warehouse, as every research question might require a slightly different selection. Finally, the dynamic nature of a longitudinal, continuous data warehouse implicates that the validity of the data may change whenever there is a data update or extension. Therefore validation should also be a continuous process.

Recommendations

Based on our experience acquired during the work performed for this thesis, we recommend several future research topics and questions that can be employed using the DTD. In addition, to facilitate the answering of these questions, various potential extensions of the DTD are suggested to further integrate data registration, extraction and interpretation.

Future research topics

- Study transfusion triggers and targets –including the evaluation of appropriate transfusion triggers and the effect of a restrictive versus a liberal transfusion policy– using all clinical information from laboratory and administrative databases, and thereby taking advantage of the presence of linked data from different registries within hospitals.

- Study rare patient outcomes such as adverse transfusion reactions in clinical subgroup analyses, and thereby taking advantage of the large number of observations in the DTD required for this kind of assessment.
- Study associations between donor, production or product characteristics (such as donor gender and product storage duration) and patient outcomes, and thereby taking advantage of the link between donor and patient data that is present in the DTD.
- Externally validate models such as clinical indication selection models and prediction models in different centers or years, and thereby taking advantage of the availability of longitudinal data from multiple hospitals.
- Identify recipient subgroups for which data-driven research is needed, in particular within the increasingly prominent group of recipients with medical indications; collaborate with clinicians and policy makers.

Extending the DTD

Develop a system for continuously monitoring data validity; this should be a transparent, automated system that offers insight for potential users of the DTD into its data validity after every update.

- Expand the DTD with information on patients who did *not* receive blood transfusions; this allows benchmarking of transfusion rates and comparing profiles of transfused recipients to non-transfused recipients. These data cannot be extracted by default as this dataset comprises almost the complete hospital registration. Therefore this expansion is optional for specific studies with well-defined patient groups.
- Expand the scope of variables with more detailed patient information, such as medical history, medicine use, vital signs (pulse, temperature and blood pressure) and mortality data; this might be done for a limited number of patient groups for the purpose of a specific study.
- Incorporate a (semi-)automated algorithm or system for the identification of transfusion indications.
- Align data extraction, interpretation and analysis with transfusion data warehouses in other countries to improve comparability and allow data synthesis and international comparison.

Finally, which one of these extensions is most valuable should be a topic of research in itself. Such an evaluation should anticipate future studies to be conducted using the data warehouse.

Concluding remarks

We aimed to contribute to a more efficient cycle of recording, collecting and using transfusion data for research that will benefit transfusion recipients and clinical practice. Future research topics that remain relevant for clinical practice include the identification of patient groups who receive blood transfusions, determining donor and product characteristics that affect patient outcomes, benchmarking blood use in hospitals and monitoring transfusion triggers and targets. The Dutch Transfusion Data warehouse project is especially suitable for answering these questions as it provides linked, multisource blood transfusion data which are supplemented every year with recent donor and recipient data. To improve the usability of the DTD, we recommend consideration of standardizing up-to-date validation of the data, supplementing the data warehouse with information on non-transfused patients, and implementing additional identification of clinical indications for transfusion.

References

1. Sweeney JD. Control of blood utilization. Transfusion and Apheresis science 2008, 39(2), 139-144.
2. Tinmouth A, MacDougall L, Fergusson D, Amin M, Graham ID, Hebert PC,Wilson K. Reducing the amount of blood transfused: a systematic review of behavioral interventions to change physicians' transfusion practices. Archives of internal medicine, 205;165(8):845-852.
3. Cohn CS, Welbig J, Bowman R, Kammann S, Frey K, Zantek N. A data-driven approach to patient blood management. Transfusion, 2014;54(2):316-322.
4. Powell AE, Davies HTO, Thomson RG. Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls. Quality and safety in health care; 2003;12(2):122-128.
5. Edgren G, Rostgaard K, Vasan SK, et al. The new Scandinavian Donations and Transfusions database (SCANDAT2): a blood safety resource with added versatility. Transfusion 2015;55(7):1600-1606.
6. Kleinman S, Busch MP, Murphy EL, Shan H, Ness P, Glynn SA, and The National Heart, Lung, and Blood Institute Recipient Epidemiology and Donor Evaluation Study (REDS-III). The National Heart, Lung, and Blood Institute Recipient Epidemiology and Donor Evaluation Study (REDS-III): a research program striving to improve blood donor and transfusion recipient outcomes. Transfusion 2014,54: 942–955.
7. Chassé M, McIntyre L, Tinmouth A, et al. Effects of blood donor characteristics in transfusion recipients: protocol of a framework to study the blood donor–recipient continuum. BMJ open 2015,5(1):e007412.

Appendices

Appendix A. Discussion of a more extended list of validity outcomes

1. Concordance with (annual) report

The agreement between the numbers found in the database and the annual report of the Blood bank was >98.7% for number of products (varying slightly for the different product types) and 99.96% for number of transfusions.

2. Linkage of data sources within the data warehouse

Linkage of transfused products to products issued by the blood bank was possible for 99.96% of all transfusions, using the identification number of the end product. Vice versa, 97.65% of products issued could be linked to actual hospital transfusions (indicating the spilling rate). Initially, only 96.727% of the products could be linked to their donation(s). We traced this back to a post-hoc modification in the coding of the product identification number at the blood bank, leading to different codes existing in the blood bank and the hospital system for the same product. When the coding was adjusted, the proportion linked products increased to 99.996%.

3. Identity

Every blood product should be uniquely identified by the combination of the donation code and the product code. In the blood bank data, a small percentage of duplicated products was found for RBC products of 0.005%; for FFP and PLT this was 0%. For one hospital, product codes were not available, therefore the broader product type was used. Based on donation code and product type, initially 1.00% of products were duplicated. It turned out that most (71.7%) of these duplicates were split products, which explains why the donation code and product type were similar. A potential pitfall is the double registration of events, for example multiple procedures that actually occurred must be differentiated from duplications in procedures registered for another purpose (e.g. financial registration). Therefore duplicated procedures (i.e. within the same patients at the same time) were removed, resulting in 0% duplicated procedures.

4. Completeness

Most important variables are present and non-missing: for the blood bank data (donor identification code, date of birth, gender, hemoglobin value, product expiration or production date) completeness of at least 98.8% and for hospital data (patient identification code, date of birth and gender) at least 99.99% completeness.

For the outcome regarding completeness of diagnoses, the distribution of number of pending diagnoses per transfusion was found to range from 0 diagnoses to up to 15. The percentage of transfusions that fall within the start and end date of at least one diagnosis was 98%. This

implicates that it will be necessary to make a selection of those diagnoses in the future if we want to determine the main indication for a transfusion.

5. Uniformity

Diagnoses and procedures were recoded into a uniform system, resulting in a linkage percentage of diagnosis codes with the reference table of 96.1%. Hb level was was registered with the precision of 1 significant decimal for >98.6% for the hospitals and 99.76% for the blood bank data.

6. Time patterns

The time patterns in number of donations, products and donors) reveal no unexpected trends, as the observed decrease is in line with the known nationally decreasing trend in RBC use. The trends in blood use by product type confirm this and also show a high relative decrease for FFP products, for example from 2010 to 2011. This decrease in FFP can be explained by the introduction of ROTEM (a method of hemostasis testing in whole blood) for thoracic surgery, and different guidelines and consensus. In the time period concerned, use of ROTEM followed a reverse trend, increasing where FFP use decreased.

In 2010 the percentage of transfused products that could not be linked to products issued was exceptionally high (2.2% versus 0.07% in other years). This percentage could be lowered to 0.17% by including blood bank data from the previous year 2009 (the unlinked products were mainly frozen plasma products that were issued in the year before). This resulted in an annual linkage percentage of 99.8% or higher.

7. Plausibility

Accurate date and time values are crucial in order to study cause-effect relationships, such as transfusion triggers and pre and post transfusion targets. A problem occurs if the registered (e.g. transfusion) time actually records the moment that a product or service (e.g. the blood product) was requested instead of administered. Generally, hemoglobin (Hb) should increase after transfusion. To check this, the difference was computed between the last Hb before and first Hb after transfusion (only Hb measurements within one day before or after the transfusion were considered). A clinically significant Hb change was defined as an increase or decrease of 8.8% relative to the first Hb measurement. This cut-off point was defined using the formula for the critical change: $2.77 * \sqrt{(CVa2 + CVi2)}$ [27]. Assuming an analytical variation (CVa) of 1.5% and an intra-individual biological variation (CVi) of 2.8% [28], the cut-off value lies at 8.8%. Although we would expect that Hb increases after transfusion, it turned out that 40% did not clinically change and 6% even decreased. Recipients with a decrease in Hb were further examined and 87% of these patients had a diagnosis indicating high bleeding risk such as the diagnosis acute bleeding, justifying the validity of a decreasing Hb value.

8. Event attributes

At the blood bank, blood products made from platelets (PLT) are produced by pooling the PLT of five donations and one FFP unit together. These pooled products should therefore be linked to five or six donations, which was the case for 100%. Another attribute of a transfusion event is a hospitalization; it was found that 99.16% of all patients were also admitted (of which 23.64% were day admissions), indicating that the remaining transfusions were given in an outpatient ward. Finally, an estimate of the proportion of potentially missing information on for example transfusions was given by the percentage of patients that were transferred to another hospital. According to the 'discharge destination' variable, 6% of patients were, at the end of their admission, sent to another hospital.

9. Consistency of hospitals within data warehouse

The validity outcomes for both hospitals were expected to be quite similar as both are teaching hospitals. The two hospitals had indeed very similar outcomes, supporting the validity of the findings.

10. Concordance with literature

Comparison of our data with previous literature on the distribution of blood products over age and gender categorized by product type [29] revealed that the distributions were quite similar, but that platelet use has shifted towards older patients. This can be explained in part by the ageing of the population, but also by changes in policy in the past ten years: thorax surgery has increased its platelet use, and also treatment of haematological disorders has become more intensive, including higher platelet use. As both heart disease and haematological disorders are more prevalent in men, there is a peak in platelet use for men aged 60-80 years. Because relatively more platelets are transfused to older patients, platelet use for children makes up a smaller part of the total use.

11. Concordance with expert feedback

Expert feedback was asked regarding the accuracy of Hb measurements. The outcomes of Step 7 (Plausibility) were presented to two clinical chemists from the participating hospitals, in order to evaluate whether these numbers seem plausible. The experts concluded that the percentage with an unexplained decrease is below 1% of all transfusions, which is acceptable. The finding that in patients with high bleeding risk, Hb value sometimes decreases and sometimes increases is also plausible; with acute bleeding it is more difficult to measure the Hb, which might lead to too much blood being given.

12. Concordance with other databases

What previous transfusion data warehouse studies have reported in terms of data validation varies greatly. The most extensive list of validation outcomes were reported by the SCANDAT study [18,19], therefore, these outcomes are shown next to the validity outcomes of the DTD (Table 3.3). SCANDAT and DTD show similar results regarding the high external

concordance of the data with external statistics and the fact that both studies identified missing data by investigating time patterns. Different is the proportion of hospitalized patients, which might be due to different registration of patients between the countries (as we found a consistently higher hospitalization rate for both of the DTD hospitals included). The estimated proportion of patients with incomplete information due to transference from our hospitals included to another hospital was up to 6% for the DTD. This might actually be an underestimation, because this 6% does not include patients who were hospitalized elsewhere prior to being hospitalized in hospitals analyzed, and given the findings that in SCANDAT 8.9% of recipients received a blood transfusion in two or more local registers.

Other transfusion database studies reported only a few outcomes: the linkage rate of transfusions to donors between 92%-99% [18,19,20,21,22] and, vice versa, estimates of wastage of blood products (i.e., issued but not transfused) of 1.3% and 7.7% [19,20]. The percentage missing values was also reported by some studies: clinical variables were missing for 13% (post-transfusion Hb) [23], 14% (ASA code) [22] and 20% (specialty)[23], the latter interestingly varying between specialties from 2% to 47%.

Appendix B. R code of the selection algorithm

The priority and test data files are available upon request.

```
# load files
priority <- read.csv("prioritization_specialisms.csv", sep=";") # load
prioritization
data <- read.csv("testdata.csv", sep=",") # load data: each row represents
a diagnosis or procedure linked to a transfusion cluster

# the algorithm function
select_indication <- function(transfusion.cluster, data, priority) {

  # select transfusion cluster
  clus <- data[data$tra.clus == transfusion.cluster, ]

  # count the number of diagnoses and procedures for this transfusion
  cluster
  number_of_proc <-
  length(unique(clus[!is.na(clus$proc.clus),"proc.clus"]))
  number_of_diag <-
  length(unique(clus[!is.na(clus$diag.clus),"diag.clus"]))

  # create an empty matrix
  algorithm_result <- vector(length=3)
  algorithm_result[1] <- transfusion.cluster

  ## procedures:
  # if no procedures: note NA
  if (number_of_proc == 0) {
    algorithm_result[2] <- NA
  }

  # if 1 procedure: select this procedure
  else if (number_of_proc == 1) {
    algorithm_result[2] <- clus$proc.clus[!is.na(clus$proc.clus)]
  }

  # if >=2 procedures:
  else if (number_of_proc >= 2) {
    if (length(which(clus$date.diff.proc == 0))) {
       algorithm_result[2] <- clus$proc.clus[which(clus$date.diff.proc ==
       0)]
```

```
  } else if (length(which(clus$date.diff.proc == -1))) {
     algorithm_result[2] <- clus$proc.clus[which(clus$date.diff.proc ==
     -1)]
  } else if (length(which(clus$date.diff.proc == 1))) {
     algorithm_result[2] <- clus$proc.clus[which(clus$date.diff.proc ==
     1)]
   } else if (length(which(clus$date.diff.proc >= -7 &
   clus$date.diff.proc < 0))) {
     cl <- clus[which(clus$date.diff.proc >= -7 & clus$date.diff.proc <
     0), ]
     algorithm_result[2] <-cl$proc.clus[which.max(cl$date.diff.proc)]
  } else {
    algorithm_result[2] <- NA
  }
}

## diagnoses:

# if a procedure has already been selected, note NA
if (!is.na(algorithm_result[2])) {
  algorithm_result[3] <- NA
}

else if (is.na(algorithm_result[2])) {
  # if no diagnosis: note 'Neonatology'
  if (number_of_diag == 0) {
    algorithm_result[3] <- "Neonatology"
  }

  # if 1 diagnosis: select this diagnosis
  else if (number_of_diag == 1) {
    algorithm_result[3] <- clus$diag.clus[!is.na(clus$diag.clus)]
  }

  # if >=2 diagnoses:
  else if (number_of_diag >= 2) {
    prior <- priority[priority$specialism %in% clus$diag.spec, ]
    minspec <- prior$specialism[which.min(prior$order.diagnoses)]
    clus_multdia <- clus[clus$diag.spec %in% minspec, ]
    algorithm_result[3] <-
    clus_multdia$diag.clus[which.min(clus_multdia$date.diff.diag)]
  }
}
```

```
  return(algorithm_result)
}

# store all results in one dataframe
do_all_clusters <- function(data, priority) {
  cells <- data.frame()
  for (transfusion.cluster in unique(data$tra.clus)) {
    cells_for_this_clus <- select_indication(transfusion.cluster, data,
    priority)
    cells <- rbind(cells, cells_for_this_clus)
  }
  colnames(cells) <- c ("tra.clus","procedure.algo","diagnosis.algo")
  return(cells)
}

# apply algorithm
print(do_all_clusters(data, priority))
```

Summary

Blood transfusion is an important medical treatment for many and diverse patients groups, saving lives but sometimes also causing adverse transfusion reactions in transfusion recipients. For this reason blood use should ideally be as low as possible. The fact that significant differences exist in the amount of blood used between countries, hospitals and even within hospitals, indicates that there is room for improvement. Moreover, there are likely to exist unrecognized risk factors in donors and blood products that might affect patient outcomes. And not only patients are affected by the way in which blood is used; it also has consequences for blood donors, doctors, hospitals, the blood bank and policy makers. In order to study these various aspects and the interplay between them, data on the complete transfusion chain are needed. Therefore we set up the *Dutch Transfusion Data warehouse* (DTD), in which data from the national blood bank and a (growing) number of Dutch hospitals are linked. As the data are extracted from electronic health records which are primarily registered for clinical use, a systematic annotation and interpretation for research purposes is lacking. Therefore, when analyzing data from the DTD the aspects as described in **Chapter 1** must be taken into account: the data are observational, multisource, intended to be nationally representative, longitudinal and continuously updated. This thesis has two main focus points: the methodological challenges involved in the collection, validation and interpretation of the data (**Chapters 2-5**), and the actual application of the data in analyses concerning donors and recipients (**Chapters 6-7**).

The design of the DTD is described in **Chapter 2**. The collection of the data started in the blood bank with data on donors (e.g., age, blood groups, antibodies), products (type of product, processing, storage time), which were linked to data from the participating hospitals (e.g. patient diagnosis, surgical procedures, laboratory parameters, number of transfusions administered). These data have a broad range of applications, four of which are illustrated in Chapter 2: identifying risk factors, predicting future blood use, benchmarking blood use, and optimizing process efficiency. For example, insight in donor- and product-related risk factors for recipient outcomes can help make transfusion more tailored and –by avoiding unnecessary transfusions– further diminish the number of transfusion reactions in patients. Before the data can be analyzed however, we must first ensure that the data quality is sufficient for a correct interpretation.

A structured stepwise approach to validate the data is developed in **Chapter 3**, which addresses *external validity* (e.g. concordance with external reports, previous studies and expert feedback) and *internal validity* (e.g. completeness, uniformity and plausibility). Part of the data present in the DTD at this time is validated, which resulted in a structured overview of the different data validity aspects. This allowed improvement of these aspects through further processing of the data and in some cases adjustment of the data extraction process. A crucial part of the data warehouse are diagnostic and procedural data which specify the type of patients and their clinical indications for a blood transfusion. The validation showed that completeness of de diagnosis variable was high: almost every transfusion could be linked to at least one diagnosis. In fact, the majority of transfusions could be linked to multiple diagnoses.

This however poses a new challenge of identifying the one specific diagnosis or procedure that most likely necessitated the transfusion.

Therefore **Chapter 4** describes the development of an algorithm to identify –out of all diagnostic and procedural data available– the most likely indication for transfusion. The algorithm was evaluated against a gold standard based on expert review of a sample of medical records. In a second step, information on misclassification was used to fine-tune the initial algorithm. The final algorithm was able to predict the majority of cases correctly (about 75%). Although this score is substantially better than a random guess, efforts to improve the predictions may be worthwhile for example by taking into account more detailed patient information. Also, before implementation of the algorithm, the obtained results should be externally validated in independent hospital datasets. When more hospitals are included in the DTD, this not only extends opportunities for data validation, but it will also improve the representativeness of the data.

As the DTD is intended to be nationally representative and new hospitals are included continuously, in **Chapter 5** several strategies for selecting hospitals for inclusion in the data warehouse are compared. The main result was that the selection strategy of maximum variation between hospitals (in terms of number of beds) is optimal for predicting blood use in the Netherlands. In practical terms this would mean for the DTD that especially hospitals at the ends of the spectrum (the smallest and largest hospitals) have the highest added value for the representativeness of the data. It should be noted however that with an increasing number of hospitals included, the differences between the selected selection strategies decrease.

In Chapters 6 and 7 donor and patient data are analysed. While the number of hospitalized patients in Dutch hospitals has been increasing since 1997, as described in **Chapter 6** the demand for red blood cell units (RBCs) has simultaneously decreased. This implies a considerable change in transfusion practice towards on average fewer blood transfusions per patient. In order to explain the RBC decrease, various patient groups (surgical, medical, obstetrical, specific age groups) were retrospectively studied in relation to RBC use between 1996 and 2005. The use of RBCs changed from being predominantly given to surgical patients to being given largely to medical patients (a relatively stable low percentage went to obstetrical patients). Changes were more marked in the higher age groups. Also a trend was observed towards the use of only one or two RBC units during a transfusion episode rather than three or more. These results suggest a more restrictive transfusion policy for surgical patients as well as an increase in medical indications for transfusion. This fits well with the current focus towards more cost-effective transfusion policies.

Donor data are employed in **Chapter 7** to develop a prediction model for anti-Rhesus D (RhD) donors. Anti-RhD plasma donors provide rhesus antibodies necessary for RhD-injections, which are required for RhD-negative women pregnant with a RhD-positive child in order to prevent hemolytic disease of the newborn. Due to the success of the RhD prevention program, the number of naturally immunized women has decreased, thereby also reducing the number of potential donors and threatening the availability of the RhD-injections. Data on Dutch anti-RhD donors in 1994-2013 were used to simulate the donor population

size and age composition for various donor recruitment scenarios. It was predicted that with a continuous influx of 27 new donors per year and a donor stopping rate of 10% per year, the population size will stabilize at 195 donors, with 2.3% of donors stopping annually due to reaching the donor age limit. With this relatively simple model we can describe and predict the size of the anti-RhD donor population and the impact of ageing sufficiently accurately.

Finally **Chapter 8** reflects on the value of transfusion data and potential improvements to further increase the utility of these data. Recommendations for research topics and possible extensions of the DTD offer a perspective on future applications of the data following the process from collection to reflection.

Samenvatting

Bloedtransfusie is een niet weg te denken medische behandeling voor een groot aantal uiteenlopende patiëntgroepen. De toediening van bloed redt levens maar leidt soms ook tot nadelige bijwerkingen bij transfusieontvangers. Daarom zou je ideaal gezien het bloedgebruik zo laag mogelijk willen houden. Het feit dat er tussen landen, ziekenhuizen en zelfs binnen ziekenhuizen aanzienlijke verschillen zijn in bloedverbruik, geeft aan dat er ruimte is voor verbetering. Bovendien zijn er mogelijk nog risicofactoren bij donors en bloedproducten die van invloed kunnen zijn op patiëntuitkomsten. De manier waarop bloed wordt gebruikt heeft niet alleen consequenties voor patiënten, maar ook voor bloeddonors, artsen, ziekenhuizen, de bloedbank en beleidsmakers. Om deze verschillende aspecten en de wisselwerking te bestuderen, zijn data nodig die de complete transfusieketen van donor tot bloedproduct tot patiënt bestrijken. Daarom hebben we het *Dutch Transfusion Data warehouse* (DTD) opgezet, waarin data van de nationale bloedbank worden gekoppeld aan data van een (toenemend) aantal Nederlandse ziekenhuizen. Omdat deze data worden geëxtraheerd uit elektronische medische dossiers die zijn bedoeld voor klinisch gebruik en administratie, ontbreekt een systematische annotatie en interpretatie voor onderzoeksdoeleinden. Bij de analyse van data uit het DTD moet daarom rekening worden gehouden met een aantal aspecten zoals beschreven in **Hoofdstuk 1**: de data zijn observationeel, afkomstig van meerdere bronnen, bedoeld om nationaal representatief te zijn, longitudinaal en regelmatig aan updates onderhevig. Dit proefschrift richt zich op twee belangrijke aandachtspunten: de methodologische uitdagingen die komen kijken bij het verzamelen, valideren en interpreteren van de data (**Hoofdstuk 2-5**), en de daadwerkelijke toepassing van de data in analysemodellen over donors en patiënten (**Hoofdstuk 6-7**).

Het opzetten van het DTD wordt beschreven in **Hoofdstuk 2**. De dataverzameling begon bij de bloedbank met data over donors (zoals leeftijd, bloedgroepen, antistoffen), producten (type, bewerking, opslagtijd), welke vervolgens werden gekoppeld aan patiëntdata uit de deelnemende ziekenhuizen (zoals diagnose, chirurgische procedure, laboratoriummetingen, aantal toegediende transfusies). Deze data hebben een breed toepassingsgebied, waarvan er vier zijn weergegeven in Hoofdstuk 2: het identificeren van risicofactoren, het voorspellen van toekomstig bloedgebruik, het benchmarken van bloedgebruik (onderling vergelijken van ziekenhuizen) en het optimaliseren van de efficiëntie van processen. Zo kan bijvoorbeeld inzicht in donor- en product-gerelateerde risicofactoren bijdragen aan beter afgestemd transfusiebeleid en –door het vermijden van onnodige transfusies– het aantal transfusiereacties bij patiënten verder verminderen. Voordat de data echter kunnen worden geanalyseerd, moeten we er allereerst voor zorgen dat de datakwaliteit voldoende is.

Een gestructureerde aanpak om de data te valideren wordt beschreven in **Hoofdstuk 3**. We onderscheiden *externe validiteit* (bijvoorbeeld overeenstemming van de data met externe rapporten, eerdere studies en expert meningen) en *interne validiteit* (bijvoorbeeld compleetheid, uniformiteit en plausibiliteit). Een deel van de gegevens aanwezig in het DTD op het moment van onderzoek wordt gevalideerd, met als resultaat een gestructureerd overzicht van de verschillende validiteitsaspecten. Dit biedt aanknopingspunten om de datakwaliteit te

verbeteren door verdere verwerking van de gegevens en eventuele aanpassing van de gegevensextractie. Een cruciaal onderdeel van de data zijn diagnose- en proceduregegevens die specificeren om welk type patiënten met welke indicaties voor bloedtransfusie het gaat. Uit de validatie bleek dat de compleetheid van de diagnosevariabele hoog was; bijna elke transfusie kon worden gekoppeld aan tenminste één diagnose. Sterker nog, de meeste transfusies konden aan meerdere diagnoses worden gekoppeld. Dit stelt ons voor een nieuwe uitdaging, namelijk om te bepalen welke diagnose of procedure de primaire reden (indicatie) voor transfusie was.

Daarom staat **Hoofdstuk 4** in het teken van het ontwikkelen van een selectiealgoritme om –uit alle beschikbare diagnostische en procedurele gegevens– op een automatische manier de meest waarschijnlijke indicatie voor transfusie te identificeren. Het algoritme werd geëvalueerd langs een gouden standaard op basis van expert review van een steekproef van medische dossiers. In een tweede stap werden gegevens over misclassificatie gebruikt om het oorspronkelijke algoritme te verbeteren. Het uiteindelijke algoritme is in staat om de transfusie-indicatie in ongeveer driekwart van de gevallen correct te selecteren. Deze score is substantieel beter dan op basis van een willekeurige keuze te verwachten zou zijn, maar het kan de moeite waard zijn om de voorspellingen verder te verbeteren door bijvoorbeeld meer en gedetailleerdere patiëntinformatie te gebruiken in het algoritme. Bovendien moet er, voordat het algoritme kan worden geïmplementeerd in het DTD, externe validatie plaatsvinden door het algoritme in onafhankelijke datasets te testen. Wanneer er meer deelnemende ziekenhuizen worden opgenomen in het DTD, biedt dit niet alleen meer mogelijkheden voor validatie, maar ook voor het verhogen van de mate van representativiteit van de data.

Omdat het DTD nationaal representatief beoogt te zijn en er continu nieuwe ziekenhuizen worden geïncludeerd, worden in **Hoofdstuk 5** verschillende strategieën voor het selecteren van ziekenhuizen vergeleken op de resulterende representativiteit. Het belangrijkste resultaat hiervan was dat met de selectiestrategie van maximale variatie tussen ziekenhuizen (wat betreft aantal bedden) het best kan worden voorspeld wat het bloedgebruik voor heel Nederland is. Praktisch gezien zou dit voor het DTD betekenen dat met name ziekenhuizen aan de uiteindes van het spectrum (de kleinste en grootste ziekenhuizen) toegevoegde waarde hebben voor de representativiteit van de data voor heel Nederland. Hierbij moet overigens worden opgemerkt dat bij een toenemend aantal geïncludeerde ziekenhuizen de verschillen tussen de onderzochte selectiestrategieën afnemen.

In Hoofdstuk 6 en 7 worden donor- en bloedgebruikdata geanalyseerd. Zoals beschreven in **Hoofdstuk 6** neemt al sinds 1997 het aantal patiënten dat wordt opgenomen in het ziekenhuis toe, terwijl de vraag naar rode bloedcelproducten (RBCs) juist afneemt. Dit impliceert een aanzienlijke verandering in de transfusiepraktijk naar gemiddeld minder bloedtransfusies per patiënt. Om de RBC-afname te verklaren werden verschillende patiëntengroepen (chirurgische, medische, obstetrische, specifieke leeftijdsgroepen) retrospectief onderzocht in relatie tot RBC gebruik in de periode tussen 1996 en 2005. In de onderzoeksperiode is het gebruik van RBC veranderd van grotendeels chirurgische naar voornamelijk medische patiënten (een relatief stabiel laag percentage ging naar verloskundige patiënten). Veranderingen waren meer uitgesproken in de hogere leeftijdsgroepen. Ook werd

een trend geobserveerd in het gebruik van slechts één of twee RBC producten per transfusie-episode in plaats van drie of meer. Deze resultaten suggereren een restrictiever transfusiebeleid voor chirurgische patiënten en een stijging van medische indicaties voor transfusie. Dit sluit goed aan bij de huidige focus op een kosteneffectief transfusiebeleid.

Donorgegevens worden in **Hoofdstuk 7** gebruikt om een voorspellingsmodel te maken voor anti-Rhesus D (RhD) donors. Anti-RhD plasmadonors voorzien in antistoffen voor de RhD-injecties ('rhesusprik'), welke noodzakelijk zijn voor RhD-negatieve zwangere vrouwen die in verwachting zijn van een RhD-positief kind om hemolytische ziekte van de pasgeborene te voorkomen. Vanwege het succes van de rhesusprik is het aantal natuurlijk geïmmuniseerde vrouwen in de loop van de jaren afgenomen en daarmee ook het aantal potentiële donors, wat mogelijk de beschikbaarheid van de rhesusprik in gevaar brengt. Gegevens over Nederlandse anti-RhD donors in 1994-2013 werden gebruikt om de grootte en leeftijdsverdeling van de donorpopulatie te simuleren voor verschillende scenario's van donorwerving. De schatting is dat met een continue instroom van 27 nieuwe donors en een verlies van 10% van de donors per jaar, de grootte van de populatie zal stabiliseren op 195 donors, waarbij jaarlijks 2,3% van de donors stopt vanwege het bereiken van de leeftijdsgrens voor het donorschap. Met dit relatief eenvoudige model kunnen we de grootte van de anti-RhD donorpopulatie en de effecten van vergrijzing voldoende accuraat beschrijven en voorspellen.

Tenslotte wordt in **Hoofdstuk 8** gereflecteerd op de waarde van transfusiedata en wat er nog kan worden verbeterd om deze data zo nuttig mogelijk te gebruiken. Aanbevelingen voor onderzoeksonderwerpen en mogelijke uitbreidingen van het DTD bieden perspectief op toekomstige toepassingen van de gegevens na het hele proces van verzameling tot interpretatie.

Dankwoord

Beste Mart,
De parallel met de muziek is makkelijk te maken: klassiek versus jazz, gebaande paden versus improvisatie. Van jouw creatieve kant heb ik geleerd om het maximale uit de data te halen, mooie (R-)grafiekjes te maken en om overal mogelijkheden te zien. Graag wil ik je bedanken voor je benaderbaarheid, de muzikale intermezzo's en de grapjes tussendoor.

Beste Kit,
Dank voor alle inhoudelijke discussies, die me uitdaagden om altijd scherp en kritisch te blijven en net wat verder te denken. Ik heb hier veel van geleerd, met name wat betreft het aanbrengen van structuur, vasthouden van hoofdlijnen en doorhakken van knopen.

Beste Erik,
Ook al kwam jij pas in de tweede helft (mei 2014) bij mijn promotietraject als co-promotor, jouw aanwezigheid, inzicht, rust en humor zijn heel belangrijk geweest. Dank hiervoor.

HTA-afdeling en PROTON-II projectgroep, Ardine, Welling, Anoukh, Giske, Marie-Josee, Henk, Peter, Rianne, Martine, Babette en alle andere collega's: dank voor de prettige omgeving, interactieve meetings en feedback. Extra woord van dank voor Peter; het heeft me veel geholpen om met jou te kunnen sparren.

Ik wil graag alle c*o-auteurs* bedanken voor het meedenken en -schrijven. In het bijzonder Aukje en Floris voor de klinische blik en het uitvoeren van dossieronderzoek; zonder jullie geen gouden standaard.

Hier ook een (symbolisch) woord van dank voor alle *deelnemers* aan het Dutch Transfusion Data warehouse project: ziekenhuizen, klinisch chemici, onze datamanagers Sandor en Janneke, patiënten en donors die toestemming hebben verleend op het KAF-formulier; allen onmisbaar voor dit proefschrift.

Aan alle (oud-)kamergenoten waaronder Anneke en Natasha toen ik net begon, en daarna Romin, Suzanne, Jolien, Shona, Pauline, Fleur, Sander, Faas, Giske, Timo, Kevin, Carline en Anne-Karien: naast een superfijne werkomgeving, was (en is) het leuk om met jullie koffie, wandelrondjes buiten, borrels, etentjes, spelletjes voetbal, pubquiz (team superkip), concerten, klimmen, etc. te doen!

Moniek, Jantien, Jurrijn, Marjolein en Els, als oudste vriendengroep hebben jullie mijn ontwikkelingen gedurende deze PhD vanaf het begin van dichtbij meegemaakt. Leuk om te zien hoe we allemaal een kant opgaan die heel verschillend is maar nog steeds heel goed bij elkaar past. Ik ben blij dat ik jullie tijdens dit hele proces had als uitlaatklep, adviesorgaan en uiteraard voor de afleiding en gezelligheid. *Michel*, bedankt voor je steun en verrassende inzichten. *Riëtte*, ik ben blij dat we de psychologie, statistiek en muziek delen.

Jurrijn, veel dank voor het ontwerpen van mijn cover; reflectie in de titel heeft nu een dubbele betekenis gekregen.

En dan mijn paranimfen:
Romin, I think that following a PhD trajectory provides excellent circumstances to get to know each other well - circumstances that have gradually shifted to the squash court, park, Muzieklokaal and any other place involved with music, beer and/or vegan food. Thanks for your support, met jou valt alles mee!

Jantien, naast al het serieuze wetenschap bedrijven, jij bij het AMC en ik bij het UMC, zijn er soms momenten nodig van onzin en de slappe lach. Onze tripjes naar Kopenhagen, Praag en Berlijn hebben mij, naast voorheen ontbrekende geografische kennis (misschien hadden we beter niet naast elkaar kunnen zitten bij aardrijkskunde), weer frisse moed gegeven om verder te gaan en ervoor gezorgd dat ik perspectief bleef houden op het einddoel (i.e. het promotiefeest).

Mijn ouders en broertje, *Bert, Liliane en Ward*: fijn dat jullie er zijn, vertrouwen hebben en begrip tonen – maar ook onbegrip, wat weer helpt bij het omzetten van onderzoeks-terminologie naar een toegankelijk verhaal.

Curriculum vitae

Loan van Hoeven was born on April 24th in Amersfoort. In 2006 she started studying Psychology at Utrecht University and classical piano at Utrecht Conservatory. After obtaining a pre-master in Philosophy, she graduated cum laude for the Psychology Master Health and Behavior at the University of Amsterdam. In her master thesis at the Medical Decision Making department (LUMC) she explored two types of decision making: intuitive (unconscious) versus rational (conscious) decision making strategies and their effect on patients' treatment decisions. In 2011 Loan worked on a study on sustainable protein consumption at Wageningen Economic Research (previously LEI), aiming to identify the motives and attitudes of consumers towards reducing meat consumption. In April 2012 she started her PhD which was a collaboration between the Julius Center and Sanquin Blood bank, located in Utrecht. Here she worked at the Health Technology Assessment department and conducted the studies resulting in this thesis. Sideprojects included reports on blood collection, screening and use in European countries for the Council of Europe. In 2015 she obtained a postgraduate Master degree in Epidemiology with the specialisation Medical Statistics.

List of publications

van Hoeven LR, Hooftman BH, Janssen MP, de Bruijne MC, de Vooght KM, Kemper PF Koopman MMW. Protocol for a national blood transfusion data warehouse from donor to recipient. BMJ Open 2016: 6(8), e010962.

van Hoeven LR, de Bruijne MC, Kemper PF, Koopman MMW, Rondeel JMM, Leyte A, Koffijberg H, Janssen MP, Roes KCB. Validation of multisource electronic health record data: an application to blood transfusion data. Submitted.

van Hoeven LR, Kreuger AL, Roes KCB, Kemper PF, Koffijberg H, Kranenburg FJ, Rondeel JMM, Janssen MP. Why was this transfusion given? Identifying clinical indications for blood transfusion in health care data. Submitted.

van Hoeven LR, Janssen MP, Roes KCB, Koffijberg H. Aiming for a representative sample: Simulating random versus purposive strategies for hospital selection. BMC Medical Research Methodology 2015: 15(1), 1.

van Hoeven LR, Koopman MMW, Koffijberg H, Roes KCB, Janssen MP. Historical time trends in red blood cell usage in the Netherlands. International Journal of Clinical Transfusion Medicine 2016: 4, 67-77.

van Hoeven LR, Berkowska MA, Verhagen OJHM, Koffijberg H, van der Schoot CE, Janssen MP. Prediction of the anti-RhD donor population size for managerial decision-making. Vox Sanguinis 2016: 11(2), 171-177.