The new Scandinavian Donations and Transfusions database (SCANDAT2): a blood safety resource with added versatility

Gustaf Edgren,[1,2] Klaus Rostgaard,[3] Senthil K. Vasan,[1] Agneta Wikman,[4] Rut Norda,[5] Ole Birger Pedersen,[6] Christian Erikstrup,[7] Kaspar René Nielsen,[8] Kjell Titlestad,[9] Henrik Ullum,[10] Mads Melbye,[3,11,12] Olof Nyrén,[1] and Henrik Hjalgrim[3]

BACKGROUND: Risks of transfusion-transmitted disease are currently at a record low in the developed world. Still, available methods for blood surveillance might not be sufficient to detect transmission of diseases with unknown etiologies or with very long incubation periods.

STUDY DESIGN AND METHODS: We have previously created the anonymized Scandinavian Donations and Transfusions (SCANDAT) database, containing data on blood donors, blood transfusions, and transfused patients, with complete follow-up of donors and patients for a range of health outcomes. Here we describe the re-creation of SCANDAT with updated, identifiable data. We collected computerized data on blood donations and transfusions from blood banks covering all of Sweden and Denmark. After data cleaning, two structurally identical databases were created and the entire database was linked with nationwide health outcomes registers to attain complete follow-up for up to 47 years regarding hospital care, cancer, and death.

RESULTS: After removal of erroneous records, the database contained 25,523,334 donation records, 21,318,794 transfusion records, and 3,692,653 unique persons with valid identification, presently followed over 40 million person-years, with possibility for future extension. Data quality is generally high with 96% of all transfusions being traceable to their respective donation(s) and a very high (>97%) concordance with official statistics on annual number of blood donations and transfusions.

CONCLUSIONS: It is possible to create a binational, nationwide database with almost 50 years of follow-up of blood donors and transfused patients for a range of health outcomes. We aim to use this database for further studies of donor health, transfusion-associated risks, and transfusion-transmitted disease.

ABBREVIATIONS: CID = component identifier; CRN(s) = civil registration number(s); DID = donation identifier; IQR = interquartile range; NRN(s) = national registration number(s); PID = personal identifier; SCANDAT = Scandinavian Donations and Transfusions.

From the [1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, and the [2]Hematology Centre, Karolinska University Hospital, Stockholm, Sweden; the [3]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; the [4]Department of Clinical Immunology and Transfusion Medicine, Karolinska University Hospital, Stockholm, Sweden; the [5]Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; the [6]Department of Clinical Immunology, Næstved Hospital, Næstved, Denmark; the [7]Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark; the [8]Department of Clinical Immunology, Aalborg University Hospital, Aalborg, Denmark; the [9]Department of Clinical Immunology, Odense University Hospital, Odense, Denmark; the [10]Department of Clinical Immunology, the Blood Bank, Rigshospitalet, University Hospital of Copenhagen, and the [11]Department of Clinical Medicine, Copenhagen University, Copenhagen, Denmark; and the [12]Department of Medicine, Stanford School of Medicine, Stanford, California.

Address reprint requests to: Gustaf Edgren, MD, PhD, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Box 281, SE-171 77 Stockholm, Sweden; e-mail: gustaf.edgren@ki.se.

The assembly of the SCANDAT2 database was made possible through grants from the Swedish Research Council (2011-30405, 2007-7469), the Swedish Heart-Lung Foundation (20090710), the Swedish Society for Medical Research (GE), and the Danish Council for Independent Research (2009B026). We would also like to acknowledge scientific support from the REDS-III program funded by the U.S. National Heart, Lung, and Blood Institute.

Received for publication September 22, 2014; revision received November 1, 2014, and accepted November 12, 2014.

doi: 10.1111/trf.12986

TRANSFUSION 2015;55:1600–1606

After decades of systematic improvement, risks of transfusion-transmitted infections are currently at a record low in the developed world.[1] Meanwhile, considering that a blood transfusion remains a near ideal mode for transmitting blood-borne disease from one individual to another, the threat of transfusion-transmitted infections must still be considered. Drawing experience from the recent emergence of West Nile virus in the US blood supply, the National Heart, Lung and Blood Institute emerging infectious disease task force recently presented a series of recommendations for the management of future threats to the blood supply.[2] However, these methods might not be sufficient to detect transmission of diseases with unknown etiologies or with very long incubation periods. Studies of disease concordance between blood donors and recipients may provide strong indications of a transmissible factor, which may be relevant for both suspected and unsuspected pathogens. As such, systematic surveillance programs may be key to ensure blood safety in the future.

We have previously described the creation and contents of the Swedish-Danish, Scandinavian Donations and Transfusions (SCANDAT) database.[3] This database, which comprises the entire computerized blood donation history of Sweden and Denmark from 1968 (Sweden) and 1980 (Denmark) until 2002, has subsequently been used for the study of disease risk in blood donors,[4,5] health consequences of blood donation,[6] transfusion transmission of disease,[7] patterns of blood use,[8] health of transfused patients,[9-13] and health-economy evaluations of blood screening.[14] The premise of the creation of the SCANDAT database was assessment of short- and long-term health effects of blood donations and blood transfusions, as well as the assessment of possible transfusion-transmitted disease. However, since the database was permanently deidentified, it was not possible to update the database or to collect further information or biologic samples from especially informative subjects.

Here we describe the creation, contents, and quality of a new and improved version of the SCANDAT database (SCANDAT2), encompassing the entire computerized blood donation and blood transfusion experience in two countries, with complete follow-up of subjects with regard to cancer, hospital care, and causes of death until 2012.

MATERIALS AND METHODS

Data sources

Continuous computerized recording of blood donation and transfusion activities began using a centralized, magnetic tape–based computer system in Sweden in 1966 and in Denmark in 1981.[15,16] Some computerized recording had been in place already in the mid-1960s in Denmark, but data from these systems were not stored on lasting media and could therefore not be retrieved retrospectively.[17] Common for all computer systems in use in both countries is that all individuals are identifiable through their unique personal identification numbers, referred to as national registration numbers (NRNs) in Sweden and as civil registration numbers (CRNs) in Denmark. These numbers, which are used in all health care contacts and serve as unique identifiers in all other health data and population registers, can thus be used to perform precise record linkages of data from many different sources.[18,19]

Data collection

The collection of data in Sweden and Denmark followed a similar process as during the creation of the original version of the SCANDAT database.[3] After acquiring approval from ethical committee (Sweden) and the Data Protection Agency (Denmark), all current transfusion medicine clinics and blood banks were contacted and asked to provide data for the update. Data were extracted from the local blood donation and transfusion databases that are kept for administrative purposes, for example, to ensure traceability of blood products and to monitor blood component production and use.

Upon reaching agreements, data managers in individual blood banks, regional blood centers, or system contractors performed database extractions locally, according to our specifications. In Sweden, this also included reimport of identifiable data from magnetic tape–based systems that had been kept by one of the major computer system contractors. We asked for data at least up to and including 2010, but data were often delivered complete until the date of data extraction. This meant that data were complete at least until 2010, but in most instances through 2012. In Sweden, encrypted data were transmitted to us through a secure file transfer system, RMFT (RepliWeb Managed File Transfer, Attunity, Inc., Burlington, MA). In Denmark, data were transmitted to us either as encrypted files over FTP service or on physical media carried by personal courier or sent as registered post.

Data processing, record linkages, and database design

As data from the different register holders were gradually delivered, we removed obviously incorrect records, for example, where there were invalid or missing dates, missing NRNs, or with missing information on primary key variables. Data were then reformatted to fit a common data structure, and once all the data had been delivered, cleaned, and reformatted, two separate but structurally identical databases were created, one in Sweden and one in Denmark. Because data had been extracted according to our detailed specifications, data processing was kept at a minimum and mostly focused on removal of duplicate

observations, as well as harmonization of coding systems for donation types, component manufacturing steps, and component types.

After the data processing in each country was completed, record linkages with official registers were performed following marginally different procedures. In Sweden, the complete database was submitted to Statistics Sweden, a government organization responsible for maintenance of a range of nationwide population registers. Here the database underwent a pseudonymization process, where all NRNs were replaced with randomly created, yet unique, identification numbers (henceforth referred to as personal identifier [PID]). The database was then linked with nationwide population registers to eliminate faulty NRNs or NRNs not corresponding to persons recorded in the national population and/or death registers as living, emigrated, or deceased. The linkages also provided information on sex, date of birth, vital status, dates of immigration and emigration, country of birth, data on migration within Sweden (i.e., between different administrative regions), several measures of socioeconomic status, and from the multigeneration register, information about first-degree familial relationships within the study population. A key file, linking NRNs to the randomly created PIDs, was then transmitted to the Swedish Board of Health and Welfare, a government agency responsible for maintaining all nationwide health data registers, for further record linkage and for storage to allow later reidentification or update of the database. Record linkages at the Swedish Board of Health and Welfare provided data from the Cancer Register, In- and Outpatient Registers, Medical Birth Register, and Cause of Death Register. The pseudonymized database, together with extracted data from the nationwide registers, was then transmitted back to our institution where it was organized into a common database structure.

In Denmark, information on sex, date of birth, vital status, country of birth, first-degree relatives, immigration, and emigration and migration within Denmark was obtained from the Civil Registration System using the CRN as key. Similarly, data were obtained from the nationwide Danish Cancer Register, Hospital Discharge Register, and Cause of Death Register. Upon completion of the data linkages, the database was pseudonymized by replacing individual CRNs with randomly created PIDs. A key file linking CRNs with PIDs was stored locally in accordance with Danish legislation.

A schematic representation of the final database design is presented in Fig. 1. Tables are linked using three unique identifiers: the PID, the donation identifier (DID), and component identifier (CID). The core of the database comprises four tables identifying persons, donations, components, and transfusions. The persons table contains baseline information on all persons in the database. The donation table holds data on all donations, including the PID of the donor, donation date, donation type, and the DID identifying that particular donation. The component table links to the donation table through the DID and also holds the CID identifying that particular component, the component type, manufacturing date, and data that allows the tracking of all donations contributing to pooled components. The transfusion table links to the donation table, through the DID; the component table, through the CID; and the persons table, through the PID. It holds the identity of the recipient, the date of transfusion, and the component type. In addition, the database also holds records on erythrocyte antigens and antibodies. Details of the nationwide population and health data registers have been described previously.[18,20]

Statistical analyses

We counted the total number of donors, donations, transfusions, and transfused patients registered in the database and produced descriptive statistics on the number of donations per donor and the number of transfusions of individual blood components, per patient. We also compared annual frequencies of number of donations and transfusions with available official statistics from 2002 and onward. Further, we also estimated the proportion of transfusions for which we could find a valid donation record; the deficit typically reflecting that initiation of registration of donations and transfusions did not always start at the same time. To describe the interrelatedness between donors and transfused patients, and hence the potential for transmission of latent agents, we also calculated the number of recipients per donor, with 5-, 10-, and 20-year follow-up from the first recorded donation.

The conduct of this study was approved by the regional ethics review board at Karolinska Institutet in Stockholm, Sweden (Reference Numbers 2009/1011, 2012/1233, 2013/37, and 2013/787) and by the Danish Data protection agency (Reference Number 2008-54-0472). All data management and statistical analysis was performed using statistical analysis software (SAS, Version 9.2 or higher, SAS Institute, Cary, NC).

RESULTS

We received data from a total of 40 local donation and transfusion databases (24 in Sweden and 16 in Denmark). After removal of duplicate records as well as records with obvious errors in key variables (e.g., dates, DIDs, or CIDs), a total of 25,640,115 donation records, 22,110,794 transfusion records, and 3,777,785 person records remained. Of these, 116,781 (0.5%) donation records, 792,000 (3.6%) transfusion records, and 85,132 (2.3%) persons records did not match with the national population or death register records due to invalid NRNs. Thus, a total of 25,523,334 donation records (18,498,546 in Sweden and 7,024,788 in

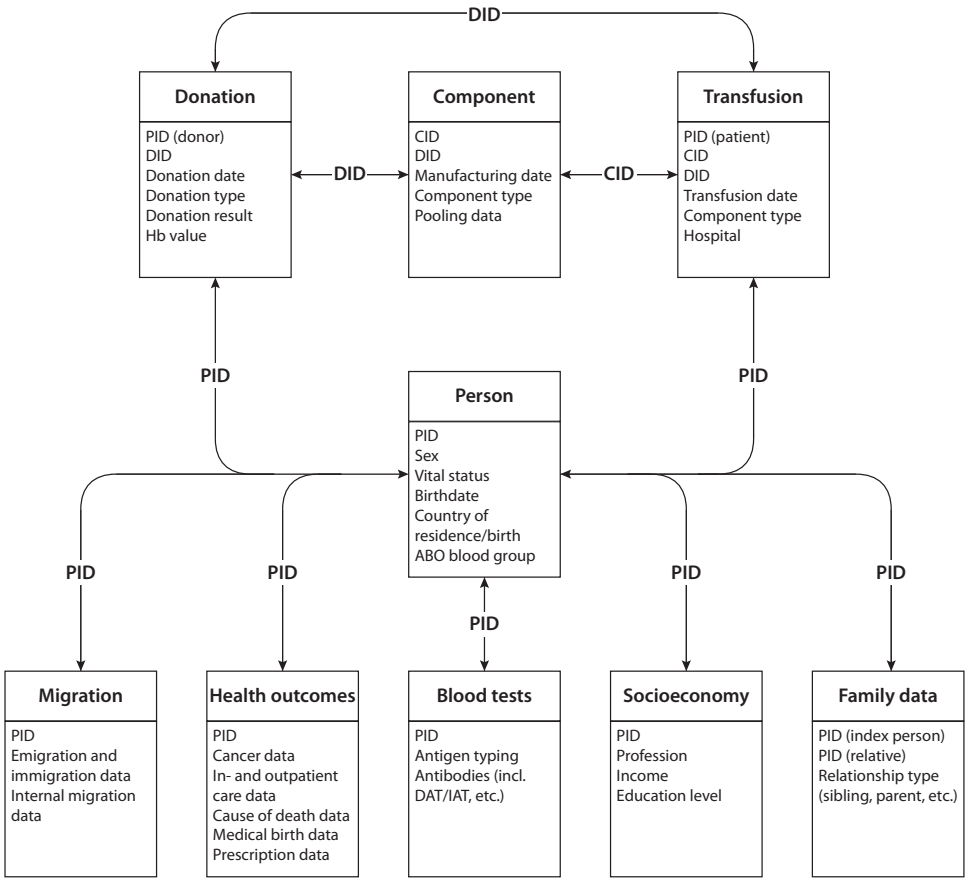

Fig. 1. Schematic description of the new SCANDAT database.

Denmark), 21,318,794 transfusion records (13,675,667 in Sweden and 7,643,127 in Denmark), and 3,692,653 persons records (2,442,719 in Sweden and 1,249,934 in Denmark) remained. Of the persons with valid identity, 1,625,325 made at least one donation (1,071,473 in Sweden and 553,852 in Denmark), 2,180,414 received at least one transfusion (1,454,574 in Sweden and 725,840 in Denmark), and 113,086 (3.1%) appeared both as donors and as transfusion recipients. Of persons who were both donors and recipients, 85% donated blood only before their first electronically registered transfusion.

The annual number of donations and transfusions, as well as the proportion of transfusions for which we were able to find a matching donation, are presented in Fig. 2A (Sweden) and Fig. 2B (Denmark). In both countries, the annual number of donations and transfusions per year increased throughout the study period, with the gradual computerization. Registration was more than 95% complete from 1996 in Sweden and from 1998 in Denmark. In Sweden, there was a decrease both in the donation and transfusion counts and in the proportion of transfusions for which we could find a matching donation around 1987. This resulted from the irreparable damage of at least one data tape.[3] Similarly, in Denmark the proportion of transfusions with matching donations was lower than expected both in 1996 and 1997 and in 2002 and 2003, where the former was mainly the result of transfusion registration starting before donation registration in the capital regions and the latter probably from temporary incompatibility during the introduction of the ISBT-128 system. In total, though, we were able to find a matching donation for 96% of all transfusions (96% in Sweden and 94% in Denmark). In both Sweden and Denmark, comparisons were made with official statistics on annual number of blood donations and transfusions for the data that was not included in the previous version of the SCANDAT database and revealed a high concordance, with the total number of donations and transfusions in our database differing from the officially compiled statistics by less than 3% in Sweden and less than 2% in Denmark (data not shown).

Throughout the study period, 47% of registered donors were women, who contributed 37% of all donations. The overall median age at first registered donation was 30.0 years (interquartile range [IQR], 22.9–40.5). Over a median follow-up of 15.3 years (IQR, 8.3-23.7 years) the blood donors made a median of eight donations (IQR, 3-20). The total follow-up in the donor cohort was 27.0

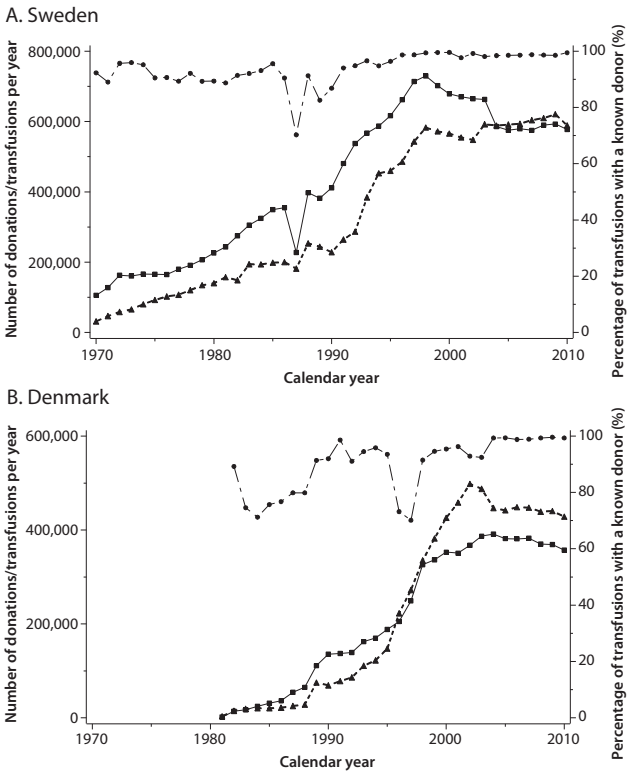

Fig. 2. Annual number of donations (■) and transfusions (▲) together with the proportion of transfusions that are traceable to their respective donation (●) presented per year for Sweden (A) and Denmark (B).

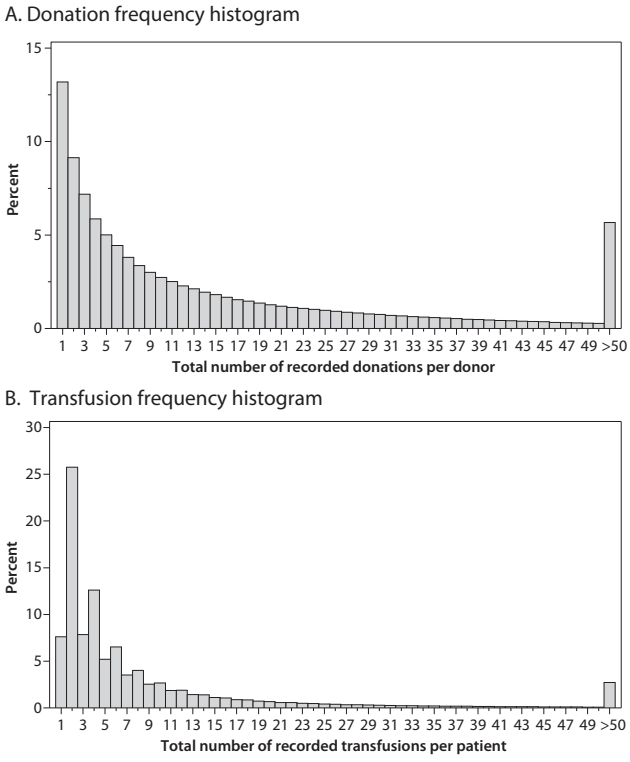

Fig. 3. Histograms of total number of donations per donor (A) and total number of transfusions per recipient (B).

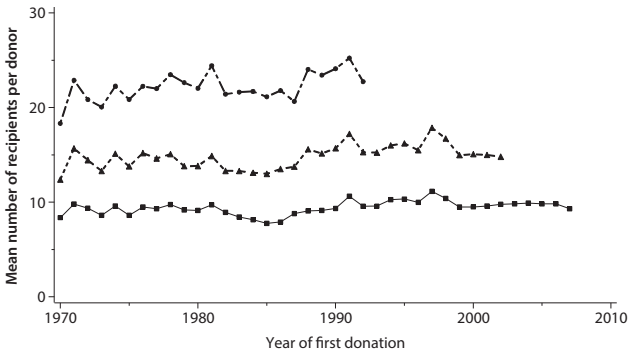

Fig. 4. Mean number of recipients per donor by year of first donation, presented with 5 (■), 10 (▲), and 20 (●) years of follow-up from the date of first donation.

million person-years. Of all transfused patients 56% were female, and women received 46% of transfused components. The median age at the first recorded transfusion was 69.9 years (IQR, 55.1-79.8 years). Over a median follow-up of 3.9 years (IQR, 0.7-9.7 years) the transfused patients received a median of four (IQR, two to nine) transfusions. The total follow-up in the recipient cohort was 14.4 million person-years.

Histograms illustrating the number of donations and transfusions per person are presented in Fig. 3. In all instances, the distributions were skewed with the majority of donors and recipients making only a limited number of donations or receiving only a limited number of transfusions, respectively. Interestingly, in both Sweden and Denmark the number of transfusions followed a pattern where patients were mostly transfused in pairs of two (i.e., 2, 4, or 6 units).

Figure 4 shows the mean number of recipients per donor with 5, 10, and 20 years of follow-up. Overall, during the first 5 years from the first recorded donation donors gave blood to a mean (±SD) of 9.7 (±9.2) recipients. Over 10 and 20 years of follow-up each donor donated to 15.4 (±16.6) and 23.1 (±29.5) recipients, respectively. Patterns were very similar in the two countries, but with slightly higher recipient counts in Sweden, and remained comparatively constant over time.

DISCUSSION

Here we describe our experience with the reassembly of a new version of the SCANDAT database (SCANDAT2) with up to 44 years of follow-up of blood donors and transfused patients. Importantly, contrary to the previous version of the database, we now have preserved identification possibilities of both donors and recipients, permitting both

updating the database and identifying especially informative subjects for further collection of information or samples. The main finding is thus the feasibility of creating such a database from routinely collected administrative data. Overall, we deem that the quality of the data is very high with only 0.5% of donation and 3.6% of transfusion records having missing or invalid identification numbers, respectively, and as many as 95% of all transfusions being traceable to its donor(s).

As with the previous version of SCANDAT, some caveats must be considered. First, while data quality seems very high and that there is a high degree of concordance with official statistics, in the absence of readily accessible and reliable data on administered transfusions in patient medical records, we were not able to conduct any person-level validation study. Second, although we now have another 10 years of exposure data and follow-up, allowing truly long-term analyses, left truncation is still a relevant problem. However, availability of detailed data on residence of all subjects ensures that we can still reliably ascertain periods during which follow-up is complete for all subjects. Third, and perhaps most importantly, while availability of very detailed, national health data registers allows the assessment of associations with a wide range of health outcomes, accounting fully for the "healthy donor effect" and for the indication for the blood transfusion is still difficult.[21] As such, inference on the health effects of repeated blood donation or transfusion may be limited.

In addition to the caveats listed above, there were also some gaps in the data that warrant mention. The dip in the donation and transfusion frequencies as well as in the traceability of transfused units in the Swedish data in the mid 1980s resulted from the loss of at least one data tape and illustrates the importance of reliable backup and the use of lasting computer media. At the same time, we experienced surprisingly—and impressively—little problems with the data from even earlier years. We also saw two periods of lower traceability in the Danish data. The first, in the mid 1990s, was the inevitable effect of electronic transfusion registration in some registers preceding donation registration. A second deficit, in 2003, reflected the lack of a uniform system during the adoption of ISBT codes around 2003. This, in turn, resulted in an inability to trace pooled blood components within the data sets received.

The primary purpose of SCANDAT2 is to provide a versatile tool with high statistical resolution for the assessment of threats to the blood supply, including possible donation-associated risks, risks of transfusion-transmitted disease, and transfusion safety in a more general sense. In this regard even simple tabulations such as those presented here provide information relevant to policy and decision makers. For instance, the continued tendency for transfusion of blood products in pairs of two—even recently—suggests that administrative routines may still override the factual needs of the patient; even though there has been a general decline in the use of blood products in recent years, more can be done to tailor the blood therapy for each patient. Also, while the demographic composition of the donor population presumably has changed over the past decades because of increasingly strict health criteria,[4] the number of recipients per donor has remained remarkably constant during the same period. This indicates a relatively stable donation behavior across different generations of blood donors, relevant to the evaluation of donor recruitment and retention efforts. Insofar as the original database has been used for all of these purposes,[4-13] we believe that our concept of long-term hemovigilance can be further systematized and expanded. We are therefore currently developing methods for an agnostic surveillance system where we search for disease concordance between donors and recipients, without specific hypothesis. We believe that such a tool, ideally coupled with a frequently updated database and with other independent linked donation-transfusion databases, also outside of Scandinavia, could strengthen transfusion safety in the future. As such, it is encouraging to see ongoing projects aiming to establish similar databases outside of Scandinavia, for example, the REDS-III study in the United States.[22]

In conclusion, here we describe the re-created, new version of the SCANDAT database. It is our firm belief that the database will form a useful tool for the assessment of both current and emerging threats to the blood supply.

ACKNOWLEDGMENTS

We are indebted to all the blood banks in Sweden and Denmark for both collecting and contributing data to this study.

CONFLICTS OF INTEREST

The authors have disclosed no conflicts of interest.

REFERENCES

1. Alter HJ, Klein HG. The hazards of blood transfusion in historical perspective. Blood 2008;112:2617-26.
2. Glynn SA, Busch MP, Dodd RY, et al. Emerging infectious agents and the nation's blood supply: responding to potential threats in the 21st century. Transfusion 2013;53: 438-54.
3. Edgren G, Hjalgrim H, Tran TN, et al. A population-based binational register for monitoring long-term outcome and possible disease concordance among blood donors and recipients. Vox Sang 2006;91:316-23.
4. Edgren G, Tran TN, Hjalgrim H, et al. Improving health profile of blood donors as a consequence of transfusion safety efforts. Transfusion 2007;47:2017-24.

5. Edgren G, Hjalgrim H, Rostgaard K, et al. Risk of gastric cancer and peptic ulcers in relation to ABO blood type: a cohort study. Am J Epidemiol 2010;172:1280-5.
6. Edgren G, Reilly M, Hjalgrim H, et al. Donation frequency, iron loss, and risk of cancer among blood donors. J Natl Cancer Inst 2008;100:572-9.
7. Edgren G, Hjalgrim H, Reilly M, et al. Risk of cancer after blood transfusion from donors with subclinical cancer: a retrospective cohort study. Lancet 2007;369:1724-30.
8. Kamper-Jorgensen M, Edgren G, Rostgaard K, et al. Blood transfusion exposure in Denmark and Sweden. Transfusion 2009;49:888-94.
9. Hjalgrim H, Edgren G, Rostgaard K, et al. Cancer incidence in blood transfusion recipients. J Natl Cancer Inst 2007;99: 1864-74.
10. Kamper-Jorgensen M, Ahlgren M, Rostgaard K, et al. Survival after blood transfusion. Transfusion 2008;48:2577-84.
11. Edgren G, Kamper-Jorgensen M, Eloranta S, et al. Duration of red blood cell storage and survival of transfused patients (CME). Transfusion 2010;50:1185-95.
12. Norda R, Andersson TM, 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. J Trauma Acute Care Surg 2012;72:954-60.
13. Shanwell A, Andersson TM, Rostgaard K, et al. Post-transfusion mortality among recipients of ABO-compatible but non-identical plasma. Vox Sang 2009;96:316-23.
14. Kamper-Jorgensen M, Hjalgrim H, Edgren G, et al. Expensive blood safety initiatives may offer less benefit than we think. Transfusion 2010;50:240-2.
15. Hogman CF, Ramgren O. Computer system for blood transfusion service. Transfusion 1970;10:121-32.
16. Hogman CF, Cassemar B. A national blood transfusion system integrating a central computer service with hospital-based minicomputer routines. Vox Sang 1981;40: 222-7.
17. Freiesleben E, Jensen KG, Tamborg O. [Registration of blood donors and electronic data processing]. Ugeskr Laeger 1965;127:1171-6.
18. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, et al. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. Eur J Epidemiol 2009;24:659-67.
19. Pedersen CB. The Danish Civil Registration System. Scand J Public Health 2011;39:22-5.
20. Thygesen LC, Ersboll AK. Danish population-based registers for public health and health-related welfare research: introduction to the supplement. Scand J Public Health 2011;39:8-10.
21. Edgren G, Hjalgrim H. Epidemiological considerations for the use of databases in transfusion research: a Scandinavian perspective. Curr Opin Hematol 2010;17:596-601.
22. Kleinman S, Busch MP, Murphy EL, et al. 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;54:942-55.