
Prepublication history for this paper is available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2015-010962).

ABSTRACT

Introduction: Blood transfusion has health-related, economical and safety implications. In order to optimise 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.

Study design and methods: Quantitative data on blood donors (eg, 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 (eg, transfusions administered, patient diagnosis, surgical procedures, laboratory parameters), which are extracted from hospital electronic health records.

Applications: Expected scientific contributions are illustrated for 4 applications: determine risk factors, predict blood use, benchmark blood use and optimise process efficiency. For each application, examples of research questions are given and analyses planned.

Conclusions: The DTD project aims to build a national, continuously updated transfusion data warehouse. These data have a wide range of applications, on the donor-production side, recipient studies on blood usage 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. To date, it is widely accepted that blood transfusions can be lifesaving and can be used for the treatment of various diseases. However, since blood transfusions may also have serious side effects, there is still much debate on optimal transfusion triggers. There is growing but inconclusive evidence that a restrictive transfusion policy is more beneficial for patients than a more liberal policy (exceptions might be patients with cardiac disease or oncological surgery). The large variation that exists in the use of blood products between countries, between hospitals and even within hospitals 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 the efficiency and safety of the donor–product–recipient relationship, more data are needed.

Even though several individual hospitals and blood banks analyse data on donors and transfusion recipients, 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. The Recipient Epidemiology and Donor Evaluation Study (REDS-III) programme in the USA is currently preparing a similar blood donor and transfusion recipient database. Finland established a recipient database starting in 2002, covering in the year 2007 70% of all blood units delivered for all potentially transfused patients. Recently a Canadian donor–recipient study was initiated, containing data from hospitals in a specific region. 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 hospitalisation.

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

Other results of transfusion data warehouse initiatives include the development of a model to predict the impact of demographic changes on the demand of RBCs. Such a model may guide donor recruitment requirements. Moreover, benchmarking events have been organised, for example, in Finland for different transfusion practices such as orthopaedics, gynaecology, haematology and heart surgery. Benchmarking discussions have led to the adoption of best practices in several cases, reflected in the reduction of differences in blood use.

The Dutch PROTON database included hospital transfusion data starting in the year 1996. 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 the donor to the recipient. This means that the DTD facilitates research on blood usage in hospitals; 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 for comprehensive studies on blood transfusion in the Netherlands. The four main applications of this data warehouse are:

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

To illustrate how the DTD initiative will be used for these applications, we 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.

**STUDY DESIGN AND 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. The starting point was the previously conducted PROTON study, consisting of a single collection of blood transfusion data in the Netherlands from 1996 to 2006. This data set is further extended with additional recipient, donor and product data.

In the Netherlands the blood supply is organised at a national level by Sanquin which is the sole supplier, enabling a centralised 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 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 (EHRs), including patient characteristics, hospitalisations, diagnoses, procedures, blood products received, blood groups, laboratory parameters and transfusion reactions (table 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 International Society of Blood Transfusion (ISBT) product code. 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 started...
from 2010 and will include future transfusions as well. The current number of donors in our database is \( \sim 500,000 \), 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 the 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 \( \sim 1,100,000 \) 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.

### Data quality

Extracting and combining large amounts of data from hospital and blood bank electronic systems is challenging: often the data have to be split into different tables (eg, by year, department or aggregation level), which 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 will ask the participating centres 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

---

**Table 1** Overview of donor and blood product data collected in the blood bank

<table>
<thead>
<tr>
<th>Donor</th>
<th>Donor number</th>
<th>Donor identification number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of birth</td>
<td>Date of birth</td>
<td>Gender</td>
</tr>
<tr>
<td>RhD blood group</td>
<td>RhD blood group</td>
<td>Kell blood group</td>
</tr>
<tr>
<td>Donor entry date</td>
<td>Date of registration at the blood bank</td>
<td>Date of first donation</td>
</tr>
<tr>
<td>Weight</td>
<td>Donor weight</td>
<td>Height</td>
</tr>
<tr>
<td>Number of donations</td>
<td>Total number of donations since 2007</td>
<td>Number whole blood donations</td>
</tr>
<tr>
<td>Number plasma donations</td>
<td>Total number of plasma donations since 2007</td>
<td>Other donations</td>
</tr>
<tr>
<td>Stopping code</td>
<td>Stopping code</td>
<td>Stopping reason</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Donation</th>
<th>Donation date</th>
<th>Donation identification number (unique for each donation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donation type</td>
<td>Type of donation (whole blood/plasmapheresis/erythrocytapheresis/plateletpheresis)</td>
<td>Donation volume</td>
</tr>
<tr>
<td>Haemoglobin level</td>
<td>Haemoglobin level (in mmol/L)</td>
<td>Platelet count</td>
</tr>
<tr>
<td>Donation location</td>
<td>Blood centre location</td>
<td>Donation duration</td>
</tr>
<tr>
<td>Apheresis machine</td>
<td>Model of the machine used for apheresis</td>
<td>Blood pressure</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Product</th>
<th>Product code</th>
<th>Product code (specifying product type, location of the blood bank, split product or not; according to ISBT 28 Standard Specification)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product modifiers</td>
<td>Optional attribute of a product (eg, cytomegalovirus (CMV) negative/positive)</td>
<td>Pool DIN</td>
</tr>
<tr>
<td>Expiration date</td>
<td>Date on which the product expires</td>
<td>Erythrocyte antibodies</td>
</tr>
<tr>
<td>Platelet phenotype</td>
<td>Human platelet antigen (HPA) phenotype</td>
<td>Date of pooling</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transport</th>
<th>Transport date</th>
<th>Date of transport of the blood product</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institute</td>
<td>Destination of the blood product</td>
<td>Return date</td>
</tr>
<tr>
<td>Return code</td>
<td>Return code specifying the reason for returning the blood product</td>
<td></td>
</tr>
</tbody>
</table>
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 harmonise 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 (ie, 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 prioritisation 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 centre, 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 laws. 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 a 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 maximised by excluding privacy sensitive information such as name and postal code.
Organisation structure
The DTD project team, consisting of experienced researchers in the areas of transfusion medicine, data modelling and healthcare 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. The main objective of the board is to guarantee 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 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 section) are linked to these steps, showing which data are necessary for each application. The main contribution of the data warehouse is to allow insight into 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 wide range of purposes. To illustrate expected scientific contributions, we describe exemplary studies that could be conducted with the DTD data set.

Application 1: risk factors
Example research question: What is the effect of donor characteristics and season on the risk of (febrile) non-haemolytic transfusion reactions (FNHTR) experienced by recipients?
Non-haemolytic transfusion reactions are relatively common, especially among haematology patients, with median reported rates for FNHTR of 4.6% for platelets and 0.33% for RBCs. 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-haemolytic reactions with season and with certain donor characteristics (age, sex, blood group, donation frequency). Donation frequency, for example, is hypothesised to affect iron storage, and might also affect patient outcomes. The primary outcome is the risk of non-haemolytic transfusion reactions. Secondary outcomes are: risk of infections, other transfusion reactions, survival and duration of hospitalisation.

Application 2: predict future blood products needed
Example research question: What is the expected use of blood (medical vs 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

---

Figure 1 Framework of the blood transfusion chain. Each part of the chain can be linked to one of the four applications.
made by distinguishing between surgical and medical use of RBCs, as well as academic, teaching and general hospitals. Observed trends in the past will be extrapolated using a regression model. Furthermore, corrections for growth and ageing of the general population can be incorporated into the predictions of the amount of blood products required.

**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 haematocrit and preoperative or admission haemoglobin. A multilevel random-effects model can be specified with the following levels: hospital type, hospital and patient. This allows the estimation of the variation in blood use between hospitals compared with 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 antibodies and ultimately 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 into the logistical requirements and limitations, costs and (health) effects of various preventive matching schemes. In the ongoing BloodMatch study, 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 RBC stocks in the blood bank and hospitals, as well as the requirements for typing of the donor base in order to fulfil the demand for typed RBCs. The findings will allow balancing various aspects of the blood transfusion chain and therefore provide the means for a global optimisation 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 into donor-related and product-related risk factors for recipient outcomes, blood transfusions can be more tailored (minimising risks) and unnecessary transfusions avoided, further reducing transfusions 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 (optimise the availability and minimise wastage of blood products). For healthcare institutions, DTD enables insight into efficient and safe use of blood products. Moreover, by participating in the project, hospitals can have better control on 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 to 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, 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 haemoglobin 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. Whereas the level of detail in the indicators themselves is of less importance (especially in complex practices such as heart surgery and haematology), the discussion between clinical experts might provide novel insights, solutions to existing problems and evolution 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 hospitalisation 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 focuses on donors’ health using donor hospital information and already has national coverage. In future it will be possible to expand the data warehouse with additional variables, either permanently or temporarily, 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 organisational 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, so that 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 centres, the Dutch Association of general hospitals and specialised institutions, and the centre of expertise for standardisation and eHealth Nictiz, and, for example, in the USA by the Centers for Medicare and Medicaid Services, promoting meaningful use of certified EHR technology. If uniform registration will be successfully implemented in hospitals, standardised 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 the standardisation or maximum likelihood methods.

CONCLUSION
The DTD contributes to the optimisation 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.

Author affiliations
1 Transfusion Technology Assessment Department, Sanquin Research, Amsterdam, The Netherlands
2 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
3 Department of Public and Occupational Health, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands
4 Department of Clinical Chemistry and Haematology, University Medical Center Utrecht, Utrecht, The Netherlands
5 Department of Transfusion Medicine, Sanquin Blood Bank, Amsterdam, The Netherlands

Contributors MPJ, MMWK and MCdB secured the funding of this study. BHH and PK obtained ethical and privacy approval. BHH and LRvH drafted the manuscript. MCdB, PK, MPJ, MMWK and KMKvD critically reviewed the manuscript. All authors approved the final version.

Funding This study was funded by Sanquin Blood Supply (PPOC-11-042).

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Our project involves building a data warehouse for use by researchers. Researchers planning a project can gain access to the data of the warehouse by completing a data request form. The project advisory board, consisting of representatives of all involved disciplines, will determine whether the request should be granted.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

REFERENCES


