Modelling blood safety

Mart Janssen

Modelling blood safety

Utrecht, Universiteit Utrecht, Faculteit Geneeskunde

Dissertation University Utrecht – with a summary in Dutch
Proefschrift Universiteit Utrecht – met een samenvatting in het Nederlands

ISBN	978-90-393-5401-8
Cover design	Eijkpunt Grafische Vormgeving, Bilthoven, the Netherlands
Print	Gildeprint Drukkerijen B.V., Enschede, the Netherlands

Modelling blood safety

Het modelleren van bloedveiligheid
(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. J.C. Stoof,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
dinsdag 19 oktober 2010
des middags te 4.15 uur

door

Martinus Petrus Janssen

geboren op 21 februari 1964 te Nijmegen

Promotoren:

Prof.dr. B.A. van Hout
Prof.dr. G.J. Bonsel

Co-promotor:

Dr. C.L. van der Poel

Sanquin Blood Supply Foundation is gratefully acknowledged for funding the majority of the research described in this thesis and for providing financial support for the publication of this thesis.
The Netherlands Organization for Health Research and Development is also appreciatively acknowledged for funding part of the research (Chapters 1, 2 and 4, ZonMw grant reference 945-06-603).
Additional financial support for the publication of this thesis was kindly provided by the Julius Center for Health Sciences and Primary Care, Novartis Vaccines and Diagnostics, and Ortho Clinical Diagnostics.

Voor mijn vaders:

Pa
Cor
Joop

TABLE OF CONTENTS

APPENDICES

LIST OF ABBREVIATIONS

ABO	ABO blood group system
AEVPI	Attributable expected value of perfect information
AIDS	Acquired immunodeficiency syndrome
ARP	At risk period
AVOI	Attributable value of information
AVR	Active viral replication
BCC	Blood collection centre
BCU	Bacterial culturing
BPI	Blood products issued
BPT	Blood product transfusion
BTC	Blood transfusion chain
CC	Confidence contour
CEA	Cost-effectiveness analysis
CET	Cost-effectiveness threshold
CI	Confidence interval
CLB	Central laboratory (Centraal Laboratorium van de Bloedtransfusiedienst van het Nederlandse Rode Kruis)
DE	Direct estimator of survival probability
DOC	Direction of change
EMEA	European medicines agency
EVPI	Expected value of perfect information
EVPPI	Expected value of partial perfect information
EVSI	Expected value of sample information
FFP	Fresh-frozen plasma
HAV	Hepatitis A virus
HBsAg	Hepatitis B surface antigen
HBV	Hepatitis B virus
HCC	Hepatocellular carcinoma
HCMV	Human cytomegalovirus
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
HTLV	Human T-lymphotropic virus (type I or II)
ICD-9	9th edition of the International Statistical Classification of Diseases and Related Health Problems, published by the WHO
ICER	Incremental cost-effectiveness ratio
ID	Individual donation
IR	Incidence rate
ISBT	International society of blood transfusion
KM	Kaplan-Meier survival probability estimator
LR	Likelihood ratio
LRF	Log reduction factor
MDE	Monotonized direct survival probability estimator
MP	Mini-pool
MPN	Most probable number
NAT	Nucleic acid test
NMB	Net monetary benefit
PARVO	Parvo virus B19
PDMP	Plasma derived medicinal product
PLT	Platelet
PROTON	Profiles of transfusion recipients (PROfiel van Transfusie ONtvangers)
PRT	Pathogen reduction technology
QALY	Quality adjusted life year
QoL	Quality of life
RBC	Red blood cell
RMSE	Root mean squared error
SCANDAT	Scandinavian donations and transfusions database
SD	Standard deviation
SE	Standard error
SK	Spearman–Kärber
SRC	Standardised regression coefficient
$TCID_{50}$	Median tissue culture infective dose
TRALI	Transfusion related acute lung injury
TTA	Transfusion technology assessment
UIR	Undetected incidence rate
vCJD	Variant Creutzfeldt–Jakob disease
VOI	Value of information
WHO	World Health Organization
WP	Window period

INTRODUCTION

“All truths are easy to understand once they are discovered; the point is to discover them.”

Galileo Galilei

Objectives

This thesis describes the development and application of methods and models to support decision making on safety measures aimed at preventing the transmission of infections by blood donors. Apart from the direct interest of selecting best practices, the thesis aims to increase awareness of the opportunities which quantitative decision models offer in this context.

Background

Blood and blood transfusion associated risks

Blood transfusion has a long history. The most often quoted first recording of a blood transfusion was in the late 15th century, where the blood of three young boys was given to pope Innocent VIII. Without success: all boys died, as did the pope. Since then, a tremendous amount of research has been performed on the therapeutic use of (human) blood and risks associated with blood transfusion.

Blood is increasingly recognized as a complex mixture of proteins and blood cells which fulfil various critical functions within the human body. Technological developments allow extraction of particular elements of blood to be used for various specific medicinal purposes. What may not be common knowledge is that blood is not transfused in the form that it is obtained from the donor, but split into components (erythrocytes, thrombocytes and plasma) and that one of those components, blood plasma, is fractionated further to extract various proteins. What is probably common knowledge is that different individuals have different blood constituencies which can lead to incompatibility issues. The most well known assortment of human blood types is the ABO blood group system. Blood group incompatibility, the fact that individuals with type A blood can only receive blood from donors of type A and type O, is still one of the larger blood transfusion related risks.[1,2] Mismatch is a type of accident that can occur as a result of an error in determining the patients' blood type (either in the hospital or at the blood bank), or as an error in administration: the wrong blood is transfused to the (wrong) patient.

The transmission of blood borne infections – be it bacterial, viral, parasitic or of any other origin – poses another, in most countries far smaller risk. Risks of in-hospital processes are difficult to control as there are many potential sources for error. In contrast, incorporating screening tests – which are available for most common blood transmittable viruses – seems relatively easy. Another effective prevention strategy, independent from screening of blood once donated, is that of pre-selection of donors. The blood bank follows a strict selection regime and applies a continuously updated list of exclusion criteria for regular as well as for new donors.

Implementation of blood transfusion safety measures

Examples of screening tests and donor selection can easily be found in recent history. After the transfusion-associated infection with the Human Immunodeficiency Virus (HIV) was discovered in 1982, questioning was introduced on behavioural risks and subsequent deferral if self-reported risky behaviour was present. This reduced the risk of HIV transmission significantly.[3] Only a few years later, in 1985, a screening test for HIV (anti-HIV test) was developed and implemented in the USA and in Europe.[2] Screening for hepatitis B (HBV) was already introduced in Germany in the 1960s (alanine transaminase [ALT] testing), but more broadly implemented was the screening for hepatitis surface antigen (HBsAg), which was introduced in the 1970s.[2,4] The discovery of the hepatitis C (HCV) virus in 1988 and the subsequent development of serological screening assays for antibodies to this virus reduced the risk of transfusion related transmission of HCV significantly.[5] The tests mentioned above are all indirect tests, as they measure the response of the human body to a viral infection, the production of anti-bodies or antigens. The time range between infection and (detectable) response differs according to viral agents and persons; this range is called the window period, a concept critical to this thesis. Nowadays the presence of the virus itself is usually determined by Nucleic Acid Tests (NAT). Most western countries apply such tests at present to detect these viruses.[2] In the Netherlands in addition to these viruses also tests are performed for syphilis, human T-lymphotropic virus (HTLV), malaria, Parvovirus B19 (PARVO) and the human cytomegalovirus (HCMV).[6] Screening tests are rarely 100% accurate and there may often be potential for improvements which may be rather costly.

Next to screening and donor selection there are other measures which may reduce the risks of infection transmission. In the early 1990s it was recognized that infectious diseases and some pathogens are more concentrated in white blood cells (leukocytes) than in the rest of the blood product. As a result, some countries implemented leukoreduction (or leukodepletion) in the late 1990s.[2] The presumed increase in infectivity of leukocytes was only one of the concerns. Other considerations concerned the potential (and undesirable) immunization against HLA antigens present on the donor leucocytes. Also, one feared the suppression of immune responses, which may for instance result in Transfusion Related Acute Lung Injury (TRALI).[7] In the Netherlands general leukodepletion of blood products was implemented in 2001.

As often seen in other areas, after one threat is controlled, priority is given to the next one. When in the early 1990s the viral risks seemed under control, the risk of bacterial contamination became (more) apparent and the first reporting on culturing of platelets for the detection of bacteria in the USA occured.[8] The Netherlands was among the first countries to introduce universal screening of platelets for bacteria in 2001.[9]

New and renewed discussions on the implementation of blood safety interventions arise whenever new evidence on infections or new strategies becomes available. A recent example of such discussions concerned variant Creutzfeldt-Jakob Disease (vCJD) which has been shown to be transmitted by blood transfusion, and for which blood screening tests have been developed.[10] In the Netherlands there have currently been three confirmed clinical cases of vCJD caused by blood transfusion.[11] A second example concerns the implementation of pathogen reduction technologies (PRT), which is also

referred to as pathogen inactivation. This is a treatment method for blood products that significantly reduces infectivity of any viral or bacterial contamination irrespective of its origin.[12] In both examples questions arise on the necessity of such measures on top of the existing safety pyramid.

Blood safety, decision making and modelling

For most of the currently implemented blood safety measures the course of events was as follows: 1) signalling of an adverse event associated with blood transfusion, which could be interpreted as caused by a specific agent; 2) development of an agent-specific screening test or other safety intervention; 3) implementation of the safety intervention into routine practice. This sequence of events seems inevitable as part of a general belief that no effort to avoid transfusion-related infections is too high.[2] The introduction of safety measures themselves seem to be infectious: once one country adopts a newly developed technology, most (adjacent) countries follow the example, partly as a result of peer pressure and more or less clear economic interests. However, with the ability to produce ever more sensitive tests and more advanced safety measures, the question arises whether such interventions are actually worthwhile. This issue was explicitly addressed in the published Dutch governmental strategy for our national medical blood provision, which expressed the need for optimal rather than maximal blood safety.[13]

For the previously mentioned examples of vCJD and PRT implementation is far from self-evident if the balance is made of pros and cons. For the first example (a screening test for vCJD) the number of transmitted infections prevented is expected to be truly negligible (as there are only very few cases) whereas the cost of screening is substantial and the consequences of disease transmission detrimental. For the second example (applying PRT) the obtained effect *is* substantial, as are the additional costs for each blood product treated. This leads to the question whether the expenditure can be justified by the risk reduction(s) obtained.

To support decision making processes in such contexts, costs and effects of the safety interventions need to be analyzed. While medical knowledge by default is obtained carrying out careful real life experiments (called randomized clinical trials), this approach is infeasible in blood transfusion safety. The main reason is the fact that the risks are very low, which renders events to become (very) rare. This poses limitations on the practicality of such experiments. In addition, for experiments addressing the effectiveness of safety interventions there are also ethical concerns.[14] This is where mathematical modelling is put forward to proceed. Models rest on an aggregate mathematical description of the *'real world'* such that conclusions concerning the modelled system can be drawn. However, from this definition it becomes clear that models serve a much wider range of applications than the evaluation of the cost-effectiveness of safety interventions alone.

Modelling the blood transfusion chain

Analyzing blood safety interventions requires modelling of the blood transfusion chain (BTC). The BTC refers to the interrelation between on one hand the blood donors and on the other the blood transfusion recipients, with in between the blood bank. A schematic of the BTC is shown in Figure 1. From this figure the position of the blood bank becomes clear: it is responsible for delivering high quality (which implies safe) blood products to

the hospitals, but is dependent on the availability and characteristics of its donors, which it draws from the general population. Also, as the blood component or blood plasma derived medicinal product is merely a supplement to the patient treatment process in hospital, the output of the BTC is revealed in the patient, and as such out of direct (observational) reach of the blood bank. However, for a proper control of the BTC, a closing loop from hospital patient outcomes back to the inputs of the BTC process is required. This is why the Transfusion Technology Assessment (TTA) group of the Julius Center in 2005 applied for a grant (which was obtained) to acquire data on the profile of the transfusion recipient. This study, the PROTON study ('PRofielen van Transfusie ONtvangers') was recently completed. The recipient data together with data on the donor population which is available from the national Dutch blood bank (Sanquin Blood Supply Foundation) allows for a major increase in the precision of survival estimates as needed in various models of the BTC, for instance on the cost-effectiveness of blood safety interventions. The main ingredients for any model concerning the BTC safety interventions will include donor epidemiology, blood bank interventions (effectiveness and costs), and the costs and effects associated with adverse events patients.

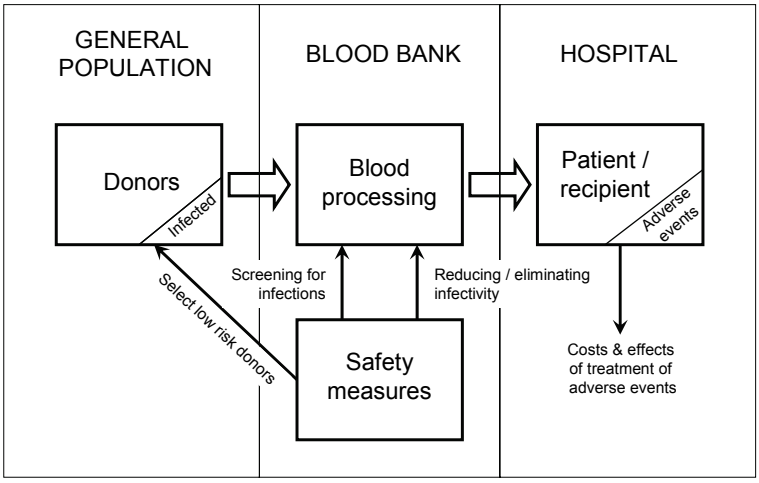


Figure 1: The blood transfusion chain (BTC)

Scope and outline

This thesis consists of a set of papers which have in common that they either (a) aim to support decisions pertaining to blood safety, and/or (b) concern the application of quantitative models and methods which were developed for that purpose. Main outcomes studied in these models are either the cost-effectiveness of safety interventions or the transmission risk of infectious agents.

The thesis consists of three parts:

In *Part A* descriptive statistics and survival modelling of blood transfusion recipients will be addressed. These are indispensable data to assess the benefits of blood safety interventions. In the Netherlands donor epidemiology can be studied using reliable data, since blood transfusion is centralized through the Sanquin organization (formerly CLB).

Data on donor and donation characteristics are collected in a standard fashion by Sanquin. The outline of the PROTON project and some descriptive epidemiological key data on Dutch blood transfusion recipients are provided in **Chapter 1**. The fact that this study was needed illustrates the hitherto lack of attention for the effectiveness and cost-effectiveness of blood transfusion, and, as part of that, of the safety interventions involved. Descriptive statistics of the Dutch blood transfusion recipient in terms of gender, disease and survival per component type were unavailable before. In **Chapter 2** a detailed study is performed on the estimation of recipient survival after transfusion.

Part B concerns the cost-effectiveness of several blood safety interventions, the uncertainty in the estimated outcomes and the implications of this uncertainty for decision making. Two cost-effectiveness analyses are presented in which two alternative implementations of additional safety interventions are compared. In **Chapter 3** the cost-effectiveness of bacterial culturing is compared to that of pathogen inactivation. The cost-effectiveness of additional HBV NAT screening, applied to individual or combined blood samples respectively, is analyzed in **Chapter 4**. The value of reducing the uncertainty of model outcomes to avoid - ultimately - erroneous decisions can be determined with Value Of Information analysis (VOI). In the past the VOI in a specific case was expressed in monetary terms only. We developed the attributable value of information concept, which enables quantification of VOI in terms of the outcomes of the intervention under consideration. In health economic applications the principal outcomes are generally costs and quality adjusted life years (QALYs). **Chapter 5** concerns the development of this new paradigm for value of information analysis.

Part C concerns the evaluation of risks of plasma derived medicinal products (PDMPs). In **Chapter 6** the structure of a generic probabilistic model for analyzing risks of PDMPs is described. This model was developed to allow analyses of the effectiveness of various risk reduction strategies on a national scale. The model renders an improvement to past practice where such risks were determined by merely stacking worst case assumptions. The most important driver of the risks of PDMPs appears the viral inactivation capacity of the production process. **Chapter 7** describes the development of a regression model dedicated to the evaluation of the robustness of the viral inactivation capacity of the production process against various process conditions. In the past the robustness was estimated using a linear regression model on viral reduction estimates based on serial dilution assay readings. The new model regresses on the serial dilution assay readings themselves. This is yet another example of an improvement of existing practice. The newly developed model greatly improved the quality of virus validation robustness studies. **Chapter 8** concerns the analysis and decision-making of observed infections in the donor population. It describes statistical tools and methods that were developed to signal changes or outliership of infectious disease incidence rates in the donor population of specific blood collection centers.

The focus of this thesis is on development and application of quantitative models to support decision making on safety measures against blood transmittable infections. The thesis concludes with a general discussion on the application of mathematical models within blood safety, with a focus on limitations, observed caveats, and potential future developments.

References

1. Hillyer CD, Josephson CD, Blajchman MA et al. *Bacterial contamination of blood components: risks, strategies, and regulation: joint ASH and AABB educational session in transfusion medicine*. Hematology Am Soc Hematol Educ Program 2003;575-89.
2. Allain JP. *Transfusion risks of yesterday and of today*. Transfus Clin Biol 2003;10:1-5.
3. Busch MP, Young MJ, Samson SM et al. *Risk of human immunodeficiency virus (HIV) transmission by blood transfusions before the implementation of HIV-1 antibody screening. The Transfusion Safety Study Group*. Transfusion 1991;31:4-11.
4. Gerlich WH, Caspari G. *Hepatitis viruses and the safety of blood donations*. J Viral Hepat 1999;6 Suppl 1:6-15.
5. Dodd R, Kurt Roth W, Ashford P et al. *Transfusion medicine and safety*. Biologicals 2009;37:62-70.
6. *Sanquin Scientific Report 2008*. Amsterdam: Sanquin Blood Supply Foundation, 2009.
7. Health Council of the Netherlands: Committee on New-Variant Creutzfeldt-JakobDisease and Leukodepletion. *Leukodepletion of blood products.* Health Council of the Netherlands. 2000. The Hague, Health Council of the Netherlands.
8. Yomtovian R, Lazarus HM, Goodnough LT et al. *A prospective microbiologic surveillance program to detect and prevent the transfusion of bacterially contaminated platelets*. Transfusion 1993;33:902-9.
9. de KD, Curvers J, de Kort WL et al. *Effects of skin disinfection method, deviation bag, and bacterial screening on clinical safety of platelet transfusions in the Netherlands*. Transfusion 2006;46:476-85.
10. Peden AH, Head MW, Ritchie DL et al. *Preclinical vCJD after blood transfusion in a PRNP codon 129 heterozygous patient*. Lancet 2004;364:527-9.
11. *Voor derde keer patiënt aan variant Creutzfeldt-Jakob overleden.* RIVM. 2-2-2009. http://www.rivm.nl/persberichten/2009/variant_creutzfeldt_jakob_2009.jsp
12. Epstein JS. *Alternative strategies in assuring blood safety: An Overview*. Biologicals 2010.
13. *Ministerieel Plan Bloedvoorziening 2006: Een beleidsvisie voor de lange termijn*. 2006.
14. Kellett JM. *Drug trial ethics*. J Med Ethics 1994;20:270.

PART

A

Epidemiological Description of Transfusion Recipients

The PROTON study: PRofiles of blOod producT transfusiON recipients in the Netherlands

Barbara A. Borkent-Raven
Mart P. Janssen
Cees L. van der Poel
Wim P. Schaasberg
Gouke J. Bonsel
Ben A. van Hout

Vox Sanguinis 2010; 99(1):54-64
DOI: 10.1111/j.1423-0410.2010.01312.x

CHAPTER 1

“Not everything that can be counted counts and not everything that counts can be counted.”

Albert Einstein

Background

Transfusion recipient data are needed for correct estimation of cost-effectiveness in terms of recipient outcomes after transfusion. Also, such data are essential for monitoring blood use, estimation of future blood use and benchmarking.

Study design and methods

A sample of 20 of 93 Dutch hospitals was selected. Datasets containing all blood product transfusions between 1996 and 2006 were extracted from hospital blood bank computer systems, containing transfusion date, blood product type and recipient characteristics such as gender, address, date of birth. The datasets were appended and matched to national hospitalization datasets including primary discharge diagnoses (ICD-9). Using these data, we estimated distributions of blood recipient characteristics in the Netherlands.

Results

The dataset contains information on 290,043 patients who received 2,405,012 blood products (1,720,075 RBC, 443,697 FFP, 241,240 PLT) from 1996 to 2006. This is 28% of total blood use in the Netherlands during this period. Comparable diagnosis and age distributions of all hospitalizations indicate included hospitals to be representative, per hospital category, for the Netherlands. Of all red blood cells (RBC), fresh-frozen plasma (FFP) and platelets (PLT), respectively 1.7%, 2.5% and 4.5% were transfused to neonates. Recipients of 65 years or older received 57.6% of RBC, 41.4% of FFP and 29.0% of PLT. Most of the blood products were transfused to patients with diseases of the circulary system (25.1%) or neoplasms (22.0%).

Conclusions

Transfusion data from a limited sample of hospitals can be used to estimate national distributions of blood recipient characteristics.

Introduction

Quantitative information on the fate of blood products issued (BPI) to hospitals is needed to analyse the cost-effectiveness of blood safety interventions such as nucleic acid amplification testing (NAT), leucocyte depletion or pathogen reduction. To know 'where the blood goes' is necessary for estimating the beneficial effects of safety interventions in terms of health gain in the recipients. Information on blood donors, donations and BPI is carefully registered by blood establishments in many countries, as in the Netherlands. Additionally, hospitals must register personal and clinical information on recipients of the blood products, in line with EU regulations. In this article, we report on the PRofiles of blOod producT transfusiON recipients (PROTON) study that resulted in the first national dataset including characteristics of blood product recipients in the Netherlands. Apart from restricting privacy regulations on matching individualized datasets, the absence of a national personal identification number in health care datasets (as in Scandinavian countries) is prohibitive.[1] We show alternative methods to match data. The resulting dataset enables answering several relevant questions on today's blood supply and its future. One application is to assess the effects of envisioned or already implemented blood safety measures. Economic evaluations, such as cost-effectiveness analyses, are increasingly considered mandatory by governmental and regulatory bodies.[2,3] Costs can be calculated from blood establishment data, but the effects of avoiding transfusion complications depend on age, morbidity and survival of the blood product recipients.[4] In this article, the current distribution of blood products over age, gender and diagnosis of recipients is estimated. These data can be directly incorporated in models for cost-effectiveness analysis of blood safety measures.

Methods

Definitions

For the aim of this study, we define a blood product transfusion (BPT) as the event at which one blood product is actually transfused. Pooled platelet transfusions (from five buffy coats) as well as split product (neonatal) transfusions are counted as one BPT. The blood product recipient profile consists of recipient gender, age at time of the BPT and primary discharge diagnosis code (ICD-9) of the hospitalization during which the BPT took place.

Primary data collection

From included hospitals, approval and commitment of the boards and professionals were obtained, as well as approval of Medical Ethical Committees. The sampled hospitals (for sampling procedure, see below) extracted microdata on all BPTs between January 1st 1996 and December 31st 2006 from their hospital blood bank computer systems, as far as data were available. All types of blood products were included, both apheresis and whole-blood derived. Participating hospitals were requested to provide the following data: type of blood product (RBC, FFP or PLT), date of transfusion, recipient gender and date of birth and address of the recipient.

Discharge diagnosis (ICD-9) microdata were obtained from the National Medical Registry (in Dutch: Landelijke Medische Registratie) that contains records of all hospitalizations in the Netherlands from 1996 to 2005 (coverage 99% in 1998). These data are stored at Statistics Netherlands (in Dutch: Centraal Bureau voor de Statistiek), the national governmental bureau of statistics of the Netherlands. The hospitalization data include unique person identification numbers. Full personal identification data of all Dutch citizens are also stored at Statistics Netherlands, including personal identification number, postal code, house number, gender and date of birth.

Sanquin, the only national public blood component provider, provided the total annual numbers of BPIs for the years 1996 to 2006, categorized into RBC, FFP and PLT, and the annual number of BPIs for specific hospitals.

Hospital selection

Given the time and budget constraints of the study, it was decided to base the national estimates of recipient profiles on a sample of hospitals rather than all 93 hospitals in the Netherlands. A hospital sample covering about 30% of the total blood use in the Netherlands was aimed at. For all hospitals, the number of hospital beds in 2006 was known.[5] The number of BPI (categorized into RBC, FFP and PLT) in 2006 was provided by Sanquin. Analysing the average blood use per hospital bed revealed significant differences between three principal hospital categories: academic, general and specialized cancer hospitals. From these categories, we drew a random sample: five out of eight academic hospitals and 14 out of 85 general hospitals. As there are only two specialized cancer hospitals, we aimed at including both these cancer hospitals in the study. Selected hospitals were invited to provide data for the study. If a hospital appeared to be unable to provide the requested data, another hospital from the same category was selected at random.

Dataset matching procedure

BPT data obtained from hospital blood bank laboratories were encrypted within the hospital and personally transported to Statistics Netherlands, to ensure patient privacy protection. In a protected environment, the data were decrypted, matched and analysed, following the strict privacy regulations in the Netherlands. Data needed conversion into a general format, as a result of differences in the various laboratory computer systems that are used in the hospitals. Next, a two-stage matching procedure was performed. First, the BPT data were matched to personal identification numbers by postal code, house number, gender and date of birth. Second, BPTs were matched to a hospitalization with a diagnosis code (ICD-9), if the personal identification number matched and the date of transfusion was between the hospital intake and discharge dates. Figure 1 shows the scheme of matching the datasets. Some BPTs were not matched to a diagnosis, but an earlier BPT in the dataset of the same recipient was available. In such cases, it was assumed that the latter BPT was given during a hospitalization with the same diagnosis as the previous one.

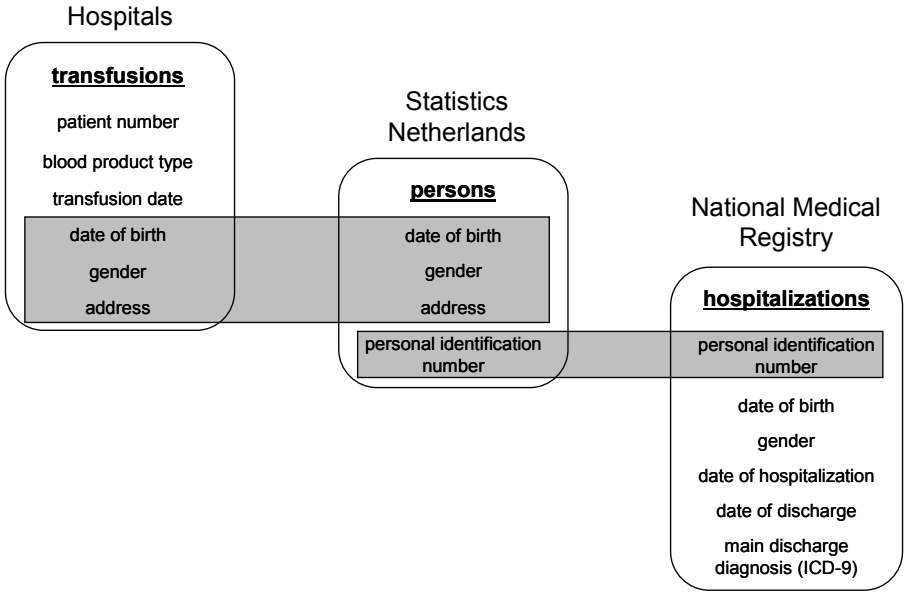


Figure 1: The PROTON recipients approach of matching existing datasets
The two matching steps are marked by the grey rectangles. Blood product transfusion datasets are downloaded from hospital bloodbank information systems, encrypted and brought to Statistics Netherlands, the national governmental statistics office that includes governance of the National Medical Registry.

Validation of sampling and matching procedures

Nationwide hospitalization data from the National Medical Registry were used to compare the distribution of patient age and diagnosis (ICD-9) of all hospitalizations (whether or not transfused) in the included hospitals and in the remaining hospitals, categorized into general and academical hospitals. When comparing these distributions, hospitalizations were counted as 'included' from the date that the transfusions of the hospital were included in the PROTON dataset. Specialized cancer hospitals were not taken into consideration here. Diagnosis (ICD-9) and age distributions of all hospitalizations in the Netherlands were compared, to evaluate whether the patient populations in the included academic and general hospital samples respectively can be considered representative for all academic and general hospitals in the Netherlands. Distributions of age, gender and type of blood product in completely matched records were compared to distributions of incomplete records to investigate whether unmatched records caused any bias.

Data analysis

To describe the profile of Dutch BPT recipients, we extrapolated the data from the included hospitals to the national level. We created weights for each BPT in the PROTON dataset by dividing the total numbers of BPIs (provided by Sanquin) per hospital category, product type and year by the corresponding numbers of BPTs in the PROTON dataset. As both cancer hospitals were included, the BPTs in those hospitals were weighted as 1. All distributions shown in this article are created using these weights.

The weights were used to estimate distributions of number of BPTs per recipient per blood product type. Furthermore, weighted counts were made of BPTs over age and gender, main diagnosis groups and main diagnosis groups according to age, all

categorized per blood product type. To improve readability of the plots concerning the relationship between recipient age and diagnosis, the graphs were smoothed using a kernel density estimator.[6]

Software

Random numbers for hospital sampling were generated using Excel (version 2003; Microsoft Corporation, Redmond, WA, USA). Data matching with the hospitalizations in the National Medical Registry was performed using SAS Enterprise Guide (version 4.1; SAS Institute Inc., Cary, NC, USA). All other data management and analysis was performed using STATA / SE (version 9.2 for Windows; StataCorp LP, College Station, TX, USA). Graphs were created using R (version 2.8.1, 2008, The R Foundation for Statistical Computing, Vienna, Austria).

Results

Collected data

Initially, 20 selected hospitals were invited to participate in the PROTON study. Of these hospitals, seven could not participate because of hospital staff time constraints and / or limitations of the computer systems. Every time a hospital appeared to be unable to join the study, another hospital from the same category was selected at random and invited. In total, five of eight academic hospitals, 14 of 85 general hospitals and two of two cancer hospitals provided data on blood products transfused between 1996 and 2006. One cancer hospital became part of an included academic hospital because of merger. Some hospitals were unable to extract data covering the whole study period, mostly as a result of IT changes. The resulting coverage by the PROTON dataset of all BPI from Sanquin is given in Table 1. In the dataset of one hospital, no distinction could be made between FFP and PLT, so only its RBC data were included. In addition to the nationwide ICD-9 diagnoses from the National Medical Registry, six hospitals provided their own diagnosis data by ICD-9 code.

The total PROTON dataset contains information on 290,043 patients who received 2,405,012 blood products (1,720,075 RBC, 443,697 FFP, 241,240 PLT) during the study period 1996–2006. For the whole study period, 28% of the total BPI in the Netherlands is covered by the PROTON dataset. Data from the National Medical Registry concerning hospitalizations in 2006 were not yet available. Of the BPTs in the PROTON dataset over the period 1996–2005, 87% could be matched to a diagnosis. Table 2 gives the matching percentages to diagnoses and the retrospective years of follow-up allowed by data provided by the included hospitals.

Table 1: Coverage of blood product transfusions (BPTs) in the PROTON dataset as compared to annual blood products issued (BPI) by Sanquin[a]

Year	# BPTs in PROTON database			# BPI by Sanquin in the Netherlands			Fraction of BPI covered by PROTON dataset		
	RBC	FFP	PLT	RBC	FFP	PLT	RBC	FFP	PLT
1996	141,675	37,204	18,251	715,366	106,972	46,519	20	35	39
1997	148,859	41,328	20,106	718,785	121,316	48,449	21	34	41
1998	159,075	47,397	22,124	713,896	111,215	49,242	22	43	45
1999	151,129	40,595	20,665	680,000	105,000	47,600	22	39	43
2000	161,466	40,883	21,317	635,731	99,576	42,797	25	41	50
2001	154,326	41,399	22,217	602,098	100,793	43,329	26	41	51
2002	161,212	42,527	23,380	626,661	104,683	43,167	26	41	54
2003	169,129	42,834	23,843	617,015	111,600	47,620	27	38	50
2004	170,229	39,341	24,007	595,090	97,200	52,680	29	40	46
2005	157,641	38,102	23,008	569,879	93,838	47,831	28	41	48
2006	145,334	32,087	22,322	556,509	92,380	51,869	26	35	43
Total	**1,720,075**	**443,697**	**241,240**	**7,031,030**	**1,144,573**	**521,103**	**24**	**39**	**46**

[a]An unknown fraction of the BPIs in the Netherlands is not transfused to patients, among others because of outdating.

Table 2: Hospital data and matching to ICD-9 primary discharge diagnosis of blood product transfusions (BPTs) in the PROTON dataset

Hospital number	Category	Years of follow-up	# patients in dataset in 1996-2006	# BPT in dataset in 1996-2006	mean # BPT per recipient in 1996-2006	% BPT matched to diagnosis in 1996-2005
1	cancer	11.0	6,651	55,020	8.3	87
2	academic/cancer	11.0	39,935	414,542	10.4	79
3	academic	11.0	30,291	289,695	9.6	92
4	academic	11.0	28,342	285,189	10.1	84
5	academic	11.0	27,514	333,606	12.1	99
6	academic	9.3	21,898	174,562	8.0	98
subtotal	*academic/cancer*		*154,631*	*1,552,614*	*10.0*	*88*
7	general	11.0	18,519	135,689	7.3	88
8	general	8.2	14,409	90,215	6.3	89
9	general	11.0	13,061	72,355	5.5	92
10	general	8.7	12,233	90,731	7.4	75
11	general	11.0	13,485	80,873	6.0	81
12	general	11.0	22,644	158,985	7.0	83
13	general	11.0	6,601	31,510	4.8	76
14	general	11.0	7,667	46,605	6.1	84
15	general	2.5	1,743	7,403	4.2	32
16	general	11.0	10,864	61,171	5.6	90
17	general	7.9	4,349	23,115	5.3	91
18	general	11.0	6,661	37,482	5.6	99
19	general	4.6	2,458	12,393	5.0	70
20	general	1.7	718	3,871	5.4	86
subtotal	*general*		*135,412*	*852,398*	*6.3*	*83*
Total			**290,043**	**2,405,012**	**8.3**	**87**

Validation of sampling and matching procedures

Overall patient age and diagnosis distributions (with or without transfusions) in the included academic and general hospitals were compared to the distributions in the remaining academic and general hospitals of the Netherlands. Data were derived directly from the National Medical Registry, including all 25 million hospitalizations of patients in the observation period. The mean age of patients was 49.1 in the included and 48.2 in the remaining general hospitals, and 44.3 in the included and 42.9 in the remaining academic hospitals. The distribution of patients over four age groups and seven largest diagnosis groups is given in Table 3. As the distributions are quite similar, the sample is regarded to be representative for all hospitals in the Netherlands, while distinguishing between academic and general hospitals. Distributions of age, gender and type of blood product in completely matched records were similar to the distributions in incomplete records. This indicates that matching was unbiased with respect to age, gender and type of blood product.

Table 3: Baseline characteristics of included (n = 20) and remaining (n = 73) hospitals in the Netherlands in 1996–2005

Data were derived directly from the National Medical Registry, including all 25 million hospitalizations of patients in the observation period, whether transfused or not.

Baseline characteristics	Academic hospitals		General hospitals	
	Distribution in PROTON (%)	Distribution in remaining hospitals (%)	Distribution in PROTON (%)	Distribution in remaining hospitals (%)
age				
0 years	6	7	6	6
1-16 years	13	15	8	10
17-40 years	22	22	21	21
41-64 years	34	33	31	30
65 years or older	25	24	34	33
*primary discharge diagnosis (ICD-9)**				
Neoplasms	19	20	12	11
circulatory system	16	17	15	14
nervous system and sense organs	10	11	11	11
musculoskeletal system and connective tissue	8	6	12	13
digestive system	8	7	9	9
injury and poisoning	7	9	6	7
pregnancy, childbirth and puerperium	6	7	8	8

***** Note that not all ICD discharge diagnosis codes are included here, and therefore the totals do not add up to 100%

Distributions of BPT dose

Figure 2 shows the annual transfusion dose for each product type per specific recipient. On average, blood recipients received 5.8 BPTs per calendar year. RBC recipients received 4.8 RBC units on average per calendar year. FFP recipients and PLT recipients received on average 5.2 units of FFP and 3.6 units of PLT respectively per calendar year.

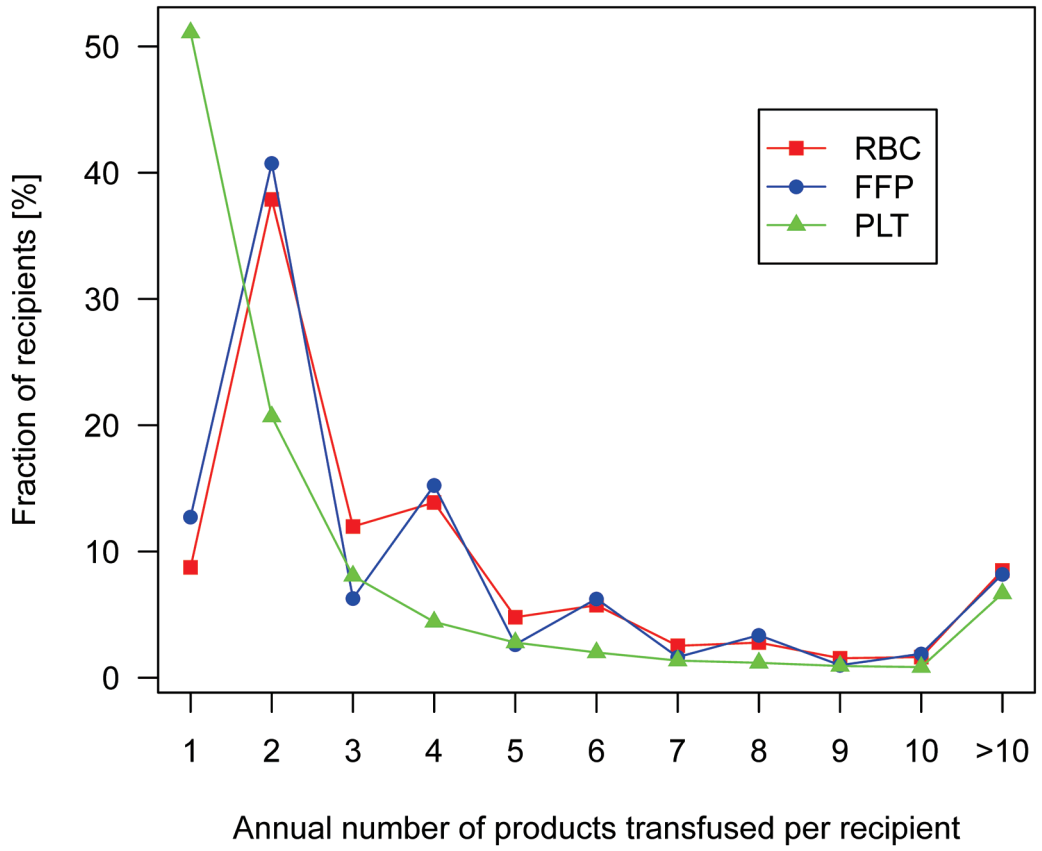


Figure 2: Distribution of annual transfusion dose per recipient for each blood product type in 1996–2006

Distributions of BPT characteristics

Figure 3 shows recipient age and gender distributions for the three blood product types. Of all RBC, 50% were transfused to men. Of all FFP and PLT, respectively 59% and 60% were transfused to men. Furthermore, 1.7% of RBC, 2.5% of FFP and 4.5% of PLT were transfused to children of age 0. Children between 1 and 16 years of age received 1.9% of RBC, 4.3% of FFP and 10.1% of PLT. Recipients between 17 and 40 years of age received 11.3% of RBC, 18.9% of FFP and 17.6% of PLT. Recipients between 41 and 64 years received 27.6% of RBC, 33.0% of FFP and 38.7% of PLT. Recipients of 65 years or older received 57.6% of RBC, 41.4% of FFP and 29.0% of PLT.

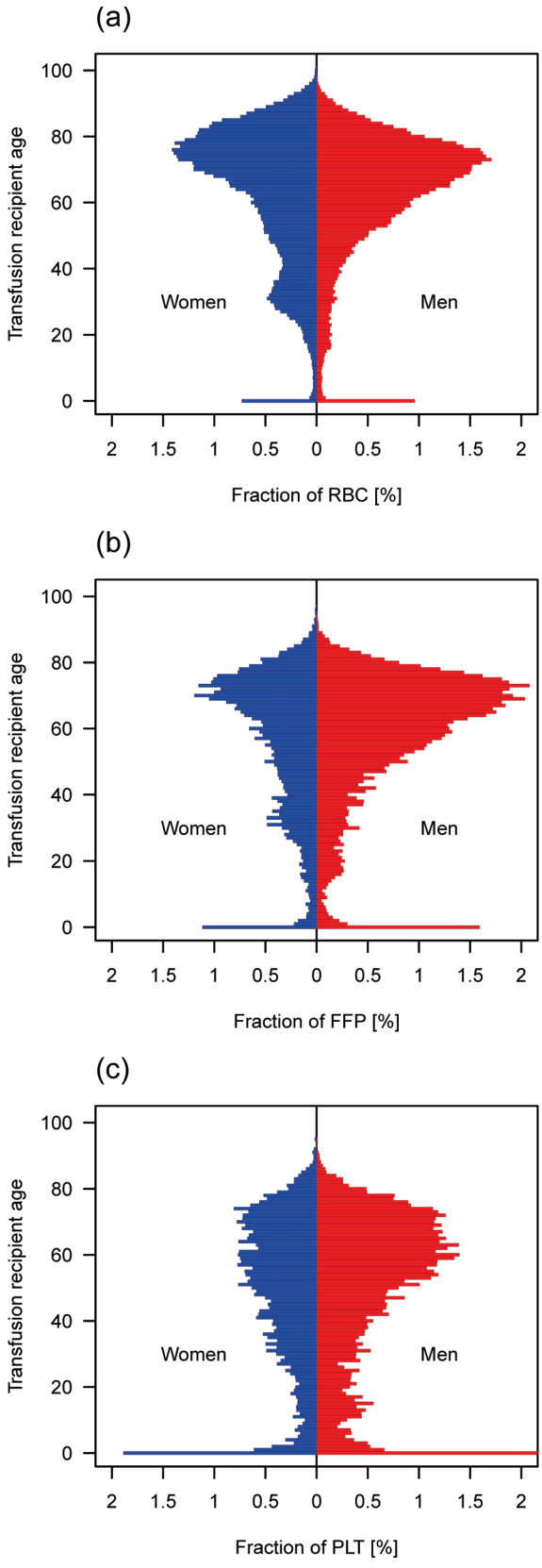


Figure 3: Age distributions of blood product recipients during the years 1996 to 2006
For the graph of FFP, recipients who received over 100 products were left out. The table from which these graphs were created can be obtained by contacting the authors. (a) Age distribution of RBC recipients. (b) Age distribution of FFP recipients. (c) Age distribution of PLT recipients.

In Table 4 the distributions of BPTs over the main diagnosis (ICD-9) groups are shown. For all three blood product types, the diagnosis group that used most blood products was the group of patients with diseases of the circulatory system (ICD-9 codes 390–459) or with neoplasms (ICD-9 codes 140–239), including haemato-oncological diseases. The ICD-9 diagnosis group called 'blood and blood-forming organs' includes patients with anaemia, coagulation defects or purpura.

Table 4: Diagnosis distributions of blood product transfusion recipients during the years 1996 to 2005

	Percentage transfused products			
ICD-9 discharge diagnosis, main group	**RBC**	**FFP**	**PLT**	**Total**
circulatory system	21.5	47.8	21.0	25.1
neoplasms	22.2	12.7	41.2	22.0
injury and poisoning	10.5	9.4	4.3	9.9
digestive system	9.8	9.0	4.4	9.4
blood and blood-forming organs	8.6	4.8	8.7	8.0
musculoskeletal system and connective tissue	5.3	1.2	0.6	4.4
genitourinary system	3.8	2.2	1.0	3.4
pregnancy, childbirth and puerperium	3.7	2.3	1.4	3.3
symptoms, signs and ill-defined conditions	3.3	2.2	3.5	3.1
respiratory system	2.4	1.1	1.5	2.1
infectious diseases	1.3	2.1	2.4	1.5
endocrine, nutritional, metabolic, immunity	1.1	0.7	1.0	1.1
congenital anomalies	0.8	2.1	2.0	1.0
perinatal period	0.9	0.6	1.9	0.9
nervous system and sense organs	0.3	0.6	0.5	0.4
skin and subcutaneous tissue	0.4	0.1	0.1	0.3
mental disorders	0.1	0.1	0.0	0.1
other	4.2	1.2	4.5	3.8

Figure 4 shows the relationship between age and ICD-9 diagnosis of recipients of RBC, FFP and PLT respectively. In all categories, most blood products were transfused to elderly patients with circulatory system diseases or with neoplasms. It appears that PLT recipients, mostly allocated to neoplasms (including haemato-oncological diseases), are more equally distributed over age. Elderly men received more blood products than women of the same age, especially far more plasma products were transfused to men than to women. This is mainly caused by men having a higher probability of being hospitalized for circulatory diseases (in the National Medical Registry, they have 1.5 times more hospitalizations than women). On the other hand, 72% of the RBC with a recipient age between 25 and 35 years was transfused to women, mainly related to childbirth. Detailed data on the age and diagnosis distribution of RBC, FFP and PLT are given in the Appendix.

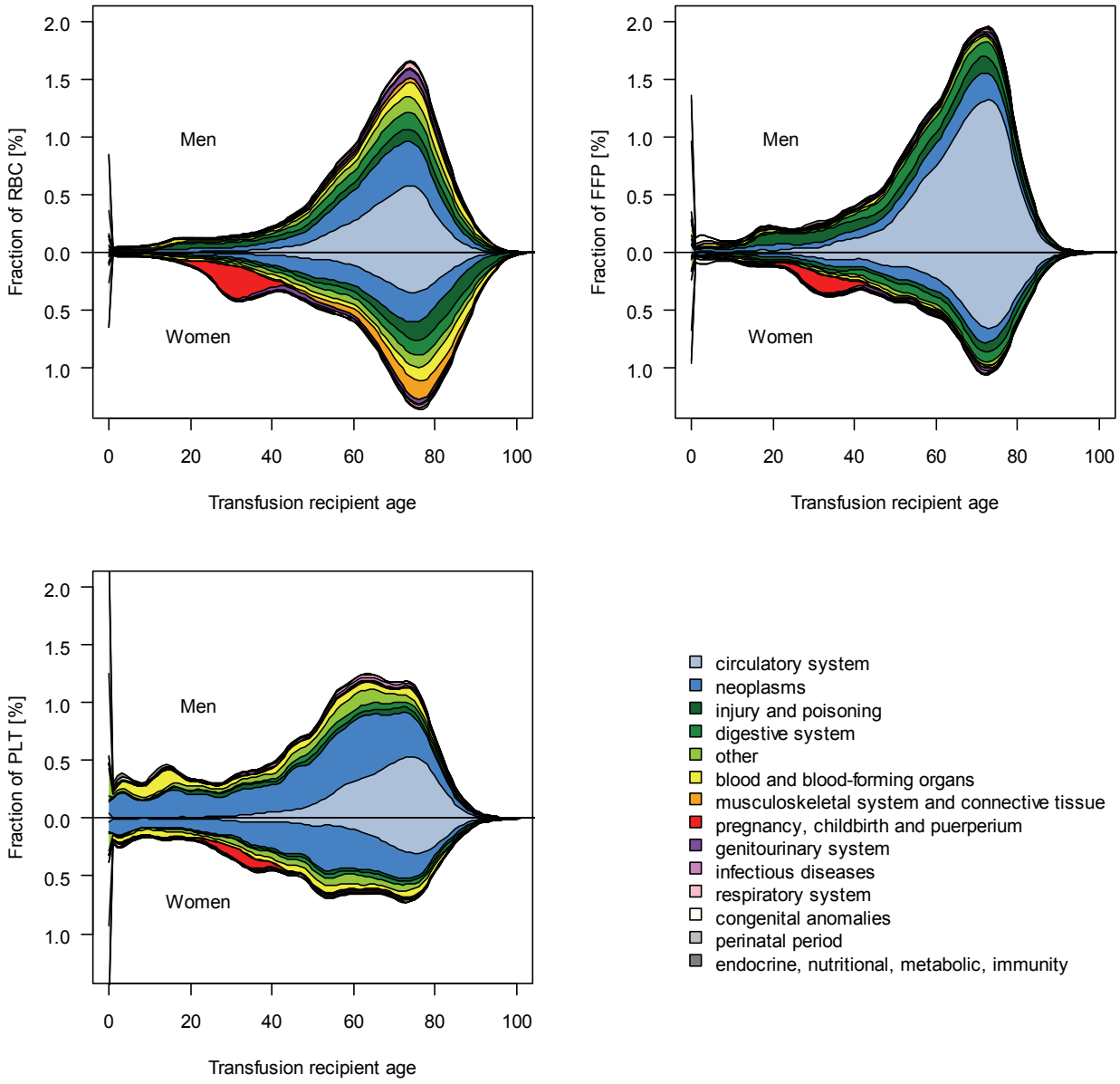


Figure 4: The relationship between age and diagnosis of various blood components
The graphs are smoothed to improve readability. (a) Age and diagnosis of RBC recipients. (b) Age and diagnosis of FFP recipients. (c) Age and diagnosis of PLT recipients.

Discussion

Applications of the PROTON dataset

We describe a method for collating nationally representative data on the distribution of BPT to recipients, including recipient diagnosis, age and gender. The dataset was created without the presence of a nationally governed personal identification number system for all inhabitants in the Netherlands, as is the case in Scandinavia.[1] BPT data are required for the correct analyses of cost-effectiveness of blood safety measures in terms of recipient outcomes after BPT. For such analyses, survival of blood product recipients after transfusion is essential. This survival can now be studied by matching the PROTON dataset to national mortality data as governed by Statistics Netherlands. The PROTON data are also useful for monitoring blood use and, in combination with demographic data, for estimating future blood use. With an ageing population, it is important to predict future need of blood products to enable anticipatory strategies in donor recruitment, as most blood is given to elder patients. Modelling scenario's on possible future medical or policy changes towards 'optimal blood use' can further refine such forecasts. In addition, participating hospitals can use the PROTON data for comparisons with their own data for benchmarking purposes on optimal blood use.

Considerations on the results

RBC (in the Netherlands 270–290 ml) and FFP (in the Netherlands 300 ml) are most often transfused in doses of 2 or 4 or 6 units. Apparently, clinicians often consider one unit to be insufficient as an adult therapeutic dose. As compared to platelets, where one adult therapeutic dose represents five whole-blood donations (buffy-coat method), RBC and FFP units are derived from one whole-blood donation (500 ml). One whole-blood donation, representing about 10–13% of total blood volume of the donor, may therefore not yield enough substitution in case of a bleeding or anaemic patient.

A considerable fraction of blood products (2%) are transfused to neonates. This group yields the highest positive effects of blood safety interventions, in terms of life years gained, because of their relatively long life expectancy. Thus, transfusion recipients with a very good prognosis positively influence the cost-effectiveness of new blood safety measures.[4]

Comparison with other studies

The only comparable dataset that contains more BPT data than ours is the Scandinavian Donations and Transfusions database (SCANDAT) dataset, which includes 11.7 million transfusions given between 1966 and 2002 in Sweden and Denmark.[1] The researchers of SCANDAT published an article about post-transfusion survival rates, but not about the distribution of blood products over age or diagnosis of transfusion recipients.[7,8] It would be interesting to compare the distributions described in this article with those from SCANDAT, especially because there is a large difference in RBC use: 50.5 RBC were used per 1000 inhabitants in Sweden in 2004 and 72.9 per 1000 in Denmark (2000–2002) against 36.6 per 1000 inhabitants in the Netherlands in 2004.[8,9]

Other studies on recipient distributions resulted in smaller and more regional datasets than ours. Regan shows age distributions of recipients in five hospitals in Oxford, UK, in which recipients of FFP and PLT are relatively older than in the Netherlands.[10] Greinacher et al. shows the age distribution of RBC recipients in a region of Germany.[11] There, 52% of RBC is transfused to patients of 60–74 years of age, against 34% of RBC transfused in the Netherlands, while 16% of RBC is transfused to recipients above 74 years of age against 32% in the Netherlands. Cobain et al. published a review in which recipient data from four countries are shown: England, USA, Australia and Denmark.[12] In this study, age distributions are reported for RBC, FFP and PLT transfused in the North of England, where the FFP recipients are older than those in our study. Blood use in the USA is summarized for three age groups. The FFP recipients and PLT recipients in our study appear to be younger than FFP and PLT recipients in the USA during 1989–1992. Cobain et al. also show three age groups for Western Australia, and these distributions are similar to ours. Only the FFP recipients in the Netherlands are somewhat younger. Age and diagnosis distributions are reported for a county in Denmark. The Danish RBC and PLT recipients are relatively old compared to ours. Apparently, age distributions of blood product recipients vary across countries. Comparing the diagnoses, in the Netherlands a smaller fraction of RBC and FFP is transfused to patients with diseases of the digestive system: 10% of RBC and 9% of FFP against 16% and 21% respectively in Denmark. Further, it is reported that 66% of the PLT are transfused to patients with neoplasms, while in the Netherlands only 41% of the PLT are transfused to that patient group. On the other hand, 9% of the PLT in the Netherlands are transfused to patients with diseases of blood or blood-forming organs against 5% of the PLT in Denmark.

These comparisons show that there are considerable differences in the profile of blood product recipients in different countries, probably reflecting differences in clinical practice. The PROTON and SCANDAT datasets may therefore not be representative for other countries, as the clinical blood use and the allocation of BPT to different patient categories may differ considerably, as is also illustrated with the Council of Europe survey data. Comparing data from different countries in a benchmark could highlight potential areas where further optimization of blood use might be possible. During our collaboration with the 20 participating hospitals, it appeared that they appreciate to compare the PROTON data to their own for benchmarking purposes towards optimal blood use.

Conclusion

The PROTON approach provides a practicable way to obtain national data without nationwide coverage, using a sample of hospitals to estimate national distributions of blood recipient characteristics. Using these distributions will improve future cost-effectiveness analyses of blood safety interventions in the Netherlands, as their effects are achieved in recipients of BPTs.

Acknowledgements

We acknowledge the Netherlands Organization for Health Research and Development (ZonMW project 945-06-603) for funding this research. We thank all hospitals that participated in the PROTON study: Academisch Medisch Centrum, Amsterdam; Antoni van Leeuwenhoekziekenhuis (NKI), Amsterdam; Tergooiziekenhuizen, location Blaricum; IJssellandziekenhuis, Capelle aan de IJssel; Hagaziekenhuis, locatie Leyenburg, Den Haag; Van Weel-Bethesdaziekenhuis, Dirksland; Albert Schweitzerziekenhuis, Dordrecht; Nij Smellinghe, Drachten; De Gelderse Vallei, Ede; Leids Universitair Medisch Centrum, Leiden; Academisch Ziekenhuis Maastricht, Maastricht; Sint-Antoniusziekenhuis, Nieuwegein; Waterlandziekenhuis, Purmerend; Erasmus Medisch Centrum, Rotterdam; Sint-Franciscusgasthuis, Rotterdam; Ruwaard-van Puttenziekenhuis, Spijkenisse; Sint-Elisabethziekenhuis, Tilburg; Tweestedenziekenhuis, Tilburg/Waalwijk; Universitair Medisch Centrum Utrecht, Utrecht; Isalaklinieken, Zwolle. We thank Edie de Bruin, Fred Gast, and all other people we worked with at Statistics Netherlands for matching the transfusion data and for enabling us to work with the matched datasets through remote access.

References

1. 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.
2. Farrugia A. *The mantra of blood safety: time for a new tune?* Vox Sang 2004;86:1-7.
3. Custer B, Hoch JS. *Cost-effectiveness analysis: what it really means for transfusion medicine decision making*. Transfus Med Rev 2009;23:1-12.
4. Borkent-Raven BA, Janssen MP, van der Poel CL et al. *Cost-effectiveness of additional hepatitis B virus nucleic acid testing of individual donations or minipools of six donations in the Netherlands*. Transfusion 2009;49:311-9.
5. Deuning CM. *Beddencapaciteit ziekenhuis 2006, in: Volksgezondheid Toekomst Verkenning, Nationale Atlas Volksgezondheid*. 19-12-2006. RIVM, Bilthoven.
6. Parzen E. *On Estimation of a Probability Density Function and Mode*. The Annals of Mathematical Statistics, 1962;33:1065-76.
7. Kamper-Jorgensen M, Ahlgren M, Rostgaard K et al. *Survival after blood transfusion*. Transfusion 2008;48:2577-84.
8. Kamper-Jorgensen M, Edgren G, Rostgaard K et al. *Blood transfusion exposure in Denmark and Sweden*. Transfusion 2009;49:888-94.
9. van der Poel CL, Janssen MP, and Borkent-Raven BA. *Report on the collection, testing and use of blood and blood components in Europe in 2004;* *http://www.edqm.eu/medias/fichiers/NEW_2004_Report_on_the_.pdf*. 2007. Council of Europe.
10. Regan F. *New donors, new recipients--who gets what in 2003?* Transfus Clin Biol 2004;11:53-6.
11. Greinacher A, Fendrich K, Alpen U et al. *Impact of demographic changes on the blood supply: Mecklenburg-West Pomerania as a model region for Europe*. Transfusion 2007;47:395-401.
12. Cobain TJ, Vamvakas EC, Wells A et al. *A survey of the demographics of blood use*. Transfus Med 2007;17:1-15.

Appendix

Table A: Distribution of RBC over age and diagnosis of recipients

Description	Age												
	0	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	>=91	Total
neoplasms	0.0	0.2	0.3	0.5	1.0	2.1	3.9	5.7	5.9	2.3	0.2	22.2
circulatory system	0.0	0.0	0.0	0.2	0.4	1.1	2.9	6.4	8.3	2.1	0.1	21.5
injury and poisoning	0.0	0.1	0.4	0.6	0.6	0.7	1.0	1.7	2.7	2.2	0.5	10.5
digestive system	0.0	0.0	0.1	0.2	0.5	0.9	1.5	2.0	2.7	1.6	0.2	9.8
blood and blood-forming organs	0.0	0.2	0.4	0.3	0.4	0.6	0.8	1.2	2.5	1.9	0.3	8.6
musculoskeletal system and connective tissue	0.0	0.0	0.1	0.1	0.1	0.3	0.7	1.3	1.9	0.7	0.0	5.3
other codes	0.1	0.1	0.1	0.1	0.2	0.3	0.6	0.9	1.1	0.5	0.1	4.2
genitourinary system	0.0	0.0	0.0	0.1	0.3	0.5	0.5	0.8	1.0	0.4	0.0	3.8
pregnancy, childbirth and puerperium	0.0	0.0	0.1	1.6	1.8	0.1	0.0	0.0	0.0	0.0	0.0	3.7
symptoms, signs and ill-defined conditions	0.0	0.0	0.0	0.1	0.1	0.3	0.5	0.8	0.9	0.4	0.0	3.3
respiratory system	0.0	0.0	0.0	0.0	0.1	0.2	0.4	0.5	0.7	0.3	0.0	2.4
infectious diseases	0.0	0.0	0.0	0.1	0.1	0.2	0.2	0.3	0.2	0.1	0.0	1.3
endocrine, nutritional, metabolic, immunity	0.0	0.0	0.0	0.0	0.1	0.1	0.2	0.2	0.3	0.2	0.0	1.1
perinatal period	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9
congenital anomalies	0.4	0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.8
skin and subcutaneous tissue	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.4
nervous system and sense organs	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.3
mental disorders	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
Total	1.5	1.0	1.9	4.0	5.9	7.4	13.2	22.0	28.6	12.9	1.5	100.0

NOTE: All values given in %

Table B: Distribution of FFP over age and diagnosis of recipients

Description	Age											
	0	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	>=91	Total
circulatory system	0.1	0.1	0.2	0.6	1.5	3.0	7.5	15.3	16.4	2.5	0.0	47.2
neoplasms	0.0	0.4	0.5	0.7	1.1	1.7	2.8	3.6	2.5	0.6	0.0	14.0
injury and poisoning	0.1	0.2	0.9	1.2	1.3	1.1	1.4	1.9	1.9	0.5	0.0	10.3
digestive system	0.0	0.1	0.2	0.2	0.6	1.7	2.2	2.2	2.0	0.7	0.1	9.9
congenital anomalies	1.1	0.6	0.2	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	2.4
infectious diseases	0.1	0.3	0.2	0.2	0.2	0.3	0.3	0.4	0.2	0.1	0.0	2.3
symptoms, signs and ill-defined conditions	0.1	0.1	0.0	0.1	0.1	0.3	0.3	0.6	0.4	0.2	0.0	2.2
genitourinary system	0.0	0.1	0.1	0.1	0.2	0.4	0.3	0.4	0.5	0.1	0.0	2.1
blood and blood-forming organs	0.0	0.3	0.3	0.1	0.3	0.5	0.2	0.2	0.2	0.0	0.0	2.1
pregnancy, childbirth and puerperium	0.0	0.0	0.0	0.8	1.2	0.0	0.0	0.0	0.0	0.0	0.0	2.0
respiratory system	0.0	0.0	0.0	0.1	0.1	0.1	0.3	0.2	0.3	0.1	0.0	1.2
musculoskeletal system and connective tissue	0.0	0.0	0.2	0.1	0.1	0.1	0.2	0.2	0.2	0.0	0.0	1.2
other codes	0.1	0.0	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.0	0.0	1.0
endocrine, nutritional, metabolic, immunity	0.1	0.0	0.0	0.2	0.1	0.1	0.1	0.0	0.1	0.0	0.0	0.7
perinatal period	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7
nervous system and sense organs	0.0	0.0	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.6
skin and subcutaneous tissue	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
mental disorders	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
Total	2.4	2.3	3.1	4.5	7.0	9.6	15.9	25.3	25.1	4.8	0.2	100.0

NOTE: All values given in %

Table C: Distribution of PLT over age and diagnosis of recipients

Description	Age											
	0.0	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	>=91	Total
neoplasms	0.2	3.2	3.0	3.1	4.4	6.3	9.0	7.5	4.0	0.5	0.0	41.2
circulatory system	0.1	0.1	0.1	0.3	0.7	1.4	3.5	6.5	7.2	1.0	0.0	21.0
blood and blood-forming organs	0.1	1.1	1.6	0.6	0.6	1.2	1.3	0.8	1.1	0.2	0.0	8.7
other codes	0.2	0.2	0.3	0.3	0.5	0.7	1.1	0.8	0.3	0.1	0.0	4.5
digestive system	0.1	0.1	0.1	0.1	0.3	0.8	0.9	0.9	0.7	0.3	0.0	4.4
injury and poisoning	0.0	0.1	0.3	0.4	0.5	0.6	0.6	0.7	0.7	0.2	0.0	4.3
symptoms, signs and ill-defined conditions	0.1	0.1	0.2	0.2	0.3	0.6	0.8	0.6	0.4	0.1	0.0	3.5
infectious diseases	0.1	0.2	0.2	0.2	0.3	0.3	0.4	0.4	0.2	0.0	0.0	2.4
congenital anomalies	1.3	0.4	0.1	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	2.0
perinatal period	1.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.9
respiratory system	0.0	0.1	0.1	0.1	0.1	0.2	0.4	0.3	0.2	0.0	0.0	1.5
pregnancy, childbirth and puerperium	0.0	0.0	0.0	0.5	0.8	0.0	0.0	0.0	0.0	0.0	0.0	1.4
endocrine, nutritional, metabolic, immunity	0.1	0.3	0.2	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	1.0
genitourinary system	0.0	0.0	0.0	0.0	0.1	0.1	0.2	0.2	0.2	0.1	0.0	1.0
musculoskeletal system and connective tissue	0.0	0.0	0.1	0.0	0.1	0.0	0.2	0.1	0.1	0.0	0.0	0.6
nervous system and sense organs	0.1	0.1	0.1	0.0	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.5
skin and subcutaneous tissue	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
mental disorders	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	4.4	6.1	6.3	6.1	9.0	12.4	18.7	19.1	15.3	2.7	0.1	100.0

NOTE: All values given in %

Estimating survival after transfusion for the cost-effectiveness evaluation of blood safety interventions

Mart P. Janssen
Barbara A. Borkent-Raven
Eric A. Cator
Cees L. van der Poel
Wim P. Schaasberg
Gouke J. Bonsel
Ben A. van Hout

Submitted for publication

CHAPTER

2

“The manipulation of statistical formulas is no substitute for knowing what one is doing.”

Hubert M. Blalock Jr.

Background

For some medical interventions, like blood transfusion, patients are exposed repeatedly to low risks of transmission of infectious diseases. Survival after transfusion (SAT) is required to assess the cost-effectiveness of safety interventions reducing these risks. SAT differs from normal survival analysis as patients usually obtain multiple transfusions and therefore patients' survival is counted multiple times: once for each transfusion. Because of this dependency the commonly used Kaplan-Meier (KM) method for survival estimation might not be appropriate. The aim of this study is to compare patient survival (after first transfusion) and SAT and to evaluate whether the KM estimator is suitable for estimating SAT.

Methods

Two different methods for SAT estimation are evaluated and applied to estimate SAT of red blood cells, platelets, plasma transfused to patients of the University Medical Center of Utrecht (UMCU). The methods used are direct estimation and estimation using the KM method. Transfusions given in the UMCU between 1995 and 2003 were collected allowing SAT estimation over 13.6 years. In addition, a simulation study was performed to compare the performance of the two methods.

Results

Both methods provide comparable estimates for SAT. However, as a result of a change in patient survival over time, differences in SAT for plasma were found when applying different methods. Also, confidence intervals obtained by the standard KM procedure deliver overly optimistic confidence intervals.

Conclusions

There is a marked difference between patient survival (after first transfusion) and SAT for each of the blood components considered. In general the KM method provides a correct estimate for SAT itself, but the confidence intervals obtained with standard procedures are incorrect.

Introduction

Blood safety has been a major issue in the 1980's and 1990's. Blood safety policy has largely been driven by liability issues and the wish to render blood products as safe as possible.[1] More recently, cost constraints triggered the wish for more rational decision making in blood safety. It may be argued that other health care can be financed alternatively using the same resources.[2,3] In the Netherlands the government strives towards optimal instead of maximum blood safety.[4] Such policy decisions require outcomes from cost-effectiveness analyses (CEA) of blood safety interventions. In such analyses the impact of adverse events on the transfusion recipients in terms of costs and effects are assessed and balanced against the costs and effects of the (to be) implemented safety intervention(s). The effect on the transfusion recipient is expressed in terms of (quality adjusted) life years lost. Therefore, transfusion recipient survival is critical when assessing CEA of blood safety interventions.

Patient survival is the probability of a patient being alive at a specified time since a marked event. There is a simple relation between a patients' survival probability and his life expectancy: the latter is the area below the survival probability curve. The higher the survival probability, the higher the life expectancy. In the blood transfusion literature patient survival is most commonly reported as the survival of a particular patient after his or her first transfusion (SFT). For CEAs however, the life years lost after any of the patients' transfusions are relevant. The life years lost due to an adverse – let's presume lethal – transfusion to a patient is equal to the life expectancy of that patient at the time of the transfusion. If that patient would receive multiple transfusions at different time points, the average number of life years lost (given that one of these transfusions would indeed be lethal) would be the average of the life expectancies of that particular patient at the times he received these transfusions. We define survival after transfusion (SAT) as the survival of a transfusion recipient since transfusion, without referring to any specific transfusion. This survival could also be interpreted as the survival of the transfused product instead of that of the patient. With SAT we can adequately describe the effect of adverse transfusion events on patients.

For the estimation of SAT we can use observed time intervals from transfusion until death (or censoring). When data is (right) censored, survival is most commonly estimated using the Kaplan-Meier (KM) product limit estimator.[5] Censoring occurs when follow-up of a patient is incomplete and it remains unknown whether a patient has died or not. However, this method requires observed time intervals considered to be independent.[6] The question that arises is whether the KM method can be applied to estimate SAT. This is questionable, as the observed time intervals often concern transfusions that were given to the same individuals, and are therefore connected by an identical endpoint (the time of patients' death or censoring).

To answer this question, we simulated a transfusion recipient population and applied two different methods for SAT estimation. In addition, we applied both methods to estimate SAT of transfusions given in the UMCU hospital between 1995 and 2003.

Methods

Estimating survival after transfusion (SAT)

Survival is commonly calculated using the product limit estimate proposed by Kaplan and Meier.[7] As discussed in the introduction, this method is invalid when there is a dependency between the observed follow-up times. For such circumstances, Kaplan and Meier propose direct estimation of the survival probability.[6] The survival probability is in that case is calculated as the fraction of observed deaths amongst all observations that might have survived until that time point. The advantage of the direct estimation (DE) method is that it provides an unbiased estimator of survival.[6] The drawback is that as a result of the diminishing amount of data used for estimating survival for longer survival times, its variance will generally be bigger than that of the KM estimator. Another drawback is that that the estimated survival will not necessarily be decreasing over time: one might find that is more likely to survive until time $t+\delta$ than it is to survive until time t. To overcome this problem the direct survival curve can be monotonized. This is a mathematical procedure that forces continuous decline of a function and can be shown to provide a better survival estimate.[8] Confidence intervals for the estimated survival probabilities are obtained by bootstrapping.[9] The DE confidence intervals are derived from the monotonized DE estimate. A formal description of both estimators is given in the Appendix A.

Simulation of a transfusion recipient cohort

A number of fictive cohorts of transfusion recipients with a known survival probability and transfusion regime were simulated. For such cohorts the theoretical SAT can be determined. This allows comparing accuracy and variation of both estimation methods. For instance, for a cohort of transfusion recipients with a SFT of $(1\text{-}t/T)^{\alpha}$ (where $t<=T$) which is subject to a constant transfusion rate per unit time, it can be derived that the SAT is equal to $(1\text{-}t/T)^{(1+1/\alpha)}$. Also, more complex cohorts were simulated where there was an associated between the transfusion intensity and patient survival to mimic transfusion practice more realistically. The simulations performed are described in detail in Appendix B.

Assessing SAT of blood components

All blood transfusions in the Netherlands are recorded such that identification of the donor who provided blood to any specific transfusion recipient is possible, and vice versa. This is a legal requirement in the Netherlands which allows notification of recipients who have in the past received blood from a donor that is found to be infected with an infectious disease. In such cases previous donations might also have transmitted infections. For each transfused product, component type, date of transfusion and a unique hospital patient identification number (PID) are recorded. The hospital registration system therefore allows retrieval of past transfusions. Transfusion recipient data was extracted from the hospital databases from the period January 1st 1991 to December 31st 2003. At the CBS (Statistics Netherlands) a death register of all Dutch citizens is available from 1995 onwards. Transfusion and death register data were matched using

the patients' date of birth, sex and address. This information allows unique identification of the patient (in most cases). This enables retrieval of either the date of death in case the patient is deceased, or a last date for follow-up in case the patient was alive at the time data retrieval (August 4, 2008). To ensure the privacy of the patients, all identifying information was encrypted before being transferred to the CBS. In addition, all analyses were performed in a secure environment and the results checked by CBS staff to make sure there was no risk of disclosure of individual patients' information.

SAT of various blood components admitted to patients in the UMCU was estimated with both the KM and the DE methods.

Software used

Data management was performed using SAS Enterprise Guide (Version 4.1, SAS, Cary, NC, USA; Statistical analyses were performed with Stata/SE (Version 9.2 for Windows, StataCorp LP, College Station, TX, USA); Derivation of the theoretical SAT was performed with Mathematica (Version 7, Wolfram Research, Champaign, IL, USA); Survival graphs were produced with Excel (Version 2002, Microsoft Corporation, Redmond, WA, USA); The simulations were performed in R (Version 2.10.2, The R Foundation for Statistical Computing, Vienna, Austria).

Results

UMCU Transfusion recipient population

From 1995 through 2003, 24,859 patients received 252,339 transfusions during 39,893 hospitalization periods. The numbers of components transfused in this period were 158,969 red blood cell (RBC) units, 60,153 fresh frozen plasma (FFP) and 33,217 platelet (PLT) units. Male patients received 58% of all components. The male-female distribution is about the same for RBC units (57% male), FFP units (58% male) and PLT units (60% male). A total of 23,115 (93%) patients were matched using sex, address and date of birth. By August 4, 2008, the date at which the deceased status was obtained from the death register, 10,348 of these patients were deceased. There were no dependencies on age, sex and number transfusions received between the matched and non-matched individuals. Both 93% of males and females were matched and both of matched and unmatched patients 52% were males.

Transfusion recipient survival

In Figure 1 both SAT and SFT is given for patients receiving RBCs, FFPs and PLTs respectively. Note that the SFT is provided for one year less than this SAT. This is because for the estimation of SFT we discarded patients that were transfused in 1995. As most patients receive all transfusions within one year, discarding transfusion recipients from this first year ensures that almost all remaining first transfusions in actual fact are first transfusions.

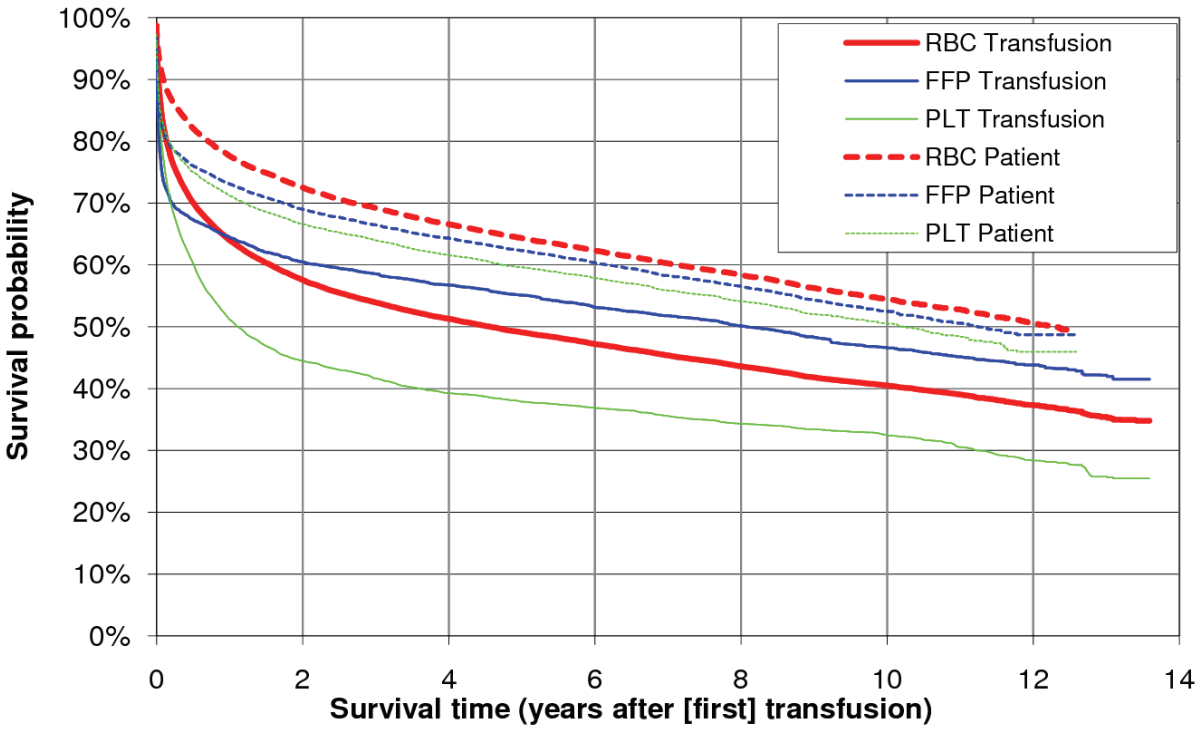


Figure 1: Patient survival after first transfusion (SFT, indicated as 'Patient') and survival after transfusion (SAT, indicated as 'Transfusion') per type of blood component.

SAT is less than SFT for all products. There is an average (over time declining) difference of about 15% in survival probability between SAT and SFT for RBCs, a difference around 7% for FFPs, and a difference around 20% for PLTs.

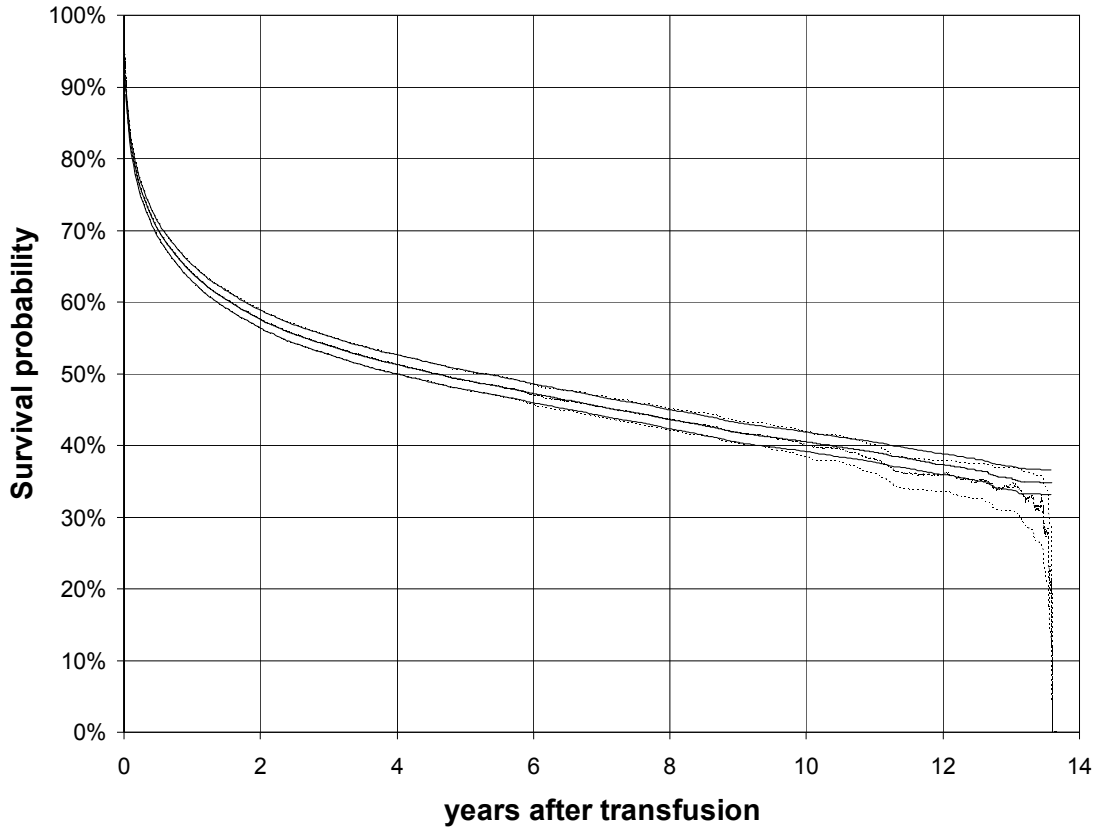
In Figure 2 SAT estimates for RBCs, FFPs and PLTs are given using KM and DE methods for survival estimation. The indicated 95% CIs are obtained by bootstrapping. Figure 2 shows that the KM and DE methods provide comparable estimates for the first six years of follow-up for all transfused components. For RBCs (Figure 2a) the results are almost identical for the first ten years follow-up (both the estimate itself and the CI). For the last four years follow-up, the estimates remain (almost completely) within each others 95% CIs.

For FFP the similarity between estimation methods remains up until 5 years follow-up. From thereon there is an increasing difference between both survival estimates, with a lower survival for the DE method. From twelve years follow-up onwards the 95% CIs are almost non-overlapping, suggesting a statistically significant difference in survival estimation. Another thing that becomes clearly visible here is the effect of the monotonizing procedure on the survival estimate just after 10 years follow-up: the monotonized estimate does not follow the inclination of the directly estimated survival probability.

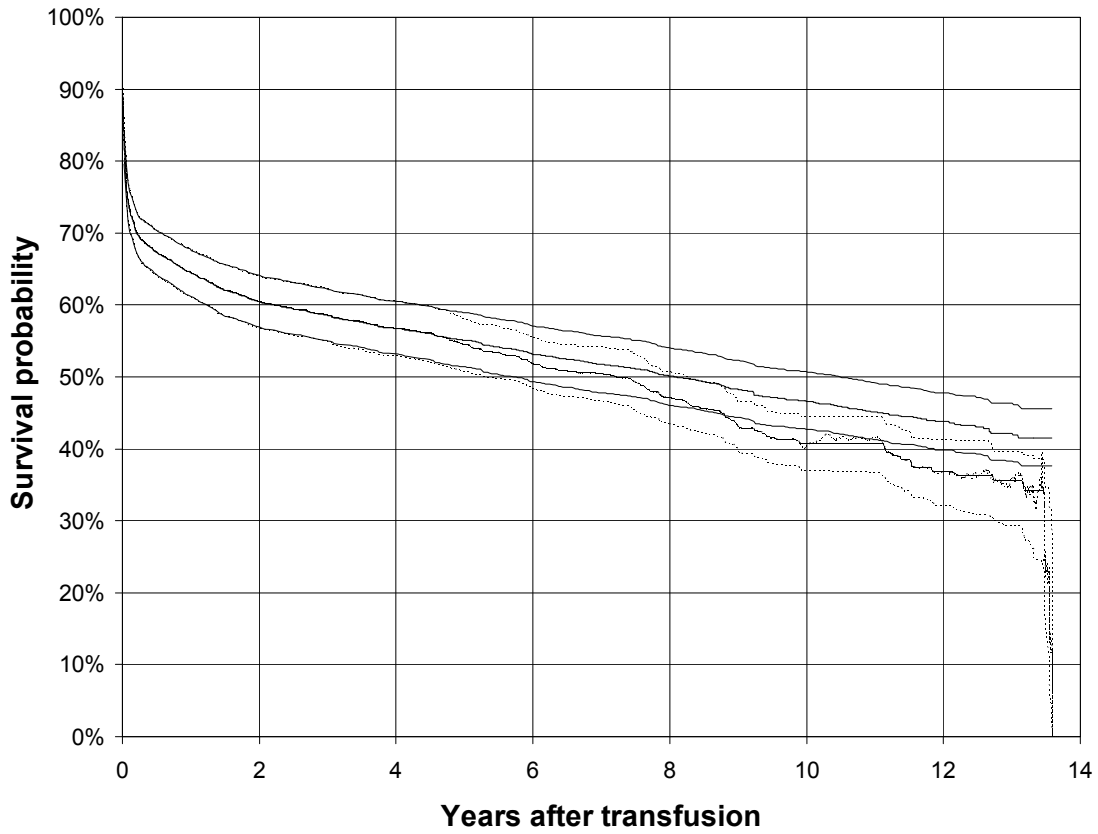
The survival estimates for PLT are again in good agreement for the first eight years of follow-up. From there on there is a small deviation, but the survival probability estimates remain in each other's 95% CIs. For the PLT estimates the difference between the raw and monotonized DE estimates becomes apparent for long-term follow-up.

All Figure 2 graphs clearly illustrate the larger CIs for long-term follow-up for the DE method as compared to the KM method.

(a)


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Survival probability
0
2
4
6
8
10
12
14
years after transfusion

(b)


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Survival probability
0
2
4
6
8
10
12
14
Years after transfusion

(c)

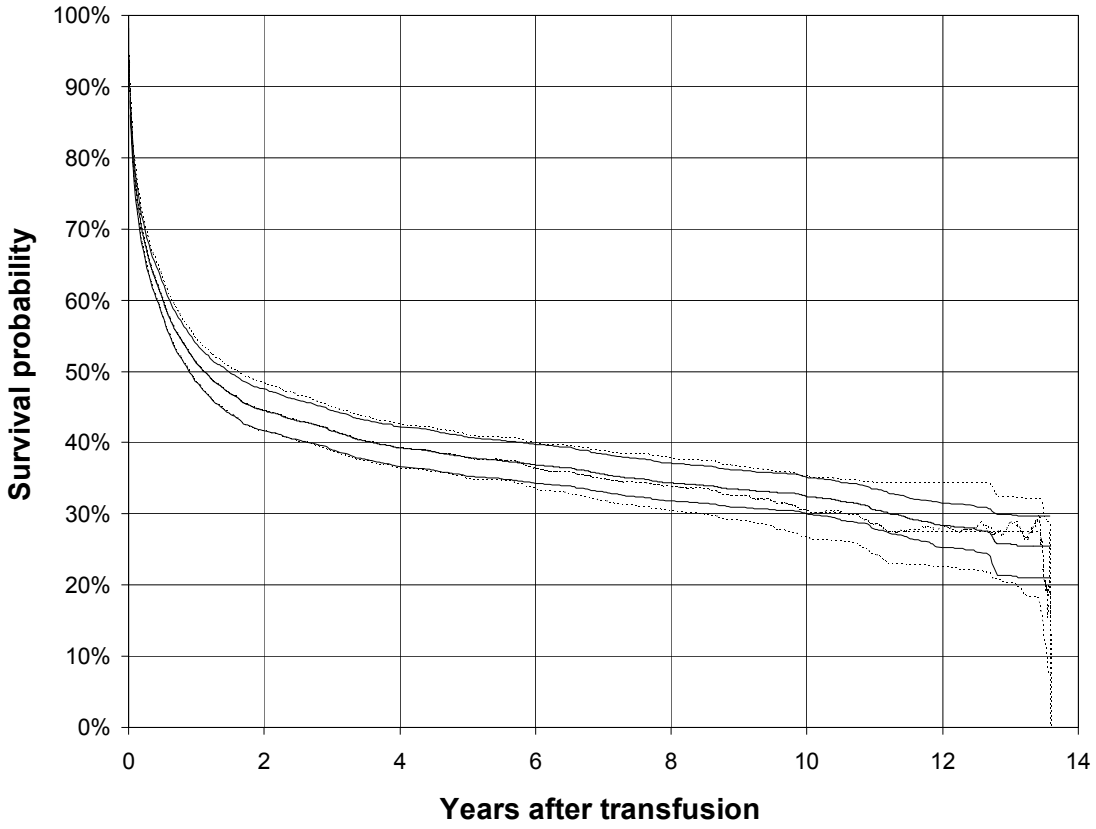


Figure 2: Survival after transfusion (SAT) of RBCs (a), FFPs (b) and PLTs (c) with 95% CI using the KM (—) and DE (---) methods. Both the raw and monotonized DE estimates are given.

Simulating transfusion recipient survival

Simulations were performed to study differences between the KM and DE estimates. In the simulations both patient survival and transfusion characteristics were varied. The simulations show that both estimators provide similar results even for a limited number of observations, and even in case of a dependency between patient survival and transfusion intensity. In case there are many observations both estimates provide near identical results. SAT confidence intervals were obtained by a bootstrapping procedure.[9] When bootstrapping observations, resampling of patients with their associated transfusions and not directly resampling of transfusions is required to obtain correct SAT CIs. This is required to maintain the dependence between transfusions given to a single patient. As standard KM CI estimation occurs without accounting for this dependence, such CI estimates will generally deliver overconfident results.

The simulation and results are described in Appendix B. The source code for the simulations is available from the authors upon request.

Discussion

There is a substantial difference between SFT and SAT for all types of blood components transfused. One would expect SAT to be less than SFT as all observed survival intervals of transfusions after the first transfusion will be less than that of the first transfusion. However, as most transfusions are given within one year after the first transfusion, this does not account for the eminent difference between SFT and SAT. From the literature it is known that SFT decreases with the number of components received.[10] As SAT can (roughly) be interpreted as survival weighted by the number of components received, it is clear that this effect will be the source of the difference between SAT and SFT. From Figure 1 it is clear that the difference between SFT and SAT is the strongest for PLTs, and the weakest for FFPs. This indicates that the association between the number of transfusions and death rate is the strongest for PLTs and the weakest for FFPs.

Survival analysis is used for a wide range of problems with many different applications. In most cases when repetitive treatment is involved one is interested in the effect of treatment. In such cases a multi-state modeling approach is appropriate.[11] However, in our specific application we are not interested in the effectiveness of treatment but purely in the survival of the patient after (any given) transfusion. Various survival models can be used to obtain different transfusion survival characteristics. For instance, an analysis of patient survival after the first, second, third, fourth, etcetera (n^{th}) transfusion. This information in combination with for instance age and clinical indication could be used for establishing a prognosis of patient survival. A description of such modeling exercise is outside the scope of this paper.

The results show that both the KM and DE estimation methods provide similar survival curves for the SAT of both RBCs and PLTs. This result was confirmed by various simulation studies. For FFP units however, there is a marked difference between the results from these two methods. Further analysis of the FFP data showed that the distribution of the number of units transfused changed over time, but, more importantly, that patient survival increased over the years. Especially the differences found in long-term survival using different methods are caused by the change in short-term patient survival over time. Where the KM estimate at a particular follow-up time is based on the average hazard rate of earlier follow-up times, the DE method is based on the survival of all patients that (potentially) survived for the duration of the follow-up time considered. As such, the use of the KM method is to be preferred over the DE method.

Our simulation studies demonstrated what was confirmed by the analyses of the transfusion data: both methods are valid for estimating SAT. Mathematical proof of the fact that both estimation methods provide similar results was derived, but is not provided here. The simulation studies illustrate that despite the fact that transfusions are grouped per patient, SAT can be correctly estimated using the KM method. However, this grouping of transfusions does affect the estimation of its associated CIs: for a proper estimation of the SAT CIs bootstrapping of patients (with associated transfusions) is required. Standard KM CI estimates will generally deliver overly optimistic CIs.

In our paper we illustrate that SAT can be estimated using both the DE and KM methods, with each method having its own (dis)advantages. For both methods CIs can be obtained by bootstrapping patients. The necessity of determining SAT was revealed

when performing cost-effectiveness analyses of for blood safety interventions, but in fact SAT is in fact relevant when analyzing the cost-effectiveness of any safety intervention. The only prerequisite is that patients have multiple exposures over a prolonged period of time and that each exposure is associated with a controllable (small) risk.

Acknowledgements

This work was funded by the Netherlands organization for health research and development (ZonMw, project 945-06-603). We would also like to acknowledge Leslie Beks for collecting and processing patient and transfusion databases from our hospital.

References

1. Allain JP. *Transfusion risks of yesterday and of today*. Transfus Clin Biol 2003;10:1-5.
2. Gibson JL, Martin DK, Singer PA. *Evidence, economics and ethics: resource allocation in health services organizations*. Healthc Q 2005;8:50-9, 4.
3. van Hulst M, De Wolf JT, Staginnus U et al. *Pharmaco-economics of blood transfusion safety: review of the available evidence*. Vox Sang 2002;83:146-55.
4. *Ministerial bloodsupply plan 2007*. 2007. The Hague, the Netherlands, Dutch Ministry of Health, Welfare and Sport.
5. Aalen OO, Borgan O, Gjessing HK. *Survival and Event History Analysis: A Process Point of View*. 1st ed. New York: Springer, 2008.
6. Kaplan EL, Meier P. *Nonparametric estimation from incomplete observations*. Journal of the American Statistical Association 1958;53:457-81.
7. Clark TG, Bradburn MJ, Love SB et al. *Survival analysis part I: basic concepts and first analyses*. Br J Cancer 2003;89:232-8.
8. Robertson T, Wright FT, Dykstra RL. *Order Restricted Statistical Inference*. Chichester: John Wiley & Sons, 1988.
9. Efron B, Tibshirani RJ. *Introduction to the Bootstrap*. 1st ed. Boca Raton, USA: Chapman & Hall/CRC, 1994.
10. Wallis JP, Wells AW, Matthews JN et al. *Long-term survival after blood transfusion: a population based study in the North of England*. Transfusion 2004;44:1025-32.
11. Andersen PK, Keiding N. *Multi-state models for event history analysis*. Stat Methods Med Res 2002;11:91-115.

APPENDIX A: DESCRIPTION OF THE KM, DE AND MONOTONIZED DE ESTIMATORS

Suppose we have observed X transfusions, given at times $(d1_1, d1_2, \ldots, d1_X)$ in an interval $[0; D]$, to N patients (in general it is possible that one patient receives several transfusions). The time at which the patient who received the transfusion either died or was censored is given by $(d2_1, d2_2, \ldots, d2_X)$. Whether the patient died or was censored is given by outcomes $(o_1, o_2, \ldots, o_X)$, where $o_i = 1$ means that the patient receiving transfusion i died. There exists a $D_1 \geq D$ such that for all patients, we know whether they have died or not before time D_1.

Let $S(t)$ be the probability that a patient will have a lifetime exceeding t after obtaining a transfusion. The survival function $S(t)$ can be obtained by either the Kaplan-Meier (KM) method or by direct estimation of the survival (DE).

Define for each $i \in \{1, \ldots, X\}$, the observed time until death or censoring as $t_i = d2_i - d1_i$. Let the ordered subset $T_d \subset \{t_1, t_2, \ldots, t_X\}$ denote all the observed times of transfusions of patients who died. Note that if two transfusions have the same observed time, this time is included only once in the set T_d. The Kaplan-Meier survival estimate is calculated by the following formula:

$$\hat{S}(t) = \prod_{s \in T_d \,:\, s<t} \left(1 - \frac{\sum_{i=1}^{X} 1_{\{o_i=1\}} 1_{\{t_i=s\}}}{\sum_{i=1}^{X} 1_{\{t_i \geq s\}}} \right) \qquad \text{(A1)}$$

Here $1_{\{\mathrm{B}\}}$ is the indicator function which takes the value 1 if condition B is true, and 0 otherwise.

The direct survival estimate is calculated by the following formula:

$$\hat{S}(t) = \frac{\sum_{i=1}^{X} 1_{\{t_i \geq t\}} 1_{\{d1_i \leq D_1 - t\}}}{\sum_{i=1}^{X} 1_{\{d1_i \leq D_1 - t\}}} \qquad \text{(A2)}$$

Here we use the fact that for all transfusions that took place before D_1 - t, it is known whether or not the patient survived longer than t.

By definition, a survival function has to be a monotonic decreasing function of time. The Kaplan-Meier estimate is such that monotonicity is implicit. However, for the direct estimator this is not the case, which means that the estimate may increase over time. To correct for this, a monotonization step can be added that transforms the survival estimator to a strictly decreasing function. In the monotonization step the integral of $\hat{S}(t)$ is taken. Of this integral the least concave majorant is taken, which is the smallest concave shape that fits over the integrated function. This majorant function is subsequently differentiated to produce the updated estimate $\hat{S}_m(t)$, the monotonized version of $\hat{S}(t)$.

APPENDIX B: DESCRIPTION OF SIMULATION OF RECIPIENT COHORTS

Survival after transfusion (SAT)

Consider a recipient population with survival probability $S(t)$, indicating the probability of an individual being alive at time t. Suppose this population had a probability $B(t)$ of obtaining a transfusion at time t. The probability of a transfusion actually occurring at time t requires the recipient to be alive at time t, and would therefore be equal to $B(t)S(t)$. The survival probability of a recipient at time t after having received a transfusion at time x would be $S(x+t)/S(x)$. The survival after transfusion (SAT) of a cohort of recipients from this population will be equal to the survival after any timepoint x, weighted by the probability of transfusion at timepoint x:

$$SAT(t) = \frac{\int_{x=0}^{T_{\max}} B(x)S(x)\frac{S(x+t)}{S(x)}dx}{\int_{x=0}^{T_{\max}} B(x)S(x)dx} = \frac{\int_{x=0}^{T_{\max}} B(x)S(x+t)dx}{\int_{x=0}^{T_{\max}} B(x)S(x)dx} \quad \text{(B1)}$$

Where:
$S(0) = 1$ and $S(t) = 0$ for $t \geq T_{\max}$

Theoretical SAT for two example recipient populations

With Equation (B1) for any recipient population and transfusion regime the SAT can be calculated. This was done for the recipient population mentioned in the paper with a probability of survival of $S(t) = (1 - t/T_{\max})^{\alpha}$ (where $0 \leq t \leq T_{\max}$) which is subject to a constant transfusion rate per unit time (transfusion intensity) c. For such a population it can be derived that the $SAT(t) = (1 - t/T_{\max})^{(1+1/\alpha)}$. For this population the average number of transfusions per recipient (the denominator of Equation (B1)) is equal to $E(n) = \alpha c T_{\max}/(1+\alpha)$.

A more complex recipient population is one where recipient survival is equal to $e^{-\alpha t}$ for $0 \leq t \leq T_{\max}$ and 0 for $t > T_{\max}$ and where the recipient hazard rate (α) is a stochastic variable. In our example population α varies between $\alpha_{\min}$ and $\alpha_{\max}$ with a linear diminishing probability density. The probability density function for α is then equal to $P(\alpha) = 2(\alpha_{\max} - \alpha)/(\alpha_{\max} - \alpha_{\min})^2$. For a recipient from this population with hazard rate α, subject to a constant transfusion intensity $c{\cdot}\alpha$, can be derived that $SAT_{\alpha}(t) = (1 - e^{-\alpha(T_{\max} - t)})/(1 - e^{-\alpha T_{\max}})$. The expected number of transfusion given to this recipient is $E_{\alpha}(n) = c(1 - e^{-\alpha T_{\max}})$.

Now with the known distribution of α and the known SAT and expected number of transfusions for any recipient with a known α, $S(t)$ and $SAT(t)$ of the population as a whole can also be calculated:

$$S(t) = \int_{\alpha=\alpha_{\min}}^{\alpha_{\max}} S_\alpha(t)P(\alpha)d\alpha = \frac{2e^{-(\alpha_{\max}+\alpha_{\min})T_{\max}}\left(e^{\alpha_{\min}t} + e^{\alpha_{\max}t}\left((\alpha_{\max}-\alpha_{\min})t-1\right)\right)}{t^2(\alpha_{\max}-\alpha_{\min})^2} \quad \text{(B2)}$$

$$SAT(t) = \frac{\int_{\alpha=\alpha_{\min}}^{\alpha_{\max}} SAT_\alpha(t)E_\alpha(n)P(\alpha)d\alpha}{\int_{\alpha=\alpha_{\min}}^{\alpha_{\max}} E_\alpha(n)P(\alpha)d\alpha} = \frac{\int_{\alpha=\alpha_{\min}}^{\alpha_{\max}} (1-e^{-\alpha(T_{\max}-t)})(\alpha_{\max}-\alpha)d\alpha}{\int_{\alpha=\alpha_{\min}}^{\alpha_{\max}} (1-e^{-\alpha T_{\max}})(\alpha_{\max}-\alpha)d\alpha} = \quad \text{(B3)}$$

$$= \frac{2T_{\max}^2\left(e^{-\alpha_{\max}t} - e^{-\alpha_{\min}t}\left((\alpha_{\max}-\alpha_{\min})t-1\right)\right) - 2t^2\left(e^{-\alpha_{\max}T_{\max}} + e^{-\alpha_{\min}T_{\max}}\left((\alpha_{\max}-\alpha_{\min})T_{\max}-1\right)\right)}{t^2T_{\max}^2(\alpha_{\max}-\alpha_{\min})^2 - 2t^2e^{-(\alpha_{\max}+\alpha_{\min})T_{\max}}\left(e^{-\alpha_{\min}T_{\max}} + e^{-\alpha_{\max}T_{\max}}\left((\alpha_{\max}-\alpha_{\min})T_{\max}-1\right)\right)}$$

Simulation of survival after transfusion

A simulation of a cohort of recipients as described above with an exponential survival and a stochastic hazard rate was performed. There was a continuous inflow of 20 new recipients per year in the recipient population which appear at Poisson distributed intervals. The constant transfusion intensity (c) was equal to 3. The recipient hazard rate (α) varied between 0.05 and 1.5 per year. The maximum survival of the recipients ($T_{\max}$) was 10 years. We observe transfusions and deaths in this population over a period of five years.

In Figure B1 the mean recipient survival in the cohort is shown as well as the distribution of the individual recipient survival. The curves for the distribution of individual recipient survival are equally spaced on the cumulative density function of the hazard rate α. The average recipient receives 2.8 transfusions. On average only 83% of the recipients receive 3 transfusions, 14% receive 2 transfusions, and only 3% of the recipients receive 1 transfusion.

Figure B2 shows various characteristics of one thousand cohort simulations. On average there are 100 recipients in the cohort who receive on average 277 transfusions. The proportion of observations that is censored is 38%, which refers to the proportion of times since transfusion. In Figure B2 also the distribution of the number transfusions per recipient in the simulated cohort is given. Note that the proportion of recipients that receive only 1 transfusion is much higher than the theoretical proportion. This is caused by the fact that in the simulation 37% of the recipients are already in the cohort when the interval of observation is started. Also, a part of the 37% of the recipients in the cohort that are censored will obtain subsequent transfusions beyond the end of the observation interval. This limitation of the observation window will bias the distribution of the number of transfusions per recipient strongly to the left.

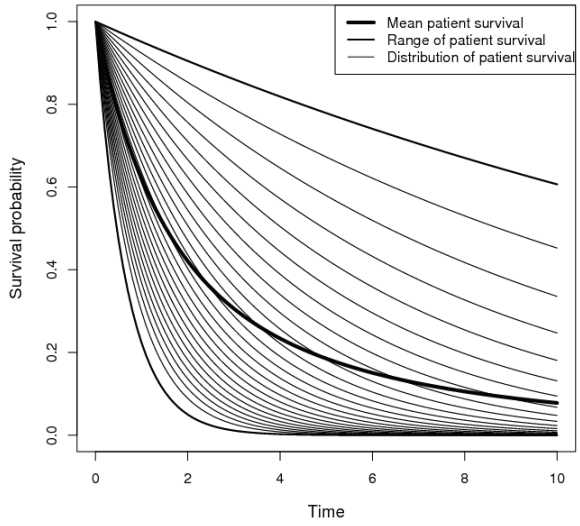


Figure B1: Mean recipient survival in the simulated cohort and the distribution of individual recipient survival.

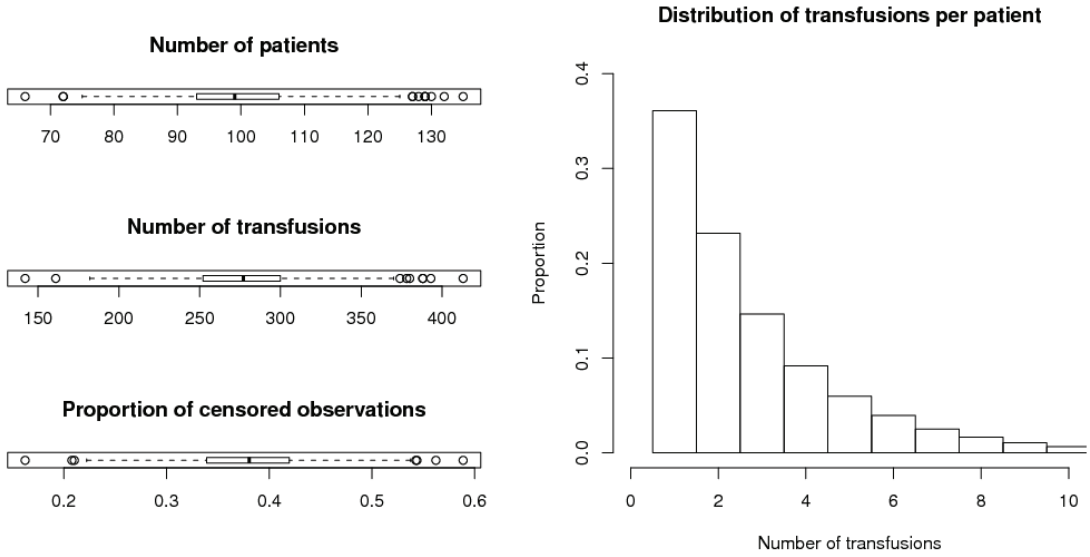


Figure B2: Various simulation characteristics: distribution of number of observed recipients and transfusions, the proportion of censored observations (transfusions for which the recipient was still alive at the end of the observation period) and the distribution of number of transfusions per recipient.

For each of the simulated cohorts SAT was estimated by means of the KM method, the DE method and the monotonized DE method. Figure B3 shows the 2.5, 50 and 97.5 percentiles of the estimated survival probability at various times after transfusion. In addition, the theoretical SAT for the recipient cohort (Formula B3) is provided. The results show that the median estimates for all three estimates coincide with the theoretical SAT, at least for the first 4 years. Also, the 95% CIs are similar for all three

methods, although those for the KM estimates seem to be the smallest and those for the non-monotonized the largest.

For each of the simulated recipient cohorts the following characteristics were determined:

- the survival probability, estimated by the KM method;
- the 95% CI of the KM estimate using greenwoods formula;
- the 95% CI by bootstrapping transfusion intervals: for each bootstrap the whole set of observed time-intervals from transfusion to recipients' death or censoring were resampled, and
- the 95% CI by bootstrapping transfused recipients: for each bootstrap the whole set of observed recipients were resampled. For each resampled recipient all associated transfusion time-intervals and outcomes (censoring or death) are added to the collection of bootstrapped time-intervals and outcomes.

In Figure B4 the 2.5, 50 and 97.5 percentiles of the simulated KM estimators are shown. In addition, the median values for each of the CI estimates over the 1000 simulated cohorts given. These were calculated for 100 equally spaced times after transfusion on the interval 0 to 5 years. Figure B4 shows that, as expected, the 95% CI of the bootstrap on transfusions is similar to the 95% CI of the KM estimator by greenwoods formula. However, the 95% CI of the simulated cohort is much wider and is very well represented by the 95% CI obtained by the recipient bootstrap. This illustrates the fact that with bootstrapping recipients the variability in the KM estimate related to the recipient cohort is correctly represented. Also it shows that the CIs of SAT are incorrectly estimated when ignoring the effect of clustering of transfusions at recipient level.

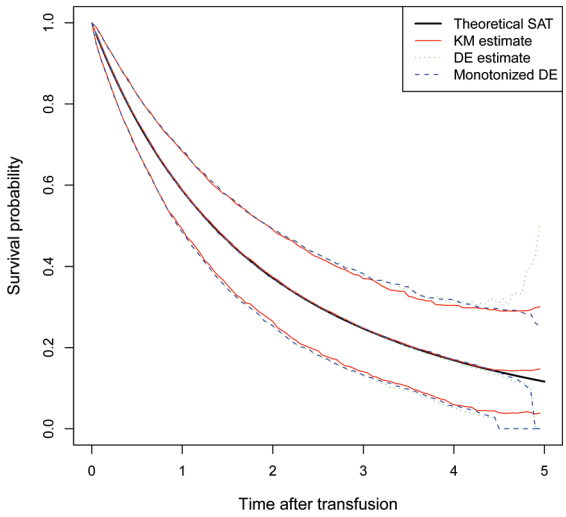


Figure B3: The 2.5, 50 and 97.5 percentiles of SAT for various survival estimation methods from a simulation of 1000 recipient cohorts.

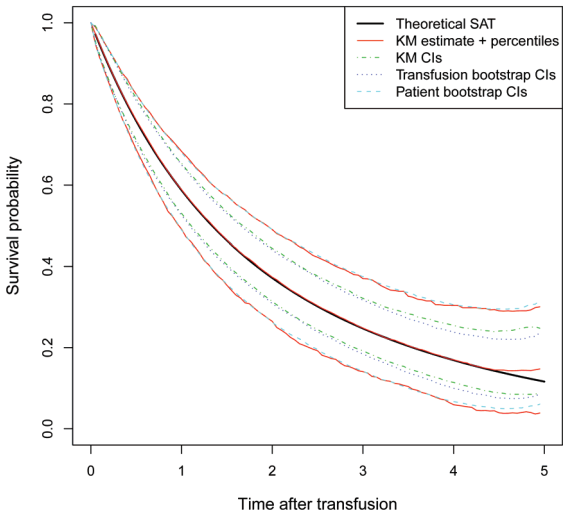


Figure B4: 2.5, 50 and 97.5 percentiles of the KM survival estimator cohorts and the median value of various 95% CI estimates from 1000 simulated cohorts.

PART

B

Cost-effectiveness and Value of Information

Costs and benefits of bacterial culturing and pathogen reduction in the Netherlands

Mart P. Janssen
Cees L. van der Poel
Erik Buskens
Luc G.A. Bonneux
Gouke J. Bonsel
Ben A. van Hout

Transfusion 2006; 46(6):956-965
DOI: 10.1111/j.1537-2995.2006.00828.x

CHAPTER

3

“Make everything as simple as possible,
but not simpler”

Albert Einstein

Background

Bacterial contamination is a life threatening risk of blood transfusion, especially with platelet transfusions. Bacterial culturing (BCU) of platelets as well as pathogen reduction (PRT) reduce the likelihood of such contamination. We analyze the cost-effectiveness (CE) of these interventions after the introduction of the diversion pouch during blood collection.

Study design and methods

The balance between costs and benefits of preventing adverse events due to platelet transfusion is assessed with a mathematical decision model and Monte Carlo simulations. Model parameters were obtained from literature and from Dutch Sanquin blood banks. The balance between costs and benefits is assessed in terms of costs per quality adjusted life year (QALY).

Results

The costs per 100,000 platelet concentrates in the Netherlands are estimated at $3,277,032 (€2,520,794) for BCU and at $18,582,844 (€14.294.495) for PRT. In comparison to the situation without BCU and PRT, costs per QALY are estimated at $90,697 (€69,767) for BCU (95%CI: $18,149-$2,088,854) and at $496,674 (€382,057) for PRT (95%CI: $143,950-$8,171,133). The ratio of differences in costs and QALYs between BCU and PRT (the relative cost-effectiveness) is estimated at $3,596,256 (€2,766,351; 95%CI: $1,100,630-$24,756,615). Large uncertainty in sepsis complication rates and platelet recipient survival exist, causing large uncertainties in the absolute CE for both interventions.

Conclusions

As a result of the unknown probability of sepsis complications and platelet recipient survival the CE ratios of BCU and PRT in the Dutch setting are highly uncertain. Despite these large uncertainties it can be concluded that BCU is without doubt more cost-effective than PRT.

Introduction

With the progression of technologies the risks of blood transfusions have been reduced considerably over the past few decades.[1] In the 1980s the risk of viral infection by blood transfusion exceeded 1 in a 1000 transfusions; nowadays this risk is more than a thousand times lower.[2] For labile blood products the focus of prevention for viral transmission was on blood screening techniques. This approach was highly successful, but leaves residual risks which are presently primarily determined by bacterial sepsis acquired through contaminated platelets.[2-4] Several techniques are available to prevent bacterial contamination from blood products, in particular the use of a diversion pouch bacterial culturing (BCU) and pathogen reduction (PRT).[5-7] In November 2001 bacterial culturing (BCU) of 100% of platelet concentrates was introduced in the Netherlands. With BCU, aerobic and anaerobic samples of the finished pooled platelet product are kept at 35°C in co-culture with storage of the buffy-coat derived platelet pools, and bacterial contamination is detected through measurement of CO_2.[5,8] In July 2004 this approach was supplemented by the introduction of a diversion pouch at blood collection. This allows separation of the first 20 to 30 ml of blood which is most likely to be contaminated by skin flora.[6]

Another approach for reducing bacterial risk of platelets transfusion is pathogen reduction, which recently became available for treatment of platelet products.[7] The most developed method for photochemical inactivation of bacteria is achieved by addition of a synthetic psoralen and illumination with UVA light.[9,10] This method is considerably more expensive than BCU but offers the additional benefit of reducing viral residual risks. The Dutch Health Council recently advised awaiting the results of more clinical studies, in particular on the added safety in an already safe system of blood provision, before the introduction of PRT.[11]

The Dutch Health Authorities have raised the issue of optimal versus maximal blood safety. With PRT likely to be more effective but also more costly, it may well be that – even when PRT would have maximum benefits – the balance between costs and benefits would still be unacceptable when considering the current Dutch situation using the diversion pouch. This idea is analyzed in this study and in doing so we also assess the CE of the introduction of BCU: a decision that has been taken previously without an extensive evaluation of costs and benefits.

Materials and Methods

The analysis concerns the comparison of two interventions reducing the risk of contaminated platelet transfusions (BCU and PRT) in a setting where the diversion pouch is part of standard practice and neither BCU nor PRT are introduced. However, we employ BCU data of 2 years of screening in the Netherlands with diversion pouch to assess the frequency of bacterial contamination and the sensitivity of BCU. Costs and benefits are analyzed using a mathematical model that brings together data on the probability of blood contamination by bacteria and the main blood transmissible viruses; data on the consequences of contamination and data on the costs of the interventions. The analysis is performed from a direct cost perspective, which means that only the

direct costs required for medical treatment are considered. There are three separate situations that are analyzed: the baseline situation with diversion pouch (1), the situation where bacterial culturing on 100% of the platelet products is being applied next to the use of a diversion pouch (2), and where a pathogen reduction technique is applied next to the use of a diversion pouch (3). The CE of the last two alternatives are compared to the baseline situation and to each other. The latter relative CE of PRT to BCU, is expressed in the ratio of differences in costs and differences in QALYs between BCU and PRT.

Cost-effectiveness

The estimates of costs are limited to the costs of the risk reduction techniques and the direct medical costs associated with the consequences of infections. The same infections are associated with estimates of life years lost and quality of life lost, expressed in terms of quality-adjusted life years.[12] Cost-effectiveness is expressed in terms of costs per QALY. Future costs and effects are discounted using a discount rate of 4% in accordance with guidelines for CE analyses.[13] Results are expressed in terms of point estimates assuming 100,000 platelet transfusions per year.

The WHO recently reported that each life year is valued at around three times the annual earnings.[14] Therefore an indication for a CE threshold would be three times the gross domestic product per head. For the Netherlands the threshold would be US$110,000 per QALY and is comparable to the value for the USA (US$120,000).

Uncertainties are addressed by uni- and multivariate sensitivity analyses. The results of the multivariate sensitivity analyses are used to calculate uncertainty margins surrounding the outcomes.

Contamination probabilities

A critical parameter of the cost-effectiveness model is the probability of contamination. The estimated probability of an infectious platelet transfusion for viruses can be derived from the measured donor incidence rates and a virus specific window period.[15] Table 1 shows the estimated residual risk of viral contamination per platelet product for HIV, HCV, HBV, and HTLV-1/2 in the Netherlands.[16]

Hepatitis B viremia and antigen are often only transiently present in the donor blood, which leads to underestimation of the actual incidence rate. A method for estimating a correction factor for the measured incidence rate based on donation intervals described in literature.[17,18] On basis of donation intervals from the Sanquin blood banks the incidence rate was estimated to be a factor 3.0 higher than the recorded incidence rate. "Platelet concentrates" in the Netherlands are prepared according to the buffy-coat method from five whole blood donations. It should be noted that in the Netherlands only repeat donations are used for the preparation of blood products.

Table 1: Contamination risks of platelet blood products

A	B	C	D= C/365	E= B x D	F= 5 x E
Type of contamination	**Mean incidence among repeat donors 1997-2003[16] (per million donor yrs)**	**Detection window[19] (days)**	**Window Risk**	**Probability of infected donation in The Netherlands (per million)**	**Probability of infected pooled (*n*=5) platelet transfusion in The Netherlands (per million)**
HIV infection	5.5	11	0.03	0.2	0.8
HBV infection*	32 (11)	59	0.16	5	26
HCV infection	2.6	10	0.03	0.1	0.4
HTLV I/II infection	1.5	51	0.14	0.2	1.0
Bacterial contamination of platelets					3,700

***** Adjusted for non-detection. (The actual measured incidence rate is given in brackets)

Since November 2001, in the Netherlands the actual bacterial contamination in platelet concentrates is measured in 100% of the products. In July 2004 also a diversion pouch at blood collection was introduced. The diversion pouch separates the first 20 to 30 ml of all donations, which are most likely to be contaminated by skin flora. The risk of bacterial contamination of platelet concentrates – as presented in Table 1 – is based on these measures selecting the results of 111,111 tested platelet products, collected after the introduction of the diversion pouch. As point estimate, we used a rate of 0.42% with an uncertainty margin of 0.32% to 0.56%.[20]

Over the years 1998 to 2001 the Sanquin Blood Supply Foundation received on average reports of 1 HBV infection and 5 complications from bacterial contamination per year resulting from hemovigilance reported by Dutch hospitals. As underreporting is likely, the reported incidents in combination with the annual number of platelet transfusions in the Netherlands (50,000) are used to estimate the lower bounds of the complication rates used in our model.

Consequences of contaminated donor blood

Complications of transfusion can arise immediately following the transfusion, in case of a bacterial contamination, or at a later stage, in case of a transmission of a chronic viral infection. A bacterial contamination becomes clinically apparent as sepsis. While the literature offers various estimates of the incidence of sepsis, no reliable data is available on the likelihood of sepsis after transfusion of contaminated platelets.[2,21-26] Only recently an article was published where an estimate was given for the probability of sepsis from contaminated platelet transfusions ranges from one in ten to two in five (10% to 40%).[27] We used a point estimate of 10% and applied a wide margin of 1% to 40% to express the uncertainty surrounding this estimate. The lower limit of 1% represents the probability at which the model predicts the 5 cases of sepsis as reported by hemovigilance in the past. The upper limit (40%) equals the upper limit from the range estimate for the probability of sepsis from contaminated platelet transfusions but also corresponds to the percentage of contaminated transfused pooled platelets that did not cause clinical complications in clinical practice (44%).[8,27] The uncertainty of the

estimation was modeled using a Weibull distribution.[28] When considering bacterial sepsis literature reports a total of 22 deaths in 118 cases, resulting in a case-fatality rate of 19%.[23-26] We applied a Beta distribution based on these figures to express its uncertainty.[28]

The consequences of transmitted viral infections are difficult to estimate as most viral infections have a variable, host dependent prognosis. Several papers dealt with this issue by models on transmitted chronic viral infections.[29-35] In our model we made the conservative assumption that viral contamination will lead to transmission, cause disease and will bring about associated costs and loss of health. Even though this is a worst case scenario, it is in line with the thought that we assess maximum benefits from PRT.

Effectiveness of interventions considered

Validation tests of the bacterial culturing system indicated a high sensitivity: defined as the proportion of contaminated products that are actually identified as being contaminated.[36,37] Its use in a production environment under a wide spectrum of bacterial contaminations might show less favorable results, as was illustrated by recent experience from routine practice in the Netherlands.[8] In routine practice 56% of 184 contaminated platelet products (mainly diphtheroid rods) were found to be transfused before detection by the culturing system. Even though none of these products lead to complications, two non-detected contaminated platelet products lead to severe septic incidents.[8] The rate of non-detected contamination of platelets resulting in sepsis events is therefore 1%. Because the actual amount of unobserved contaminated platelets is unknown and only those detected through the hemovigilance system are accounted for, in our model we allowed for a factor of 10 for underreporting of septic incidents or detection of contaminated products. We therefore modeled a sensitivity of the bacterial culturing system of 90%, with an uncertainty range of 9%.

PRT has demonstrated the reduction of contamination with at least a log 4 factor for a wide range of bacteria, viruses and parasites.[7] However, PRT has shown to fail reduction of bacteria in platelet concentrates as well.[38] Corresponding with the idea that we will analyze the CE of PRT at its best, we presumed the sensitivity of PRT to be 100%, an extreme assumption in favor of the PRT treatment.

Costs and effects

The cost estimates per platelet concentrate for bacterial culturing and pathogen reduction are presented in Table 2. Costs associated with BCU have been derived from operational data of the Sanquin blood banks. The operational costs for PRT were estimated by Sanquin and were based on the price as indicated by the manufacturer ($100 per pooled platelet concentrate). All costs were converted at a rate of 1.3 US$ per EURO. The cost of personnel, housing and maintenance for PRT production are estimated at $19 per product. In addition, the cost of production loss due to the limited shelf life (19.8% at 5 days outdating[39] which is current practice), the production loss caused by the pathogen reduction process itself (10%) and the increased platelet usage in clinical practice (ranging from 0% to 30%) was accounted for.[10,40] All costs reflect costs in the year 2002.

On basis of national hospital records the average cost of sepsis incidents have been calculated to be $7,000 per incident.[41] An estimate of the upper limit to the cost of sepsis of $18,000 was derived considering all costs of all events where sepsis was diagnosed.

Based on these figures, we estimate the direct costs of sepsis-treatment at $12,500 with an uncertainty margin of $7,000 to $18,000.

The costs of treatment of viral diseases require longer time horizons and depend on patient prognosis. In a recent publication estimates of the annual cost of viral diseases ranged from US$1,000 to US$50,000 per year depending on the type of infection.[33] An earlier CE study on platelet use indicated a maximum discounted cumulative lifetime cost of US$150,000 (US 1996 dollars).[31] Given the annual figures this might seem low, but periods of high costs are accompanied with high mortality rates. These costs are also likely to occur in the remote future where discounting leads to lower values.

In our model we made the assumption that the discounted total lifetime cost of a viral infection would be $250,000, irrespective of the type of viral disease transmitted and age of the infected patient. We used a range of $125,000-$375,000 to express our uncertainties.

In the Netherlands platelet recipients are on average 50 years of age.[42] Patient survival is paramount in determining the CE, especially when the life expectancy is low as is the case with platelet recipients. Literature indicates that the 5 year survival of platelets is only 21%.[43] We calculated the life expectancy of platelet recipients on basis of the results from Wallis and extended the life expectancy on basis of our national health statistics and in accordance with the age distribution on the Dutch platelet recipient distribution.[42-44] Data from the EuroQol group indicate that the average 50 year old person has a quality of life index of about 0.75 decreasing to a value of about 0.65 at the age of 85.45 Using these figures, the expected discounted number of QALYs of a Dutch platelet recipient is estimated at 4.4 years, ranging between 2 and 6 years. This estimate is used for the number of life years lost in case a patient obtains an infection and dies from sepsis.

From other CE studies it was derived that the reduction of cumulative QALY adjusted life years due to viral infections is less than 5%.[31,33] Therefore we presumed the utility of patient obtaining a viral infection to be 95% (ranging between 90% and 100%).

Sensitivity analysis

On the basis of the assumptions as outlined above, event trees were constructed describing incidents and outcomes after blood transfusion for both treatment alternatives and the reference situation (no additional treatment). Additionally, uni- and multivariate sensitivity analyses were carried out to assess the robustness of the results. The univariate sensitivity shows to what extent the estimates of costs per QALY change when the underlying estimates are subsequently changed within the margins of uncertainty. It is analyzed whether one might draw different conclusions when varying them within these margins. In the multivariate sensitivity analysis costs and effects are assessed when all estimates are varied at the same time drawing at random from the distributions reflecting the uncertainties (Monte Carlo simulation). Table 2 summarizes all point estimates, the associated uncertainty margins and the distributions used in the sensitivity analyses. In case of normal distributions, the lower and upper limits of the uncertainty margins correspond with the lower and upper 95% limits. In case of Pert distributions, the lower and upper limits are the real lower and upper limits. A Pert distribution is a truncated Beta-distribution that is characterized by a lower limit, an upper limit and a most likely value.[28] When the distances between the most likely value

and the upper and lower limit are identical (as is the case here) the shape resembles that of a normal distribution with truncated upper and lower bounds.

Table 2: Model and range parameters

Model parameter	Parameter value	Range	Distribution type
Median recipient age (years)	50	0-90	Dutch platelet
Mean quality of life adjusted life expectancy (QALYs)	4	2-6	recipient
Probability of bacterial contamination	0.42%	0.32% - 0.56%	Beta
Probability of severe sepsis given bacterial contamination	10%	1% - 40%	Weibull
Probability of death given severe sepsis	19%	13% - 27%	Beta
Sensitivity bacterial culturing	90%	81% - 99%	Pert
Sensitivity pathogen reduction	100%	-	
Probability of viral infection (all infections)	$3x10^{-5}$	$1x10^{-6}$ - $6x10^{-5}$	Beta
Cost of sepsis treatment ($)	12,500	7,000 - 18,000	Pert
Workup cost bacterial culturing ($)	31	28 - 35	Pert
Workup cost pathogen reduction ($)	178	152 - 205	Pert
Increased usage of PRT platelets	15%	0% - 30%	Uniform
Total discounted cost of treatment of patient with viral infection ($)	250,000	125,000 - 375,000	Pert
Utility of patient with viral infection	0.95	0.9 - 1.0	Pert
Discount rate (%)	4%	-	

Results

Three scenarios are analyzed after the introduction of the diversion pouch: no additional treatment (1), bacterial culturing (2) and pathogen reduction (3). Figure 1 shows the event tree for the reference situation (without additional treatment) and its associated outcome probabilities together with the main outcomes of all three scenarios.

It is estimated that in 100,000 platelet transfusions BCU reduces the number of adverse events (clinical sepsis or transmission of viral diseases) from 45 to 7. In case of perfect protection by PRT, this number would reduce it to zero. The additional costs of BCU are estimated at $3,277,032; the additional cost of PRT at $18,582,844. This means that the cost of BCU and PRT equal $86,079 and $411,850 respectively per event prevented. When comparing PRT with BCU it is estimated that PRT will prevent an additional 3 sepsis events, 1 sepsis deaths and 3 viral infections leading to a cost per additional prevented event of $2,170,959.

Event tree					Probability (No additional treatment scenario)	Cases in 100,000 transfusions		
						No additional treatment	Bacterial Culturing	Pathogen Reduction
Trans-fusion	Sterile 0.996			→	0.996	99,577	99,958	100,000
		Sub-clinical		→	0.004	381	38	0.0
	Bact. contam. 0.0042	0.90		→	3×10^{-4}	34	3	0.0
		Sepsis 0.10	Death 0.19		8×10^{-5}	8	1	0.0
		HIV 8×10^{-7}			8×10^{-7}	0.1	0.1	0.0
		HBV 3×10^{-5}			3×10^{-5}	3	3	0.0
→	Viral contam. 0.9999	HCV 4×10^{-7}			4×10^{-7}	0.04	0.04	0.0
		HTLV-I/II 1×10^{-6}			1×10^{-6}	0.1	0.1	0.0
					Total number of events	45	7	0
					Additional treatment costs (US$)	0	3,277,032	18,582,844
					Cost per event prevented (US$)		86,079	411,850
					Cost per additional event prevented (US$)			2,170,959

→ means that the branch is continued at the viral contamination part of the event tree

Figure 1: Event tree model, outcome frequencies and costs per treatment option per 100,000 platelet transfusions.

When associating the costs per event prevented with estimates of the QALYs lost due to these events, estimates are obtained of the cost per QALY gained. Figure 2a shows the results in terms of the relative CE of both treatments as compared to the "no additional treatment" alternative. In this graph on the vertical axis the difference in costs are plotted against the difference in QALYs on the horizontal axis. The upper cloud of points represents the outcomes of 1000 random draws from the multivariate sensitivity analysis regarding the comparison between PRT and "no additional treatment". The lower cloud represents these when comparing BCU with "no additional treatment". The central estimates are that BCU gains 31 (95%CI: 2-103) QALYs against a net cost of $2.801.202 (95%CI: $1.731.058-$3.343.982) and that PRT gains 35 (95%CI: 2-114) QALYs against a net cost of $17.349.086 (95%CI: $14.447.548-$20.420.219). For each of the situations, point estimates and 95% confidence intervals are drawn in the figure. Also the CE threshold value is shown.

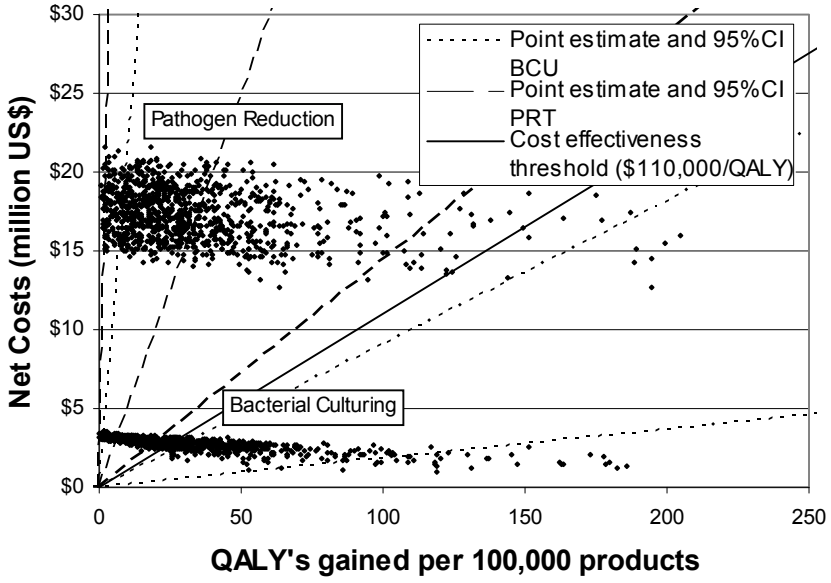


Figure 2a: Relative CE of BCU and PRT versus "no additional treatment"

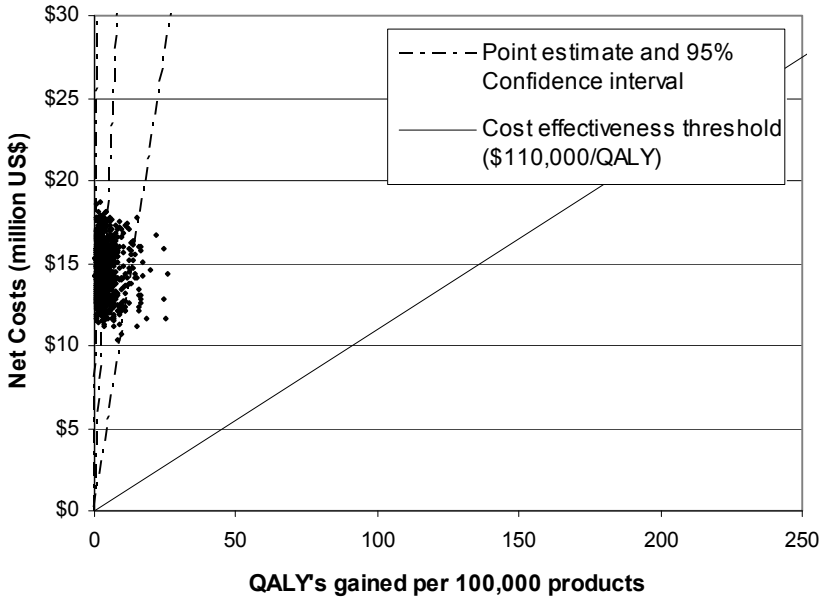


Figure 2b: Relative cost-effectiveness of PRT versus BCU

The costs per QALY are estimated at $90.697 (95%CI: $18.149-$2.088.854) for BCU and at $496.674 (95%CI: $143.950-$8.171.133) for PRT. Figure 2b shows the results when comparing PRT with BCU. Here differences in cost between PRT and BCU are plotted against differences in QALYs, using the same data as shown in Figure 2a. The central estimate is a cost per QALY of $3.596.256 (95%CI: $1.100.630-$24.756.615).

In Table 3 the outcomes of the univariate sensitivity analysis are given, where it is analyzed to what extent the costs effectiveness ratios of BCU versus "no additional treatment" and PRT versus BCU change when changing the subsequent point-estimates to the upper and lower limits of their uncertainty ranges. Table 3 shows that the main

parameters affecting the CE are: the probability of sepsis given bacterial contamination, the patients' quality adjusted life expectancy, the probability of death given sepsis and the probability of bacterial contamination. The model outcome is relatively insensitive to changes in all other model parameters. The table also shows that the relative CE of PRT to BCU is most sensitive to the probability of sepsis given bacterial contamination and the test sensitivity of the BCU system.

Table 3: Percentage change in CE-ratios at the outer limits of the margins of uncertainty

Model parameter	Range	BCU versus "No workup"		PRT versus "No workup"		PRT versus BCU	
		Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit
Probability of sepsis given bacterial contamination (%)	1 - 40	925	-88	703	-77	306	-72
Mean discounted quality adjusted life expectancy (years)	2 - 6	118	-27	118	-27	118	-27
Probability of death given severe sepsis (%)	13 - 27	47	-30	46	-30	38	-27
Probability of bacterial contamination (%)	0.31 - 0.56	38	-29	33	-25	27	-22
Workup cost pathogen reduction ($)	152 - 205	0	0	-16	16	-19	19
Sensitivity bacterial culturing (%)	81 - 99	13	-11	0	0	-43	324
Workup cost bacterial culturing ($)	28 - 35	-12	12	0	0	2	-2
Cost of sepsis treatment ($)	7,000 - 18,000	7	-7	1	-1	0	0
Probability of viral infection (all infections)	$1x10^{-6}$ - $6x10^{-5}$	0	0	5	-6	20	-19
Total discounted viral infection treatment cost ($)	125,000 - 375,000	0	0	2	-2	2	-2
Utility of patient with viral infection (QALY)	0.9 - 1.0	0	0	-2	2	-13	18

Discussion

The results from this study lead to the conclusion that in the Netherlands the CE of BCU is superior to PRT. The main reason is the fact that bacterial culturing significantly reduces the incidence of sepsis events, which is the major contributor to the overall risk of platelet transfusions. Moreover, it does so against relative low costs. Although PRT may reduce this probability even further, the increase in costs are well above the costs that are considered acceptable.[46]

The main uncertainties with respect to the absolute CE concern the probability of bacterial contamination related complications and the expected health gain. These uncertainties however affect both treatment options equally. So even though the absolute CE of BCU can be questioned (there is a 53% chance that it will be above the CE threshold), it will by far be more cost-effective than PRT as is shown by the relative CE graph shown in Figure 2b. As awareness of bacterial contamination of platelets increases and hemovigilance systems generate more data, more knowledge about the sepsis

probability will become available which will enable better estimates for the CE of these safety measures.

The modeled probabilities of sepsis (See Figure 1: 34+8=42 per 100,000 transfusions) and death (8 per 100,000 transfusions) through bacterial contamination are in line with values reported in literature (40 to 100 and 7 to 13 per 100,000 for sepsis and death respectively).[24,47] It should be noted that the modeled values reflect the incidence rates after the introduction of the diversion pouch in the Netherlands, where the contamination and incident probability would be expected to be a factor two lower than the published incidence rates.[20,48] Possible other causes for differences in incidence rates might be: geographic and demographic differences (1) , differences in donor collection procedures (e.g. arm cleansing) (2), the fact that the estimated sepsis probability does not reflect the Dutch setting (3), or the reported sepsis and death incidence rates underestimate true incidence rates (4).

The number of sepsis events after introduction of the bacterial culturing system as predicted by the model is 1 in 24,000 transfusions. The number of sepsis events reported after the introduction of the bacterial culturing system, but before the introduction of the diversion pouch was 1 in 14,000.[8] Therefore the projected number of sepsis events is about half the number of events found before the introduction of the diversion pouch and corresponds to the expected number of cases.

As a result of a high bacterial contamination probability (0.42%) there is a high number of sepsis events predicted by our model. There is a fairly large discrepancy between the number of estimated sepsis events from Figure 1 (42 per year) and the number of sepsis events as reported by hemovigilance (5 per year). The hemovigilance reporting is likely to underestimate the true number of sepsis events as patients requiring platelet transfusions are generally severely ill and have sepsis from many other causes. Data from our academic hospital show that 10% of non-discharged platelet recipients will not survive one week past the date of transfusion. However, even if the actual sepsis probability would be much lower than the sepsis probability modeled, this would result in a higher estimate for the CE ratio, but will not affect the relative CE of BCU over PRT as it will affect the benefits for both interventions.

The experience with the BacT/Alert system in a "release as negative to date" schedule has shown that a large number of contaminated pooled platelets (56%) are transfused before detection.[8] This shows that the system of bacterial culturing is not at all flawless. However, contamination detected after delivery to the hospital, concerns bacteria that are slow growing and generally less pathogenic. Apparently, these do not lead to complications and the system seems to be adequately safeguarding transmission of bacterial contamination. Even though in some cases the system does fail, as shown by the cases reported back through the hemovigilance system, it remains by far a more cost-effective than PRT in preventing transmission of bacterial contamination through platelet products.

Table 3 clearly illustrates the effect of the high uncertainty with respect to the likelihood of sepsis complications on the CE of both BCU and PRT. Our analysis show that as long as the probability of sepsis exceeds 1 in 2800 transfusions, the CE of BCU is expected to be below the $110,000 per QALY threshold. Contrarily, there is not one variable which – within its margins of uncertainty – would lead to an estimate of the costs per QALY of PRT under $110,000 or under $1,000,000 if compared to BCU. As

such, the conclusion that PRT is not very cost-effective in comparison to BCU can be called robust. The reason is that whatever assumption is made, there is little doubt that BCU is a major step forward in comparison to "no additional treatment". After BCU, which reduces 90% or more of the bacterial contamination risk, there is hardly any residual risk left. Even though PRT reduces risk completely (in our theoretical model that is), the additional effectiveness of PRT is relatively small in comparison to the additional costs. Table 3 also shows that the relative CE is primarily influenced by the probability of sepsis given bacterial contamination and the sensitivity of BCU. This is caused by the fact that by reducing this probability the relative gain from PRT will diminish and as a result, the CE ratio will increase. An increase of the sensitivity of BCU will directly cause an increase of the relative CE ratio.

Bell *et al.* estimated the CE of PRT for platelet products. Here the baseline analysis showed a CE ratio of PRT in the order of US$500,000 to US$2,000,000 per QALY depending on the patient type.[33] This is well in line with our point estimate of $496.674 and a 95% confidence interval ($143.950-$8.171.133). In contrast to this study, we did not model any potential future scenarios. For example long-term toxic side-effects of PRT as a result of treatment were not considered, but might influence the acceptability this technology. We also did not model any effects of an unknown hepatitis C virus like agent, nor did we model any effects of an unknown non-enveloped virus that is relatively unaffected by PRT.

In our analysis we did not consider the irradiation procedure for immuno-suppressed patients. We did not consider this specific subgroup as they only consume 15% of all platelet concentrates in the Netherlands. Possible protection against CMV virus was also not assessed as 100% of platelets are leukocyte depleted to a level less than 106 cells per product. The Canadian CMV consensus concluded that leukocyte depletion reduces CMV risk by transfusion to an undetectable level.[49]

The costs and effects associated with viral infections in our model could be considered crude or overly simplistic, especially as advanced models exist that describe disease progressions and associated costs for these diseases in detail. However, we applied uncertain but conservative estimates for both costs and effects with respect to these diseases and found that these assumptions do not at all affect the outcome of the CE comparison of BCU and PRT, as is shown in the sensitivity analysis. Therefore, the model used is sufficiently adequate to serve its purpose.

One additional beneficial effect of bacterial culturing that is not included in our analyses is the shelf life extension of bacterially screened platelets. It has been shown that with BCU the platelet shelf life can be extended from 5 to 7 days which will cause a significant reduction of the production loss.[50] Dutch data on the benefits of shelf life extension are not available yet, but it is clear that this will further increase the CE of BCU. The cost of PRT platelets originate primarily from the cost of the PRT process and therefore it is possible to assess the effect of extending the platelet shelf life for PRT platelets. Our model indicates that a reduction of outdating from 20% to 5% will reduce the CE ratio of PRT by 10%. This will not affect any of the conclusions drawn earlier.

In the Netherlands the vast majority (86%) of platelet products are buffy-coat derived platelet pools.[16] This situation is different in the USA where the majority of platelet products are apheresis products. If costs (and methods) of platelet production in the USA would be comparable to those in the Netherlands, then BCU would be cost-effective only if the rate of sepsis through transfusion of apheresis products would exceed

1 in 3000 transfusions. As one of the reasons for applying apheresis in preparing is the reduction of the contamination rate, this is not likely to be the case.

The conclusion towards the CE of BCU is based upon CE criteria applied in general health care setting where prevention or cure of disease is being judged. In case of treatment of blood products however the setting is different as these risks are iatrogenic. In the environmental domain the Dutch government has maintained a risk policy since the 1980s that "no individual should be exposed to an activity imposing a risk of dying greater than one in a million (10^{-6}) per year".[51] Although this standard is not applicable to the safety of medical interventions nor to blood products but for the general population, it does explicitly quantify a negligible risk level for an individual. It is clear that in the situation where no additional treatment is applied the risk of death through bacterial contamination will exceed this level, as the estimated probability of death through bacterial sepsis after bacterial culturing will be in the order of 10^{-5} per pooled platelet transfusion. Whether the introduction of bacterial screening provides a sufficient level of safety will depend on the safety level that is considered appropriate and can only be demonstrated if more insight into the incidence of transfusion related sepsis events is obtained by hemovigilance programmes. In case of treatment with PRT the required level of safety could possibly be met irrespective the sepsis incidence rate As no legislation with regard to acceptable contamination levels or acceptable risk of contaminations exists, here the dilemma between maximal (PRT) and optimal (BCU) risk reduction is at hand. Further research and political guidance is required to back up such decisions.

Acknowledgements

The authors would most kindly like to thank from the Sanquin blood supply foundation A. Los for his help in obtaining donation frequency information and D. de Korte PhD for his useful comments on this manuscript. We would also like to thank the referees for their constructive comments, which significantly improved the quality of our paper.

References

1. Goodnough LT, Brecher ME, Kanter MH et al. *Medical progress - Transfusion medicine (First of two parts) - Blood transfusion*. New England Journal of Medicine 1999;340:438-47.
2. Hillyer CD, Josephson CD, Blajchman MA et al. *Bacterial Contamination of Blood Components: Risks, Strategies, and Regulation: Joint ASH and AABB Educational Session in Transfusion Medicine*. Hematology (Am Soc Hematol Educ Program) 2003;575-89.
3. Goodman C, Chan S, Collins P et al. *Ensuring blood safety and availability in the US: technological advances, costs, and challenges to payment--final report*. Transfusion 2003;43:3S-46S.
4. Allain JP. *Transfusion risks of yesterday and of today*. Transfus Clin Biol 2003;10:1-5.
5. de Korte D, Marcelis JH, Verhoeven AJ et al. *Diversion of first blood volume results in a reduction of bacterial contamination for whole-blood collections*. Vox Sang 2002;83:13-6.
6. Wagner SJ. *Transfusion-transmitted bacterial infection: risks, sources and interventions*. Vox Sang 2004;86:157-63.
7. Blajchman MA, Goldman M, Baeza F. *Improving the bacteriological safety of platelet transfusions*. Transfus Med Rev 2004;18:11-24.
8. te Boekhorst PA, Beckers EA, Vos MC et al. *Clinical significance of bacteriologic screening in platelet concentrates*. Transfusion 2005;45:514-9.
9. van Rhenen D, Gulliksson H, Cazenave JP et al. *Transfusion of pooled buffy coat platelet components prepared with photochemical pathogen inactivation treatment: the euroSPRITE trial*. Blood 2003;101:2426-33.
10. McCullough J, Vesole DH, Benjamin RJ et al. *Therapeutic Efficacy and Safety of Platelets Treated with a Photochemical Process for Pathogen Inactivation: The SPRINT Trial*. Blood 2004.
11. Health Council of The Netherlands. *Pathogen reduction in blood products*. Publication no. 2003/16. 2003.
12. Gold MR, Siegel JE, Russel LB, Weinstein MC. *Cost-effectiveness in health and medicine*. New York: Oxford University Press, 1996.
13. Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes;Second edition*. New York: Oxford University Press, 2003.
14. Sachs JD. *Macroeconomics and Health: Investing in Health for Economic Development*. World Health Organization 2001, 2001.
15. Schreiber GB, Busch MP, Kleinman SH et al. *The risk of transfusion-transmitted viral infections. The Retrovirus Epidemiology Donor Study*. N Engl J Med 1996;334:1685-90.
16. Sanquin Blood Supply Foundation. *Sanquin Annual report*. 2003.
17. Korelitz JJ, Busch MP, Kleinman SH et al. *A method for estimating hepatitis B virus incidence rates in volunteer blood donors*. Transfusion 1997;37:634-40.
18. Glynn SA, Kleinman SH, Schreiber GB et al. *Trends in incidence and prevalence of major transfusion-transmissible viral infections in US blood donors, 1991 to 1996. Retrovirus Epidemiology Donor Study (REDS)*. JAMA 2000;284:229-35.
19. Dodd RY, Notari EP, Stramer SL. *Current prevalence and incidence of infectious disease markers and estimated window-period risk in the American Red Cross blood donor population*. Transfusion 2002;42:975-9.
20. de Korte D, Curvers J, de Kort WLAM et al. *Effects of skin disinfection method, deviation bag and bacterial screening on clinical safety of platelet transfusions in the Netherlands.* Transfusion (in press) 2005.
21. Morrow JF, Braine HG, Kickler TS et al. *Septic reactions to platelet transfusions. A persistent problem*. JAMA 1991;266:555-8.
22. Yomtovian R, Lazarus HM, Goodnough LT et al. *A prospective microbiologic surveillance program to detect and prevent the transfusion of bacterially contaminated platelets*. Transfusion 1993;33:902-9.
23. Kuehnert MJ, Roth VR, Haley NR et al. *Transfusion-transmitted bacterial infection in the United States, 1998 through 2000*. Transfusion 2001;41:1493-9.

24. Ness P, Braine H, King K et al. *Single-donor platelets reduce the risk of septic platelet transfusion reactions*. Transfusion 2001;41:857-61.
25. Perez P, Salmi LR, Follea G et al. *Determinants of transfusion-associated bacterial contamination: results of the French BACTHEM Case-Control Study*. Transfusion 2001;41:862-72.
26. Stainsby D, et.al. *Serious Hazards of Transfusion; Shot Annual Report 2003*. 2004.
27. Brecher ME, Hay SN. *Bacterial contamination of blood components*. Clin Microbiol Rev 2005;18:195-204.
28. Evans M, Hastings NAJ, Peacock JB. *Statistical Distributions - 3d Edition*. 3 ed. John Wiley & Sons, 2000.
29. Jackson BR, Busch MP, Stramer SL et al. *The cost-effectiveness of NAT for HIV, HCV, and HBV in whole-blood donations*. Transfusion 2003;43:721-9.
30. AuBuchon JP. *Cost-effectiveness of new blood safety technologies*. Dev Biol Stand 2000;102:211-5.
31. Lopez-Plaza I, Weissfeld J, Triulzi DJ. *The cost-effectiveness of reducing donor exposures with single-donor versus pooled random-donor platelets*. Transfusion 1999;39:925-32.
32. Pereira A. *Cost-effectiveness analysis and the selection of blood products*. Curr Opin Hematol 2000;7:420-5.
33. Bell CE, Botteman MF, Gao X et al. *Cost-effectiveness of transfusion of platelet components prepared with pathogen inactivation treatment in the United States*. Clin Ther 2003;25:2464-86.
34. Pereira A. *Health and economic impact of posttransfusion hepatitis B and cost-effectiveness analysis of expanded HBV testing protocols of blood donors: a study focused on the European Union*. Transfusion 2003;43:192-201.
35. Pereira A, Sanz C. *A model of the health and economic impact of posttransfusion hepatitis C: application to cost-effectiveness analysis of further expansion of HCV screening protocols*. Transfusion 2000;40:1182-91.
36. Brecher ME, Hay SN, Rothenberg SJ. *Validation of BacT/ALERT plastic culture bottles for use in testing of whole-blood-derived leukoreduced platelet-rich-plasma-derived platelets*. Transfusion 2004;44:1174-8.
37. Brecher ME, Hay SN, Rose AD et al. *Evaluation of BacT/ALERT plastic culture bottles for use in testing pooled whole blood-derived leukoreduced platelet-rich plasma platelets with a single contaminated unit*. Transfusion 2005;45:1512-7.
38. Knutson F, Alfonso R, Dupuis K et al. *Photochemical inactivation of bacteria and HIV in buffy-coat-derived platelet concentrates under conditions that preserve in vitro platelet function*. Vox Sang 2000;78:209-16.
39. Hofland M and Verhoeven AJ. *[cost reduction through improved stock management of trombocyte products]*. NVB-Bulletin . 1-3-2004.
40. Picker SM, Speer R, Gathof BS. *Evaluation of processing characteristics of photochemically treated pooled platelets: target requirements for the INTERCEPT Blood System comply with routine use after process optimization*. Transfus Med 2004;14:217-23.
41. Polder JJ, Takken J, Meerding WJ, Kommer GJ, and Stokx LJ. *Costs of illness in The Netherlands (Kosten van ziekte in Nederland)*. [RIVM-report number: 270751005]. 2002. Rijksinstituut voor Volksgezondheid en Milieu, Bilthoven.
42. van der Poel CL, van Noord PA, Smit S et al. *[Epidemiology of blood transfusions]*. Ned Tijdschr Geneeskd 1999;143:1639.
43. Wallis JP, Wells AW, Matthews JN et al. *Long-term survival after blood transfusion: a population based study in the North of England*. Transfusion 2004;44:1025-32.
44. Statistics Netherlands (CBS). *Health statistics: deaths by causes*. www.cbs.nl. 2001. The Hague.
45. Prevolnik Rupel V. *20th Plenary meeting of the EuroQol group*. Republic of Slovenia; Ministry of Health, 2005.
46. Sachs JD. *Macroeconomics and Health: Investing in Health for Economic Development*. World Health Organization 2001, 2001.
47. Brecher ME, Hay SN. *Improving platelet safety: bacterial contamination of platelets*. Curr Hematol Rep 2004;3:121-7.

48. McDonald CP, Roy A, Mahajan P et al. *Relative values of the interventions of diversion and improved donor-arm disinfection to reduce the bacterial risk from blood transfusion*. Vox Sang 2004;86:178-82.
49. Blajchman MA, Goldman M, Freedman JJ et al. *Proceedings of a consensus conference: prevention of post-transfusion CMV in the era of universal leukoreduction*. Transfus Med Rev 2001;15:1-20.
50. Munksgaard L, Albjerg L, Lillevang ST et al. *Detection of bacterial contamination of platelet components: six years' experience with the BacT/ALERT system*. Transfusion 2004;44:1166-73.
51. Ministry of VROM. *Handling Risks (Omgaan met risico's)*. 1989. Zoetermeer.

Cost-effectiveness of additional HBV NAT testing of individual donations or minipools of six donations in the Netherlands

Barbara A. Borkent-Raven
Mart P. Janssen
Cees L. van der Poel
G. Ardine de Wit
Gouke J. Bonsel
Ben A. van Hout

Transfusion 2009; 49(2):311-319
DOI: 10.1111/j.1537-2995.2008.01968.x

CHAPTER

4

“Models are to be used, but not to be believed.”

Henri Theil

Background

To further reduce the risk of hepatitis B virus (HBV) transmission by blood transfusion, nucleic acid testing (NAT) can be employed. The aim of this study is to estimate the incremental cost-effectiveness ratio (ICER) in the Netherlands of employing a triplex NAT assay aimed at HBV nucleic acid detection in individual donations (ID-NAT) or in minipools of 6 donations (MP-6-NAT), compared to a triplex NAT assay in minipools of 24 donations (MP-24-NAT).

Methods

A mathematical model was made of the whole transfusion chain from donors to recipients of blood in the Netherlands. The annual number of avoided HBV transmissions was estimated with the window-period incidence model. The natural history of a HBV infection in recipients is described by a Markov model.

Results

The ICER of adding HBV MP-6-NAT or HBV ID-NAT in the Netherlands is €303,218 (95% confidence interval [CI], €233,001-€408,388) and €518,995 (95% CI, €399,359-€699,120) per quality-adjusted life year, respectively. The ICER is strongly dependent on the age of the transfusion recipient.

Conclusions

The cost-effectiveness of additional HBV NAT is limited by the limited loss of life caused by HBV transmission. Despite a higher effectiveness, HBV ID-NAT is less cost-effective than MP-6-NAT due to higher costs. A future equivalent participation of immigrants from HBV-endemic countries in the donor base renders HBV NAT only slightly more cost-effective.

Introduction

Although all blood transfusions are associated with a residual risk, over the past decades the risk of transmission of viruses has decreased considerably.[1] Nevertheless public concern about transmission of human immunodeficiency virus (HIV), hepatitis B virus (HBV), and hepatitis C virus (HCV) keeps up the pressure to implement extra interventions making blood transfusions even safer.[2] To spend health care budgets efficiently, Dutch public health policies aim at optimal instead of maximum blood safety.[3] This implicates that costs and effects of new screening methods for blood products should be taken into account when deciding whether or not to introduce new (screening) methods.

Additional to serologic screening of all blood donations for anti-HIV-1/2, anti-HCV, hepatitis B surface antigen (HBsAg), anti-HTLV-1/2, and treponemal antibody, HCV nucleic acid testing (NAT) and HIV NAT in minipools of 48 donations were implemented in the Netherlands in 1999 and 2000, respectively. The introduction of HCV NAT and HIV NAT was not based on cost-effectiveness analyses, but it was a result of regulatory requirements for plasma pools for fractionation.

Owing to CE certification requirements for in vitro diagnostics for blood screening the previously published platform for NAT minipool screening, which forms the base of the current program, needs replacement.[4] In Europe only two triplex NAT platforms, both designed for simultaneous NAT of HIV, HBV, and HCV are CE certified. This certification is mandatory for blood donor testing.[5] One platform (Procleix Tigris, Chiron, Emeryville, CA) is primarily designed and CE marked for triplex NAT on individual donations (ID-NAT). The other platform (cobas s 201, Roche, Indianapolis, IN) is primarily designed and CE marked for triplex NAT in minipools of 24 donations (MP-24-NAT) or minipools of 6 donations (MP-6-NAT). This implies that NAT will be performed in pools of at most 24 donations. The MP-24-NAT format is not likely to detect HBV infections in addition to those already detected by present serologic screening with a highly sensitive HBsAg test (Prism, Abbott, Abbott Park, IL).[6] ID-NAT has a higher effectiveness, but is also more expensive than MP-NAT. In the light of cost containment in health care, a cost-effectiveness analysis of installing triplex NAT in a pool size sufficiently small to detect additional HBV infections is performed, either in the MP-6-NAT or ID-NAT format.

Some European countries also screen for antibodies to hepatitis B core antigen (anti-HBc), previously implemented in the United States as a surrogate marker for non-A, non-B hepatitis and more recently as an additional test for the detection of cryptogenic HBV infections. In the Netherlands the request for anti-HBc testing was declined by the minister of health given the poor cost-effectiveness estimates and considerable loss of donors. Therefore, anti-HBc testing is not included in this analysis.

The incremental cost-effectiveness ratio (ICER) can be used as a tool to support the decision on whether newly developed safety measures should be introduced. The ICER reflects the ratio of the additional effects (life-years saved) of a new blood test to its additional costs.[7] This number can be used as an objective measure to compare safety measures. The aim of this study is to estimate the ICER of HBV MP-6-NAT and ID-NAT compared to triplex MP-24-NAT in the Netherlands, a country of low endemicity for hepatitis B.

Materials and Methods

Costs and effects

The additional costs are the costs of testing subtracted by the expected costs related to HBV infection that are avoided as a result of the additional screening. The incremental testing costs for both MP-6-NAT and ID-NAT versus the reference of MP-24-NAT were calculated by Sanquin for the formal annual budget proposals to the Ministry of Health. These encompass all incremental costs within Sanquin, including a.o. reagents, disposables, personnel, and investments. The costs of an HBV infection include direct medical costs and indirect costs due to inability to work.

The effects of implementing MP-6-NAT or ID-NAT are measured in quality-adjusted life-years (QALYs) gained by prevented HBV transmissions to transfusion recipients. QALYs are calculated as the product of life-years experienced after an intervention (here blood transfusion) and the quality of life (QoL) of the recipient during every subsequent year, expressed as a number between 0 (death) and 1 (normal QoL). Hence, life-years are weighted or indexed. The main reason to incorporate QALYs instead of life-years is that an improvement of QoL is also relevant when calculating the effect of avoiding an infection. Because mortality due to HBV transmission is not 100 percent, consequently, it is possible that the number of QALYs lost due to infection will be larger than the number of life-years lost. We do not take into account the existing QoL of transfusion recipients, being aware that in some recipients the contraction of HBV infection only further reduces already compromised QoL. We assume that the resulting overestimation of QoL effects of avoiding HBV infections is counterbalanced by the observation that part of the existing diseases for which blood donations are given progress worse if a co-infection with HBV is present.

Note that we do not consider other causes of HBV infections, but only the incremental infections due to blood transfusion. The low number of HBV infections possibly transmitted by blood transfusion prohibits a clinical experiment measuring the efficacy of the intervention; therefore, a modeling approach is used. A mathematical model of the transfusion chain from donor to recipient is developed to estimate the number of infections avoided by MP-6-NAT or ID-NAT. A Markov model is used to calculate the expected medical costs and the expected number of QALYs lost by one HBV infection.

The number of avoided infections

During early infection, the viral load may be lower than the detection limit of the screening test. The period during which this occurs is called the window period (WP). The effect of MP-6-NAT or ID-NAT is based on the number of infected donations that fall in the WP of the serologic HBsAg test, but not in the WP of the NAT test. This number can be determined using the "window period incidence model".[8] The length of the WP multiplied with the incidence rate among repeat blood donors and the number of donations per year yields the number of window donations per year.

In the Netherlands, donors do not give blood for transfusion at their first attendance. At first visit they are only interviewed and tested. The measured incidence rate must be corrected for the estimated number of incidents not detected by donor screening, which depends on the interdonation intervals among Dutch donors.[9,10]

Two WPs need to be considered in the case of HBV: in the ramp-up phase just after infection and in the early reconvalescence when HBsAg disappears and anti-HBs does not yet neutralize the virus. The length of the WP is determined by two variables: sensitivity of the test and doubling time of the virus during the acute phase and half-life of the virus during reconvalescence. The sensitivity for the current Prism HBsAg test is estimated 3000 IU per mL, which corresponds to the WP reduction experimentally found for this test.[6] The cut-off ratio of HBV NAT is 22.2 IU per mL for MP-6-NAT and 10.9 IU per mL for ID-NAT (H.T.M. Cuijpers, personal communication, 2006). The doubling time of HBV in the acute initial phase of the infection is 2.6 days and the half-life in the recovering phase is 1.6 days.[11] It is assumed that transfusing a blood component containing 30 to 200 mL of plasma and donated during the WP of the serologic HBsAg test, but positive for NAT (either MP-6-NAT or ID-NAT, i.e., containing respectively more than 22.2 or 10.9 IU/mL,) results in a HBV transmission to the recipient.

Description of the disease model

After calculating the number of avoided infections, the costs and effects of one HBV infection caused by blood transfusion are estimated. This is done by use of a Markov model, being an update of the model used to study the cost-effectiveness of HBV vaccination in the Netherlands.[12] Updates concern treatment costs, the possibility of children from 0 to 15 years old being infected by horizontal transmission and the inclusion of QoL. The Markov model describes the natural history of an HBV infection with dependence on age at infection. It consists of five health states for each of which mortality, costs, and QoL are estimated. The transition probabilities indicate the annual probability to move from one health state to another. These probabilities differ according to age. The outcomes of the model are the expected number of QALYs gained by avoiding one infection and the expected costs due to one HBV infection. A lifelong time horizon is used. The costs are expressed in 2005 euros. The model starts with an event tree for the first 6 months of the infection: the acute phase, which is shown in Figure 1. The fraction of asymptomatic patients is estimated at 90.5 percent when the patient is infected at age lower than 5 years of age, 90.2 percent when the infection occurs at ages 5 to 10, 89.7 percent for ages 11 to 15, and 66.7 percent for patients older than 15 years.[13]

It is assumed that a liver transplantation is not performed for patients younger than 5 years of age. It is assumed that, in case the virus is not cleared from the blood after 6 months, the patient becomes a chronic HBV carrier. The higher the age of the infected patient, the lower the probability to become a chronic carrier. This relationship can be described as:[14]

$$P_{chronic}(age) = e^{-0.645(age+0.5)^{0.445}}$$

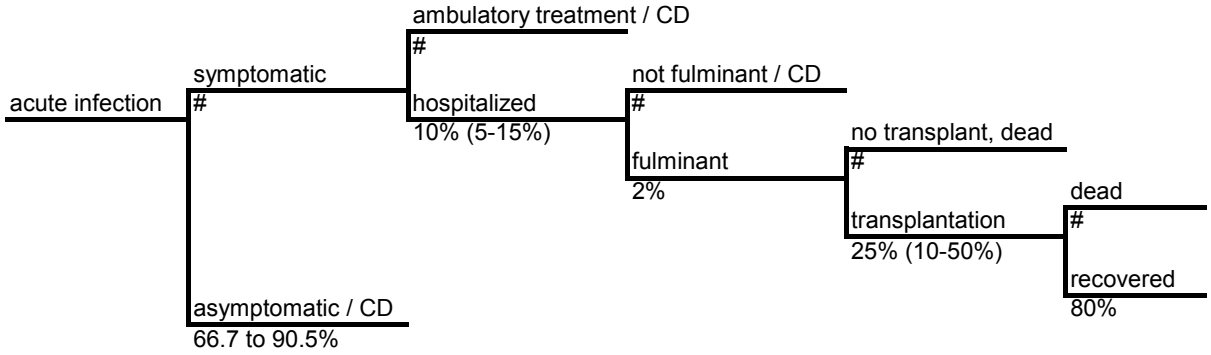


Figure 1: Event tree of the acute phase of HBV infection

A probability indicated with "#" equals 1 minus the other branch probability. Some probabilities can have different values according to the age of the person concerned (ranges are given between brackets). It is assumed that transplantation is not performed in patients younger than 5 years of age. CD = chronic disease

In the chronic stage of infection, two basic states are described by the model: there is either active viral replication (AVR) or there is no AVR, but the virus is still present in the patient's blood ("healthy" carrier). The probability of having AVR is 65 percent (range, 55%-75%) for chronic carriers. In the subsequent years, a chronic HBV carrier can reside in one of the following states: healthy carrier without active AVR, AVR, compensated cirrhosis, decompensated cirrhosis, and hepatocellular carcinoma (HCC). The costs and loss of QoL associated with these infection states are given in Table 1.

Table 1: Costs and loss of quality of life during HBV infection

Costs are given in 2005 Euro's and corrected for inflation (inflation factor 1.183; currency exchange rate €1 = $1.45 in March 2008). Quality of life is given on a scale of 0 to 1, where 1 is perfect health.

State	Costs[17]	Quality of life
ambulatory treatment in acute phase	€373	0.70[18]
acute clinical infection, non fulminant, hospitalized	€2,679	0.60
acute clinical infection, fulminant, hospitalized	€3,779	0.20[18]
Immune	€0‡	1.00
healthy, no AVR	€113‡	0.97
AVR	€1,284‡	0.95[19]
compensated cirrhosis	€1,784‡	0.74[20]
decompensated cirrhosis	€9,924‡	0.66[20]
HCC	€13,717‡	0.65[20]
interferon treatment	€5,825‡	0.91
Transplantation	€79,058	
after transplantation (first six months)	€7,025‡	0.69[20]
after transplantation (subsequent years)	€17,025‡	0.86[19]

‡ treatment costs per year

The uncertainty in costs per disease state is assumed to range from 0.5 to 3 times the point estimate. The transition probabilities for the chronic phase and their ranges are given in Table 2.

Table 2: Annual probabilities used in the Markov model for the chronic phase of HBV infection

It is assumed that transition from one state to another takes place halfway the year.

From	To	Probability (%)	Range (%)	Note
State transition probabilities				
healthy, no AVR	HCC	2.5		
AVR	healthy, no AVR	9	1-15	
AVR	compensated cirrhosis	5.5	1-8.5	from year 6ǂ
compensated cirrhosis	decompensated cirrhosis	3.75	1.5-5	from year 11
compensated cirrhosis	HCC	1.75	1-2.5	from year 11
compensated cirrhosis	dead	5.5	3-7.5	from year 16
decompensated cirrhosis	HCC	7.75	5-10	from year 11
decompensated cirrhosis	dead	22	17-60	
HCC	dead	65	40-80	
Other probabilities				
AVR-patient gets interferon		30	20-50	from year 2 to 5
interferon is succesful		35	25-40	
decomp. patient gets transplantation		12.5	5-40	
recovery after transplantation		60		

ǂ since infection

In the model it is assumed that transitions from one state to another take place halfway through a year. The treatment costs for a particular year therefore depend on the state of the patient at the start and at the end of the year. All future costs are discounted at a rate of 4 percent per year and the life-years at a rate of 1.5 percent per year, as recommended for health economic evaluations in the Netherlands.[15] The age distribution of transfusion recipients is obtained in a pilot study performed in 1997 in the region of the Utrecht blood bank.[16] This study encompassed a random sample of 1000 whole blood donations being traced back to all recipients of the thereof derived components: for example, red blood cells (RBCs), platelets (PLTs), and fresh-frozen plasma.

This study only included recipient age profiles per transfused blood product. It did not include mortality rates or disease variables. Given that short-term mortality of recipients of blood[21] may be influential to the model, we included preliminary recipient survival data from a pilot of the PROTON study on transfusion recipient profiles. From this pilot mortality rates of recipients of blood components in the University Medical Center Utrecht are available, stratified to age. The pilot study includes transfusion recipient information between 1995 and 2003, linked to mortality databases from Statistics Netherlands, which maintains the national mortality registry. Given the availability of these survival data of transfusion recipients, these were incorporated into the model for the first 5 years after transfusion. The survival rates were calculated for the transfusion recipient of age 0 and furthermore for age groups of 5 years. For subsequent years general mortality rates from

Statistics Netherlands are used. The weighted mean of the survival rate in the first 5 years after transfusion is given in Table 3, together with overall survival rates from various other articles.

Table 3: Weighted average of the survival rate after transfusion

Source	Time since transfusion				
	1 year	**2 years**	**3 years**	**4 years**	**5 years**
UMC Utrecht	67%	58%	53%	48%	45%
Tynell[22]	66%		52%		
Kleinman[23]	69%	60%	53%	50%	46%
Wallis[21]		59%			47%
Kamper-Jorgensen[24]	74%				53%

Survival rates for different age groups were comparable to those reported by Kleinman and colleagues.[23] Recipients are assumed not to be vaccinated nor to be immune for HBV infection by resolved previous infection. There is no universal vaccination for HBV for the general public,[25] and data regarding prevalence of protective anti-HBs in hospital patient populations is unknown.

Immigrant donors

Immigration from HBV-endemic countries to the Netherlands is accompanied by elevated HBV incidences,[26] while participation of immigrants as blood donors is being encouraged. This is the reason for studying the change of incidence and ICERs when the fraction of immigrants among donors becomes identical to the general Dutch population, which at present is not yet the case. Information about reported HBV cases in the Netherlands in 2005 is used.[26] Acute HBV infections have to be notified to public health authorities. The National Institute for Public Health and the Environment (RIVM) administers these reports. The country of birth of the HBV-infected persons is registered as well. In 2005, 20 percent of the incidence was registered in immigrants from HBV-endemic countries. It is assumed that this proportion is equivalent to the proportion of immigrant donors among HBV-infected donors, should the fraction of immigrant donors from endemic countries become similar to that of the general population. From 1995 to 2003, there were 57 notified cases of HBV among repeat donors in the Netherlands, of which 5 were immigrants from endemic countries.[27] This information can be used to estimate the incidence among repeat donors in case the fraction of immigrant donors becomes similar to that of the general population. Variations in the proportion of immigrants in the general population and in the donor population are modeled using suitable beta distributions.

Computational issues

The cost-effectiveness model is implemented in a computer worksheet (Microsoft Excel, MS-Excel 2002, Microsoft Corp., Redmond, WA). Simulations and sensitivity analysis are performed with an add-in for MS-Excel (@Risk Professional Version 4.5.2, Palisade Corp., Ithaca, NY).

Results

Incidence and WP

The main numerical results are summarized in Table 4. The measured incidence rate of HBV in the Dutch repeat donor population is 1.43 per 100,000 donor-years. This number must be increased with a factor 3.0 for the estimated number of incidents not detected by donor screening, which results in a corrected incidence rate of $4.3x10^{-5}$ per donor-year. This incidence rate is based on donations in the period 1989 to 2006 in the Netherlands. The number of donor expositions of blood recipients is based on the number of units supplied by Sanquin in 2006 and is estimated at 890,069 donor expositions from RBCs, PLT, and plasma transfusions per year. With the use of the sensitivities of the Prism HBsAg test, ID-NAT, and MP-6-NAT, the doubling time of HBV in the acute initial phase of the infection and the half-life of the virus in the reconvalescent phase, the length of the reduction of the WPs are computed. For MP-6-NAT the mean reduction of the WP compared to the Prism HBsAg test is 18.4 days for the acute phase and 11.3 days for reconvalescence, in total 29.7 days. For ID-NAT the mean reduction of the WP compared to the Prism HBsAg test is 21.1 days for the acute phase and 13.0 days for reconvalescence, in total 34.0 days. Thus, the probability of detecting a MP-6-NAT-positive blood donation that is given during one of the serologic WPs is 3.5 per 1,000,000 donations, preventing 3.12 (95% confidence interval [CI], 2.62-3.66) HBV cases per year. The probability of finding an ID-NAT-positive blood donation that is given during one of the serologic WPs is 4.0 per 1,000,000 donations, preventing 3.57 (95% CI, 3.01-4.19) cases per year.

The ICERs

The estimated incremental annual costs for implementing triplex MP-6-NAT or ID-NAT in the Netherlands compared to triplex MP-24-NAT are calculated by Sanquin at €959,600 and €1,877,785 respectively (H. Bos, personal communication, 2006). The Markov model in combination with the age distribution of the recipients of blood components yields expected treatment costs of €1122 (95% CI, €532-€2320) per HBV infection. The mean number of QALYs that are lost as a result of one HBV infection is estimated at 1.01 (95% CI, 0.80-1.22). The ICER of HBV NAT in the Netherlands is estimated at €303,218 (95% CI, €233,001-€408,388) per QALY for MP-6-NAT and €518,995 (95% CI, €399,359-€699,120) for ID-NAT. The ICER is also calculated for the future scenario where the fraction of immigrants from endemic countries becomes identical to the fraction of these immigrants in the general Dutch population. In the Netherlands 299 incident cases of hepatitis B were reported in 2005.[26] Of all HBV patients 20 percent are born in a country where hepatitis B is endemic, whereas 8.69 percent of all inhabitants are born in these countries.[28] It can be calculated that for the future situation the incidence of HBV among donors will increase by 8 percent and will become 4.6 per 100,000 donor-years. Therefore, if immigrant donors are participating proportionally in the donor base in the Netherlands, the ICER becomes €280,835 (95% CI, €216,015-€379,793) per QALY for MP-6-NAT and €480,742 (95% CI, €370,067-€650,374) per QALY for ID-NAT.

Table 4: Summary of main quantitative results

Description	MP-6-NAT	ID-NAT	MP-6-NAT; immigrants included	ID-NAT; immigrants included
cut-off per donation IU/ml	22.2	10.9	22.2	10.9
first window period reduction	18.4	21.1	18.4	21.1
second window period reduction	11.3	13.0	11.3	13.0
window period reduction (compared to HBsAg)	29.7	34.0	29.7	34.0
incidence of HBV among donors	4.3×10^{-5}	4.3×10^{-5}	4.6×10^{-5}	4.6×10^{-5}
number of cases prevented	3.12	3.57	3.36	3.85
cost per case prevented	€308,001	€526,383	€285,348	€487,669
Incremental Cost-effectiveness Ratio (ICER) in €/QALY	€303,218	€518,995	€280,835	€480,742

Sensitivity analysis

Model sensitivity is analyzed by regression analyses of output variables on model input variables. The model sensitivity is expressed in standardized regression coefficients (SRCs). These show that the uncertainty in costs prevented per case is primarily related to the uncertainty in cost of treatment (SRC, 89%; R^2 = 93%) and second by the recovery rate of patients who contracted an HBV infection (SRC, -25%). The uncertainty in QALYs lost is primarily determined by the transition probability from AVR to compensated cirrhosis (SRC, 26%; R^2 = 85%). The uncertainty in cost-effectiveness is primarily caused by the HBV incidence rate (SRC, -87%; R^2 = 75%) to which it is roughly inversely proportional: double the incidence rate would halve the ICER. Approximately 20 percent of the variance in the ICER is induced by the limited sample size of the transfusion recipient population (817 samples). Because the costs of treatment are negligible compared to the screening costs, the ICER is insensitive to the discount rate for cost (4% used in this analysis) and the ICER increases almost linearly with increasing screening costs. The ICERs reduce with approximately 50 percent when health outcomes are not discounted. When discount rates of 3 percent are used, like in similar studies, the ICERs grow to €550,344 per QALY for MP-6-NAT and €942,249 per QALY for ID-NAT.

Recipient age is an important variable in the cost-effectiveness of screening. To illustrate this, the ICER for both NATs is calculated per patient age (Figure 2). The ICER for MP-6-NAT ranges from approximately €21,000 in neonates to €11 million in patients 95 years of age. The ICER for ID-NAT ranges from approximately €37,000 for neonates to €19 million in patients 95 years of age. The QALYs gained in the current results are mainly due to avoiding infections in children. They have both a high life expectancy and a high probability of becoming a chronic HBV carrier, which results in cost-effective screening.

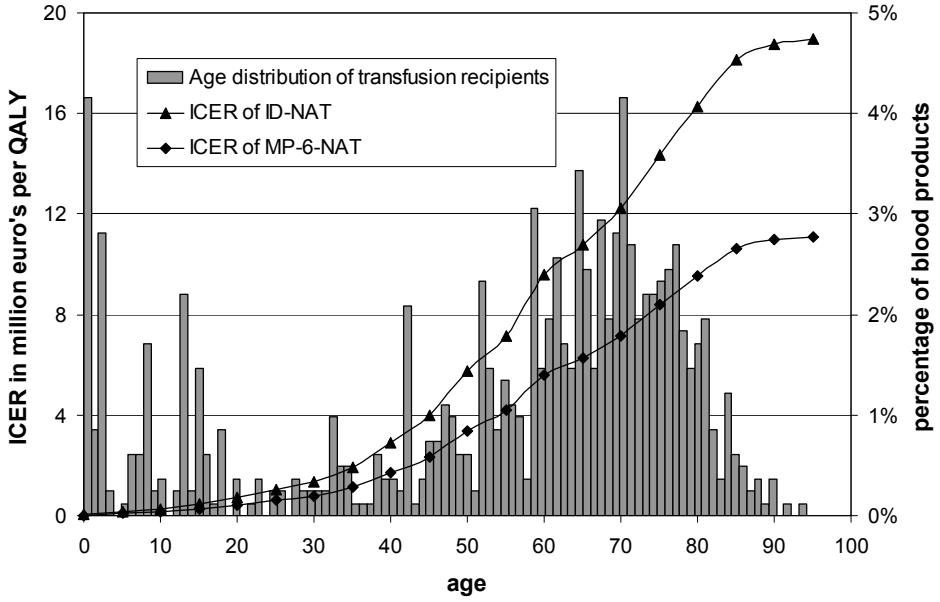


Figure 2: Sensitivity analysis of ICERs of HBV NAT to patient age at constant testing costs

Discussion

Application of the results

It is shown that in the Netherlands, implementation of triplex MP-6-NAT is more cost-effective for prevention of HBV transmission by blood components than triplex ID-NAT, even though the latter is more sensitive. The additional risk reduction achieved by the ID-NAT compared to MP-6-NAT comes at relatively too high a price. The ICER of prevention of HBV transmission by MP-6-NAT or ID-NAT compared to MP-24-NAT, the latter not being designed for additional HBV detection, amounts to €303,218 (95% CI, €233,001-€408,388) and €518,995 (95% CI, €399,359-€699,120) per QALY, respectively. This ICER can become €280,835 (95% CI, €216,015-€379,793) per QALY for MP-6-NAT and €480,742 (95% CI, €370,067-€650,374) per QALY for ID-NAT when immigrants from HBV-endemic countries join the donor population on an equivalent demographic basis. The influx of immigrants only slightly affects the ICER of the HBV NAT tests.

For application to preventive and/or curative health care interventions, several threshold values for the ICER are suggested, ranging from €20,000 to more than €100,000 per QALY.[29] In these policy views, however, producer's risks such as liability and public concern, resulting in public enquiries where blood safety is scrutinized,[2] are not accounted for. Neither are the costs thereof. An alternative, but not widely used, policy view could be analogous with the requirements for sterility testing in pharmaceutical products, for example, a probability of less than 10^{-6} of the end product being contaminated with one detectable infectious unit would be considered a "sterile" intravenous pharmaceutical. At present there is no consensus on an acceptable cost-effectiveness threshold value for blood safety measures. When we compare our ICERs of

HBV NAT with the cost-effectiveness of other blood safety measures, such as HCV NAT and HIV NAT, it appears that in the past ICERs of millions of euros per QALY did not deter decision makers from introducing such screening.[30,31] Thus, among NATs, HBV NAT is relatively more cost-effective, even in a low-endemic country such as the Netherlands. In general, interventions that are directed at the avoidance of risks, rather than the reduction of mortality or morbidity, are associated with higher ICERs. Society is willing to pay more for risk avoidance.[32]

Strengths and weaknesses of the model

In this analysis the WP incidence model for donors was used to calculate the number of avoided HBV infections. Besides during the two mentioned WPs, the HBV NAT test can also be positive for donations that are negative for the presence of HBsAg when there is still replication at low level after recovery or in case of an escape mutant of the virus. This is called an occult hepatitis B virus infection. The frequency of occult hepatitis B virus infections in the Netherlands donor population is unknown, but the number of averted contaminated blood products might be higher than estimated here. On the other hand, the level of infectivity of these products is unclear, while donations given during the WP of acute hepatitis B are known to be highly infectious.[33]

Possible secondary transmission of HBV from blood recipients to their partners is not included in this model. The incorporation of such secondary transmissions would improve the cost-effectiveness as estimated in our study. However, as the mean age of blood recipients is rather high and the recipients are sick patients, secondary transmission can be expected to be lower than secondary transmission in the general population and this risk is essentially not known.

The likelihood of chronic HBV carriership may be higher in immune-compromised transfusion recipients. This might increase the costs of treatment for HBV and loss of QALYs and could reduce the ICER of HBV NAT.

The model for age distribution and survival of blood recipients in the Netherlands is the subject of ongoing research (PROTON study). Life expectancies of patients as available for one academic hospital are used here. The life expectancy of blood recipients in the Netherlands in general might be higher, because the patient population in university hospitals is more severely ill than in general hospitals. Assigning patient categories to the health economic effects of HBV NAT, for instance, limiting the application of HBV NAT to transfusion products for pediatric transfusions only, introduces another policy view. Note that the costs per donation are likely to increase when the number of tested donations is reduced; this is not further elaborated in our model because it does not seem to be a likely scenario in the Netherlands.

Comparison with other studies

The ICER depends on the reference scenario, which is a very sensitive test for HBsAg plus triplex MP-24-NAT in the Netherlands. In countries with a less sensitive – previous generation – HBsAg test, the yield of the HBV NAT might be higher. The effectiveness of the additional HBV screening is largely dependent on the incidence rate. In regions with a high incidence, like Mediterranean European countries, the yield of the additional HBV screening will be higher. Other ICERs for HBV NAT are in the literature.[18,34-36] The ICERs presented in four referenced articles range from €4.9 million to €66 million per QALY.

These articles compare individual donor HBV NAT to HBsAg testing either by Prism or other assays. In addition, the three articles by Busch, Jackson, and Marshall compare HBV NAT in minipools of 16 to 24 donors to the HBsAg test. However, none of the studies is similar to our study design: the change from triplex MP-24-NAT (not detecting HBV) to either triplex MP-6-NAT or to triplex ID-NAT. Jackson, Marshall, and Pereira describe details on their Markov models. There are some slight differences between the models. One model contains fewer health states for HBV infection than our model[34] and one of the models yields some lower transmission probabilities.[18] Further, our analysis is based on an age distribution of transfusion recipients. Age distributions appear to be a very sensitive parameter in the model. Only the study of Pereira employs an age distribution for the recipients.[18] In the article of Pereira a similar length of WP reduction is computed.[18] The Busch, Marshall, and Jackson models yield smaller WP reductions, rendering higher ICERs.[34-36] This may explain the difference between the Busch and Pereira studies. From these studies it can be concluded that HBV MP-24-NAT in comparison to the HBsAg test is not cost-effective. Nevertheless, when triplex MP-24-NAT is already introduced due to EC regulations, the ICER of reducing the pool size to six or one appears to be more cost-effective than previously taken blood safety measures.

In conclusion, in the Netherlands, reducing the minipool size of a CE-marked triplex NAT assay from 24 into 6 donations yields a relatively cost-effective prevention of HBV transmission compared to other blood safety measures that have already been implemented. Furthermore, triplex MP-6-NAT is more cost-effective for prevention of HBV transmission than triplex ID-NAT, despite the fact that the latter assay is more sensitive. A future equivalent participation of immigrants from HBV-endemic countries in the donor base renders HBV NAT only slightly more cost-effective.

Acknowledgments

We thank H.T.M. Cuijpers from Sanquin Blood Supply Foundation for providing the cut-off ratios of some HBsAg and HBV DNA tests. We also thank H. Bos from Sanquin Blood Supply Foundation for providing the testing cost estimates used in this article. We would like to acknowledge the Netherlands Organization for Health Research and Development (ZonMw project 945-06-603) for funding this research.

References

1. AuBuchon JP. *Emily Cooley Memorial Award. Managing change to improve transfusion safety*. Transfusion 2004;44:1377-83.
2. Wilson K. *The Krever Commission--10 years later*. CMAJ 2007;177:1387-9.
3. van der Poel CL, de Boer JE, Reesink HW et al. *[Optimal versus maximal safety of the blood transfusion chain in The Netherlands; results of a conference. College for Blood Transfusion of the Dutch Red Cross]*. Ned Tijdschr Geneeskd 1998;142:285-8.
4. Cuypers HT, van DR, Viret JF et al. *European multi-centre validation study of NucliSens Extractor in combination with HCV Amplicor 2.0 assay for HCV-NAT screening of plasma pools*. Biologicals 1999;27:303-14.
5. *Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices*. 2008.
6. Koppelman MH, Assal A, Chudy M et al. *Multicenter performance evaluation of a transcription-mediated amplification assay for screening of human immunodeficiency virus-1 RNA, hepatitis C virus RNA, and hepatitis B virus DNA in blood donations*. Transfusion 2005;45:1258-66.
7. Meltzer MI. *Introduction to health economics for physicians*. Lancet 2001;358:993-8.
8. Kleinman S. *Residual risk of transfusion transmitted viral infections among seronegative donors: application of the incidence/window period model*. Dev Biol Stand 2000;102:61-5.
9. Korelitz JJ, Busch MP, Kleinman SH et al. *A method for estimating hepatitis B virus incidence rates in volunteer blood donors. National Heart, Lung, and Blood Institute Retrovirus Epidemiology Donor Study*. Transfusion 1997;37:634-40.
10. Janssen MP, van der Poel CL, Buskens E et al. *Costs and benefits of bacterial culturing and pathogen reduction in the Netherlands*. Transfusion 2006;46:956-65.
11. Yoshikawa A, Gotanda Y, Itabashi M et al. *HBV NAT positive [corrected] blood donors in the early and late stages of HBV infection: analyses of the window period and kinetics of HBV DNA*. Vox Sang 2005;88:77-86.
12. de Wit G, Kretzschmar M, Smits L, Struijs J, Postma M, van de Laar M, and Jager J. *Kosten-effectiviteit van algemene vaccinatie tegen hepatitis B – (interimrapportage)*. RIVM rapport 403 505 004. 2000.
13. McMahon BJ, Alward WL, Hall DB et al. *Acute hepatitis B virus infection: relation of age to the clinical expression of disease and subsequent development of the carrier state*. J Infect Dis 1985;151:599-603.
14. Edmunds WJ, Medley GF, Nokes DJ et al. *The influence of age on the development of the hepatitis B carrier state*. Proc Biol Sci 1993;253:197-201.
15. Rodenburg-van Dieten HEM. *Richtlijnen voor farmaco-economisch onderzoek; evaluatie en actualisatie*. RIVM rapport 24088606. 27-10-2005.
16. van Noord PAH, van der Poel CL, and van den Boom FMLG. *Epidemiologisch onderzoek bij ontvangers van bloedproducten*. Nederlands Tijdschrift voor de Geneeskunde 141[1126]. 1997.
17. Beutels Ph, Tormans G, Van Damme P et al. *Algemene vaccinatie tegen hepatitis B in Vlaanderen: kosten-effectief?* Tijdschrift voor Sociale Gezondheidszorg 1996;74:272-81.
18. Pereira A. *Health and economic impact of posttransfusion hepatitis B and cost-effectiveness analysis of expanded HBV testing protocols of blood donors: a study focused on the European Union*. Transfusion 2003;43:192-201.
19. Siebert U, Sroczynski G, Rossol S et al. *Cost effectiveness of peginterferon alpha-2b plus ribavirin versus interferon alpha-2b plus ribavirin for initial treatment of chronic hepatitis C*. Gut 2003;52:425-32.
20. Chong CA, Gulamhussein A, Heathcote EJ et al. *Health-state utilities and quality of life in hepatitis C patients*. Am J Gastroenterol 2003;98:630-8.
21. Wallis JP, Wells AW, Matthews JN et al. *Long-term survival after blood transfusion: a population based study in the North of England*. Transfusion 2004;44:1025-32.

22. Tynell E, Norda R, Shanwell A et al. *Long-term survival in transfusion recipients in Sweden, 1993*. Transfusion 2001;41:251-5.
23. Kleinman S, Marshall D, AuBuchon J et al. *Survival after transfusion as assessed in a large multistate US cohort*. Transfusion 2004;44:386-90.
24. Kamper-Jorgensen M, Ahlgren M, Rostgaard K et al. *Survival after blood transfusion*. Transfusion 2008.
25. Gezondheidsraad. *Algemene vaccinatie tegen hepatitis B*. 2001. Den Haag, Gezondheidsraad.
26. de Boer IM, op de Coul ELM, Koedijk FDH, van Veen MG, and van de Laar MJW. *HIV and sexually transmitted infections in the Netherlands in 2005*. RIVM rapport 441100024/2006. 2006.
27. van der Bij AK, Coutinho RA, van der Poel CL. *Surveillance of risk profiles among new and repeat blood donors with transfusion-transmissible infections from 1995 through 2003 in the Netherlands*. Transfusion 2006;46:1729-36.
28. *Statline, electronic database of Statistics Netherlands, www.cbs.nl*. 6-10-2007.
29. RVZ. *Zinnige en duurzame zorg*. 2006. Zoetermeer, Raad voor de Volksgezondheid en Zorg.
30. van HM, de Wolf JT, Staginnus U et al. *Pharmaco-economics of blood transfusion safety: review of the available evidence*. Vox Sang 2002;83:146-55.
31. Custer B. *Economic analyses of blood safety and transfusion medicine interventions: a systematic review*. Transfus Med Rev 2004;18:127-43.
32. Tengs TO, Adams ME, Pliskin JS et al. *Five-hundred life-saving interventions and their cost-effectiveness*. Risk Anal 1995;15:369-90.
33. Reesink HW, Engelfriet CP, Henn G et al. *Occult hepatitis B infection in blood donors*. Vox Sang 2008;94:153-66.
34. Jackson BR, Busch MP, Stramer SL et al. *The cost-effectiveness of NAT for HIV, HCV, and HBV in whole-blood donations*. Transfusion 2003;43:721-9.
35. Busch MP. *Should HBV DNA NAT replace HBsAg and/or anti-HBc screening of blood donors?* Transfus Clin Biol 2004;11:26-32.
36. Marshall DA, Kleinman SH, Wong JB et al. *Cost-effectiveness of nucleic acid test screening of volunteer blood donations for hepatitis B, hepatitis C and human immunodeficiency virus in the United States*. Vox Sang 2004;86:28-40.

Enhancing Value of Information analyses

Mart P. Janssen
H. (Erik) Koffijberg

Value in Health 2009; 12(6):935-941
DOI: 10.1111/j.1524-4733.2009.00548.x

CHAPTER

5

“To know one’s ignorance is the best part of knowledge.”

Lao Tzu

Objective

To demonstrate that it is feasible and recommendable to present value of information outcomes in terms of underlying costs and effects in addition to costs alone.

Methods

The benefits of collecting additional information on health economic outcomes before deciding on a preferred policy when evaluating alternative strategies with uncertain outcomes are quantified in a value of information analysis. In general, costs and effects are combined into one single dimension to determine the expected monetary value of information. Separate information on costs and effects is lost. This information, however, remains relevant to the decision maker. The concept of the attributable value of information is introduced where the expected value of information is expressed in terms of the expected changes in health outcomes.

Results

The attributable expected value of perfect information is calculated for a number of hypothetical examples. These examples are used to demonstrate the benefits of the new approach as well as its calculation. Benefits are: (1) insight into the expected costs and expected effects gained as a result of carrying out further research to reduce or eliminate decision uncertainty, and (2) the likelihood that the outcome of additional research will result in a change in preferred policy.

Conclusions

Decision-making may be enhanced and clarified by attributable value of information analysis in addition to current analyses. Since obtaining the attributable value of information results is straightforward, and requires only a minimal computational effort, its calculation is recommended for all future value of information analyses.

Introduction

With the continuous rise of health care costs, efficiency of health care and health care research is becoming increasingly important. Value of information (VOI) analyses allows a systematic quantification of the expected financial gain from additional information for decisions made under uncertainty. These kind of analyses are increasingly applied in health policy research and health technology assessment.[1-5] VOI outcome is proposed as a measure for setting priorities on additional research amongst research projects, as it indicates a project's potential for return on investment. However, there is a price to be paid for obtaining the VOI. In a conventional comparison of health care interventions outcomes are projected in a cost-effectiveness plane, giving information on incremental costs and effects for all interventions. Calculating the VOI is done within a Net Monetary Benefit (NMB) framework. Within this framework the setting of a particular cost-effectiveness threshold is defined which provides a means for valuating health effects in terms of their monetary equivalent. This enables a decision on the preferred intervention for any combination of intervention outcomes. This is the one with the highest monetary equivalent, i.e. the highest NMB. However, as costs and effects are combined into one single dimension, information on each of the separate dimensions is lost. This information, however, can be – and we will argue that it is – relevant to the decision maker.

We develop the concept of 'Attributable Value of Information' (AVOI). The AVOI is a straightforward extension of the general VOI concept whereby it is split into separate parts for costs and effects that add up to the total VOI. We will demonstrate and apply this concept to the measure of Expected Value of Perfect information (EVPI). The AVOI allows the decision maker to judge not only the value of additional research, but also its impact in terms of the expected shift in costs and effects as compared to the currently preferred intervention. These additional insights improve a decision maker's ability to evaluate the merits of additional research.

Methods

Standard EVPI

Measures such as the EVPI and the Expected Value of Sample Information (EVSI) may be used for research prioritization, based on comparisons between expected benefits and costs of future research projects.[6] The EVPI was introduced in the 1960s and has a straightforward interpretation: it is the difference between the expected utility of the decision when no uncertainty exists (perfect information) and expected utility of the decision based upon current evidence.[7] In a health economic context effects are usually expressed in terms of Quality Adjusted Life Years (QALYs). Subsequently, the NMB is used to determine the utility of a particular outcome of an intervention.[8] The NMB (Equation 1) is a simple linear function of the cost-effectiveness threshold (CET), costs (C in $) and effects ($Q$ in QALYs). The preferred intervention is the one with the highest

expected NMB. The EVPI as function of the CET over all interventions j with respect to the parameter space Ω of costs and effects (with realizations ω) is given by Equation 2.

$$NMB(j,\omega,CET) = CET \cdot Q(j,\omega) - C(j,\omega) \quad (1)$$

$$EVPI(j,\omega,CET) = E_{\Omega}\left[max_j\left(NMB(j,\omega,CET)\right)\right] - max_j\left(E_{\Omega}\left[NMB(j,\omega,CET)\right]\right) \quad (2)$$

For each intervention the bivariate (correlated) distribution of intervention outcomes (costs and effects) is required for the calculation of the EVPI. The second part of Equation 2 refers to the mean NMB for each of the interventions considered. Conversely, the first part of this equation refers to the mean (of the maximum) NMB for each potential combination of outcomes separately. This mimics the situation where the best decision is taken in each viable situation.

Despite the simple form of Equation 2, it does not reveal any characteristics of the EVPI. For example, it does not reveal how the EVPI is affected by differences in costs and effects between interventions and their dispersion. This complicates the understanding of the EVPI at the basic level: EVPI values cannot be interpreted in terms of costs and effects.

EVPI and the segregation of costs and effects

The transformation of intervention outcomes to a NMB is required to decide on the preferred intervention, which is the one with the highest expected NMB. In the calculation of the EVPI the outcome of either intervention is expressed in the NMB which results in a monetary equivalent in which the EVPI is expressed. However, as the NMB in Equation 2 for any given j and ω is calculated using Equation 1 with specific cost and effect outcomes, the contribution of costs and effects to the EVPI can be separated. This only requires taking the respective costs or effects associated with the intervention that maximizes the NMB of Equation 2. By doing this the attributable costs (Equation 3a) and attributable effects (Equation 3b) of the EVPI are obtained.

$$AEVPI_{Cost}(j,\omega,CET) = E_{\Omega}\left[f_c\left(max_j\left(NMB(j,\omega,CET)\right)\right)\right] - f_c\left(max_j\left(E_{\Omega}\left[NMB(j,\omega,CET)\right]\right)\right) \quad (3a)$$

$$AEVPI_{Effect}(j,\omega,CET) = E_{\Omega}\left[f_e\left(max_j\left(NMB(j,\omega,CET)\right)\right)\right] - f_e\left(max_j\left(E_{\Omega}\left[NMB(j,\omega,CET)\right]\right)\right) \quad (3b)$$

The functions f_c and f_e in Equation 3 refer to the respective underlying costs and effects associated with the NMB they are applied to. The fact that the NMB is a linear combination of costs and effects implies that the functions f_c and f_e can also be applied to the expected value of the NMB. The results are equivalent to the expected costs and expected effects respectively. Also, the attributable effects multiplied by the CET minus the attributable costs are equivalent to the EVPI. A formal derivation of the attributable EVPI is given in the Appendix.

Simulations to illustrate the additional value of the attributable EVPI

To illustrate how the attributable EVPI (AEVPI) can support decision-making three examples are given. For each example the standard EVPI and AEVPI are calculated. In addition, a step by step illustration of how the AEVPI is calculated will be provided for one of the examples. Calculations were performed using Microsoft Excel® (MS-Excel 2002, Microsoft Corp., Redmond, WA, USA) and R® (version 2.7.2, The R Foundation for Statistical Computing, Vienna, Austria). A fully accessible spreadsheet and R code containing the examples presented in this paper can be obtained from the authors upon request.

Results

Calculation of the EVPI requires a probabilistic estimation of costs and effects of two interventions and a CET. For both interventions, the distributions of costs and effects are generally presented graphically in a cost-effectiveness plane. Figure 1 shows an example of a cost-effectiveness plane with a generic decision problem concerning two interventions; intervention (1) with expected outcome P_1, and intervention (2) with expected outcome P_2. The uncertainty in the outcomes of P_1 and P_2 is visualized as a 95% confidence interval (CI) for their respective costs and effects. In this example costs and effects follow uncorrelated normal distributions. Mean costs and effects as well as the dispersion of these outcomes are known: for intervention (1) the mean costs are $1000 with a standard deviation (SD) of $600. The mean effect is 0.2 QALYs (SD=0.3 QALYs); intervention (2) has mean costs of $400 (SD=$1000) and mean effect of 0.1 QALYs (SD=0.2 QALYs). In Figure 1 a line is drawn through P_1 which represents a CET of $25,000 per QALY. Any point on this line has an NMB that is equal to that of intervention (1). Given this CET, intervention (1) is the preferred intervention since P_2 lies above the CET line, which implies that P_2 has the lesser expected NMB. The expected NMB of intervention (1) equals 0.2 QALYs x $25,000 per QALY - $1,000 = $4,000. The expected NMB of intervention (2) equals 0.1 QALYs x $25,000 per QALY - $400 = $2,100. However, although the expected NMB of intervention (1) exceeds the expected NMB of intervention (2) there is a chance, given the uncertainty in costs and effects of both interventions, that the NMB of intervention (2) is the highest. In that case intervention (2) is to be preferred. Points P_1^* and P_2^* represent the expected costs and expected effects of those outcomes that have either intervention (1) or (2) respectively as the preferred intervention. Point A indicates the mean costs and effects of the preferred intervention (irrespective of this being (1) or (2)). Point A is the average of P_1^* and P_2^*, weighting each of these with their respective probability of being the preferred intervention. The cost difference between P_1 and A represents the attributable EVPI cost, whereas the effect difference between P_1 and A represents the attributable EVPI effect. This is the graphical equivalent of Equation 3.

Example calculations

The calculation of the attributable EVPI is illustrated using the example given in Figure 1. The results are shown in Table 1. This table shows simulated costs and effects for both interventions. Samples were drawn randomly from aforementioned normal distributions. From Table 1 it can be found that the estimated average NMB on basis of 20 samples is

$4371 and $1,025 for interventions (1) and (2) respectively ('Sample average' row of columns C and F). Intervention (1) would be (correctly) considered the preferred intervention on the basis of this limited sample. The attributable EVPI cost ($AEVPI_{Cost}$) is estimated at -$314 and the attributable EVPI effect ($AEVPI_{Effect}$) at 0.084 QALYs. The attributable effect value, i.e. the cost-equivalent of the attributable effect ($AEVPI_{Effect\text{-}Value}$), is found by multiplying the attributable effect with the CET (0.084 x 25,000=$2,100). Note that the attributable effect value minus the attributable costs is indeed – apart from rounding errors – equal to the EVPI ($2410).

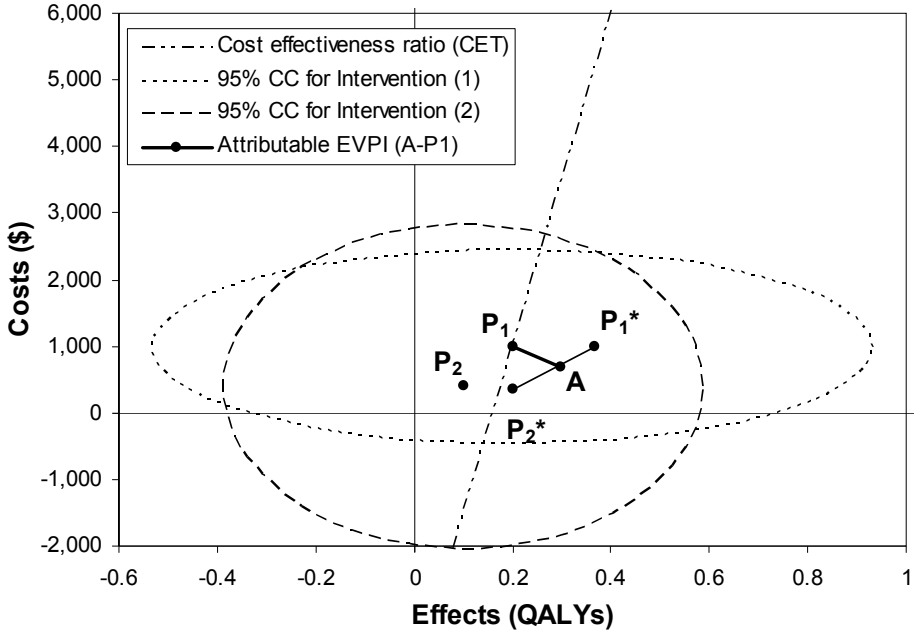


Figure 1: Attributable EVPI of Example 1

Shown are the mean costs, effects and 95% confidence contours of two interventions (P_1 and P_2), their respective expected mean costs and effects conditional on being the preferred intervention whilst knowing their exact outcomes (P_1^* and P_2^*), and the expected outcome conditional on perfect information (A). The attributable EVPI is represented by the difference between A and P_1 and consists of attributable costs and attributable effects.
Intervention 1 (P1) is normally distributed with costs $1,000 (SD=$600) and effects 0.2 QALYs (SD=0.3 QALYs). Intervention (2) (P2) is normally distributed with costs $400 (SD=$1000) and effects 0.1 QALYs (SD=0.2 QALYs). The EVPI for this decision problem is $2,755, the $AEVPI_{Cost}$ and $AEVPI_{Effect}$ are -$309 and 0.098 QALYs respectively.

It is common to represent the value of the EVPI graphically as a function of the CET. In Figure 2 this relation is shown for the EVPI and, in addition, for the attributable costs and attributable effect-values. Please note that minus the $AEVPI_{Cost}$ is shown so that the sum of the AEVPI lines equals the (total) EVPI. From Figure 2 is apparent that for a CET less than $6,000 per QALY, both the $AEVPI_{Effect\text{-}Value}$, and $AEVPI_{Cost}$ are increasing. As the $-AEVPI_{Cost}$ line is declining the $AEVPI_{Cost}$ is increasing. Here intervention (2) is the preferred intervention. At $6000 per QALY the decision maker is indifferent as the expected NMB is equal for both interventions. For larger CET values intervention (1) is preferred. At a CET of zero intervention (1) will only be preferred in case it results in lower costs. Therefore, the $AEVPI_{Cost}$ at this point is negative, and minus the $AEVPI_{Cost}$ in Figure 2 is positive. As long as there is uncertainty with respect to the sign of the relative effect outcome, the contribution of the $AEVPI_{Effect\text{-}Value}$ to the EVPI will increase with an increasing CET. This is the cause for the unlimited growth found in many EVPI curves.

Table 1: Simulated data for illustration of the attributable EVPI (AEVPI) calculation

Column:	A	B	C	D	E	F	G	H	J	K	L	M	N	P
Description:	Intervention (1) (OPI)			Intervention (2)			Preferred intervention number	Preferred intervention per sample			Differences preferred intervention per sample and overall preferred†			Direction of change
Units:	Cost ($)	Effect (QALY)	NMB ($)*	Cost ($)	Effect (QALY)	NMB ($)*		Cost ($)	Effect (QALY)	NMB ($)*	Cost ($)	Effect (QALY)	NMB ($)*	(quadrant number)
Formula:		B x CET - A			E x CET - D		If (F>C) then 2 else 1	If (G=1) then A else D	If (G=1) then B else E	If (G=1) then C else F	H - Average $Cost_{OPI}$	J - Average $Effect_{OPI}$	K – Average NMB_{OPI}	‡
	687	-0.36	-9,645	56	-0.07	-1,921	2	56	-0.07	-1,921	-923	-0.29	-6,292	↘ (Q4)
	886	0.65	15,295	-576	0.07	2,397	1	886	0.65	15,295	-93	0.43	10,924	
	964	0.65	15,306	-187	0.08	2,123	1	964	0.65	15,306	-15	0.44	10,934	
	788	0.09	1,574	1,672	0.24	4,402	2	1,672	0.24	4,402	693	0.03	30	↗ (Q1)
	1,132	0.14	2,349	583	0.41	9,664	2	583	0.41	9,664	-396	0.20	5,292	↘ (Q4)
	1,925	0.15	1,820	-93	0.09	2,431	2	-93	0.09	2,431	-1,072	-0.12	-1,940	↙ (Q3)
	423	0.67	16,297	-1,055	-0.04	74	1	423	0.67	16,297	-556	0.45	11,926	
	628	0.24	5,470	-892	0.17	5,214	1	628	0.24	5,470	-351	0.03	1,099	
	858	-0.12	-3,841	123	-0.07	-1,807	2	123	-0.07	-1,807	-857	-0.28	-6,178	↘ (Q4)
Samples:	559	0.27	6,275	234	0.39	9,495	2	234	0.39	9,495	-745	0.18	5,124	↘ (Q4)
	1,134	-0.26	-7,532	-517	-0.21	-4,763	2	-517	-0.21	-4,763	-1,496	-0.43	-9,134	↘ (Q4)
	1,516	0.17	2,748	2,228	-0.43	-13,100	1	1,516	0.17	2,748	537	-0.04	-1,624	
	859	0.30	6,575	1,899	0.04	-932	1	859	0.30	6,575	-120	0.08	2,204	
	1,084	-0.22	-6,462	-1,168	0.11	3,864	2	-1,168	0.11	3,864	-2,147	-0.11	-507	↘ (Q4)
	1,302	0.67	15,344	-1,586	-0.05	326	1	1,302	0.67	15,344	323	0.45	10,973	
	1,117	0.77	18,091	992	-0.24	-6,908	1	1,117	0.77	18,091	138	0.55	13,720	
	748	0.17	3,589	865	0.15	2,910	1	748	0.17	3,589	-232	-0.04	-783	
	-10	0.42	10,502	115	0.24	5,827	1	-10	0.42	10,502	-989	0.21	6,131	
	2,311	-0.23	-8,150	3,309	0.26	3,221	2	3,309	0.26	3,221	2,330	0.05	-1,150	↗ (Q1)
	675	0.10	1,816	3,644	0.06	-2,025	1	675	0.10	1,816	-304	-0.11	-2,555	
											$AEVPI_{Cost}$	$AEVPI_{Effect}$	**EVPI**	
Sample average:	979	0.21	4,371	482	0.06	1,025	1.45	666	0.30	6,781	-314	0.084	2,410	Q2=0%, Q1=10%, Q3=5%, Q4=30%
Exact average (n→∞):	1,000	0.20	4,000	400	0.10	2,100	1.42	691	0.30	6,755	-309	0.098	2,755	Q2=0%, Q1=11%, Q3=4%, Q4=27%

* Cost effectiveness threshold (CET): $25,000 per QALY

† Overall preferred intervention (OPI): Intervention (1), since E(NMB) intervention (1) > E(NMB) intervention (2)

‡ if (G=2) then { if (E>B) then [if (D>A) then Q1 else Q4] else [if (D>A) then Q2 else Q3] } else BLANK

Table 1 legend

In the table above, calculation of the attributable EVPI is illustrated for 20 random samples taken from two interventions with normally distributed costs and effects. Intervention (1) has mean costs of $1000 (SD $600) and a mean effect of 0.2 QALYs (SD 0.3 QALYs), intervention (2) has mean costs of $400 (SD $1000) and a mean effect of 0.1 QALYs (SD 0.2 QALYs). Based on the simulated costs (columns A and D) and effects (columns B and E), the NMB can be calculated for each sample (columns C and F).The intervention with the highest average NMB (sample average of columns C and F) is the overall preferred intervention, which in this case is intervention (1) ($4371 > $1025).The AEVPI is estimated using the preferred intervention per sample (columns H through K) and subtracting the overall preferred intervention outcomes (OPI, in this case the sample averages of intervention (1), columns A through C).The results are given in columns L through N: the average of column L is the attributable EVPI cost ($AEVPI_{Cost}$); the average of column M is the attributable EVPI effect ($AEVPI_{Effect}$); the average of column N equals the total EVPI.

In column P, the direction of change (DOC) is given. In cases where the alternative intervention is preferred, the quadrant number indicates whether this intervention is going to be: more costly and more effective (Q1), more costly and less effective (Q2, but by definition this alternative is not preferred), less costly and less effective (Q3) or less costly and more effective (Q4). In the "sample average" row, per quadrant the probability of occurrence is given. The sample average of the preferred intervention (column G) minus 1 (= 0.45) is the estimated probability that the currently preferred intervention will change after obtaining perfect information on the intervention outcomes. This probability equals the sum of the direction of change probabilities per quadrant.

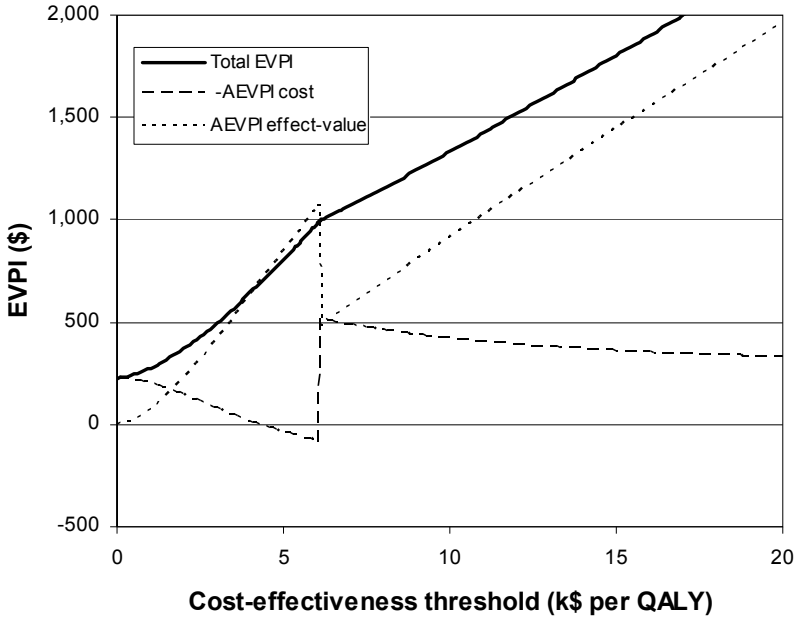


Figure 2: EVPI and Attributable EVPIs of example 1

Please note that minus the $AEVPI_{Cost}$ is shown so that the sum of the AEVPI lines are equal to the (total) EVPI.

For a CET of $6000 per QALY the preferred intervention changes. As a result, the attributable EVPIs show a discontinuity as the outcome reference point (second term in equations 2 and 3) is no longer the expected outcome of intervention (2), but that of intervention (1). For higher CETs the EVPI is dominated by the effect outcome and the $AEVPI_{Cost}$ becomes dependent on the probability that intervention (2) is more effective than intervention (1). As in this setting this probability becomes rather insensitive to changes in CET values over $6000 per QALY, the $AEVPI_{Cost}$ only slowly increases beyond this point.

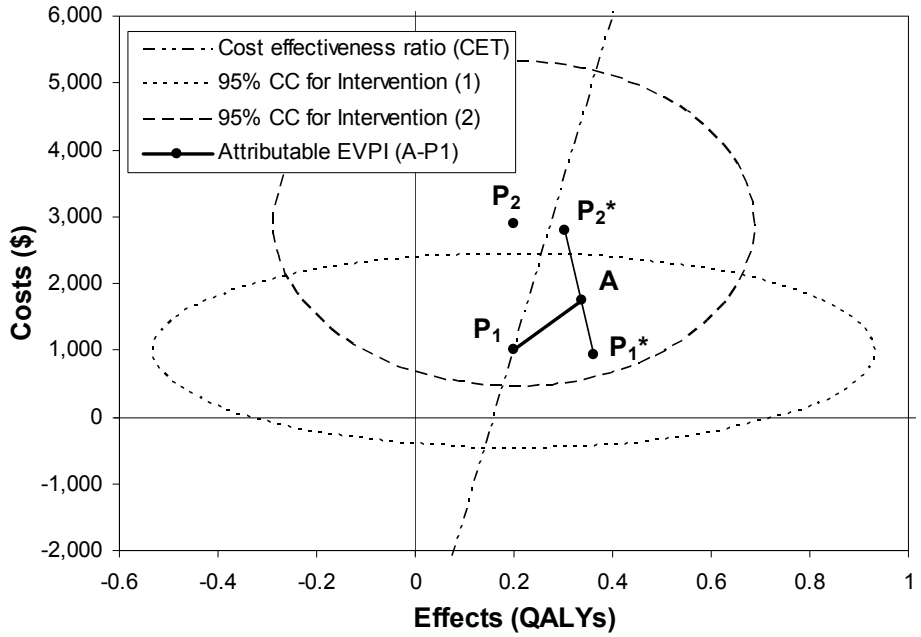


Figure 3: AEVPI Example 2

Average outcome and 95% confidence contours of Intervention 1 (P_1) with normally distributed costs \$1,000 (SD=\$600) and effects 0.2 QALYs (SD=0.3 QALYs) and Intervention (2) (P_2) with normally distributed costs \$2,900 (SD=\$1000) and effects 0.2 QALYs (SD=0.2 QALYs). The EVPI for this decision problem is \$2,755, the $AEVPI_{Cost}$ and $AEVPI_{Effect}$ are \$734 and 0.140 QALYs respectively.

Additional examples

Two more examples of a decision between two interventions with uncertain outcomes are given in Figures 3 and 4. A summary of the characteristics and the outcomes of all examples is given in Table 2. In Figure 3 (Example 2) a situation is given where intervention (2) is shifted upwards along the CET line such that the expected effect is doubled. Comparing NMB outcomes between intervention (1) and (2) in this example will therefore give the same results as when comparing NMB outcomes in Example 1. Hence, the EVPI of Example 1 is identical to that of Example 2 and cannot be used to distinguish between them.

The difference between these two situations, however, is clearly illustrated by the AEVPI. In Example 1 the AEVPI costs are -\$309 whereas in Example 2 these are \$734. From the presentation in the cost-effectiveness plane it is clear that the outcome of additional research is likely to result in a preferred strategy which has higher expected costs and effects. EVPI is sometimes presented without information on relative cost-effectiveness of interventions, especially when multiple interventions are compared.[9] On the other hand, when taking decisions on strategic research investments to improve decision making, apart from the expected value of the investment (the EVPI), the impact of the research outcomes will be of major concern to the decision maker: he or she would like to know whether the outcome of additional research is going to increase or decrease health care costs and effects, what the likelihood of such scenarios is, and what the likelihood is that the outcome of additional research is going to change current policy. The first question is answered exactly by the AEVPI as it represents the expected change

in costs and effects of additional research relative to the currently preferred intervention. The second question can be answered by looking at the 'direction of change' probabilities for each of the quadrants: in Example 1 there is 31% probability (=Q3+Q4=4%+27%, Table 2) that obtaining perfect information will result in a cost reduction, whereas in Example 2 there is a 39% (=Q1) probability that this will result in an increase in costs (Table 2). The third question is merely the sum of the probabilities of change for the three relevant quadrant directions (Q1, Q3, and Q4).

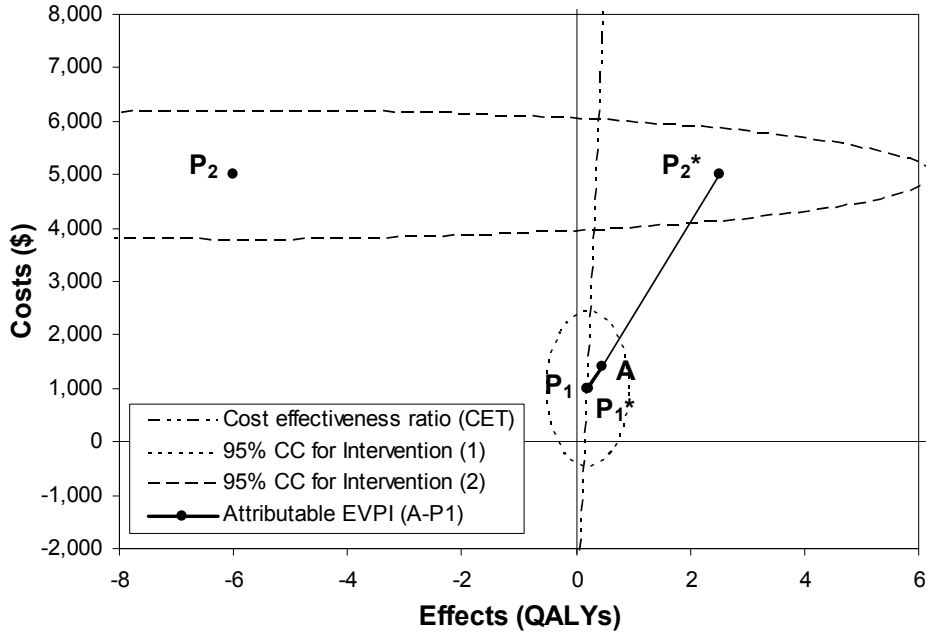


Figure 4: AEVPI example 3

Average outcome and 95% confidence contours of Intervention 1 (P1) with normally distributed costs and effects of $1,000 (SD=$600) and 0.2 QALYs (SD=0.3 QALYs) and Intervention (2) (P2) with normally distributed costs $5,000 (SD=$500) and effects -6 QALYs (SD=5 QALYs). The EVPI for this decision problem is $6,079, the $AEVPI_{Cost}$ and $AEVPI_{Effect}$ are $408 and 0.259 QALYs respectively.

Table 2: Summary of characteristics and outcomes of all examples.

Description	Outcome of Intervention 2* Average (SD)		AEVPI Cost	AEVPI Effect	AEVPI Effect-Value†	EVPI	Direction of change probabilities‡	Probability of change of preferred intervention
	Cost ($)	Effect (QALY)	($)	(QALY)	($)	($)	↖ (Q2) ↗ (Q1) ↙ (Q3) ↘ (Q4)	Q1+Q2 +Q3+Q4
Example 1 (Figure 1)	400 (1,000)	0.1 (0.2)	-309	0.098	2,446	2,755	Q2=0%, Q1=11% Q3=4%, Q4=27%	42%
Example 2 (Figure 3)	2,900 (1,000)	0.2 (0.2)	734	0.140	3,489	2,755	Q2=0%, Q1=39% Q3=0%, Q4=3%	42%
Example 3 (Figure 4)	5,000 (500)	-6.0 (5.0)	408	0.259	6,486	6,079	Q2=0%, Q1=10% Q3=0%, Q4=0%	10%

* The average cost of Intervention 1 is $1,000 (SD=$600) and the average effect is 0.2 QALYs (SD=0.3 QALYs).
† The $AEVPI_{Effect\text{-}Value}$ is the product of the $AEVPI_{Effect}$ and the CET ($25,000 per QALY).
‡ Direction of change probabilities indicate the probability that the alternative intervention will either increase costs (Q1 and Q2), decrease costs (Q3 and Q4), increase effects (Q1 and Q4), or decrease effects (Q2 and Q3).

In Example 3 (Figure 4) intervention (1) is compared to an expensive intervention with a highly uncertain effect. Intervention (2) has expected costs of $5,000 (SD=$500) and expected effects of -6 QALYs (SD=5 QALYs). As a result of the high uncertainty of the effect outcome, intervention (2) can potentially be far more effective than intervention (1). The benefits of reducing the current decision uncertainty are expressed in an EVPI of $6,078. This is more than twice the EVPI from Example 1. However, although the value of information for this situation is considerably higher, adopting intervention (2) results in an increase in treatment costs. This is expressed in the $AEVPI_{Cost}$, which in this case equals +$408 compared with -$309 in Example 1. Moreover, there is a high chance (90%) that the additional research will not change the current preferred policy at all. This poses a higher (political) investment risk. Note that at the same time the returns can also be substantially higher, which is reflected in the higher EVPI. However, the EVPI estimates presume a neutral risk attitude which might not at all be the risk attitude of the decision maker. Both the impact on treatment costs and probability of policy change are important factors a decision maker would like to take into consideration when comparing value of information outcomes.

Discussion

There are various other types of VOI analyses available in addition to the EVPI analysis as described in this paper.[6,10] An example is the 'Expected Value of Partial Perfect Information' (EVPPI).[11] Here not the benefit from eliminating uncertainty in all model parameters is obtained, but of that of a particular parameter or a set of parameters. Another measure is the analysis of the 'Expected Value of Sample Information' (EVSI).[12] In an EVSI analysis a more realistic scenario for obtaining additional information, e.g. by performing a clinical trial, is defined and gains on improved decision making are quantified. The outcomes of all current VOI analyses are hampered by the same limitations as described in this paper. Fortunately, as the concept of calculating AVOI applies to the way results are presented and not to the way they are obtained, it can be applied to any VOI analysis. Often VOI results are presented as a function of the CET (in case of EVPI of EVPPI) or as a function of the size of an additional randomized trial (in case of EVSI). This allows the decision maker to analyze the sensitivity of the VOI. Any of these standard VOI results can easily be extended with attributable VOI results. As an example Figure 2 provides attributable costs and effect-values as a function of the CET.

VOI analyses are generally performed by Monte-Carlo simulations. Estimating the AVOI is as straightforward (or challenging) as estimating any of the existing VOI measures. However, as both costs and effects are required for calculation of the NMB, which in turn is required to determine the preferred intervention, it is merely a matter of collecting the correct samples to enable calculation of the AVOI. This can be done with virtually no computational overhead to any of the existing procedures.

The AVOI provides a means to add relevant information to the current VOI, by providing the decision maker with insight into the expected outcome of additional research in terms of changes in costs and effects of future interventions. In addition, the probability of change in any particular direction can be provided, e.g. the likelihood that costs and effects will increase (or decrease) as a result of the outcomes of additional research. Also, it allows assessment of the likelihood that the additional research will

result in a change in current policy with respect to the preferred intervention. This additional information can be decisive when a selection has to be made between multiple research projects. The advantages of the AVOI in combination with the limited computational overhead required leads us to strongly recommend AVOI as a general extension to the existing presentation of VOI results.

Acknowledgements

We would like to thank Ben van Hout for his strong views on the information loss within the NMB framework which was the seed for the work presented in this paper. Also, we would like to thank the reviewers for their constructive comments which have greatly improved the quality of our paper.

References

1. Eckermann S, Willan AR. *Expected value of information and decision making in HTA*. Health Econ 2007;16:195-209.
2. Yokota F, Thompson KM. *Value of information literature analysis: a review of applications in health risk management*. Med Decis Making 2004;24:287-98.
3. Wailoo AJ, Sutton AJ, Cooper NJ et al. *Cost-effectiveness and value of information analyses of neuraminidase inhibitors for the treatment of influenza*. Value Health 2008;11:160-71.
4. Bojke L, Claxton K, Sculpher MJ et al. *Identifying research priorities: the value of information associated with repeat screening for age-related macular degeneration*. Med Decis Making 2008;28:33-43.
5. Claxton K, Ginnelly L, Sculpher M et al. *A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment programme*. Health Technol Assess 2004;8:1-103, iii.
6. Claxton KP, Sculpher MJ. *Using value of information analysis to prioritise health research: some lessons from recent UK experience*. Pharmacoeconomics 2006;24:1055-68.
7. Raiffa H, Schlaifer R. *Applied Statistical Decision Theory*. Cambridge, MA, USA: MIT Press, 1961.
8. Willan A, Briggs A. *Statistical Analysis of Cost-Effectiveness Data*. Chichester, England: John Wiley & Sons Ltd, 2006.
9. Colbourn TE, Asseburg C, Bojke L et al. *Preventive strategies for group B streptococcal and other bacterial infections in early infancy: cost effectiveness and value of information analyses*. BMJ 2007;335:655.
10. Tappenden P, Chilcott JB, Eggington S et al. *Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-beta and glatiramer acetate for multiple sclerosis*. Health Technol Assess 2004;8:iii, 1-iii,78.
11. Brennan A, Kharroubi S, O'Hagan A et al. *Calculating Partial Expected Value of Perfect Information via Monte Carlo Sampling Algorithms*. Medical Decision Making 2007;27:448-70.
12. Ades AE, Lu G, Claxton K. *Expected Value of Sample Information Calculations in Medical Decision Modeling*. Medical Decision Making 2004;24:207-27.

APPENDIX

Derivation of the attributable expected value of perfect information

Presume a situation where a decision has to be made between j interventions, each with an uncertain outcome in terms of costs C (in dollars) and effects Q (in QALYs). The outcomes are determined by uncertain parameters with known probability distributions. These parameters lie in the parameter space Ω and have realizations ω. The utility of an outcome can be expressed in terms of a Net Monetary Benefit (NMB). In the NMB the intervention effect is simply transformed in a monetary equivalent by multiplying it with a willingness to pay value, the cost-effectiveness threshold (CET). The NMB for an intervention j with cost C and effect Q equals:

$$NMB(j,\omega,CET) = CET \cdot Q(j,\omega) - C(j,\omega) \quad (1)$$

The preferred intervention is the one with the highest expected NMB. The Expected Value of Perfect Information (EVPI) can be formulated using the above definition of the NMB:

$$EVPI(j,\omega,CET) = E_{\Omega}\left[max_j\left(NMB(j,\omega,CET)\right)\right] - max_j\left(E_{\Omega}\left[NMB(j,\omega,CET)\right]\right) \quad (2)$$

Substitution of equation (1) in equation (2) provides separate terms for cost and effects for both terms in equation (2). Maximization of these terms is fulfilled by replacing the appropriate criteria inside the definition of costs and effects:

$$\begin{aligned} &EVPI(j,\omega,CET) = \\ &\quad E_{\Omega}\left[CET \cdot Q(j:max_j\left(NMB(j,\omega,CET)\right),\omega) - C(j:max_j\left(NMB(j,\omega,CET)\right),\omega)\right] \\ &\quad - E_{\Omega}\left[CET \cdot Q(j:max_j\left(E_{\Omega}\left[NMB(j,\omega,CET)\right]\right),\omega) - C(j:max_j\left(E_{\Omega}\left[NMB(j,\omega,CET)\right]\right),\omega)\right] \end{aligned}$$

Taking the expectation over each term separately and rearranging the terms leads to:

$$\begin{aligned} &EVPI(j,\omega,CET) = \\ &\quad CET \cdot \left(E_{\Omega}\left[Q(j:max_j\left(NMB(j,\omega,CET)\right),\omega)\right] - E_{\Omega}\left[Q(j:max_j\left(E_{\Omega}\left[NMB(j,\omega,CET)\right]\right),\omega)\right]\right) \\ &\quad - \left(E_{\Omega}\left[C(j:max_j\left(NMB(j,\omega,CET)\right),\omega)\right] - E_{\Omega}\left[C(j:max_j\left(E_{\Omega}\left[NMB(j,\omega,CET)\right]\right),\omega)\right]\right) \end{aligned}$$

Which we use to define the attributable cost and effect EVPI:

$$EVPI(j,\omega,CET) = CET \cdot AEVPI_{Effect}(j,\omega,CET) - AEVPI_{Cost}(j,\omega,CET) \quad (3)$$

Where:

$$\begin{aligned} AEVPI_{Effect}(j,\omega,CET) = {} & E_{\Omega}\left[Q(j:max_j\left(NMB(j,\omega,CET)\right),\omega)\right] \\ & - E_{\Omega}\left[Q(j:max_j\left(E_{\Omega}\left[NMB(j,\omega,CET)\right]\right),\omega)\right] \end{aligned}$$

$$\begin{aligned} AEVPI_{Cost}(j,\omega,CET) = {} & E_{\Omega}\left[C(j:max_j\left(NMB(j,\omega,CET)\right),\omega)\right] \\ & - E_{\Omega}\left[C(j:max_j\left(E_{\Omega}\left[NMB(j,\omega,CET)\right]\right),\omega)\right] \end{aligned}$$

PART

C

Risks of Viral Transmission by Plasma Derived Medicinal Products

A probabilistic model for analyzing viral risks of plasma-derived medicinal products

Mart P. Janssen
Jan Over
H. Theo M. Cuijpers
Cees L. van der Poel
Ben A. van Hout

Transfusion 2009; 48(1):153-162
DOI: 10.1111/j.1537-2995.2007.01493.x

CHAPTER

6

“All models are wrong, but some are useful.”

George Box

Background

The prevention of transmission of viral infections by plasma-derived medicinal products is of concern to manufacturers, legislators and patient representative groups. Recent European legislation requires a viral risk assessment for all new marketing applications of such products.

Methods

A discrete event Monte Carlo model was developed to determine the viral transmission risks of the plasma-derived medicinal products. The model incorporates donor epidemiology, donation intervals, efficiency of screening tests for viral markers, inventory hold period, size and composition of the manufacturing pool, production time, process virus reduction capacity and product yield. With the model the HIV and HCV contamination risks of a typical hypothetical plasma product were calculated and the sensitivity of the risk to various model parameters analyzed.

Results

The residual HIV/HCV risk of finished product is linear in change with viral incidence rate, and inversely linear with product yield and process virus reduction capacity. For the product analyzed in this paper, the residual risk is less sensitive to changes in screening test pool size, donation frequency and inventory hold period. There is only a limited dependency on the donation type (apheresis or whole blood donations) and a negligible dependency on the manufacturing pool size.

Conclusions

The use of probabilistic model simulation techniques is indispensable when estimating realistic residual viral risks of plasma-derived medicinal products. In contrast to traditional residual risk estimations, the probabilistic approach allows incorporation of specific manufacturing decisions and therefore provides the only feasible alternative for a correct assessment of residual risks.

Introduction

Since the transmission of various virus infections in the past, the prevention of transmission of viral infections by plasma-derived medicinal products has been a matter of concern for manufacturers, legislators and patient representative groups. Where in the past rather a range of prescriptive measures were required for a marketing license, at present European legislation requires a viral risk assessment for Human Immunodeficiency Virus (HIV), hepatitis A, B and C viruses (HAV, HBV and HCV), and Parvo B19 virus for all new marketing applications.[1] Where prescriptive measures (like requiring a number of orthogonal virus reduction steps in a process) do provide a means for standardization of safety, "risk" addresses the relevant concern directly: the likelihood of contamination of the plasma-derived medicinal product. The concept of risk provides legislators a means for comparing and harmonizing manufacturing processes in a more consistent and coherent way which in the future might resolve in a definition of acceptable risk levels for plasma-derived medicinal products.

Following the requirements of the risk assessment guideline[1] a risk model was developed which takes into account epidemiology of HIV and HCV, results of donor screening tests, donation intervals, weight (calculated from volume in mLs), inventory hold period, plasma pool size and composition, process virus reduction capacity, product yield, product filling size and further production strategies. The general structure of the model allows assessment of the risk of any plasma derivatives manufactured by plasma fractionation. However, for the analysis of the risks of viruses other than HIV and HCV an adaptation of virus-specific parts of the model will be required.

Materials and methods

The whole process from donor through plasma intake to product manufacturing needs to be considered for an assessment of the contamination risk of the final container (vial of finished product). Assumptions need to be made to model different parts of the manufacturing process. The manufacturing process is schematically shown in Figure 1.

The whole of the manufacturing process can be described as follows: there is a population that is at risk of acquiring a blood transmittable viral infection. Part of this population donates blood or plasma. The donor population is screened by questionnaires at each donation, and donations (either whole blood or apheresis donations) are tested for known viral infections. The plasma collected from the donors' blood is combined to large pools and subsequently split into separate fractions by various steps in the manufacturing process (e.g. by precipitation, chromatography, filtration, etc.) resulting in different plasma products. These products are then filled in final containers for administration to the patient. Some manufacturing steps will have a built-in potential for reducing the presence or infectivity of pathogens, other steps are specifically incorporated in the manufacturing process for this purpose.

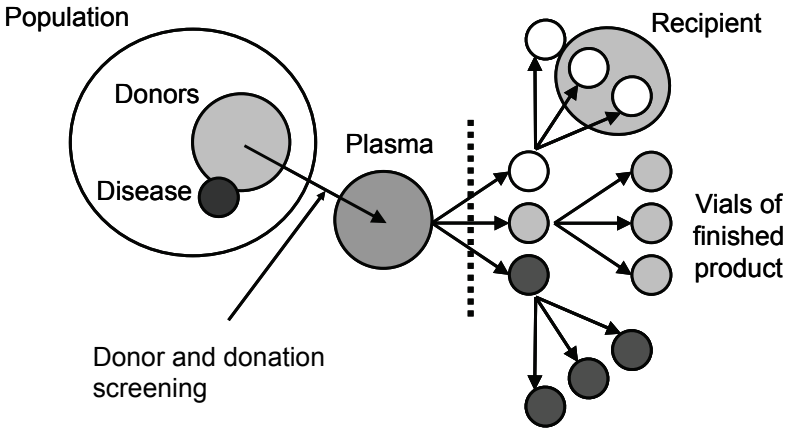


Figure 1: Schematic of plasma manufacturing process

The manufacturing pool of plasma units is used as a basis for assessing the contamination risk of the final container, as this is the base from which a particular batch of vials of finished product is manufactured. The likelihood of contamination of the manufacturing pool and the associated level of contamination can be assessed using virus epidemiology, donation pattern, viral infection characteristics, sensitivity of the screening tests in use and the manufacturing pool characteristics (pool size and composition of donation types). The contamination probability of final containers manufactured from this pool can be calculated using the processing characteristics for a specific product (pool size, virus reduction capacity of the process, and product yield), given a specific level of contamination of the plasma manufacturing pool. The combination of the two steps permits evaluation of the distribution of the probability of contamination of the final containers (see also Figure 3).

Monte Carlo simulation

The contamination risk of the final container is calculated using Monte Carlo simulation techniques. A mathematical model of the manufacturing process is developed which contains parameters that are variable. Examples of such model parameters are donation interval, donation size, manufacturing pool size, virus reduction factor and product yield. For each model parameter one random value is drawn from a probability distribution that is representative for the variability/uncertainty of this model parameter. Next, the contamination risk of the final container is calculated using the model. By repeating this procedure and collecting each of the calculated contamination risk estimates, the variability of the contamination risk estimate becomes apparent. The simulation is repeated thousands of times to obtain an appropriate representation of the dispersion of the contamination risk.

Modeling contamination of the manufacturing pool

Plasma for the manufacturing pool is obtained from donors by whole blood collection or by apheresis of plasma. As manufacturing pools can be large (tens of thousands of donations are combined for some products), storage time of units prior to initiating pooling can be considerable, and donors may contribute multiple donations to a given manufacturing pool. The time from infection of a donor to detection by donation

screening is referred to as the window period. The window period relates to the sensitivity of blood screening tests.[2,3] The window period has been shortened in recent years by the introduction of more sensitive serological tests and by nucleic acid amplification testing (NAT) for HIV and HCV. The probability of a donor being at risk of infecting the manufacturing pool at the time of donation equals the product of the length of the window period and the virus incidence rate in the donor population. This probability is used as the starting point for assessing the contamination risk of the manufacturing pool. The number of window donations in the manufacturing pool is modeled as a binomial distribution with the number of donations in the pool and the window donation probability as input parameters.

The likelihood of detecting the viral contamination of an infected donor increases with the amount of virus present in the blood. A published model using both the virus replication time and the screening test sensitivity depending on the viral load is applied to model the amount of virus present in a donation and the probability of detecting this virus by the screening test at a given point in time after the infection.[4] This model, which is described in more detail in Appendix A, enables simulation of various window donation contamination levels and their associated probability of detection.

The probability of donors returning for subsequent donations within a given time period can be estimated using donation interval data derived from the blood bank database. This information is used to estimate the likelihood of donors with non-detected contaminated window donations returning for subsequent donations at which time the infection is detected with subsequent serological or NAT screening. Manufacturing pool contamination can be prevented in case a donor returns before the previous unit is taken into processing for pooling. If an infected donor returns for a subsequent donation before the final container product is delivered to the market, contaminated products could be either removed or destroyed. Therefore, only infected blood from donors who do not or not timely enough return for subsequent donations contribute to the contamination of the manufacturing pool. Manufacturing strategy dependent decisions as applied in daily practice can easily be incorporated into the simulation model.

In Figure 2 two examples of donors that get infected with HCV before donating, and the consequence of their infection for the manufacturing pool, are illustrated. The first infected plasma unit (from donor 7) is collected long before the processing of plasma units to the manufacturing pool starts. The time between the infection and the donation is 5.9 days. The virus titer and detection probability can be determined using the model described in Appendix A. At the time of the donation the virus titer is 37 gEq/mL and the probability of detection is 5%. It is assumed (or rather, determined by chance during the simulation) that the test fails to detect the presence of the virus and the donor is allowed to continue donating blood. On the next donation the donor blood is tested again and found positive serologically. The preceding window donation is intercepted and therefore does not contribute to the risk, as the positive second test occurs before the plasma is taken into manufacturing.

The second infected plasma unit (from donor 9) is collected at the end of the plasma collection interval for a destined plasma pool (most left dark grey area). The time between the infection and the donation is 8.1 days, which implies that the virus titer at time of donation is 450 gEq/mL and the probability of detection 80%. In this case the test unfortunately fails to detect the presence of the virus and the plasma is taken into

manufacturing, even though the probability of this event to happen is one in five. The donor is not retested before the product is being manufactured and distributed.

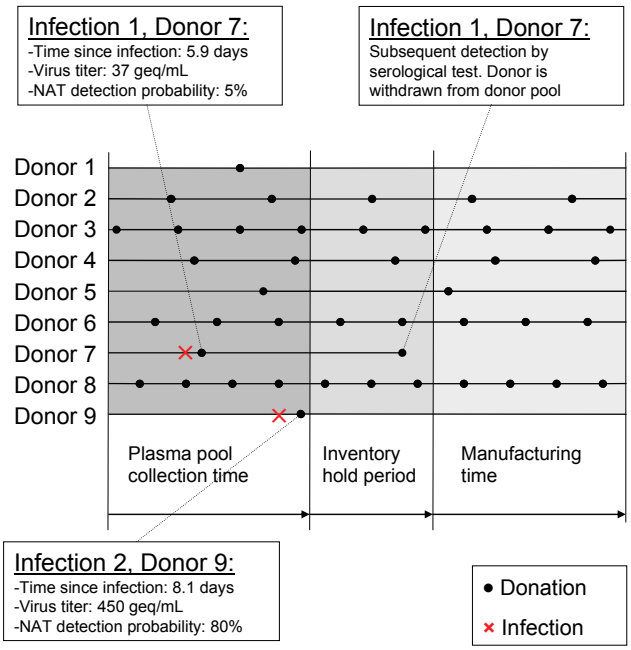


Figure 2: Simulation examples of infected donors

This figure shows some examples of how donation pattern and timing of the infection affect the probability of window donation detection. The dots on the horizontal lines indicate the timing of donations from a single donor. The crosses indicate occurrences of an infection in a donor. The first gray block indicates donations that are combined to a manufacturing pool. The second block indicates the inventory hold period of the manufacturing pool which is dictated by the size of the plasma stock. The last gray block indicates the product manufacturing time (pipeline). After a window donation in the plasma collection block, a subsequent donation in the second or third gray area enables detection of this window donation before or after manufacturing pool composition, respectively.

So, in summary, the calculation of the contamination risk of a specific product starts with modeling the contamination of the manufacturing pool (see Figure 3). The manufacturing pool is the link between virus epidemiology and testing of the incoming plasma (process input), and the final container (process output).

When size and composition of a specific manufacturing pool are known, the probability of a window donation being present can be calculated using the virus-specific window periods and the virus incidence. The actual number of window donations is simulated with the total number of donations in the manufacturing pool and the window donation probability. For each of these window donations the time of infection relative to the time of donation and the timing of the collection of the plasma for the manufacturing pool is simulated. The level of contamination and the probability of detection by the screening test can be calculated. By simulating subsequent donations as well, the possibility of late detection is also incorporated. This requires inclusion of the plasma inventory hold and of the manufacturing pipeline of the specific product in the model. Knowing the probability of detection and the level of contamination of all undetected contaminated donations, the total contamination level of the manufacturing pool is estimated.

The complete risk assessment procedure is shown in Figure 3. A detailed step-by-step description of the risk calculation sequence is given in Appendix B.

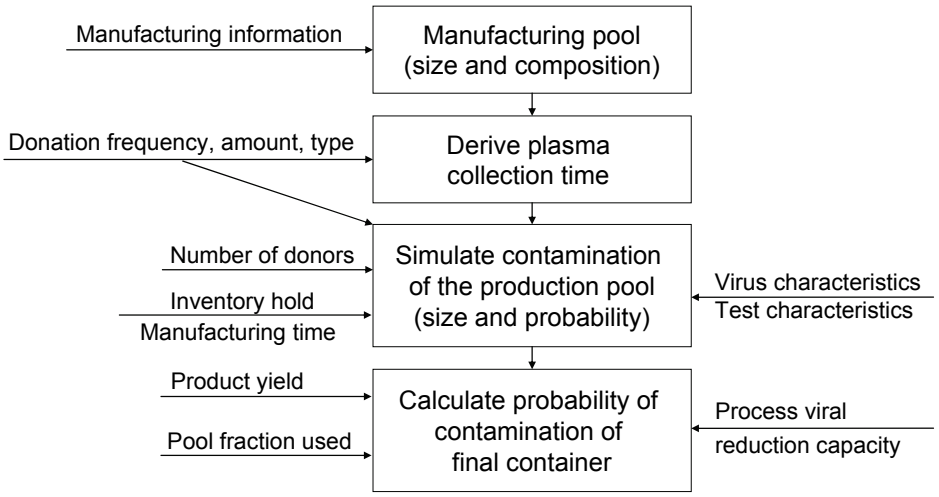


Figure 3: Risk model calculation flow chart
The central blocks represent the subsequent steps in the risk calculation procedure. The arrows to the blocks represent inputs required at each stage of the procedure.

Modeling the contamination of the final container product.

The contamination probability of a final container is calculated using a linear dose-response model with a known specific level of contamination of the plasma manufacturing pool. It is clear that the probability of contamination of a final container increases with the amount of virus present in the manufacturing pool (of plasma). The probability of contamination decreases with the number of vials that are produced from the manufacturing pool. The number of vials produced depends on the efficiency of the manufacturing process and is generally expressed in terms of a process yield factor, the number of containers of finished product that are produced per kilogram plasma processed. The removal of virus and/or reduction of infectivity that is realized during manufacturing is expressed in terms of a log reduction factor. This is the common logarithm (with base ten) of the ratio of the amount of virus and/or infectivity before and after a processing step. Log reduction factors are usually assessed by spiking studies.[5-8]

Log reduction factors play an important role in the assessment of viral contamination risks, as these factors are the primary safety features of the manufacturing process itself and are usually very effective. The European guidelines give a thorough description of the use of log reduction factors in the risk assessments justifying the use of log reduction factors separately assessed for different process steps and their subsequent addition in the final container risk assessment.[1] It is generally accepted that only log reduction factors obtained for orthogonal process steps may be added. Orthogonal in this context means that the mode of action of the virus reducing methods is different (e.g. partitioning by protein precipitation or filtration, physical methods like heat treatment, or chemical methods like solvent/detergent treatment).

The formula used for calculating the final container contamination risk for a manufacturing pool with known contamination level is given in Appendix A.

Robustness, sensitivity and uncertainty analyses

The contamination risk of the final container (or product vial) as derived by the probabilistic model is evaluated against the contamination risk as assessed by the conventional non-probabilistic approach. In the conventional deterministic approach the contamination level is calculated presuming a constant contamination level during a window donation which equals the detection level of the screening test. This estimate is presumed to provide a "worst-case" approximation of the contamination risk required by the authorities.

The model is subsequently explored by means of a multivariate sensitivity analysis. This analysis allows an evaluation of the sensitivity of the contamination risk to model parameters, but also enables disclosure of potential interactions between model parameters. In this analysis the contamination risk is evaluated for each possible combination of model parameter settings over a plausible range of values. Model parameters studied are viral incidence rate, manufacturing pool size, product yield, type of donation, size of test pool, process reduction factor, inventory hold period, and size of the donor population.

Uncertainty with respect to the contamination risk is not solely caused by variability or uncertainty related to manufacturing process specifics (like the number of donations in the manufacturing pool or the value of the virus reduction factor), but is also caused by chance (e.g. what is the contamination level of an infected donation that is not detected by the screening test). The latter is addressed separately in the sensitivity analysis.

Computational tools used

The model is implemented in an MS-Excel 2002 spreadsheet. The Monte Carlo simulations are performed with the @Risk Professional add-in (version 4.5.2) for MS-Excel. Sensitivity analysis is performed with the statistical package S-plus 6.2 Professional.

Results

The contamination risk of a non-existing but representative typical plasma-derived medicinal Product X is analyzed. Product and manufacturing characteristics for this product are given in Table 1. The probability distributions used to represent respective parameter variabilities are given in the last column of the first part of Table 1. The donation interval distributions for apheresis and whole blood donations are based on the donation intervals from Dutch donors from 2000 through 2003, and are 5.0 and 1.5 donations per annum respectively.[9] The incidence rates for HIV and HCV infection are the Dutch incidence rates in the same time period.[10]

Table 1: Baseline product and manufacturing characteristics

Representative values for a manufacturing situation are chosen here. The triangular probabilistic distribution function refers to the often used triangular shaped probability density function. The values drawn from this distribution are limited to the upper and lower range values indicated. The window period is derived from the characteristics of the screening test provided in this table in combination with the screening test detection model as described in Appendix A.

Description	Value	Range (probabilistic distribution function used)
Manufacturing pool size [kg]	15,000	14,000-16,000 (Triang.)
Product yield [vials of end product/kg plasma]	0.20	0.18-0.22 (Triangular)
Proportion apheresis donations	30%	20%-40% (Triangular)
Inventory hold period [days]	360	240-480 (Triangular)
Process reduction factor	6	5-7 (Normal, SD=0.5)
NAT Screening test pool size	48	
Number of donations (recovered / apheresis)	32,640 / 6,878	

Description	HIV	HCV
Incidence rate amongst donors [per donor year][10] (Confidence interval)	5.5×10^{-6} ($3.4\text{-}8.1 \times 10^{-6}$)	2.6×10^{-6} ($1.3\text{-}4.5 \times 10^{-6}$)
Viral doubling time [days][3]	0.85	0.62
Virus concentration at the onset of infection [gEq/mL][11]	0.05	0.05
NAT testing detection levels [50%/95% gEq/mL][12]	5/32	4/21
Window period (99.9% detection) [days]	14.7	10.2

In Figure 4 the contamination risk of a Product X vial is shown based on 4,000 simulations. There is uncertainty as to the actual value of the contamination risk as a result of the variability or uncertainty in process parameters and in the actual contamination level. The mean/median log risks are -8.2/-8.3 for HIV and -8.7/-8.8 for HCV respectively. The 95% confidence intervals are -6.3/-9.9 for HIV and -7.0/-10.3 for HCV.

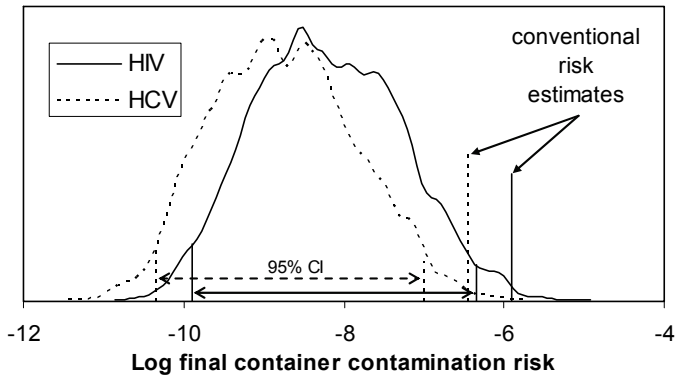


Figure 4: HIV and HCV contamination risks of final container of product X

The probabilistic final container contamination risks for HIV and HCV for Product X are shown as well as the conventional "worst-case" risk estimates

In the past, European plasma manufacturers tended to calculate the vial risk using an exposure to the 95% screening test threshold level over the full window period. This obviously leads to an overestimation of the true vial risk. Alternative assumptions can be considered appropriate (e.g. a 50% screening test threshold level as being the average of 0% and 100% screening test level), but are not in line with the "worst case" approach required by the regulatory authorities. Using this deterministic approach, the vial risks for HIV and HCV are calculated to be -5.9 and -6.4 logs respectively. The deterministic calculation of the HIV risk is shown in Table 2.

The probabilistic models indicate that the probability that the actual risk exceeds the conventional risk estimates is less than 0.5%. This means that the conventional risk estimates (in this case) are indeed "worst-case" estimates.

Table 2: Conventional deterministic calculation of the HIV risk

This table shows the conventional deterministic calculation of the HIV risk of Product X. Each of the intermediate steps in the calculation of the final container risk is presented.

Description (explanation)	Calculation	Outcome
Contamination level in donation [gEq/mL] (detection threshold x pool size)	32 x 48 =	1,536
Average donation size [mL]		380
Average viral load [gEq/mL]	1,536 x 380 =	583,022
Number of vials per batch (product yield x pool size)	0.2 x 15,000 =	3,000
Viral load per vial [log gEq]	$Log_{10}(583,022 / 3,000) = Log_{10}(194) =$	2.3
Contamination level per vial after process reduction [log gEq]	2.3 - 6.0 =	-3.7
Contamination probability [log] (incidence rate x nr of donations in production pool x window risk)	$Log_{10}(5.5x10^{-6} \times 39,518 \times 11 / 365) =$	-2.2
Risk per vial [log gEq]		-5.9

Model sensitivity was analyzed using multivariate regression analysis. To enable the construction of a regression model, the risk was calculated with viable high and low values for all main model parameters. The contamination risk was evaluated for all

possible combinations of parameter settings (high, low and standard values). This approach applied to seven main model parameters resulted in 2187 ($=3^7$) risk estimates. A multivariate linear regression model was fit to the (log transformed) model parameters. The linear model tries to predict the contamination risk at a specific setting by simply multiplying each of the individual parameter settings with a fixed regression coefficient and summarizing these individual mathematical products to a fixed constant risk. The regression coefficients representing the sensitivity of the log vial risk to the respective model parameters are shown in Table 3. The R^2 (which is a measure of appropriateness of the model) was 99.99% out of 100% indicating extremely high model conformity. This means that the risk can be predicted very well using the regression models' coefficients. All model parameters were found highly significant regression parameters.

Table 3 shows that if the incidence rate within the donor population is doubled, the vial risk for Product X increases by a factor two (=incidence rate increase times the regression coefficient). The vial contamination risk is also reduced by the same factor if either the yield or the process reduction factor is doubled. The vial risk is less sensitive to the size of the test pool, inventory hold period, and percentage apheresis donations and almost insensitive to the size of the manufacturing pool. In case the pool size is increased by a factor 10, the change in vial contamination risk is expected to be $\log_{10}(10)$x-0.05=1x-0.05 logs $=10^{-0.05}$= 89%, so an 11% decrease in contamination risk.

Table 3: Sensitivity analysis results

The coefficients represent the sensitivity of $\log_{10}$ vial contamination risk to change in the respective model parameters. With the exception of the % apheresis donations all parameters are on a $\log_{10}$ scale. For these parameters, the coefficients indicate a change in log vial risk associated with a tenfold increase of the respective model parameter.

Model parameter (values used in the regression model building)	Regression coefficient	Standard Error	Probability of rejecting param. significance
Baseline log reduction factor	-8.48	0.0021	0.000
Log incidence rate ($5.5x10^{-7}$; $5.5x10^{-6}$; $5.5x10^{-5}$)	1.00	0.0016	0.000
Log manufacturing pool size (1,500; 15,000; 150,000)	-0.05	0.0016	0.000
Log yield (0.02; 0.2; 2)	-0.99	0.0016	0.000
% Apheresis donations (0%; 50%; 100%)	-0.33	0.0032	0.000
Log test pool size (24; 48; 96)	0.66	0.0053	0.000
Log process reduction (3; 6; 9)	-1.00	0.0005	0.000
Log inventory hold period (180; 360; 720)	-0.68	0.0053	0.000

The vial risk was also assessed for a theoretical donor population donating at twice the current average donation frequency and thus requiring only half the size of the donor population. The regression results from this analysis revealed that the main regression coefficients (for incidence rate, yield, process reduction and test pool size) remained unchanged, but that the average risk reduced by 0.75 logs.

Most parameters affect the size of the vial risk. The size of the test pool (which is directly related to the screening test sensitivity) however does not only affect the size of the vial contamination risk but also its dispersion (data not shown). The more sensitive the test is, the smaller the contamination range of undetected donations will be. The standard deviations of the vial risks shown in Figure 4 are around 0.9 logs. Over the range of tested values (a factor two of the nominal values) the standard deviation of the vial contamination risk increases with 0.25 logs per log increase of the size of the test pool. This is equivalent to an increase in dispersion of the log risk of 7.5% per doubling of the test pool.

Discussion

European legislation at present requires a viral risk assessment for all new marketing applications of plasma-derived medicines. The viral risk assessment is required to consider various factors that influence the potential level of infectious virus particles in a dose of final product. Explicitly noted are the virus epidemiology in the donor population, the frequency of donations from an individual donor, the titer of viremic donations, the testing for viral markers, the inventory hold measures, the virus inactivation/removal steps and the product yield. In the model described in this paper all of these aspects are incorporated in a systematic way.

There is a vast amount of literature concerning mathematical models of HIV and HCV viral kinetics after infection and after NAT detection.[2-4,13] However for HBV, Parvovirus B19 and HAV this is less so. Modeling the contamination risk for these viruses can be done in a similar fashion as with the other viruses in this paper. However, implementation details might vary depending on the type of virus considered and testing regime applied. Also, because of lack of published data on virus behavior, expert opinion will have to be used to build virus models.

One drawback of the probabilistic approach is that the outcome is a probability density function of the product vial risk. This leaves open a number of different characterizations of the outcome and its associated interpretations:

1) mean risk: the expected average of the probability distribution
2) median risk (50% percentile): the risk that divides the probability distribution in half; half of the risks from this distribution is higher/lower than the median value.
3) n^{th} risk percentile: the probability that the risk is less than the n^{th} percentile is n%.

All of these interpretations of "risk" serve their own purpose and can be used as a representation of the final container risk in one number. The 95% percentile value can be considered a reasonably robust worst-case risk estimate.

A comparison of the probabilistic and deterministic risk estimates shows that the conventional deterministic risk estimates lie in the upper tail of the probabilistic estimates and are close to the 95% risk estimate (about 0.5 logs difference). A value in the upper tail (anything up from the 95% percentile) can be seen as a conservative estimate of the product risk. However, a significant reduction in risk is obtained when comparing the average or median probabilistic (log) with the conventional risk estimates.

The results from the probabilistic model were initially perceived unexpectedly high. One would expect the probabilistic estimates to be substantially lower than the conventional estimates as a result of the strong reduction in the estimated viral load in contaminated donations, but apparently this is not the case. However, even though the numbers are pretty close there is a large difference in the interpretation of these numbers: the conventional method does not account for any variability associated with the parameters used to estimate the contamination risk. The probabilistic model however, incorporates and reveals the effect of these variabilities on the contamination risk. Critical here is that in the conventional analysis the uncertainty with respect to the log reduction factor is completely discarded whereas this factor contributes largely to the dispersion in the probabilistic risk estimate. The expected reduction associated with the more stringent modeling of the screening assay's ability to detect virus in the donation of an infected donor is roughly canceled out by the uncertainty associated with the process reduction factor. However, it should be noted that this coincidental comparability of risk estimates may in general not occur.

One of the big advantages of the probabilistic modeling approach is that the model can be used to explore areas for improvement of safety. Here the results of the sensitivity analysis become of relevance. It is clear from Table 3 that for Product X the risk is linear with the incidence rate, and inversely linear with product yield and process reduction factor. It should be stressed that the results from this table are only valid for Product X and for the viruses considered here. With an increase in incidence rate, the sensitivity of the final container risk to incidence rate will change as a result of the fact that there is a nonlinear relation between the infection rate and the probability that the manufacturing pool is contaminated. This relation is linear as long as the number of donations times the probability of contamination is less than 1%. In case this probability is low, as for Product X, this relationship does hold over an order of magnitude as was shown in the sensitivity analysis. The linear relationship will hold for any product for the other two parameters, as this is part of the linear dose response model used for calculating the final container risk for a given manufacturing pool contamination level. It should be noted that in case a virus is (relatively) unaffected by the production process, application of the linear dose response model could lead to a situation where the risk of a final container exceeds one. The maximum risk is therefore cut-off at this value. The sensitivity of the risk to parameter changes for the other model parameters will also be dependent on product and manufacturing specifics. However, it can be expected that the dependence on manufacturing pool size will always be limited. With an n-fold increase of manufacturing pool size the probability of contamination increases (roughly by a factor n), but the risk of final container contamination decreases at the same rate, as n times as many final containers will be produced from the manufacturing pool.

The analysis performed with the reduced donor population (half the original size) showed that the increase in donation frequency significantly reduced the contamination risks of Product X for HCV and HIV. The effect of increase in donation frequency is already indicated in Table 3 where a risk reduction is observed when switching from regular donors for whole blood to high-frequency apheresis donors for plasma. The analysis of the reduced donor population presumed not only a strong increase in average donation frequency (3 to 4 times as high), but also a high donation frequency for a large

proportion of the donor population (80% of the regular donors returning within 175 days, 80% of the apheresis donors returning within 50 days). This is necessary to maintain the level of plasma intake from the donor population. The probability of detecting an infected donor in a large donor pool increases dramatically in this situation. This hypothetical scenario is therefore mimicking to some extent the effect of an extensively increased inventory hold period. In theory, an infinite risk reduction could be achieved by a full quarantine strategy, meaning that only plasma from donors that have revisited the blood center after a certain minimum time to give blood or plasma while still testing negative for viral markers is used for manufacturing. Even though the half-donor-population scenario shows the risk reducing potential of quarantine procedures, it would be unrealistic to expect such donor behavior.

Traditionally, in deterministic calculations the final container risk is calculated using a predefined contamination of the manufacturing pool (e.g. one contaminated donation). Although this "conditional risk" provides information on the robustness of the manufacturing process against potential contamination and allows comparison of manufacturing processes, it does not provide any information on the actual final container risk as this is highly dependent on the incidence rate in the donor population. Two identical plants producing identical products obtaining their plasma from donor populations with different incidence rates will produce products with different contamination risks. The probabilistic model can easily be used to provide a (probabilistic) estimate for the conditional risk. It only requires an assessment of the second stage of the risk model, as it does not require modeling of the contamination of the manufacturing pool (see Figure 3).

This paper tries to demonstrate the advantages of probabilistic modeling in the assessment of viral contamination risks of plasma-derived medicinal products. It is first of all virtually impossible to calculate a trustworthy contamination risk in a deterministic way. There is no easy way, not to say that it is impossible, to derive a general description of the effect of donation behavior, screening test performance, product manufacturing timescales and decision strategies following screening test results on the final container risk. The relationship between these effects can be very easily and realistically simulated. Also, the probabilistic approach generates insight in the uncertainty associated with the estimated product vial risk and allows a consistent and clear definition of a "worst-case" risk. Furthermore, the model can be easily adapted to assess transmission risks of other viruses and other transmissible infectious diseases, like for example vCJD. And finally, the model can be used to evaluate the impact of different model parameters on product risk and therefore allows an assessment of the (cost-)effectiveness of alternative manufacturing strategies or additional safety measures.

References

1. EMEA. *Guideline on assessing the risk for virus transmission - New chapter 6 of the note for guidance on plasma-derived medicinal products (CPMP/BWP/269/95)*. CPMP/BWP/5180/03. 21-10-2004.
2. Kleinman S, Busch MP, Korelitz JJ et al. *The incidence/window period model and its use to assess the risk of transfusion-transmitted human immunodeficiency virus and hepatitis C virus infection*. Transfus Med Rev 1997;11:155-72.
3. Glynn SA, Kleinman SH, Wright DJ et al. *International application of the incidence rate/window period model*. Transfusion 2002;42:966-72.
4. Weusten JJ, van Drimmelen HA, Lelie PN. *Mathematic modeling of the risk of HBV, HCV, and HIV transmission by window-phase donations not detected by NAT*. Transfusion 2002;42:537-48.
5. EMEA. *Note for Guidance on virus validation studies: the design, contribution and interpretation of studies validating the inactivation and removal of viruses.* CPMP/BWP/268/95. 14-2-1996.
6. Terpstra FG, Parkkinen J, Tolo H et al. *Viral safety of Nanogam, a new 15 nm-filtered liquid immunoglobulin product*. Vox Sang 2006;90:21-32.
7. Farshid M. *Viral safety evaluation of plasma-derived therapeutic products*. Dev Biol (Basel) 2004;118:11-5.
8. Darling AJ. *Considerations in performing virus spiking experiments and process validation studies*. Dev Biol Stand 1993;81:221-9.
9. *Sanquin Annual Report 2003*. 2004. Amsterdam, Sanquin Blood Supply Foundation.
10. Janssen MP, van der Poel CL, Buskens E et al. *Costs and benefits of bacterial culturing and pathogen reduction in the Netherlands*. Transfusion 2006;46:956-65.
11. Busch MP, Glynn SA, Stramer SL et al. *A new strategy for estimating risks of transfusion-transmitted viral infections based on rates of detection of recently infected donors*. Transfusion 2005;45:254-64.
12. Jongerius JM, Sjerps M, Cuijpers HT et al. *Validation of the NucliSens Extractor combined with the AmpliScreen HIV version 1.5 and HCV version 2.0 test for application in NAT minipool screening*. Transfusion 2002;42:792-7.
13. Kleinman SH, Busch MP. *The risks of transfusion-transmitted infection: direct estimation and mathematical modelling*. Baillieres Best Pract Res Clin Haematol 2000;13:631-49.

APPENDIX A: MODELS FOR VIRUS REPLICATION AND SCREENING TEST DETECTION PROBABILITY

It is presumed that after infection, the virus will reproduce at a constant reproduction rate with known doubling time. A constant reproduction rate will result in an exponential increase of viral particles in the blood. This process has been observed in several seroconversion panels and is described in the literature.[3,4] Using the doubling time model, the viral titer can be estimated as a function of time since infection:

$$C = C_0 2^{t/\lambda}$$

Where:

- C = concentration of viral particles in the donor blood at a given time t [gEq/mL]
- C_0 = C at $t = 0$ (the start of the window phase), the virus concentration at the onset of infection [gEq/mL]
- t = time since infection [days]
- λ = doubling time, describing growth rate of a virus in an infected donor [days]

The detection probability is modeled as a Gaussian distributed function of the logarithm of the contamination level.[4] This model requires two parameters to characterize the measurement sensitivity: the 50% and 95% probability of detecting the virus concentration.
The detection probability as a function of the viral titer is given below:

$$p_{\text{det}} = \Phi\left(1.65 \frac{\log(C / X_{50})}{\log(X_{95} / X_{50})}\right)$$

Where:

- p_{det} = detection probability
- Φ = cumulative standard normal probability density function
- C = viral titer [gEq/mL]
- X_{50} = viral titer with a 50% detection probability [gEq/mL]
- X_{95} = viral titer with a 95% detection probability [gEq/mL]

APPENDIX B: CALCULATION OF FINAL CONTAINER RISK

The contamination risk of a final container from a contaminated manufacturing pool is calculated using the following formula:

$$p_{fp} = \frac{np_c}{W\psi\alpha}$$

Where:

p_{fp} = probability of contamination of a vial of final product
n = number of viral particles in the manufacturing pool
p_c = manufacturing pool contamination probability
W = manufacturing pool weight [kg]
ψ = manufacturing process yield [products/kg plasma]
α = process viral reduction factor

Risk calculation sequence

The individual steps of the simulation process for calculating the final container (product vial) contamination probability are summarized as follows:

1. Starting point for the analysis is the manufacturing pool. Information on the manufacturing pool is obtained from manufacturing records. Per simulation a random set is taken from these records, being a combination of: manufacturing pool size and donation type(s).
2. The number of whole blood and apheresis donations required is calculated using the composition and size of the manufacturing pool.
3. The time required for collecting the plasma for the manufacturing pool is calculated from the number of donations and donation rate per donation type.
4. The frequency distribution of specific donation types enables the calculation of the number of donors that contribute to the manufacturing pool.
5. Per virus an 'at risk period' (ARP) is calculated. This period is the time interval in which an infected donor could contribute to infecting the manufacturing pool. The ARP is a virus-specific interval and equals a maximum interval at risk for a given infection. For HIV and HCV this interval is the time from infection to (100%) detection by the relevant screening test. Any donation in the ARP may lead to one (or more) infectious donation(s) to the manufacturing pool.
6. The viral incidence rate times the ARP yields an estimate for the probability of a donor (or a donation) being at risk of contamination at any point in time. The probability distribution of the number of 'at risk' donations in a specific manufacturing pool can be calculated with the 'at risk' probability and the number of donors, using a binomial probability distribution.

7. A specific number of donors at risk are drawn from this distribution for each simulation at random in accordance with their likelihoods corresponding to this binomial distribution.
8. For each donor 'at risk' the time of infection, time of donation, donation frequency and donated volume are taken at random from appropriate distributions.
9. The viral load(s) at time of donation is (are) calculated from the simulated timing of the donation(s).
10. The probability of detection is calculated, using the viral load(s), screening test characteristics and timing of the donations in relation to production timescales (See Figure 2).
11. Whether or not detection occurs is simulated in accordance with the calculated detection probability.
12. Each non-detected contaminated donation is multiplied by its donation volume to calculate the viral contamination level.
13. The sum of viral contaminations from non-detected contaminated donations renders the total viral contamination level in the plasma manufacturing pool.
14. The contamination probability of a final container is calculated using the total viral contamination level in the manufacturing pool, the manufacturing pool size, the process yield and the process reduction factor.
15. The final container risk is calculated by combining the risks of the intermediate products, in case an end product is produced by mixing several intermediate products.

Note

A two-stage Monte Carlo procedure is performed to calculate the final container risks. A two-stage procedure is required because not all simulated 'at risk' manufacturing pools will result in an infected manufacturing pool. This will depend on the detection of an 'at risk' donation. In the first stage the manufacturing pool is simulated (size and compilation) and the contamination probability of this specific manufacturing pool calculated. The second stage involves calculation of the risk reduction through plasma processing. This is equivalent to calculation of the conditional risk, the risk of contamination of a final container given that the manufacturing pool was contaminated. The combination of these two allows estimation of the final container contamination probability.

Regression analysis on serial dilution data from virus validation robustness studies

Nan van Geloven
Eric A. Cator
Hendrik P. Lopuhaä
Mart P. Janssen

Statistica Neerlandica 2009; 63(3):245-257
DOI: 10.1111/j.1467-9574.2009.00416.x

CHAPTER 7

“It is better to be roughly right than precisely wrong.”

John Maynard Keynes

Abstract

To ensure the safety of plasma-derived medicinal products, the Dutch Blood Supply Foundation (Sanquin) performs virus validation experiments. Data from these experiments are based on serial dilution assays. Regression analysis on assay data faces several problems: only a small number of data points are available, data contain censoring and are subject to sampling error. Furthermore, the process variability inherent to the experiments is not evident. In this paper we address these problems by introducing a regression model for serial dilution data and by analyzing how validation experiments and simulation techniques can help elucidate various sources of variability the experiments are subject to. These are then incorporated into the regression model.

Introduction

Despite a profound process of selection and screening of donors and the testing of donor blood, plasma-derived medicinal products still have a residual risk of infection. Use of medicinal products derived from human blood or plasma has led to the transmission of viruses, including viruses causing hepatitis A, B and C and Acquired Immuno-Deficiency Syndrome (AIDS). Presently, products derived from large plasma pools are submitted to several manufacturing processes, which include steps that inactivate or remove viruses. An example of such a process is pasteurization: heating plasma at 60°C during a particular period of time effectively inactivates any virus particles present.

At the Dutch Blood Supply Foundation (Sanquin) the virus reduction capacity of processing steps is evaluated by controlled scaled-down processes in an experimental laboratory setting. In these experiments, the starting material is deliberately 'spiked' with a known virus that has characteristics comparable with that of a 'human' virus of interest. The virus reduction capacity of the reduction process is evaluated by comparing the outcomes of serial dilution assays, one assay obtained before the reduction process and one thereafter. In such an assay, an original solution containing the virus is divided into samples that are increasingly diluted. The diluted samples eventually contain such a small amount of the original solution that the most diluted samples will not contain (infectious) virus particles anymore. To evaluate infectiousness, the diluted samples are added to little wells that contain susceptible cell lines. After some time (typically 1 or 2 weeks) the condition of the cells is inspected. If the cells show infection by the virus, this is scored as a positive response. A typical outcome of an assay with 12 dilution steps (with dilution factors $1/3^i$, $i = 5, 6, \ldots, 16$), performed in eightfold, is shown in Table 1. The bottom row is known as the dose-response curve and represents the total number of positive responses in the read out for each dilution step. These are used to estimate the amount virus in the original solution.

Table 1: An example of the read out of a serial dilution assay

Replicate	Dilution steps											
	5	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
1	+	+	+	+	+	-	-	-	-	-	-	-
2	+	+	+	-	-	-	-	-	-	-	-	-
3	+	+	+	+	+	+	-	-	-	-	-	-
4	+	+	+	+	-	-	-	-	-	-	-	-
5	+	+	+	+	+	-	-	-	-	-	-	-
6	+	+	+	+	+	+	+	+	-	-	-	-
7	+	+	+	+	-	-	-	-	-	-	-	-
8	+	+	+	+	+	+	-	-	-	-	-	-
Total +	8	8	8	7	5	3	1	1	0	0	0	0

A method commonly used in biological assays is that of Spearman-Kärber (SK).[1-7] This method helps estimate the median tissue culture infective dose ($TCID_{50}$), which is the sample volume that gives a response in 50% of the cell cultures. The virus reduction capacity of the reduction process is then measured by the log reduction factor (LRF) defined as the difference of the $^{10}\log(TCID_{50})$ estimates determined before and after the reduction process.

During a reduction process different process variables can influence the final inactivation or removal of viruses. A widely used reduction process is pasteurization. Three variables that may play a role in the effectiveness of this process were identified: temperature, pH value and saccharose level. These three variables have their standard settings, but the nature of the processes causes fluctuations in these variables. To provide information relating the sensitivity of a process to these fluctuations in the process conditions, and consequently allow assessing its reliability during normal usage, one may conduct a robustness study. This is required by regulation.[8] The main question in such a study is: how sensitive is the virus reduction capacity to variations in process conditions? As starting material and process equipment are very expensive, usually a fractional factorial design is used. An example of such a design used at Sanquin to analyze the pasteurization method is given in Table 2, which displays the LRFs and their standard errors (both calculated by the SK method) determined after 30 and 120 minutes of pasteurization for six settings of the three process variables. Two additional centre-point measurements (run 5 and 6) are added to the 2^{3-1} design. We will refer to this example as experiment 1152. The LRF values determined after 120 minutes of pasteurization with settings 3 and 4 correspond to data for which the information about the $TCID_{50}$ is censored. Censoring can occur in different ways. For instance, when the read out for each dilution step (each of the columns in Table 1) contains more than four positive responses. In such a case the $TCID_{50}$ is only known to lie beyond the last dilution.

Table 2: Experiment 1152, an example of a 2^{3-1} design

Run	Process variables			Time-points	
	Temperature (°C)	pH (−)	Saccharose (%)	30 minutes (LRF ± SE)	120 minutes (LRF ± SE)
1	58.5	6.6	62	2.7±0.2	3.4±0.2
2	58.5	7.4	58	3.0±0.2	4.0±0.2
3	62.5	6.6	58	3.6±0.2	≥5.1±0.2
4	62.5	7.4	62	3.3±0.2	≥4.9±0.2
5	60.5	7.0	60	3.1±0.2	4.0±0.2
6	60.5	7.0	60	3.2±0.2	4.0±0.2

Standard procedures used on data such as in Table 2 do not seem to be sufficient for determining whether a process variable significantly affects the virus reduction process. Apart from difficulties caused by the censored data, the current fractional factorial designs provide a rather small number of data points for a regression analysis, especially since the LRF estimates are recorded with considerable standard errors (SE) compared with the effect sizes. It is not clear how to deal with these SEs, which should represent the sampling (estimation) errors. Moreover, the nature of virus experiments also raises questions about the reliability of the experiments under normal settings. Tests on variability of the technical equipment and of the virus stocks are standard in the laboratory. However, their outcomes have never been looked at in the light of robustness studies.

In this paper we will address these problems: scarce, censored data subject to sampling error and process variability. Instead of imposing the general bioassay model inherent to the method of Spearman-Kärber, we use the most probable number (MPN) model.[9] The MPN model appears to be more suitable for modelling the specific assay at hand, i.e. a

serial dilution assay. The MPN method uses the complete information in the dose-response curve and allows assays with censored, possibly non-monotone data. Here we extend the MPN model to incorporate the possible influence of the process variables. Significance of factors is tested with this extended MPN model and the results are compared with the results of a more traditional regression analysis (in which the SK outcomes are analyzed using a standard linear model). Additionally, the sources of variability in the experimental outcomes are examined thoroughly. To this end, a new validation experiment was set up and performed in the laboratory. As the extended MPN model did not explain all the observed variability, an attempt is made to explain the additional variability by other possible factors such as the technical errors introduced during the sampling and diluting of the volumes. We show that the extended MPN model allows assessing the robustness of a virus-reducing step while taking the different sources of variability into account.

In the following section we explain the model and report on the significance of factors. The validation experiment, and how to assign the variability observed among the LRF values are described in the subsequent section.

The MPN model with covariates

We denote the volume of the solution we want to know the amount of virus of as V and refer to the dilution level with the indicator $i = 1, 2, \ldots, k$, where k denotes the total number of dilution levels. At each dilution level i, a volume v_i of the original solution V is tested in n_i replicates. For $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n_i$, we define

$$w_{ij} = \begin{cases} 1 \text{ if the } j^{\text{th}} \text{ well at the } i^{\text{th}} \text{ dilution level is positive (shows virus)} \\ 0 \text{ otherwise.} \end{cases}$$

Hence, the number of positive wells at the i^{th} dilution is

$$d_i = \sum_{j=1}^{n_i} w_{ij}$$

and the outcome from a dilution assay is represented by the vector $\mathbf{d} = (d_1, \ldots, d_k)$, called the 'dose-response curve'. Typically, one chooses equidistant log-volumes. Denote p_i as the probability that a particular well at the i^{th} dilution is not infected and write $\zeta_i = {}^{10}\log v_i$ as the corresponding log-volume. In biological assays one usually assumes that $1-p_i=F(\zeta_i)$ for some unknown non-increasing function F, known as the tolerance distribution, and a standard problem is to estimate the mean or median of F. The method of Spearman-Kärber is a nonparametric estimate of the mean of F, traditionally used to estimate the $TCID_{50}$.[3,4] Note that it is therefore actually used to estimate the median of F. However, the SK estimate could be of use if F is (point) symmetric, in which case mean and median coincide. The quality of the SK estimator depends on whether the log-volumes densely span the range of F, i.e. a large number of dilution steps with $F(\zeta_i) \approx 0$ and $F(\zeta_i) \approx 1$. Apart from the fact that there is no specific reason to assume F to be symmetric, assays in which either $\hat{p}_1 = d_1 / n$ differs a lot from one or $\hat{p}_k = d_k / n$ differs a lot from zero will yield a censored estimate. Other criticism on the SK method can be found in the literature.[10,11]

Alternatively, we use the MPN model to estimate the infectivity of the solutions.[9] It does not impose a symmetric relation between p_i and ζ_i (in fact, it yields the asymmetric relation (1) below), and it does not put restrictions on the d_i values. The MPN model assumes that infectious virus particles are randomly distributed throughout the solution and that each sample from the liquid, when added to the cells, is certain to exhibit growth whenever the sample contains one or more infectious particles. Then, if the original solution of V(ml) contains N such particles and at the i^{th} dilution we put v_i(ml) into a well, the probability that a single particle ends up in the sample is v_i/V. As there is assumed to be no kind of attraction or repulsion between the viruses, this holds for any virus particle, irrespective of the positions of the others. Consequently, the probability that none of the virus particles is in the sample is given by

$$\left(1-\frac{v_i}{V}\right)^N = \left(1-\frac{mv_i}{N}\right)^N \approx e^{-mv_i},$$

where $m = N/V$ denotes the mean number of infectious virus particles per unit volume and we take into account that the number of virus particles is large. This leads to the following model for the probability that a single sample at the i^{th} dilution level is not infected:

$$p_i = e^{-mv_i}. \tag{1}$$

Hence, the probability that d_i of n_i samples are positive is given by the binomial probability

$$\binom{n_i}{d_i} e^{-mv_i(n_i-d_i)}\left(1-e^{-mv_i}\right)^{d_i}. \tag{2}$$

Assuming the outcomes at each dilution to be independent, the likelihood of the outcome $\mathbf{d} = (d_1, \ldots, d_k)$ is given by

$$L(m) = \prod_{i=1}^{k} \binom{n_i}{d_i} e^{-mv_i(n_i-d_i)}\left(1-e^{-mv_i}\right)^{d_i}. \tag{3}$$

The value $\hat{m}$ that maximizes (3) is the maximum likelihood estimator for m and is called the 'most probable number'. Note that it estimates an actual number of infectious virus particles instead of a dose. An estimate for the $TCID_{50}$ could, in this model, be found from solving (1) with $m = \hat{m}$ and putting p_i equal to 0.5. This shows that the $^{10}\log(TCID_{50})$ and $^{10}\log(m)$ differ by a constant factor in this model. We will refer to $^{10}\log(m)$ as the 'titre', a term commonly used in virology to express an amount of virus.

In a robustness study one does not perform a serial dilution assay for a single solution, but for several solutions, each one obtained under a certain setting of the process conditions, which we refer to as the p covariates, $x_1, x_2, \ldots, x_p$. Suppose we have R dose response curves $\mathbf{d}_r = (d_{r1}, \ldots, d_{rk})$, for $r = 1, 2, \ldots, R$, obtained from the samples belonging to R separate serial dilution assays. Because we are interested only in determining main effects, we model the titres as depending linearly on the covariates:

$$^{10}\log(m_r) = \beta_0 + x_{r1}\beta_1 + \ldots + x_{rp}\beta_p = \mathbf{x}_r'\beta. \tag{4}$$

where $\mathbf{x}_r = (1, x_{r1}, \ldots, x_{rp})$ and $\beta = (\beta_0, \beta_1, \ldots, \beta_p)$. The total likelihood of the outcomes of a robustness study with R serial dilution assays then becomes:

$$L(\beta) = \prod_{r=1}^{R} \prod_{i=1}^{k} \binom{n_i}{d_{ri}} \exp\left(-10^{\mathbf{x}_r'\beta} v_i (n_i - d_{ri})\right) \left(1 - \exp(-10^{\mathbf{x}_r'\beta} v_i)\right)^{d_{ri}} . \quad (5)$$

Note that this likelihood corresponds to serial dilution assays where the number of dilution steps, the number of replicates and tested volumes are the same in each assay. Obviously one could vary these settings, which would lead to a similar likelihood. Maximizing (5) for β produces estimates for the effect sizes of the process parameters. Results are given in Table 3.

Table 3: Effect sizes and asymptotic and simulated *p*-values in the new model

Source	30 minutes			120 minutes		
	Effect	**Chi-square**	**Simulated**	**Effect**	**Chi-square**	**Simulated**
Overall		2.5×10^{-5}	<0.0001		0	<0.0001
Temperature	-0.29	3.4×10^{-6}	<0.0001	-0.80	0	<0.0001
pH	0.00	0.9704	0.9678	0.00	0.8589	0.8589
Saccharose	0.10	0.1198	0.1154	0.27	2.78×10^{-6}	<0.0001

As an example, the estimated effect on the test titres 30 minutes after moving from a low (58.5 °C) to centre (60.5 °C) temperature setting or from centre to high (62.5 °C) temperature setting is −0.29 (more virus reduction at a higher temperature). Moreover, results from tests on the significance of the regression coefficients and on overall model significance can be found in this table. These are derived from likelihood ratio tests based on (5). We have reported asymptotic *p*-values on the basis of the chi-squared approximation for the likelihood ratio test statistic as well as simulated *p*-values computed from 10,000 simulated dose-response curves. These were generated from binomial distributions with failure probabilities p_i given by (1), using the fitted coefficients $\hat{\beta}$ from (4).

We will compare these results with those from a more traditional approach. Following the standard regression analysis performed at Sanquin, one would use the method of Spearman-Kärber to estimate the $TCID_{50}$ of the spiked material before the reduction step and of the test samples after the step, calculate the log reduction factors and fit a linear regression model to these LRFs. Then, tests on the significance of regression coefficients may be performed by standard F- and t-tests. At the second time-point, SK will not provide suitable LRFs for this because of the censoring. In order to mirror the new model, the traditional analysis was performed using the reduction factors calculated as the difference of the titres estimated by the MPN method before and after the reduction steps. Results are given in Table 4. Note that the apparent change in the sign of the effects is caused by the fact that not the end materials, but the reduction factors (start material minus end materials) are regressed now.

A comparison of Tables 3 and 4 shows that the extended MPN model in most cases gives lower *p*-values. The two models do not agree on the significance of saccharose on both time-points. Let us take a closer look at this. At the first time-point, the LRF approximately lowers 0.3, moving from a saccharose low to a high setting, in both settings of the temperature (a clear significant factor). At the second time-point, the LRF lowers by 1.0 and 0.6 in these two cases (again using MPN estimates). Apparently, the

standard analysis can point out a significant linear effect of saccharose at the first time-point, while it cannot at the second. The extended MPN method judges conversely: the effect of saccharose at the first time-point is too small compared with the sample variation; at the second time-point it is assigned a significant effect.

Table 4: *p*-values in the traditional analysis

Source	30 minutes		120 minutes	
	Effect	***F*/*t* -test**	**Effect**	***F*/*t* -test**
Overall		0.0164		0.1591
Temperature	0.30	0.0069	0.90	0.0667
pH	0.00	1.0000	-0.10	0.7227
Saccharose	-0.15	0.0267	-0.40	0.2441

Variability of pasteurization

In the previous section, we introduced a new regression model that tested whether covariates (process parameters) significantly influenced the virus reduction capacity of pasteurization, assuming that all the variability in this pasteurization step was explained by our model. This means that the only reason we do not find the same virus reduction factors when we pasteurize twice, would be because of the randomness in the sampling of the dilution assays that were used; the actual achieved reduction factor is modelled to be equal for both pasteurization steps. This seems unrealistic: one would expect an inherent variability because of the biological nature of the virus-reducing process, and there might be other sources of variability as well.

To get a grip on this variability, a *de novo* pasteurization experiment was set up and carried out in the laboratory. With this experiment we want to assess how results vary if pasteurization runs are performed under exactly the same process settings. Figure 1 shows a schematic depiction of the design of this experiment. In the experiment six pasteurization runs are performed. In each run the amount of virus is determined at three time-points: in spiked material after 30, 60 and 120 seconds of pasteurization. In each of these 18 cases, a dilution assay is performed by examining three culture plates consisting of 96 (=8×12) wells. This yields a total of 54 dose-response curves.

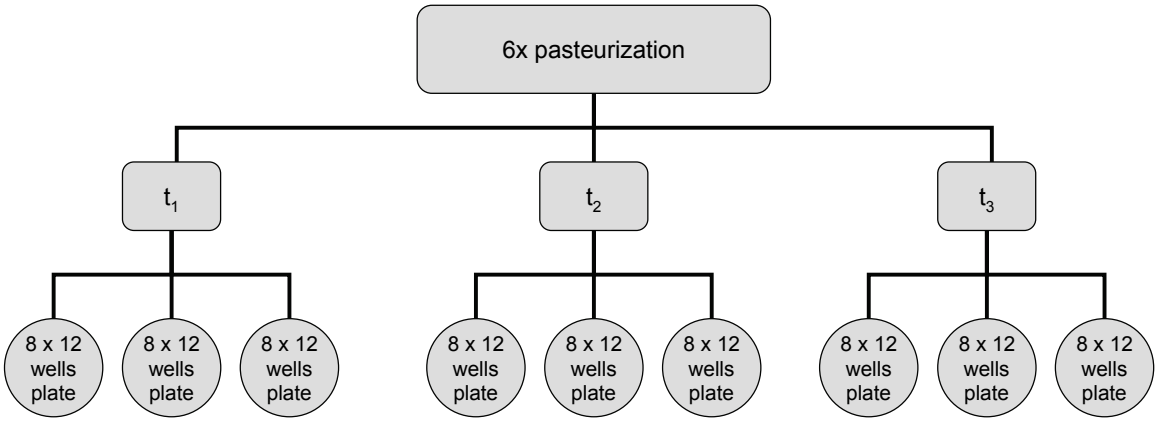


Figure 1: Design of the experiment

Modelling sources of variability

From the 54 assay outcomes we want to quantify possible sources of variability in the reduction process. This is necessary for the interpretation of the robustness studies, as we do not want to let the regression be misled by noise. But first, let us repeat and expand the notation. Again the dilution level is referred to with the indicator $i = 1, \ldots, k$, where k denotes the total number of dilution levels. In our experiment each assay consists of 12 dilutions, so $k = 12$. The runs are denoted by $r = 1, \ldots, R$, with R the total number of runs. In this experiment R = 6. The amount virus is determined at certain time-points $t \in T$, in this case $T = \{30, 60, 120\}$. The repetition of the assay in this experiment is denoted by j = 1, 2, 3. Just as before, the number of wells tested at dilution level i is n_i =8 and the volume of the original solution tested at level i is v_i. As different dilutions are made to anticipate on the decrease of amount of virus in time, the v_i may differ at the different time-points. The actual data points are the $d_{ri}^{j}(t)$ values, which give the number of wells that are scored positive in run r, at time t, at repetition j and at dilution level i. Finally, we model the mean number of infectious particles per volume $m_r(t)$ as a function of time. Now analogous to the previous section the likelihood of $d_{ri}^{j}(t)$ can be expressed as:

$$L(m_r(t)) = \binom{n_i}{d_{ri}^{j}(t)} \left(1 - \exp(-m_r(t)v_i)\right)^{d_{ri}^{j}(t)} \exp\left(-m_r(t)v_i(n_i - d_{ri}^{j}(t))\right). \tag{6}$$

If all $d_{ri}^{j}(t)$ are assumed to be independent, the total likelihood can be expressed as

$$\prod_{r,t,j,i} \binom{n_i}{d_{ri}^{j}(t)} \left(1 - \exp(-m_r(t)v_i)\right)^{d_{ri}^{j}(t)} \exp\left(-m_r(t)v_i(n_i - d_{ri}^{j}(t))\right), \tag{7}$$

and the log-likelihood (omitting the constant term) as:

$$\sum_{r,t,j,i} d_{ri}^{j}(t) \ln\left(e^{-m_r(t)v_i} - 1\right) - n_i v_i m_r(t)\,. \tag{8}$$

We can estimate the titres $^{10}\log(m_r(t))$ by maximizing the (log)likelihood function. When we do this separately for each time-point at each run, we obtain Table 5.

Furthermore, we use a likelihood ratio test based on (8) to test the hypothesis that virus reduction behaved the same in each run, i.e. $m_1(t) = \cdots = m_6(t)$, for all $t \in T$. It turns out that this hypothesis is resoundingly rejected (p-value from simulations smaller than 10^{-5}). Apparently, other significant sources of variability exist.

Table 5: Titres of the 18 test samples estimated independently

Time	Run number					
	1	**2**	**3**	**4**	**5**	**6**
t=30	3.92	4.01	4.26	3.62	3.82	3.99
t=60	2.84	2.59	3.08	2.86	2.92	2.89
t=120	1.19	1.32	1.01	1.51	1.69	1.65

Technical error

The first source of variability we could quantify was the error made when diluting the material and when titrating the material into the well (we will call this a 'technical error'). Ideally, the volume of the original material tested at dilution level i is given by

$$v_i = v_0 f^{-(c_0+i)},$$

where v_0 is the total volume titrated into a well (typically 50μl), f is the dilution factor (typically 3), and c_0+i represents the times the well at dilution level i is diluted (an assay does not have to start with sample $v_0/3$). When we take dilution errors into account, we could model the test volume in well l, $l = 1, \ldots, n_i$ at dilution level i as

$$v_{il} = (v_0 + \tau_{il}) \prod_{s=1}^{c_0+i} \left(f + \varepsilon_s\right)^{-1}.$$

Here we dropped some of the indices, but of course we take an independent realization of τ_{il} each time a titration into a well is made, and we take an independent realization of ε_s each time a dilution is made (when a dilution from level $i-1$ to level i is made, each corresponding well is titrated using the same dilution). We were able to derive direct estimates of the distributions of τ_{il} and ε_s based on the specifications of the instruments used. We obtained (in ml):

$$\tau_{il} \sim N(0, (0.0015)^2) \text{ and } \varepsilon_s \sim N(0, (0.2)^2).$$

The probability that well l at dilution level i is positive, will now be modelled by

$$\mathbf{P}(\text{positive}) = 1 - \exp\left(- m_r(t)(v_0 + \tau_{il}) \prod_{s=1}^{c_0+i} \left(f + \varepsilon_s\right)^{-1} \right). \quad (9)$$

The likelihood could be changed accordingly, but this is not a practical approach, as it becomes intractable. Simulation, however, is still very well possible. To test the hypothesis that $m_1(t) = \cdots = m_6(t)$ for all $t \in T$, we use the original likelihood ratio test statistic (based on (8)), but find the *p*-value of this test by simulating according the model using the technical error, specified by (9). This approach still leads to *p*-values smaller than 10^{-5}, suggesting that even though we incorporated the volume errors, the model does still not explain all variability.

Another way to look at the variability is by determining the root mean squared error (RMSE) of $^{10}\log(\hat{m}(t))$, the estimator for the titre $^{10}\log(m(t))$, using simulation in the different models, and then comparing that with the empirical RMSE of the six estimated titres resulting from the six runs. Because of the large variations at the third time-point, t_3 (caused by the small number of virus particles present by that time), this approach is only useful for the first two time-points t_1 and t_2. As shown in Table 6, the RMSE of the titre increases only slightly when we take the technical error into account; the empirical RMSE seems substantially bigger. We wish to point out that the estimate of the titre in each run is based on three assays. In practise, usually only one assay is used, in which case the RMSE would be √3 times as big. Furthermore, one might expect that the RMSE of the model without technical error might depend on the actual value of $m(t)$, but it turns out that this is not the case. The reason is that with a higher concentration of viruses, the sample is diluted more often, so that $m \cdot v_i$ lies around $\log(2)$ (which would

make the probability of a positive well equal to ½), and it is this product $m \cdot v_i$ that determines the Fisher information of $^{10}\log(m(t))$.[11,12]

An obvious technical error could be the exactness of the time-points. The laboratory technicians estimate that the standard error in taking out the sample at the correct time is about 1 second. However, we do not have a reasonable model to describe $m(t)$ as a function of t. An exponential decay might seem a natural choice, but all our tests rejected this model, while only looking at three time-points. Therefore, we cannot predict how an error in time would propagate into the RMSE of $m(t)$.

Table 6: Root mean squared error of estimated titre

Time	Sample	Sample+ technical	Empirical
t_1	0.073	0.093	0.196
t_2	0.072	0.084	0.144

Biological error

As we were not able to quantify (model) any other sources of variability, we model the inherent variability of the pasteurization in the following way: after pasteurization, the actual titre is given by a mean titre $^{10}\log(m(t))$ plus a random effect $\kappa(t)$. We will refer to $\kappa(t)$ as the biological error. This means that the probability that well l at dilution level i is positive, will now be modelled by (compare with (9))

$$\mathbf{P}(\text{positive}) = 1 - \exp\left(-v_{il} \cdot 10^{\log(m_r(t))+\kappa_r(t)}\right). \qquad (10)$$

Here, $\kappa_r(t) \sim N(0, \sigma_{\text{bio}}^2)$. We can use Table 6 to get a rough estimate of σ_{bio}^2, because the mean squared error of the estimator $^{10}\log(\hat{m}(t))$ is given by the sum of the RMSE^2 found in the table and σ_{bio}^2. Then we can use the empirical RMSE to get the estimate

$$\begin{aligned} t_1: \quad & \hat{\sigma}_{\text{bio}}^2 = 0.196^2 - 0.093^2 \quad \Rightarrow \quad \hat{\sigma}_{\text{bio}} = 0.173 \\ t_2: \quad & \hat{\sigma}_{\text{bio}}^2 = 0.144^2 - 0.084^2 \quad \Rightarrow \quad \hat{\sigma}_{\text{bio}} = 0.117 \end{aligned}.$$

If we insist on having the same relative deviation of the virus reduction at the two different time-points, we would get as a first indication

$$\hat{\sigma}_{\text{bio}} = 0.15 .$$

A way to check this value is to see if the model specified by (10) does explain the observed variability in the six runs. To this end, we performed the same likelihood ratio test, again determining the *p*-value by simulation, using different values for σ_{bio} (Table 7). Judging from the *p*-values in this table, it seems that 0.15 is not an unreasonable value for σ_{bio}.

Table 7: *p*-values on repeatability, including volume errors and a range of biological errors

σ_{bio}	*p*-value
0.08	0.0060
0.11	0.0640
0.15	0.3387
0.17	0.5106

The robustness study revisited

We have introduced a new regression model for the robustness experiment 1152, but so far, we have tested hypotheses on the process parameters without taking the technical and biological errors into account. The results from our new pasteurization experiment show that these two errors are significant in determining the accuracy of estimated titres, so we will repeat the significance tests on the process parameters.

In experiment 1152, the same equipment is used for diluting the material and titrating it into the micro culture plates as in the pasteurization experiment used to test the repeatability. We therefore include the same volume error estimates as before in the simulations of this experiment. This means that for each run, the dilution steps (dilution factor of 3) are simulated with a random error term ε_d, with $\varepsilon_d \sim N(0, (0.2)^2)$. The titrated volumes (0.05 ml) are simulated with an error term τ_{il}, with $\tau_{il} \sim N(0, (0.0015)^2)$.

Extrapolating the biological factor we found before is a bit more delicate. Because the example of experiment 1152 is also a pasteurization experiment using the same virus that was used in the variability experiment, we could argue that the same biological error is applicable. However, differences in the time period of pasteurization and in the starting material do exist. If one wants to make a sound estimate of the biological error inherent to a certain reduction process, one should perform a validation study similar to the foregoing of the process at hand. Although this may take some effort, assessing the repeatability of the process is indispensable for performing a sound regression on process parameters, for it prevents one from regressing on noise. With the above reservations in mind, we add a biological error $\kappa \sim N(0, \sigma_{bio}^2)$, with σ_{bio} = 0.15. Obviously, adding the error terms does not change our parameter estimates, but it does curb our tests for significance as can be seen in Table 8.

Comparing Table 8 with Table 3 shows that the regression model still makes the same choices in those parameters which it points out as significant at a 0.05 significance level, but the *p*-values are slightly more moderate.

Table 8: *p*-values with volume error and biological error added to the simulations

Source	Simulated	
	30 minutes	120 minutes
Overall	0.0327	<0.0001
Temperature	0.0052	<0.0001
pH	0.9831	0.8982
Saccharose	0.3337	0.0035

Conclusions

In the analysis of virus validation robustness studies one is faced with a limited number of data points, which contain censoring and are subject to sampling error and process variability. None of these problems could be addressed properly within a standard regression model. By extending a parametric method used for estimating the amount of virus from a dose-response curve, the problems with censored data and sampling errors were overcome. This new model does not use intermediate estimates. Using all observations instead of intermediate results provided stronger *p*-values to the relevant covariates.

A new experiment testing the process variability showed that non-negligible biological variation is inherent to the processes. Volume errors, whose impact was unknown so far, have only minor effect on the process outcomes. After quantifying the error terms, the new regression model can be fit incorporating these errors. In this way, the significance of certain factors can be established despite the inherent variability of the process at hand. Given the advantages of the new model, this is proposed for future analyses of virus validation robustness studies.

References

1. Spearman C. *The Method of "Right and Wrong Cases' (Constant Stimuli) without Gauss's Formulae*. Br J of Psych 1908;2:227-42.
2. Kärber G. *Beitrag zur kollektieven Behandlung pharmakologischer Reihenversuche*. Archiv für Experimentelle Pathologie und Pharmakologie 1931;162:480-3.
3. Church JD, Cobb EB. *On the equivalence of Spearman-Kärber and maximum likelihood estimates of the mean*. Journal of the American Statistical Association 1973;68:201-2.
4. Miller RG. *Nonparametric estimators of the mean tolerance in bioassay*. Biometrika 1973;60:535-42.
5. Hoekstra JA. *Estimation of the LC50, A Review*. Environmetrics 1991;2:139-52.
6. *Bekanntmachung über die Zulassung von Arzneimitteln: Anforderungen an Validierungsstudien zum Nachweis der Virussicherheit von Arzneimitteln aus menschlichem Blut oder Plasma*. Bundesanzeiger Nr. 84 4740-4744. 1993. Paul-Ehrlich-Institut und Bundesgesundheitsamt.
7. Nanthakumar A, Govindarajulu Z. *Fixed-Width Estimation of the Mean of Logistic Response Function Using Spearman-Karber Estimator*. Biometrical Journal 1999;41:445-56.
8. Biotechnology Working Party. *Note for guidance on virus validation studies: the design, contribution and interpretation of studies validating the inactivation and removal of viruses*. CPMP/BWP/268/95. 14-2-1996. London, EMEA.
9. Cochran WG. *Estimation of bacterial densities by means of the "most probable number"*. Biometrics 1950;6:105-16.
10. Myers LE, McQuay LJ, Hollinger FB. *Dilution assay statistics*. J Clin Microbiol 1994;32:732-9.
11. Kundi M. *One-hit models for virus inactivation studies*. Antiviral Research 1999;41:145-52.
12. Russek E, Colwell RR. *Computation of most probable number*. Appl Environ Microbiol 1983;45:1646-50.

Monitoring viral incidence rates: Tools for the implementation of European Union regulations

Mart P. Janssen
Erik A. Cator
Cees L. van der Poel
Wim P. Schaasberg
Gouke J. Bonsel
Ben A. van Hout

Vox Sanguinis 2009; 96(4):298-308
DOI: 10.1111/j.1423-0410.2009.01166.x

CHAPTER

8

“There are no facts, only interpretations”

Friedrich Nietzsche

Background and Objectives

European legislation requires manufacturers of plasma products to report epidemiological data on human immunodeficiency virus, hepatitis B virus and hepatitis C virus in donor populations. The incidence rates of such infections are directly related to the risk of infection transmission. We propose two statistical tests to evaluate these incidence rates.

Materials and Methods

Infection data of the four Dutch blood collection centres from 2003 through 2006 were analysed. For transversal comparison of centres and detection of increased incidence rates, a new statistical test was developed (outlier test). For longitudinal detection of trends in incidence rates, a generic test for trend is proposed. The power and risk of non-detection are evaluated for both tests.

Results

Application of the outlier test did not reveal any significantly increased incidence rates among centres in The Netherlands. The test for trend showed no significant increase in incidence rates in individual centres, but on national level a statistically significant increase in hepatitis C virus incidence was observed (p-value of 0.01).

Conclusions

The proposed tests allow signalling of outlier centres and trends in incidence rates both at individual centre and at national levels. Graphical support and the use of as much relevant historical data as possible is recommended. The statistical tests described are generic and can be applied by any blood establishment and plasma fractionation institute.

Introduction

Since 2005 the European Medicines Agency (EMEA) requires manufacturers of plasma products to collect epidemiological data on Human Immunodeficiency Virus (HIV), hepatitis B virus (HBV) and hepatitis C virus (HCV) infections in their donor populations.[1,2] Collecting information on the infection risk in donor populations is considered "…an essential part of the measures taken to ensure an adequate selection of donors of blood and plasma".[2] Continuous epidemiological surveillance at individual blood/plasma collection centres (BCCs) together with an annual update of the assessment is therefore required by the EMEA. Epidemiological data should to be collected on those blood-borne infectious agents for which a potential transmission by blood products is well recognized and for which routine testing of blood and plasma donations is mandatory.[2] The minimum data collected at the time of writing cover anti HIV 1/2, anti HCV and HBsAg test results of all donations. The Plasma Master File (PMF) holder, i.e. the organization responsible for the appropriate licenses for producing plasma derived medicines, is required to report the results of additional screening tests separately (e.g. NAT assays or anti HBc).[3] Next to collecting and summarizing the data, the PMF holder needs to define acceptable ranges for the epidemiological data, and criteria which identify individual BCCs that are performing beyond these acceptable ranges. These criteria have to be applied, and results have to be submitted to the regulatory authorities. Follow-up information has to be provided on the actions undertaken with regard to any deviant BCCs and the effectiveness of remedial actions have to be described. In addition to the cross-sectional analyses at BCC level, the EMEA requires the last four years of data to be analyzed for trends in infectious marker rates in the donor population.[2]

In this paper we describe a systematic approach towards implementation of the EMEA requirements into operational guidelines for the Dutch Sanquin Blood Supply Foundation. Starting from data on observed infections in the donor population, two statistical tests are described that allow the decision maker to make a formal judgment on the presence of deviant (‘unacceptable’) incidence rates (IRs). The decision maker has to specify a range of acceptability for an outlier BCC a priori. Only then a test can be performed to assess whether a BCC is exceeding this range of acceptability. In addition, significance levels for the tests have to be specified upfront by the decision maker: the significance level determines how the trade-off is made between the likelihood of deciding that there is a deviant IR when there is not, and the likelihood of deciding that there is no deviant IR when in fact there is. In addition to the statistical tests, tools are provided that aim to help the decision maker understand the implications of his decisions. As the proposed monitoring tests are generic, they can be applied by any plasma manufacturer, whether producing for the European market or not.

Materials and methods

Detection of an outlier BCC

Before a statistical test can be performed to determine whether there is an outlier BCC, a definition of an outlier BCC has to be provided. In this paper an outlier BCC is defined as a BCC of which the IR is R times (or more) higher than the mean IR of the remaining BCCs. The pre-specified factor R expresses a margin in risk acceptability, which is relative, as the anchor is the base rate of the other BCCs. The consequent risk increase (by possibly not detected infectious blood) caused by an individual BCC with an R times increased IR can be calculated. This allows the achievable risk reduction to be weighted against potential measures taken whenever this BCC is found to be an outlier. These could range from performing a root cause analysis to – at the extreme – closing down of a specific BCC. The relationship between an increase in IR in a specific BCC and its contribution to the risk increase in the overall blood supply can be described by a simple mathematical equation. Consider a reference population with a mean IR λ_r. Now presume there is one specific BCC with a proportion δ of the total population which has an increased IR of exactly $R \cdot \lambda_r$. The risk ratio (RR), which is the relative risk increase caused by that one BCC can be derived from the IR of the total population (λ_t) and the IR of the reference population (λ_r):

$$RR = \frac{\lambda_t}{\lambda_r} = \frac{R \cdot \delta \cdot \lambda_r + (1-\delta) \cdot \lambda_r}{\lambda_r} = \frac{\lambda_r \cdot (1 + (R-1) \cdot \delta)}{\lambda_r} = 1 + (R-1) \cdot \delta \qquad (1)$$

Where:
- RR = ratio of risks of total population and reference population
- λ_t = total population incidence rate
- λ_r = (non elevated) reference population incidence rate
- R = ratio between incidence rates of elevated centre and reference population (incidence rate ratio)
- δ = elevated centre proportion of the total population

From equation 1 the obvious result is found that the impact of the risk increase in the outlier BCC equals the product of the relative risk and the population size. In Figure 1 the relationship is shown between the relative IR increase in the outlier BCC (IR ratio/relative risk R), its proportion of the total population (δ) and the risk reduction obtained ($= (\lambda_t - \lambda_r)/\lambda_t$) when correcting the IR of the outlier BCC towards the reference population IR.

This figure can be used to decide what IR ratio (R) for an outlier BCC is considered unacceptable. It allows balancing the impact of risk control measures (which are associated with the size of the donor population or donor BCC) and the associated potential risk reduction achieved when intervening. It should be noted that Figure 1 is applicable to *any* subgroup within the donor population, be it a BCC or a sub-population having specific characteristics which can be associated with a deviant risk of infection.

For instance, in case a blood establishment would consist of 20 equal sized BCCs and one would decide to intervene only when a risk reduction of 5% or more could be achieved, the IR ratio (R) would be about 2.

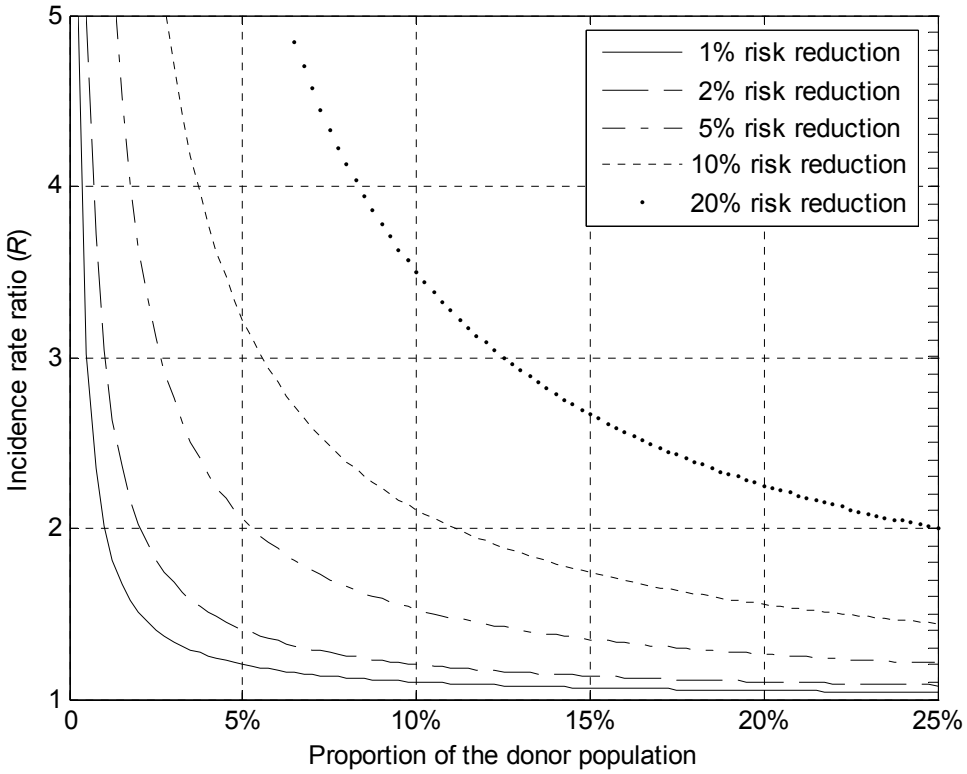


Figure 1: Achievable risk reduction as function of BCC size (proportion of donor population) and IR ratio (R) of the outlier centre.

Risk is generally defined as the probability of a pre-specified undesired outcome.[4] Both the probability and the size of the undesired outcome are often not exactly known. This situation clearly applies here: the estimated IRs are subject to statistical uncertainty. Accounting for this uncertainty is a second prerequisite to arrive from infection data to decision rules. The judgment of differences between (or changes in) IRs can be done by formal hypothesis testing. This means that a null hypothesis (H_0) is formulated, e.g. 'IRs are the same across all BCCs', in conjunction with a complementary alternative hypothesis (H_a), 'IRs are not the same across all BCCs'. Depending on the data the tester will either reject or accept the null hypothesis. Such formal approach requires the tester to decide which likelihood of being in error he is willing to accept when making a decision based on the test outcome. By convention the significance level of the test α (or type I error) relates to the likelihood of making the wrong decision when the alternative hypothesis is accepted (i.e., the null hypothesis is rejected, the test provides a false positive outcome), whereas β (or type II error) describes the likelihood of making the wrong decision when accepting the null hypothesis (a false negative test outcome). There is a trade-off to be made as an increase in α will be associated with a decrease in β and vice versa. This trade-off may vary with the level of risk to be controlled: one may for instance require less certainty when looking for an increase in HIV than for an HCV infection, which would translate to a higher setting of α for HIV.

With the proposed definition of an outlier BCC and an IR ratio (R) as a threshold, a statistical test can be applied to the observed infections to determine whether an outlier BCC is present. A very versatile model that is generally applied to data involving counts is the Poisson distribution.[5] This model is also applicable to the situation at hand, where numbers of observed infections per year are considered. Comparing Poisson rates has been long described in the literature.[6-8] However, a statistical test for detecting an increased IR amongst multiple BCCs, especially with an extremely low number of observations, requires a specific test. The test that was developed and applied in this paper is specified in Appendix A.

Within statistics, the (graphical) presentation of the probability of a positive test outcome as a function of the property to be detected, which in this case is an elevated IR in one of the BCCs, is referred to as the power curve.[9] The power curves for outlier BCC detection depend on reference population IR, IR ratio (R *x* population IR), BCC sizes and significance levels. Power curves allow the evaluation of the probability of detection of an outlier BCC with a specific IR. However, Instead of considering the probability of non-detection at one particular IR the average *risk* of non-detection over a range of (increased) outlier BCC IRs can also be considered. This makes sense as the size of a potential increase in IR is unknown. As the risk of disease transmission increases linearly with the IR, risks are calculated as the product of the IR and the probability of non-detection. The risk from an elevated IR is a function of the significance level applied. With an increase in significance level, the risk from an elevated IR will decrease, however, at the expense of an increasing number of false positive test outcomes.

The relation between significance level and risk of non-detection is calculated and presented graphically for various IRs. As these graphs show the relationship between the false positive test rate and the risk of non-detection, they allow the decision maker to select the most appropriate significance levels for the outlier tests. Significance levels might be chosen differently for different diseases. The construction of the risk curves is explained in detail in Appendix B.

Detecting a trend

There are various ways of tracking a trend in IRs. The 'Control Chart', which was developed for statistical process control, is widely applied within industrial processes and provides an easy means for detecting trends.[9] The main strengths of this technique are its simple structure and graphical presentation. A statistical test can also be used to detect a trend in IRs. Detection of change points in (Poisson) IR series has been studied extensively in the past[10] and a test for isotonic change is proposed for trend detection: isotonic change implies that in a series of subsequent observations each succeeding observation is at least as high as the previous one. The advantage of this test is that it, unlike for instance a test for linear trend, is able to detect a sudden increase in IRs e.g. a threefold IR in the last year of observation. A formal description of the test for trend is provided in Appendix A.

Power curves for the test for trend can be constructed for any predefined trend behaviour. Power curves based on linear increasing IRs are given in Appendix B. The power curves were constructed for various reference population IRs and significance

levels. Similar to the previous test, the risk of an undetected trend as a function of the significance level for various reference population IRs is provided.

Application to Dutch repeat donor infection data

Sanquin Blood Supply Foundation collects BCC-specific data on infections. Tests for outlier BCC detection and trend are applied to the HIV, HBV and HCV infections found by blood screening tests in the years 2003 through 2006.

Computational issues, tools used and source code

The derivation of the likelihood estimators described in Appendix A was performed using Mathematica® (Version 5.2, Wolfram Research Inc., Champaign, IL). The power and risk curves of the statistical tests as well as all figures were constructed using MATLAB® (Version 7.2, The MathWorks Inc., Natick, MA). The *p*-values of all statistical tests were obtained by taking 10,000 samples. The MATLAB® code for all calculations in this paper, an extended description of the outlier test and the code for performing both tests with the free software environment for statistical computing R® (The R Foundation for Statistical Computing, Vienna, Austria) are available from the authors upon request.

Results

Test for outlier BCC detection

In the Netherlands there are four regional collection BCCs. Data collection and reporting occurs per BCC on annual basis in a similar way.[11] To assess the ability of the statistical test to detect elevated IRs in the Dutch setting, the outlier test was evaluated for a setting with four BCCs with (hypothetical) IRs of 0.5, 1, 2 and 4 infections per year. For each of these reference IRs (n_r, applicable to three of the four BCCs), the risk from an undetected outlier BCC is given as a function of the applied significance level (Figure 2). The test assesses whether any one centre has a higher IR than the average of the other centres.

From Figure 2 it can be found that for a reference population IR resulting in an IR of 0.5 infections per BCC per year, the undetected increase in IR in an outlier BCC, presuming that there is one, is expected to be 3.3 infections per year when the significance level of the test is set at 5%. Increasing the significance level from 5% to 10% will reduce the risk from 3.3 to 2.8 infections per year (17% reduction). The decision maker should decide whether the risk reduction of 17% is worth the additional 5% false positive test outcomes. Reducing the significance level to 2.5% would reduce the false positive rate by a factor of 2, but would also increase the risk by 31%.

Note that the risks could also be expressed in terms of the expected number of transmitted infections per year, which could again be extended in terms health outcome and costs. This requires additional modelling of the transmission risk with the window period model and modelling of the health of the transfusion recipient population.

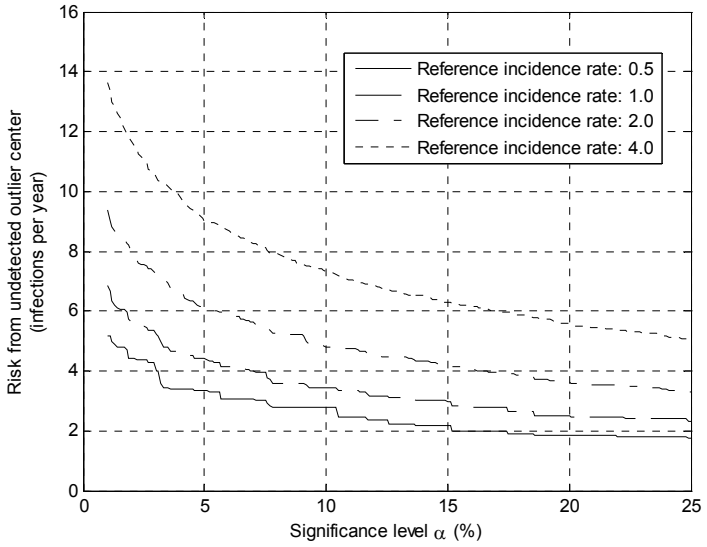


Figure 2: Risk from an undetected outlier BCC (incidence rate in infections per year) as a function of significance level α for various reference incidence rates (IRs), tested for an IR ratio (R) of 1.

Test for trend detection

For BCC IRs of 0.5, 1, 2 and 4 infections per year, the probability of trend detection of was calculated as a function of a fixed annual IR increase. In Appendix B power curves for various IRs and significance levels are given. In Figure 3 the risk from an undetected trend is given as a function of the applied significance level for various IRs. Figure 3 shows that for a BCC with an IR of 0.5 the expected undetected annual IR increase, given that there is a positive trend, is almost 1 (0.95) infection per annum when the significance level is set at 5%. Increasing the significance level to 10% will reduce this risk by 0.35 infections per year (-37%).

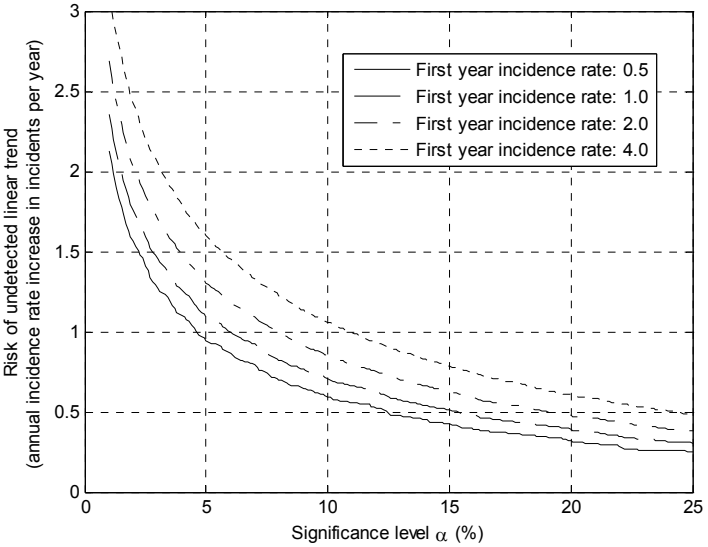


Figure 3: Risk of undetected linear trend (undetected annual IR increase) as a function of significance level α for various reference IRs (first year BCC IR).

Application of the tests to Dutch repeat donor infections

In Table 1 the reported HBV, HIV and HCV infections in repeat donors as well as donor population sizes for each of the regional BCCs of the Sanquin Blood Supply Foundation in the period 2003 through 2006 are given.[11] The statistical tests for trend and for outlier detection were applied to these data. Given that each centre constitutes about 25% of the total donor population, an increase of IR in one of these centres will impact the overall IR considerably and a considerable risk reduction can be obtained even for values of R just over 1 (refer Figure 1). The outlier test was therefore performed for an IR ratio (R) of 1. The final column of Table 1 contains the *p*-values for both the outlier detection and trend tests.

From Table 1 it can be found that none of the BCCs has an IR higher than the average of the other BCCs. However, when considering a significance level of 5% (which is commonly applied for statistical tests) there does seem to be a trend in HCV and HIV infections, which have *p*-values of 1% and 4% respectively.

The first trend concerns a national trend in HCV infections. This fact is further supported by Figure 4 where a control chart of the HCV IR over the past 12 years is shown. In this figure the average IR over the years 1995-2005 is drawn as well as the associated 95% upper control limit (UCL) and 95% lower control limit (LCL). The UCL here equals the average IR plus 1.64 times the expected standard deviation of the IR in each year.[9] As this standard deviation depends on the population size, the UCL varies over the years. In addition, for each observation above the average the probability of observing this number of events (or more) is given. This number is derived from a Poisson distribution using the average IR, population size and number of observed events in each year. The control chart indicates that given the average IR over the past 11 years, there is a 1.3% probability that 5 (or more) observations would have occurred by chance in case the IR would have remained unchanged. This observation is in line with the outcome of the statistical test, which indicates an unjust rejection of no increase in IR of 1%.

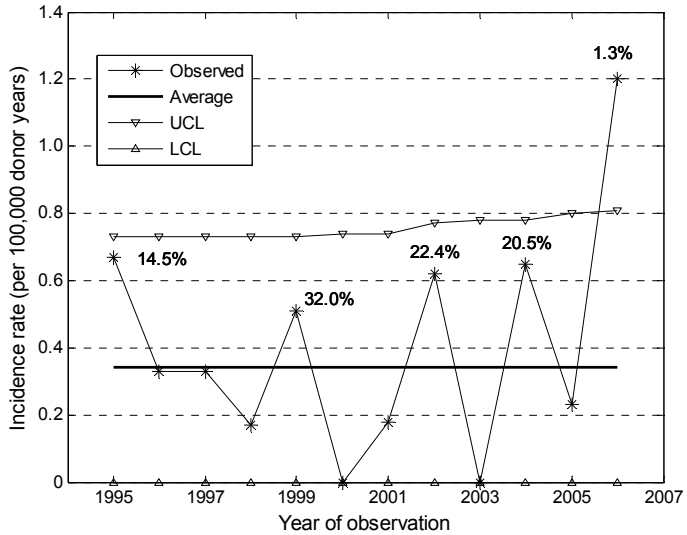


Figure 4: HCV IR Control Chart.

The second trend is associated with a steady increase in the number of observed HIV infections in BCC 4 (*p*-value of 4%). However, considering that the number of HIV infections in this BCC in 2001 and 2002 were 2 and 1 respectively, it is clear that there is no increase in IR. This outcome is rather the result of the limited amount of historical data considered.

Table 1: Analysis of Dutch HBV, HIV and HCV repeat donor infection data.

Description	Reporting year				Average number of infections	Observed incidence rate x 10^5 donor years (95% CI)[a]	*p*-value test for trend (t) or outlier (o) detection
	2003	2004	2005	2006			
HBV infections							
BCC 1	0	4	2	0	1.50	1.27 (0.51-2.58)	0.10 (t)
BCC 2	3	1	1	1	1.50	1.34 (0.53-2.71)	0.82 (t)
BCC 3	2	0	3	2	1.75	1.78 (0.76-3.43)	0.30 (t)
BCC 4	4	1	3	2	2.50	2.18 (1.09-3.82)	0.65 (t)
Total	9	6	9	5	7.25	1.64 (1.11-2.31)	0.72 (t)
Comparing BCCs using 2006 data only							0.43 (o)
Comparing BCCs using aggregated data from 2003 through 2006							0.69 (o)
HIV infections							
BCC 1	1	0	0	1	0.50	0.42 (0.07-1.31)	0.45 (t)
BCC 2	0	2	0	0	0.50	0.45 (0.07-1.38)	0.40 (t)
BCC 3	0	2	1	1	1.00	1.01 (0.31-2.36)	0.24 (t)
BCC 4	0	0	1	2	0.75	0.65 (0.16-1.69)	0.04 (t)
All BCCs	1	4	2	4	2.75	0.62 (0.32-1.06)	0.16 (t)
Comparing BCCs using 2006 data only							0.65 (o)
Comparing BCCs using aggregated data from 2003 through 2006							0.74 (o)
HCV infections							
BCC 1	0	2	0	1	0.75	0.64 (0.16-1.65)	0.29 (t)
BCC 2	0	1	0	2	0.75	0.67 (0.17-1.74)	0.08 (t)
BCC 3	0	0	1	1	0.50	0.51 (0.08-1.57)	0.17 (t)
BCC 4	0	0	0	1	0.25	0.22 (0.01-0.96)	0.14 (t)
All BCCs	0	3	1	5	2.25	0.51 (0.24-0.92)	0.01 (t)
Comparing BCCs using 2006 data only							0.92 (o)
Comparing BCCs using aggregated data from 2003 through 2006							0.77 (o)
Any infection							
BCC 1	1	6	2	2	2.75	2.34 (1.21-4.00)	0.20 (t)
BCC 2	3	4	1	3	2.75	2.46 (1.28-4.21)	0.58 (t)
BCC 3	2	2	5	4	3.25	3.30 (1.81-5.43)	0.20 (t)
BCC 4	4	1	4	5	3.50	3.05 (1.72-4.94)	0.20 (t)
All BCCs	10	13	12	14	12.25	2.76 (2.06-3.61)	0.26 (t)
Comparing BCCs using 2006 data only							0.62 (o)
Comparing BCCs using aggregated data from 2003 through 2006							0.81 (o)

Population size (x 1000)					**Average population (x 1000)**
BCC 1	124	123	112	112	118
BCC 2	124	115	110	100	112
BCC 3	99	100	103	93	99
BCC 4	125	121	110	102	115
All BCCs	472	460	434	407	433

[a] Note that the incidence rate presented is a proxy and equals the number of infections divided by the total number of donors in a given year. This 'rate of positive repeat tested donors' is one of the alternatives for incidence rate estimation given in the EMEA guideline.[2]

Discussion

In this paper two statistical tests are described that can support compliance with the requirements prescribed by the EMEA for the monitoring of epidemiological data on blood transmissible infections.[2] By applying the tests to observational data it is shown that it is feasible to monitor IRs using these tests, even in the case of a low number of observations per BCC. The tests indicate that there are no differences in IR between Dutch BCCs, but that there is a statistically significant increase in national HCV IR in the Netherlands (*p*-value of 0.01).

Alternatively, one could use control charts to monitor infection rates. Even though statistical hypothesis testing is a fundamentally superior approach, control charts do provide an intuitively appealing presentation of the data which facilitates the interpretation of the results. It is therefore recommended to provide control charts in addition to statistical test results when testing for trends.

Each test performed has a probability of giving a false positive outcome, depending on the significance level chosen. The outcome of each of the tests from Table 1 was judged on *p*-values smaller than 5%. In Table 1 the outcomes of in total 28 tests are considered. With a 5% probability of a false positive test outcome for each of these tests, the likelihood that at least one of these judgments is unjust is considerable. One could try and correct for this by adjusting the significance level for each individual test such that an overall false positive rate target is met. However, the setting of significance levels in this application is rather a matter of fine-tuning the monitoring strategies to balance true and false positive test outcomes by the decision maker. Therefore, enhancing awareness of the impact of the choice of significance levels should be the goal rather than striving for a reduction of the false positive rate itself. The power and risk curves of the tests presented in this paper are intended for that purpose.

One of the difficulties with the incidence monitoring strategies is (fortunately) the scarcity of infections. Even in the Dutch situation having BCCs with over a hundred thousand donors, the number of observed events is small. Only by pooling the data (looking at all BCCs) sufficient power is obtained to detect a statistically significant trend in the number of HCV infections. It is therefore recommended to use as much relevant historical evidence as possible when comparing BCCs, as this improves the power to detect deviations.

As was shown in this paper, limiting the monitoring data to consider a historical perspective of only three years can lead to erroneous conclusions. On the other hand, factors that will influence the number of observed infections will change over time. This includes not only the IR itself, but also the sensitivity of the screening tests used, the size of the donor population (which is correct for), the donor population characteristics (e.g. compilation, donation frequency) and the quality of the data collection process. This generally means that the longer the historical perspective, the less informative the data will be with respect to the current situation. Therefore, a balance has to be found between validity and quantity of the data available to obtain the best quality inference on basis of the information available. The party most equipped to make these judgments would be the blood establishment.

There is an advantage of looking at the data at the broadest level possible, e.g. national level, to increase the power of detecting deviations. This strategy, however, will not reveal potential concentrated outliers, which is the primary objective of the incidence monitoring guideline. Monitoring smaller BCCs will result in fewer observations and therefore more uncertainty associated with the IR estimated. This means that larger deviations are required before these can be identified as such. It is, however, possible to estimate the IR for any BCC – regardless how small the BCC size – when using the time since last infection. Given a maximum acceptable IR and a significance level, it is possible to determine a 'time between infections interval' for each BCC. In case the time since last infection for a particular BCC would be less than this interval, the IR at that BCC would exceed the set IR threshold. This technique, known as reliability testing, is widely applied in industry.[9,12]

In our example we used the rate of positive repeat tested donors as a proxy for the IR. These proxies are by far more easily obtainable but will be slightly different from the 'true' IRs, which take into account the donation history of each individual donor. When using the proxies consistently for comparing centres or looking for trends, unless donation patterns vary significantly, the findings based on these proxies can be expected to be identical to those based on exact IRs. Likewise, we use IRs as a proxy for the risk of infectious disease transmission. For HIV and HCV there is a direct relationship between the observed number of infections and the residual risk of infection. For HBV the relation between the observed infections and the IR is less straightforward as a result of the fact that the HBV viremia is transient and can occur unnoted in between donations.[13] However, the HBV infection risk will increase with an increasing IR as well. As long as there is no interaction between the probability of obtaining the infection and the donation interval (which determines the likelihood of an infection being observed), the observed number of HBV infections will be a proper marker for the risk of infection. Such an interaction is indeed considered very unlikely.

Despite the transparency of the EMEA guideline, and the clear goal it serves, this expert-based guideline in some respects is not operationally elaborated. Consequently, a number of issues that have to be resolved before a set of risk management decision rules can be constructed which for any data collected for EMEA submission provide an unequivocal judgment on outliership or change over time. The main question is: "Which contamination risk is considered (un)acceptable?". Another EMEA guideline, on risks of plasma-derived medicines, requires manufacturers to analyze the contamination risk of their final products.[14] Even though this guideline requires quantification of risk, unfortunately, it does not provide any indication as to what an acceptable contamination risk of a final product is. A pre-defined product safety level would, in combination with production (viral inactivation capacity of various processing steps, batch sizes, quarantine periods) and donor population characteristics (number of donors and their respective donation frequencies), allow specification of an acceptable number of contaminated donations found in screening tests.[15] Specifying acceptable risk levels, however, is a difficult and disputable task. This is mainly because of the fact that even though "risk" itself can be defined quite clearly, it has multi-facetted attributes which will be valued differently by different individuals and parties.[16] Risks are only acceptable when there is a positive trade-off: in blood transfusion the trade-off is that the blood can

be lifesaving for patients. The acceptability of risk therefore strongly depends on the perceived balance between risks and benefits by the decision maker. Given that there is a BCC with an elevated IR, the requirement for a safety intervention will for instance depend on: the size of the increase in IR (What is its effect on the overall transmission risk?); the safety interventions applied (How effective is a foreseen safety measure?, What are the associated costs?); the overall level of risk (Is there need for intervention at all?); the size of the BBC (How will safety measures affect the availability of the blood supply?); and the type or location of BCC (What will be the message to the donor- and general population?). It is clear that no unified solution can be provided to answer all these questions. What can be done, however, is to provide tools that can determine whether there *is* a BCC with an increased IR and whether there *is* a positive trend in IR. Such tools respond directly to the requirements formulated in the EMEA guideline.[2]

The risk curves presented in this paper allow an interpretation of the benefits from increasing the significance levels of the various tests. However, even with these curves available it is difficult to balance these benefits against the false positive rate as the likelihood and size of an increased IR remains unknown. Nonetheless, being responsible for the process of incidence monitoring, management is now confronted with explicating a decision on the level of risk they are unwilling to accept, at which level of (un)certainty and at which cost (of action). With the monitoring tools described in this paper, these can be formalised by: (1) the selection of an (un)acceptable IR increase for an outlier BCC, (2) the selection of a significance levels for the outlier test and the test for trend. The graph on achievable risk reduction aids in defining appropriate IR ratio levels. The risk graphs for the statistical tests support the selection of appropriate significance levels. The decision maker might want further information on the relation between the increase in IR and the consequences in terms of a number of expected disease transmissions. In addition, he might want these expressed in terms of loss of health and associated costs. However, once the testing parameters are set, applying the statistical tests is straightforward and allows full compliance with EMEA regulations.

References

1. Seitz R, Heiden M, Nubling CM et al. *The harmonization of the regulation of blood products: a European perspective*. Vox Sang 2008.
2. The European Agency for the Evaluation of Medicinal Products. *Guideline on epidemiological data on blood transmissible infections*. EMEA/CPMP/BWP/125/04. 20-1-2005.
3. *Directive 2002/98/EC of the European Parliament and of the Council of 27 January 2003 setting standards of quality and safety for the collection, testing, processing, storage and distribution of human blood and blood components and amending Directive 2001/83/EC*. Official Journal of the European Union 2003;L33:30-40.
4. Cox LA. *Risk Analysis: Foundations, Models and Methods*. Assinippi Park, Massachusetts, USA: Kluwer Academic Publishers, 2001.
5. Pawitan Y. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford, GB: Oxford University Press, 2001.
6. Birnbaum A. *Statistical Methods for Poisson Processes and Exponential Populations*. Journal of the American Statistical Association 1954;49:254-66.
7. Daly L. *Simple SAS macros for the calculation of exact binomial and Poisson confidence limits*. Comput Biol Med 1992;22:351-61.
8. Ng HK, Tang ML. *Testing the equality of two Poisson means using the rate ratio*. Stat Med 2005;24:955-65.
9. Kenett R, Zacks S. *Modern Industrial Statistics*. Pacific Grove, CA, USA: Brooks/Cole Publishing Company, 1998.
10. Perry MB. *Robust change detection and change point estimation for poisson count processes*. 28-5-2004. Florida State University Dissertation.
11. *Sanquin Annual Reports: 2003, 2004, 2005, 2006*. Amsterdam: Sanquin Blood Supply Foundation, 2007.
12. Cushing M. *Mathematica templates for sequential reliability/maintainability test design and analysis*. TR-687. 1-6-2001. Maryland, AMSAA.
13. Korelitz JJ, Busch MP, Kleinman SH et al. *A method for estimating hepatitis B virus incidence rates in volunteer blood donors. National Heart, Lung, and Blood Institute Retrovirus Epidemiology Donor Study*. Transfusion 1997;37:634-40.
14. The European Agency for the Evaluation of Medicinal Products. *Guideline on assessing the risk for virus transmission - New chapter 6 of the note for guidance on plasma-derived medicinal products (CPMP/BWP/269/95)*. EMEA/CPMP/BWP/5180/03. 21-10-2004.
15. Janssen MP, Over J, van der Poel CL et al. *A probabilistic model for analyzing viral risks of plasma-derived medicinal products*. Transfusion 2008;48:153-62.
16. Fischhoff B, Lichtenstein S, Derby SL, Slovic P, Keeney RL. *Acceptable Risk: A Critial Guide*. Cambridge, USA: Cambridge University Press, 1981.

APPENDIX A: DESCRIPTION OF TREND AND OUTLIER DETECTION TESTS

In this appendix two tests (a test for outlier detection and a test for trend) are described. As both test are likelihood ratio tests, this appendix will start with a brief introduction to likelihood ratio testing.

Likelihood ratio testing

A likelihood-ratio (LR) test is a statistical test in which a ratio between the maximum likelihood of a result under two different hypotheses is calculated. The value of this ratio is indicative for which of the hypotheses is most likely. The two hypotheses are generally referred to as the null-hypothesis (H_0) and the alternative hypothesis (H_a). The numerator of the LR test corresponds to the maximum probability of an observed result under the alternative hypothesis. The likelihood ratio is a random variable with a probability distribution. The distribution of 2log(LR) will tend to be a chi-squared distribution for large sized samples. This fact can be used to determine an associated *p*-value, the likelihood of unjustly rejecting H_0. However, as in the case of monitoring infections only few observations are available, *p*-values are obtained by means of simulation. This is done by generating samples from H_0 and calculating the LR for each of these samples. The *p*-value now becomes the fraction of samples with a LR higher than the LR found using the observed data.[5]

Detecting an outlier BCC using a likelihood ratio test

Presume n BCCs with D_j donors where the average IR per donor in each BCC equals λ_j. Assuming that the number of infections in each BCC is Poisson distributed, the donor population IR per BCC can be estimated by maximizing the likelihood using the observed number of infections N_j in each BCC j with donor population size D_j:

$$L(\lambda) = \prod_{j=1}^{n} \frac{\left(\lambda_j D_j\right)^{N_j}}{N_j!} e^{-\lambda_j D_j}, \tag{A.1}$$

where $\lambda = (\lambda_1, \ldots, \lambda_n)$ is the array of annual donor population IRs per BCC. The likelihood ratio test statistic T now becomes:

$$T = \frac{\sup_{\lambda \in H_\alpha} L(\lambda)}{\sup_{\lambda \in H_0} L(\lambda)} \tag{A.2}$$

Where, for fixed $R \geq 1$, the null and alternative hypotheses become:

$$H_0 = \left\{ \lambda \in [0,\infty)^n : \lambda_{(n)} \leq \frac{R \sum_{i=1}^{n-1} \lambda_{(i)} D_{(i)}}{\sum_{i=1}^{n-1} D_{(i)}} \right\} \tag{A.3}$$

$$H_a = \left\{ \lambda \in [0,\infty)^n : \lambda_{(n)} > \frac{R\sum_{i=1}^{n-1} \lambda_{(i)} D_{(i)}}{\sum_{i=1}^{n-1} D_{(i)}} \right\}$$

Here, $\lambda_{(1)}$ indicates the smallest of the set $\{\lambda_1, \ldots, \lambda_n\}$, $\lambda_{(2)}$ indicates the second smallest, etcetera.

Detecting a trend using a likelihood ratio test

The test for trend is based on a LR test whereby the null- and alternative hypotheses are formulated as:

$$\begin{aligned} H_0 : \quad & \lambda_i = \lambda_0 \qquad \text{for } i = 1, \ldots, n \\ H_a : \quad & \lambda_i \geq \lambda_{i-1} \qquad \text{for } i = 2, \ldots, \tau \\ & \lambda_{\tau+1} > \lambda_\tau \\ & \lambda_i \geq \lambda_{i-1} \qquad \text{for } i = \tau + 2, \ldots, n \end{aligned}$$

For H_0 the mean IR is found the most likely value for λ_0. For H_a the convex lower bound is the most likely value for λ_i. The convex lower bound is the largest convex function that interconnects the cumulative number of observed IRs. The formulation of the likelihood and its test is identical to that of the test for outlier detection as formulated in equations (A.1) and (A.2).

APPENDIX B: VARIOUS POWER CURVES AND THEIR CONVERSION TO RISK CURVES

In this appendix power curves for the test for outlier BCC detection and the test for trend are given. Also, the rationale and construction of the risk curves as given in the paper are described.

Detecting an outlier BCC

There are four regional BCCs in the Netherlands. Data collection and reporting occurs per BCC on annual basis. To assess the ability of the statistical test to detect elevated IRs in the Dutch setting, the test was evaluated for a setting with four BCCs for IRs of 0.5, 1, 2 and 4 infections per BCC per year. For each of the reference IRs n_r (which is applicable to three BCCs), the probability of a positive test outcome was calculated as a function of the IR of the fourth (outlier) BCC. Each graph in Figure B1 this probability is given for a different significance level α. For each of the reference IRs two lines are shown: the probability of a positive test outcome for the outlier BCC (solid) and the probability of a positive test outcome for any of the BCCs (dashed). The latter is the total probability of a positive test outcome, and as such the sum of 'true' and 'false' positives.

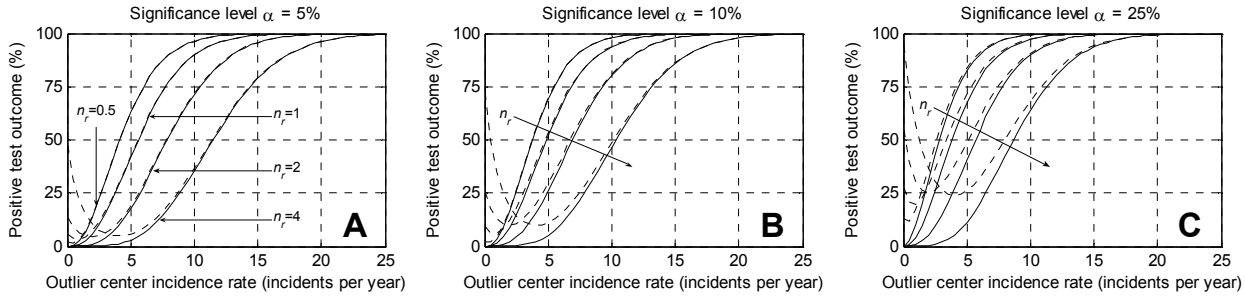


Figure B1: Probability of a positive test outcome as a function of outlier BCC incidence rate (IR), reference IR (n_r) for various significance levels (α). Tests are performed for an IR ratio (R) of 1. The dashed line indicates the probability of a positive test outcome; the solid line indicates the probability of a positive test outcome where the BCC of which the IR is varied is actually identified as outlier.

An increasing significance level of the test (a higher value for α) will cause the test to produce a positive test outcome more easily. This will reduce the likelihood of a false negative test outcome, but will at the same time increase the likelihood of a false positive test outcome. This effect is clearly shown in Figure B1: graph C (testing at a significance level of 25%) shows that with a reference IR (n_r) of 4 the probability of a positive test outcome for an outlier BCC IR of 10 is about 60%. When the test is performed at a 5% significance level (graph A), this probability is only about 40%. At the same time, the false positive test outcome rates are 5% and 25% respectively when the outlier BCC IR is 4.

The high 'false' positive test outcome probabilities for low outlier BCC IRs are in fact true positives. This seemingly contradictory result is caused by the fact that whenever the outlier IR becomes less than the reference IR, the average IR of any two reference BCCs plus the 'outlier' BCC *is* lower than that of the remaining other reference BCC. In that case all reference BCCs become 'outlier' BCCs. As expected, the minimum positive test outcome probabilities are therefore located at the reference IR and the value of this minimum equals the significance level of the test. The graphs show that this is not the case for the lowest reference IR (n_r=0.5). This is a result of the limited range of possible outcomes associated with low IRs.

Detecting a trend

In Figure B2 the probability of a positive outcome of the test for trend as a function of an increase in IR is given for different reference IRs and significance levels. On the x-axis of Figure B2 the annual IR increase is given. The reference IR is the IR in the first year of observation.

Most of the general observations from Figure B1 are applicable to Figure B2 as well. In Figure B2 the effect of increasing the significance level of the test (a higher value for α) on the probability of a positive test outcome is clearly illustrated. For all but the lowest reference IR the probability of a positive test outcome can be found to be close (or equal) to the significance level wherever the x-value equals zero (no trend). The relative sensitivity of the test increases with an increase in reference IR.

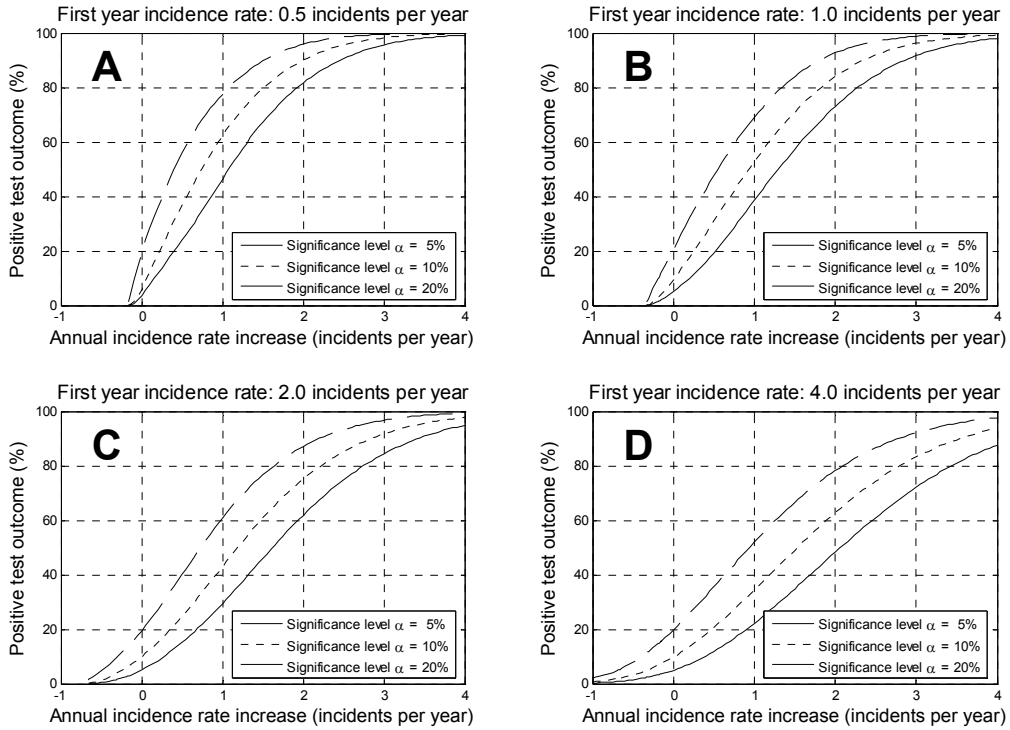


Figure B2: Probability of a positive test outcome as function of annual IR increase, reference IR (first year IR) and significance level α.

From power to risk curves

The power curves from Figure B1 and B2 show the probability of a positive test outcome as a function of the outlier BCC IR (B1) or the annual IR increase (B2). The test has a probability of a positive test outcome depending on the significance level and the deviation the test is intended to detect. Even though the latter is unknown, the decision maker has to decide on the significance level a priori. The residual risk of an undetected transmission is a function of the IR: the higher the IR, the higher the risk of transmission. This implies that the risk from a 2% probability of not detecting an IR of 5 is identical to the risk from a 1% probability of not detecting an IR of 10.

The average risk from an IR increase with probability *p* and probability density function *D* equals:

$$R_{ir} = p \int_{ir=ir_0}^{\infty} D(ir) \cdot ir \cdot \left(1 - p_{\alpha}(ir)\right) d(ir) \qquad \text{(B.1)}$$

Where:

R_{ir}	=	risk from an undetected increased IR
p	=	probability of the increased IR
$D(ir)$	=	probability density function of the increased IR
ir	=	increased incidence rate
ir_0	=	threshold IR above which the test outcome should become positive
$p_{\alpha}(ir)$	=	probability of positive test outcome as a function of significance level α and the increased incidence rate (ir)

The likelihood of an increased IR is unknown, let alone its probability distribution. Therefore, $D(ir)$ is presumed to be constant and represented by a uniform distribution over $[ir_0, \infty]$. However, infinity as the right limit renders $D(ir)$ (and therefore R_{ir}) to become zero. One could also set $D(ir)$ to 1, which turns R_{ir} into a measure of cumulative risk. This provides a reasonable base for comparing risks, but one that is rather difficult to interpret directly. Alternatively one could set $D(ir)$ to a value that does allow a direct interpretation to R_{ir}. As a first step the expected value of an undetected IR (*UIR*) is calculated where it is presumed that there is an undetected IR increase of unknown magnitude. Considering every IR equally likely and α is set to 5%, the *UIR* equals

$$UIR = \frac{E\big(x \cdot f(x)\big)}{E\big(f(x)\big)} = \frac{\int_{ir=ir_0}^{\infty} ir \cdot \big(1 - p_{5\%}(ir)\big)\, d(ir)}{\int_{ir=ir_0}^{\infty} \big(1 - p_{5\%}(ir)\big)\, d(ir)} \tag{B.2}$$

Apart from the probability of an increased IR (p) this formula is equivalent to formula B.1 when α = 5% and $D(ir) = \frac{1}{\int_{ir=ir_0}^{\infty} \big(1 - p_{5\%}(ir)\big)\, d(ir)} = D_{5\%}$

Using this value for $D(ir)$ the interpretation of R_{ir} becomes: *'the risk from an IR increase* ($p \cdot D_{5\%}$) *of which the probability of an undetected increased IR, when testing with a significance level of 5%, is equal to 1'*. This definition for $p \cdot D(ir)$ is used for all risks curves in this paper:

$$R_{ir} = \frac{\int_{ir=ir_0}^{\infty} ir \cdot \big(1 - p_{\alpha}(ir)\big)\, d(ir)}{\int_{ir=ir_0}^{\infty} \big(1 - p_{5\%}(ir)\big)\, d(ir)} \tag{B.3}$$

One might argue that considering *any* IR increase equally likely is an unrealistic assumption. However, the probability of a positive test outcome will rapidly become 100% when the IR (or the annual IR increase in case of trend detection) increases. This implies that this assumption will only ensure equal consideration of all increases that could potentially be missed by the test. As the range of incomplete detection is an order of magnitude or (in most cases) less, this assumption does hold in practice. In case a decision maker finds another distribution of $D(ir)$ more appropriate, the calculated risk can be adjusted accordingly.

GENERAL DISCUSSION

“Nothing is more difficult, and therefore more precious, than to be able to decide.”

Napoleon Bonaparte

General discussion

Decision making on blood safety interventions requires a valid comparison of outcomes following the options at stake. Few would argue that this could be achieved by randomized controlled trials (RCT's) in view of the extreme rarity of events, the complex relations involved, the demanding ethical environment, and the time it would take before conclusions can be drawn. This thesis argues the case to apply a modelling approach instead, allowing for a comprehensive analysis of the entire blood transfusion chain (BTC), from donor to recipient.

In this general discussion, first a description of the elements of the BTC and associated modelling aspects will be given. Next, the advancements from the work described in the papers from this thesis will be discussed. Subsequently, critical aspects and key benefits of modelling in general will be summarized, followed by some examples of current blood safety models with their limitations and room for improvement. Finally, the role of models within the decision making process will be discussed, followed by a general conclusion.

The components of the blood transfusion chain and modelling their relations

The blood transfusion chain starts with a donor providing his of her blood, hence with descriptive epidemiology of the donor and of donation behaviour. The risk of transmission of any infectious disease by blood transfusion has a linear association with its incidence rate and the cost-effectiveness of any associated safety intervention is proportional to the inverse of this rate. The probability of a blood sample containing infectious particles bears relation to several features of the donor. Additionally, donation behaviour affects the likelihood that an infected donor is identified at a next visit to the blood bank and this may especially be of interest for the analysis of plasma and plasma derived medicinal products (PDMPs).

Donor selection rests on a complex combination of considerations, as does testing for infectious diseases at the blood bank. Too strict donor selection potentially endangers the total volume blood collected, and some criteria are also ethically sensitive. When considering testing, account has to be taken of the so called 'window period' which is the time between the actual infection of a donor and the moment that this infection can be detected by the currently applied screening test.

The next chain concerns the way the donor blood is processed by the blood bank. Ideally, this is done in such a way that clinical effectiveness is optimized with minimal risk of adverse events for future recipients. Combining blood from different donors will increase the probability of contamination of a final product roughly by the number of products that are combined (provided that the initial risk is low). Combining products implies an increase in risk for any particular product, but *not* an increase in risk for the recipient population as a whole. Contrary, splitting of donor blood *will* increase the risk of disease transmission for the population (that is, as long as splitting does not eliminate infectivity), as it leads to a multiplication of the number of recipients that are exposed. This clarifies the cause for concern with respect to PDMPs where blood (fractions) from

up to 20.000 donors are combined for collective processing, and subsequently divided into a large number of small units for clinical use. The infectivity of a particular blood partition depends on its ability to bind infectious actors: some partitions are more infectious than others. However, blood components are generally modelled as equally infectious (100% infectivity).

Administration to the patient is the next chain. Blood products are administered either through intravenous injection (blood components and PDMPs) or intramuscular injection (PDMPs only). First of all, the administration despite all precautions, is prone to error and itself represents an infectious risk.[1] From a modelling point of view not only exposure in terms of the number of products a patient receives is relevant, but so is the timing of these transfusions. For example, consider the case of a patient who lifelong receives 240 transfusions of which one will result in immediate death. Now imagine these 240 transfusions either administered within the first year or evenly distributed over 20 years to come. The expected loss of health due to the infectious agent is very different for each of these instances. The latter can be regarded a Russian roulette game that might result in administering the infected transfusion last (conditional on survival). The expected loss of life of the latter case is therefore substantially less than the first case.

The final chain concerns disease models. Such (Markov) models predict disease progression in an infected patient and the patients' quality of life as the disease progresses. They are used to determine the cumulative effect (in terms of costs and health outcomes) of adverse events.

Given the complexity of the BTC, a fair estimation of the impact of safety measures is difficult to obtain. The rather mechanistic and reasonably well-understood nature of the BTC, the fairly accurate and complete data on most components of the BTC, and the suitability of the entire process for description by mathematical models enable an evaluation of risks and health economical aspects *in silico*.

A review of the papers presented in this thesis

In this thesis various methods and models are described to either model or evaluate blood safety. The paper on the PROTON study (Chapter 1) provides descriptive statistics on transfusion recipients. This paper forms an exception to all other papers, as it does not model a relation between BTC components, but provides epidemiological data that are indispensable for any computation of outcomes through changes in the BTC. Chapter 1 largely confirmed existing knowledge concerning blood transfusion recipients. The added value is that it puts numbers to various transfusion recipient characteristics. It also showed a gender imbalance with respect to the use of FFP and PLT blood components: in the Netherlands men receive 1.5 times more of such components than do women. We do not know whether this reflects an inequality in diseases requiring BT, or inequality of treatment.

Chapter 2 concerns the estimation of survival after transfusion. In this paper it is shown that the standard Kaplan-Meier survival estimation method can be applied to calculate survival after transfusion. This fact is not self-evident as transfusion usually is not a single dose single event, but patients generally obtain multiple transfusions. First, there is a relationship between the number of transfusions that a patient receives and his mortality rate, which is evidently a dependency between transfusions given to any single

individual. This dependency however does *not* affect the estimation of survival after transfusion as there is no dependency between the timing of the various transfusions a patient obtained. Second, there *is* a dependency between transfusions given to any particular patient in the sense that the patient is the unit of observation. Therefore, the variance of the survival estimate will be influenced by the number of patients observed rather than by the number of transfusions observed. This means that the variance estimates of the standard survival methods fail (these are downward biased). This chapter concerns a statistical estimation problem that is very specific to the analysis of the cost-effectiveness of blood safety interventions. Nevertheless, it is a very important detail as survival after transfusion is the main determinant of cost-effectiveness estimates. It also shows how important critical review is whenever applying existing models in a non-standard setting. We had many discussions on the applicability of the standard Kaplan-Meier method, and in the end turned out to be both right and wrong.

A straightforward and rather simplistic mathematical model underlies Chapter 3, which compares the cost-effectiveness of risk reduction by bacterial culturing with that of pathogen reduction. A number of simplifying assumptions were made for this model, not all of which were realistic. Like for example the assumption that pathogen reduction is 100% effective. This simplification was made to reduce the complexity of the model and to avert the efforts of obtaining more detailed information on the degree of pathogen reduction. The price paid was an overestimation of the cost-effectiveness of pathogen reduction. However, this bias is irrelevant with respect to the decision concerning which of the two interventions is most cost-effective. It proves the point that "the better is the enemy of the good". In this paper the concern as to the relevance of the cost-effectiveness threshold proposed for medical interventions in this context is raised. This point will be discussed further on in this chapter.

Chapter 4 concerns the cost-effectiveness of additional HBV screening. Here the age distribution of transfusion recipients, as well as the incremental cost-effectiveness ratio as function of age is shown (Chapter 4, Figure 2). This graph highlights how strongly the incremental cost-effectiveness ratio depends on recipient age. On average the safety intervention can be considered reasonably cost-effective, however, for elder patients this is definitely not the case. In this chapter it is also shown how sensitivity analyses can provide valuable insights into the cost-effectiveness of blood safety interventions. By doing so it again touches on the discussion on optimal blood use, or rather, on the limits of the acceptability of costs for blood safety.

In Chapter 5 we describe an enhancement of the value of information concept, whereby, unlike in a conventional VOI analysis, the decision and the valuation of the (clinical) outcomes are separated. This enables a new perspective on the VOI outcome: namely terms of the expected costs and expected effects. This allows a decision maker to balance the expected outcome of acquiring additional information in terms of clinical outcomes as well as a valuation in monetary terms. Also, this concept provides the entrance to applying VOI in settings with complex decision criteria, for instance in a situation where preference is determined by a combination of both cost-effectiveness and budget impact.

In Chapter 6 a probabilistic model for analyzing viral risks of PDMPs is described. This model was specifically developed to comply with new EU regulations on PDMPs. Risk models used previously were unable to adequately incorporate donor characteristics. For simplicity such calculations in the past started with presuming a contaminated plasma

pool. A more realistic assessment of the risk of contamination requires both modelling of donor behaviour (in terms of donation frequencies) and modelling of the sensitivity of screening tests applied. This is a typical example of a model where the critical features and their interrelationship are easily described, but where the construction and computation of the model output is a challenge, as this requires a discrete event or probabilistic modelling approach. This is most likely what hindered the implementation and use of such a model in the past.

Chapter 7 concerns regression analysis of serial dilution assay data. A new regression model was developed which allows direct estimation of the parameters of interest. In the past a two-step approach was used where first the observations were converted into reduction factors, which where then used in a linear regression model. Not only did the new approach provide a more valid model for estimation, it also changed the significance of regression parameters for the example the model was applied to. The application of the new regression model can be seen as a significant step forward in the quality of viral validation studies performed at Sanquin.

Chapter 8 illustrates the development and application of a true decision-making tool for the BTC. Statistical tests and models were developed and applied to help interpret observed infections in the donor population. In principle these statistical tests can be applied as such. However, we went one step further and mapped costs versus benefits of applying a particular decision threshold using so-called risk curves. However, further work needs to be done before these can be put into practice. This work would involve incorporation of risk outcomes, but also elicitation of decision maker preferences and beliefs concerning the likelihood of the actual increase of various incidence rates.

Blood safety models: room for improvement

When reading the papers in this thesis one may become aware that a modellers' job is never done: models can always be upgraded in terms of additional model functionality, whenever new data becomes available, or as a result of new insights with respect to the system they are designed to represent. However, to undertake such efforts it must be clear upfront that there will be a substantial pay-off as any effort related to the extension or rework of any of these models will be considerable.

On the other hand, there are a number of areas were there is certainly room for improvement to models currently applied within blood safety. The first concerns the generally applied window period model. This model presumes that infections are not detected in any donations made within the window period and that beyond the window period infections are always detected. Furthermore, infections are presumed to be 100% infectious. It is obvious that the likelihood of detecting an infection will increase gradually over time, as infectivity will increase with time since infection, with a final detection probability that will be dependent on the sensitivity of the test applied. An existing algorithm which incorporates these effects, taken from earlier studies, is part of our model used to calculate the risk of PDMPs.[2] Similar models could also be applied for the assessment of the risk of disease transmission by blood components.

Another area for improvement is that of the incidence estimation for hepatitis B infections. The commonly used model divides the infected population into three categories: (1) individuals who are chronic carriers of the disease and will always be

detected by the screening test (5%); (2) individuals who show a primary antibody response and will never be detected with a HBsAg test (25%); (3) individuals who have transient antigenemia and are therefore only detectable for a limited time period (70%).[3] The model presumes that a significant proportion of the donor population can not be detected by HBsAg screening at all, and the likelihood of detection for the largest proportion of donors is dependent on their donation interval. This results in a significant underestimation of the HBV incidence rate when based on observed incidents alone. With the sensitivity of current screening test and knowledge on HBV infections this model seems outdated. An improved and more realistic estimate of the HBV incidence in the donor population which combines population dynamics, test sensitivity and virus characteristics, could be easily established.

Incidence rate estimation amongst repeat donors is currently limited to a two-year time horizon.[4] This limitation can cause a significant upward bias in repeat donor incidence rate estimates. This estimate can be easily improved by using all donation data available. However, this might impact the validity of comparing incidence rate data across countries as such data are unlikely to be available for analysis in many countries. Therefore a balance has to be sought between statistical precision and practicability.

The data obtained in the PROTON study allow subgroup analyses based on transfusion recipient age, gender, disease, number of transfusions, combination of blood products obtained or on any combination of these factors. This enables identification of patient clusters where the use of blood safety interventions is not (cost-)effective at all. However, such analyses are only useful if the results can lead to a differentiation in blood safety for different patient groups. This would require the development of a decision framework which includes political, ethical, practical, legal as well as economic aspects of blood safety.

Models used to analyze the cost-effectiveness of blood safety interventions tend to involve merely the path from donor to transfusion recipient. However, an infected donor can, and in some cases is even likely to, cause further transmission of the disease. Modelling of such secondary transmissions will add to the (cost-)effectiveness of safety interventions. However, this would require social network modelling, specific to the transfusion recipient population. As discussed in this thesis, the mapping of the transfusion recipient is the level of knowledge that is currently available. The mapping of social networks of transfusion recipients, with the aim of modelling secondary transmission of infectious diseases, is one of the areas for future research in this field.

Many infectious diseases are efficiently transmitted by blood transfusion. Therefore, secondary transmissions are particularly important when considering emerging infections, where blood transfusion might play a critical role in the spread of the disease in the general population.

Critical aspects and key benefits of modelling

Models aim to help structuring a problem in such a way that account is taken of all aspects which are important and that any uncertainties about outcomes are expressed in such a way that decisions can be made. There generally is a whole range of issues involved in the process of developing mathematical models.

First of all, models have to be as simple as possible without losing relevance to the system they are designed to represent. Even though this sounds easy, it is one of the

more difficult tasks in model development. Usually the modeller starts off with a complex mental picture of reality. Next, this mental picture is reduced to the simplest form possible. The art of modelling now lies in the skill of preserving all essential aspects of the original system and their interrelationships whilst converting it into a limited number of mathematical expressions. The benefits of simple models are ample: not only are they easier to explain and communicate, they are also easier to validate and less prone to (implementation) errors.

A second limitation of models is that they require a certain capability for abstract or conceptual thinking to be understood. This implies that it's natural appeal and usability is not evident at all. This should be a point of continuous attention to any modeller, especially when communicating models or their outcomes to a wider audience.

Also, the process of model development is all but standard and as such difficult to protocol and verify. Nevertheless, there are guidelines for the appraisal of (for instance health economic) models.[5-7] These require models to at least: be replicable, have face validity and be open to external scrutiny. Moreover, with the increasing range of application, in many other areas – the list of which is endless – guidelines on good practice concerning the development of mathematical models can be found in the literature.[8-12] Translational use of such guidelines may improve the quality of models developed.

And after one has developed a model, there is always the nagging question whether it is actually valid. Whilst models that predict the weather can be validated and potentially adapted every day, no (or only limited) such possibilities exist for blood safety models. This is because when modelling risks, almost per definition there is only very little data available for model validation. Therefore, model validation has to rely strongly on critical (peer) review.

Models themselves can be considered a risk as well. In literature this risk, referred to as 'model risk', is defined as the risk of developing a model that does not fit the system that is supposed to represent.[13,14] In the end it is often the qualification and experience of the model developer(s) that determines the quality of a model. This makes transparency and room for external validation key to model acceptance. These should always be aware of not applying the Truman approach, referring to his famous quote "*If you can't convince them, confuse them.*"

Regardless of the reservations one may have about models, there are many advantages associated with their development and there is no field of research where they are not intensively applied.[15,16] The development process itself requires the modeller to become familiar with various aspects of a system by means of inquiries, and as such results in an enhancement of the understanding of the system by its operator. This easily occurs when modelling a complete system (like for instance the blood transfusion chain) whereby an overview of all aspects of a system is collected, but can just as easily occur when modelling a small detail of a system.

Once developed, models allow us to simulate complex systems, make extrapolations, perform sensitivity analysis and answer various what-if questions. By analyzing and interpreting analysis results an in-depth understanding of the system can be obtained that was simply impossible before.

Models in decision making

Various models that were developed in the context of blood safety have been presented in this thesis. The incentive for model development was either to improve current practice (e.g. apparent limitations), to enable further model development (e.g. methods development), or to derive specific characteristics (e.g. the cost-effectiveness for various testing alternatives). When specific characteristics are required, usually a model for decision making support is developed. However, it is clear that in this context not only by definition the scope of a model is limited. A model is merely a reflection of the real world (e.g. a system) within a particular setting, and with preset boundaries. The decision maker, who bears responsibility on the functioning of this system, is generally an actor within a complex field of internal and external influences, all again within a wider societal context (Figure 1). The decision maker will therefore in general not be able to take a decision based on model outcomes alone, but rather by optimizing the decision within the full context of his or her awareness.

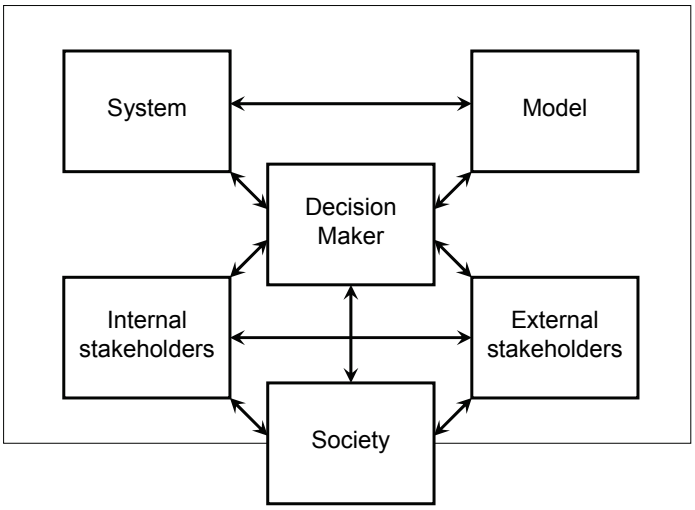


Figure 1: Interactions between a (blood transfusion) system, the system model, the decision maker and various stakeholders, all part of society.

This explains why 'optimal blood safety', as referred to in the governmental policy statement on the Dutch blood provision, is extremely difficult to obtain.[17] In addition, a 'true' (mathematical) optimum requires a formal definition and valuation of benefits and losses. It is clear that such definitions are difficult to formulate when considering the full decision making context as depicted in Figure 1, if not impossible as their interpretation are subjective by definition.[18]

As models will never be able to cover all aspects relevant to the decision maker, it is their role is to inform the decision maker, not to replace the decision making process. For example, one may want to consider that a perceived lack of safety affects attitudes towards blood donation or leads to sub-optimal use of blood by physicians or negative responses by blood recipients. The question is whether the management of such perceptions and the consequences should to be addressed in a model. Theoretically, the

answer is preferably yes. However, such an approach would need estimates that might be impossible to obtain and predictions that would be surrounded with tremendous uncertainties, overshadowing all that *is* known. This illustrates the fact that there are limits to the usefulness of quantitative models, especially in highly uncertain and complex settings.[19]

Still, one of the reasons for developing cost-effectiveness and risk models in blood safety is that they can be used as a standard for comparing interventions. Not only do they enable comparing different strategies for the implementation of tests for a specific disease, they also allow comparing tests for different diseases and even allow a comparison of safety interventions across areas of application. Part of this standardization may also be that one defines a threshold to discern what is acceptable and what is not.

In the Netherlands there is some guidance as to what is considered to be an acceptable expense for health gain in a healthcare setting.[20] Expenses for blood safety clearly exceed these expenses by orders of magnitude.[21] However, if compared to expenses made for other safety interventions, they seem far less extravagant.[20] It now becomes debatable as to which of these standards is applicable.

This dilemma can be illustrated by means of a small example. Presume that there is a new surgical intervention which is extremely effective and life saving. However, the intervention is rather costly and its cost-effectiveness ratio is twice as high as the cost-effectiveness threshold considered acceptable. It now seems odd to decide *not* to perform this intervention because of its too high cost-effectiveness ratio, whereas one is quite willing to spend far *more* money for the safety of the blood used for the surgical procedure that is currently performed than for the health gain by the intervention itself. Spending this money on the new procedure would evidently result in more lives saved. The apparent conflict originates from reasoning that it is not rational to accept inefficient use of health care budgets. However, even though not performing the new surgical intervention would lead to more deaths, one will never be able to name the patient that would not have died during surgery in case the new procedure had been applied. The outcome is merely a statistic. However, an infectious disease transmitted by blood transfusion is all but a statistic: it is a patient with a name and a face. Not only will such a transmission of disease – given that all other causes can be excluded – be related to a past transfusion, but also to a particular donor that was infected at the time of donation. It is an event that is adverse to the intention of the undergone treatment, and more, it is an event that might have been prevented. It is our desire – and even more so that of a decision maker responsible for blood transfusion safety – to prevent such a scenario from happening that renders the higher price to be paid acceptable. Therefore, the cost-effectiveness of blood safety interventions should be measured against standards for safety interventions instead of standards for health care interventions.

When looking for a general standard for blood safety one may want to realize that formulation of such a standard is a fiction. Not only the perspectives might differ, like in the previous example, but the risks may also involve many different outcomes that will be interpreted and valued very differently by different stakeholders. In case of the risk of HIV transmission by blood transfusion for instance, the maximum number of transmissions per infected individual is known and limited. For vCJD however, certain doom scenarios with a large number of disease transmissions cannot be excluded. Imagine the impact of a lifelong frequent plasma donor revealing a vCJD infection. One

may want to prevent such doom scenarios at any price, implying that its effects are weighed by their perceived impact as well. It is clear that such considerations are not, or at least far less, at stake when considering competing familiar health care interventions. As such, a general guideline for an acceptable safety level is infeasible.

Arguing the above, one may still want to try and quantify what ones priorities are such that rational decisions can be taken based on what one defines as being acceptable. When one does not, uncertainty may result in a tendency for decision makers who are accountable for their decisions to act precautionary: "rather safe than sorry". Such tendencies can be particularly appealing in a political setting where costs of safety will be posed on the general public, whereas benefits of avoiding these risks are on the side of the decision maker. The 'marketing of fear' can be instrumental in such decision making processes and will irrevocably lead to non-optimal decision making.

Discussions on what is acceptable in terms of costs per QALY and costs per life year gained may take a long time and might even be irresolvable. One may want to help guide such discussions by defining an intermediate threshold and next explore the boundaries of which risks are acceptable, and under which restrictions. Consensus on an acceptable safety level would alleviate the decision making process. Decisions with respect to new safety interventions would have a justifiable base of reference and the decision making would be transparent. Also, taking a stand on what is an acceptable risk might stimulate the consideration or development of alternative strategies for managing such risks. If we are able to lay down in Dutch law acceptable levels for flood risks, why not do the same for blood risks?[22] Such an attempt, at least for the transmission risks of well known viruses, should be feasible as these risks and benefits are well understood and quantified. This might not result in an 'optimal' blood safety level per se, but would result in a transparent and disputable decision process and defendable decisions concerning blood safety interventions. The results from the PROTON study form an indispensable support for such a decision process.

Conclusion

Models fulfil the need to improve insight and are able to deliver, be it at a price: they force reality into a rigid mathematical structure. Good models however, force focus on the essential aspects of a system which allows deriving valid conclusions. They offer the opportunity to develop novel insights, explore alternatives and view outcomes from new perspectives. They deliver an irreplaceable and unequivocal contribution to the quality of decision making within blood safety.

A final remark about the use of models may be that even when one doesn't develop models in a way as presented in this thesis, one still uses models. Any structured way of thinking considering a policy problem follows a similar process – but now not formalized – to come to conclusions and as such uses a model. Such informal models are the real black boxes: they are impossible to criticise or verify and rest on authority alone. Scary.

References

1. Goodnough LT. *Risks of blood transfusion*. Anesthesiol Clin North America 2005;23:241-52.
2. Weusten JJ, van Drimmelen HA, Lelie PN. *Mathematic modeling of the risk of HBV, HCV, and HIV transmission by window-phase donations not detected by NAT*. Transfusion 2002;42:537-48.
3. Korelitz JJ, Busch MP, Kleinman SH et al. *A method for estimating hepatitis B virus incidence rates in volunteer blood donors. National Heart, Lung, and Blood Institute Retrovirus Epidemiology Donor Study*. Transfusion 1997;37:634-40.
4. The European Agency for the Evaluation of Medicinal Products. *Guideline on epidemiological data on blood transmissible infections*. EMEA/CPMP/BWP/125/04. 20-1-2005.
5. Weinstein MC, O'Brien B, Hornberger J et al. *Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies*. Value Health 2003;6:9-17.
6. NICE. *Guide to the Methods of Technology Appraisal*. N0515. 2004. National Institute for Clinical Excellence.
7. Philips Z, Bojke L, Sculpher M et al. *Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment*. PharmacoEconomics 2006;24:355-71.
8. Sheiner LB, Steimer JL. *Pharmacokinetic/pharmacodynamic modeling in drug development*. Annu Rev Pharmacol Toxicol 2000;40:67-95.
9. McMahon A, Heathcote J, Carey M, and Erskine A. *Guide to Good Practice for the Development of Conceptual Models and the Selection and Application of Mathematical Models of Contaminant Transport Processes in the Subsurface*. National Groundwater & Contaminated Land Centre report NC/99/38/2. 2001. Bristol, Environmental Agency.
10. *Good Practice Guide for Atmospheric Dispersion Modelling*. ME number: 522. 2004. Wellington, Ministry for the Environment.
11. Refsgaard JC, Henriksen HJ. *Modelling guidelines--terminology and guiding principles*. Advances in Water Resources 2004;27:71-82.
12. Esward TJ and Sokhan VP. *Mathematical Modelling for Metrology on the Nanoscale*. DEM-ES 005. 2005. Teddington, National Physics Laboratory.
13. Derman E. *Model risk*. 1996. Goldman, Sachs & Co. Quantitative Strategies Research Notes.
14. Taleb NN. *The Black Swan: The Impact of the Highly Improbable*. London: Penguin Books, 2007.
15. Hosking RJ, Venturino E. *Aspects of Mathematical Modelling: Applications in Science, Medicine, Economics and Management*. 1st ed. Basel: Birkhauser Verlag AG, 2008.
16. Gershenfeld N. *The nature of mathematical modeling*. Cambridge: Cambridge University Press, 2003.
17. VWS. *Ministerieel Plan Bloedvoorziening 2009-2011; Een beleidsdocument voor de middellange termijn*. 19-11-2008. Den Haag.
18. Saaty TL. *Decision making with the analytic hierarchy process*. International Journal of Services Sciences 2008;1:83-98.
19. Walker WE, Harremoes P, Rotmans J et al. *Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support*. Integrated Assessment 2003;4:5-17.
20. *Zinnige en duurzame Zorg*. 2006. Zoetermeer, Raad voor de Volksgezondheid en Zorg (RVZ).
21. van Hulst M, De Wolf JT, Staginnus U et al. *Pharmaco-economics of blood transfusion safety: review of the available evidence*. Vox Sang 2002;83:146-55.
22. *Waterwet*. Staatsblad 2009;490:1-44.

SUMMARY / SAMENVATTING

Summary

This thesis describes the development and application of methods and models to support decision making on safety measures aimed at preventing the transmission of infections by blood donors. Apart from the direct interest of selecting best practices, the thesis aims to increase awareness of the opportunities offered by quantitative decision models in this context.

When referring to safety measures within blood safety one can think of screening tests for blood donors, quarantine periods for blood plasma, or methods for detecting or eliminating contaminations during processing. To support decision making processes in such contexts, costs as well as effects of safety interventions need to be analyzed. Although medical knowledge by default is obtained by carrying out careful real life experiments, this approach is infeasible in blood transfusion safety. The main reason is the fact that the risks involved are very low, which means that events are (very) rare. This poses limitations on the practicality of such experiments. This is where mathematical modelling is put forward to proceed. Models rest on an aggregate mathematical description of the *'real world'* such that conclusions concerning the modelled system can be drawn.

Analyzing blood safety interventions requires modelling of the blood transfusion chain (BTC). The BTC refers to the interrelation between on one hand the blood donors and on the other the blood transfusion recipients. The main ingredients for any model concerning the BTC safety interventions will include donor epidemiology, blood bank interventions, recipient characteristics, and the costs and effects associated with adverse events in patients.

The thesis consists of a collection of papers which (a) aim to support decisions pertaining to blood safety, and/or (b) concern the application of quantitative models and methods which were developed for that purpose. Main outcomes studied in these models are either the cost-effectiveness of safety interventions or the transmission risk of infectious agents. The thesis consists of three parts:

In *Part A* descriptive statistics and survival modelling of blood transfusion recipients are addressed.

Chapter 1 describes the outline of the PROTON study. In this study a random sample of 20 Dutch hospitals was taken which covered 28% of all blood transfused in the Netherlands between 1996 and 2006. This allows the construction of descriptive statistics of the Dutch blood transfusion recipient in terms of gender, disease and survival per type of blood component, which were previously unavailable. Of all red blood cells (RBC), fresh-frozen plasma (FFP) and platelets (PLT), respectively 1.7%, 2.5% and 4.5% were transfused to neonates. Recipients of 65 years or older received 57.6%, 41.4% and 29.0% of all RBC, FFP and PLT transfusions respectively. Most blood products were transfused to patients with diseases of the circulatory system (25.1%) and neoplasms (22.0%).

In Chapter 2 a detailed study is performed on how to estimate recipient survival after transfusion. In this chapter it is shown that the standard Kaplan-Meier survival

estimation method can be applied to calculate survival after transfusion despite the fact that patients can receive multiple transfusions, which implies that there is a dependency between observed follow-up times. This dependency however does not affect the survival after transfusion estimate as there is no dependency between the timing of the various transfusions a patient obtained. However, there is a dependency between transfusions given to any particular patient in the sense that the patient is the unit of observation. Therefore, the variance of the survival estimate will be influenced by the number of patients observed rather than by the number of transfusions observed. This means that the variance estimates of the standard survival method fail for this application, and that these are biased downward.

Part B concerns the cost-effectiveness of several blood safety interventions, the uncertainty in the estimated outcomes and the implications of this uncertainty for decision making.

In Chapter 3 the cost-effectiveness of bacterial culturing (BCU) is compared to that of pathogen inactivation (PRT) for detecting bacterial contamination in platelets. The costs per 100,000 platelet concentrates in the Netherlands are estimated at €2,520,794 for BCU and at €14,294,495 for PRT. In comparison to the situation without BCU and PRT, costs per quality-adjusted life year (QALY) are estimated at €69,767 for BCU (95%CI: €13,960-€1,606,811) and at €382,057 for PRT (95%CI: €110,731-€6,285,487). The ratio of differences in costs and QALYs between BCU and PRT (the relative cost-effectiveness) is estimated at €2,766,351 (95%CI: €846,639-€19,043,550). Large uncertainty in sepsis complication rates and platelet recipient survival (at the time of publication) exists, causing a large uncertainty in the absolute cost-effectiveness for both interventions. These render the cost-effectiveness ratio estimates of BCU and PRT highly uncertain. However, despite these large uncertainties, which affect the cost-effectiveness BCU and PRT similarly, it can be concluded that BCU is without doubt more cost-effective than PRT.

In chapter 4 the cost-effectiveness is determined of screening for hepatitis B using a nucleic acid screening test methodology (HBV NAT) in addition to the existing serological screening test. This is done for individual testing (ID) or combined blood sample testing in pools of six (MP-6). The incremental cost-effectiveness ratio (ICER) of adding HBV MP-6-NAT or HBV ID-NAT in the Netherlands is €303,218 (95%CI: €233,001-€408,388) and €518,995 (95%CI: €399,359-€699,120) per QALY, respectively. For both strategies the ICER is strongly dependent on the age of transfusion recipients. The cost-effectiveness of additional HBV NAT is limited as a result of the relatively small number of life years lost by HBV transmissions. Despite a higher effectiveness, HBV ID-NAT is less cost-effective than MP-6-NAT due to its higher costs.

In Chapter 5 an enhancement of the value of information (VOI) concept is described, whereby, unlike in the conventional approach, there is a separation between the decision and the valuation of the (clinical) outcomes. This novel attributable VOI concept enables a new perspective on the VOI outcome: namely in terms of the expected costs and effects. This allows a decision maker to balance the expected outcome of acquiring additional information in terms of clinical outcomes as well as in terms of its monetary equivalent. This concept also provides the entrance to applying VOI in settings with complex decision criteria, for instance in a situation where preference is determined by a combination of both cost-effectiveness and budget impact. As attributable value of

information results can be obtained by merely reorganizing the already available data, this extension is recommended for all value of information analyses.

Part C concerns the evaluation of risks of plasma derived medicinal products.

In Chapter 6 a generic probabilistic model for analyzing risks of plasma derived medicinal products is described. Existing, previously used risk models were unable to adequately incorporate donor characteristics. The new model allows an analysis of the effectiveness of various risk reduction strategies. Sensitivity analyses show that the contamination risk of a finished product increases linear with viral incidence rate in the donor population, and is inversely related to product yield and to the viral reduction capacity of the production process. The contamination risk is less sensitive to changes in the pool size of the screening test, the average donation frequency, and the inventory hold period. There is only a limited dependency on the donation type (apheresis or whole blood donations) and a negligible dependency on the size of the manufacturing pool. The new probabilistic model renders a vast improvement to past practice where not only donor characteristics were ignored, but also contamination risks were determined by merely stacking worst case assumptions.

Chapter 7 describes the development and application of a regression model dedicated to the evaluation of the robustness of the viral inactivation capacity of the production process against various process conditions. In the past, the robustness was determined using a linear regression model on viral reduction estimates based on serial dilution assay readings of individual experiments. The new model regresses on all the serial dilution assay readings directly. It overcomes conceptual problems with the earlier approach and provides stronger *p*-values to the relevant covariates. The new approach is recommended for the analysis of all future robustness studies.

Chapter 8 concerns the analysis of observed infections in the donor population and subsequent decision making. It describes statistical tools and methods developed to signal changes or outliership of incidence rates in the donor population of specific blood collection centres. Analyses did not reveal any significantly increased incidence rates among donor centres in The Netherlands, nor were any significant increases in incidence rates in individual centres observed. However, on national level a statistically significant increase in hepatitis C virus incidence was observed (*p*-value of 0.01). The statistical tests described are generic and can be applied by any blood establishment or plasma fractionation institute.

In this thesis a whole range of models and their application have been described: from models for donor epidemiology through the processing of blood plasma to the analysis of transfusion recipient survival and the evaluation of costs and effects of safety interventions. All of these models aim to enhance insights and are able to deliver, be it at a price: they force reality into a rigid mathematical structure. However, when forcing focus on the essential aspects of a system they still allow deriving valid conclusions. Models offer the opportunity to develop novel insights, explore alternatives and view outcomes from new perspectives, and therefore deliver an irreplaceable and unequivocal contribution to the quality of decision making within blood safety.

Samenvatting

Dit proefschrift beschrijft de ontwikkeling en toepassing van methoden en modellen ter ondersteuning van de besluitvorming ten aanzien van interventies gericht op het voorkomen van de overdracht van infecties door bloeddonors. Naast het directe belang om de huidige stand der techniek te illustreren, is het proefschrift erop gericht om het bewustzijn van de mogelijkheden die kwantitatieve beslismodellen in deze context te bieden hebben te vergroten.

Bij bloedveiligheidsinterventies kan gedacht worden aan bijvoorbeeld screeningtests voor bloeddonoren, quarantaineperiodes voor lang houdbare producten, of methoden om contaminatie tijdens de verwerking op te sporen of te elimineren. Ter ondersteuning van besluitvormingsprocessen in deze context moeten zowel de kosten als de effecten van de veiligheidsinterventies worden geanalyseerd. Terwijl medische kennis normaliter wordt verkregen door het zorgvuldig uitvoeren van klinische experimenten, is een dergelijke aanpak bij bloedveiligheid niet mogelijk. De belangrijkste reden is dat de betreffende risico's zeer gering zijn, waardoor de ongewenste uitkomsten (zeer) zeldzaam zijn. Dit levert beperkingen op voor de uitvoerbaarheid van dergelijke experimenten. Een aanpak met wiskundige modellering biedt dan uitkomst. Modellen zijn gebaseerd op een wiskundige beschrijving van de 'echte wereld', zodanig dat valide conclusies over het gemodelleerde systeem getrokken kunnen worden.

Het analyseren van bloedveiligheidsinterventies vereist het modelleren van de bloedtransfusieketen (BTC). De BTC verwijst naar de onderlinge relatie tussen enerzijds de bloeddonors en anderzijds de ontvangers van bloedtransfusies. De belangrijkste ingrediënten voor een model met betrekking tot de veiligheid van BTC interventies omvatten donorepidemiologie, bloedbankinterventies, kenmerken van de transfusieontvanger, en de aan bijwerkingen in patiënten gerelateerde kosten en effecten.

Dit proefschrift bestaat uit een verzameling van publicaties die (a) erop gericht zijn om beslissingen in verband met bloedveiligheidsinterventies te ondersteunen, en/of (b) betrekking hebben op het toepassen van kwantitatieve modellen en methoden die voor dat doel werden ontwikkeld. Belangrijkste uitkomstmaten betreffen de kosteneffectiviteit van bloedveiligheidsinterventies en het risico van overdracht van infectieuze agentia.

Het proefschrift bestaat uit drie delen:
Het eerste deel betreft een aantal beschrijvende statistieken en het modelleren van de overleving van bloedtransfusieontvangers.

Hoofdstuk 1 beschrijft de eerste uitkomsten van de PROTON studie. In dit onderzoek werd een aselecte steekproef van 20 Nederlandse ziekenhuizen genomen die 28% van alle bloedtransfusies in Nederland tussen 1996 en 2006 omvatte. Dit maakt beschrijvende statistieken van de Nederlandse bloedtransfusieontvanger mogelijk in termen van geslacht, ziekte en overleving per type bloedproduct, die voorheen niet beschikbaar waren. Van alle rode bloedcellen (RBC), vers bevroren plasma (FFP) en bloedplaatjes (PLT), is respectievelijk 1,7%, 2,5% en 4,5% getransfundeerd aan pasgeborenen. Ontvangers van 65 jaar of ouder ontvingen respectievelijk 57,6%, 41,4%

en 29,0% van alle RBC, FFP en PLT transfusies. Het merendeel van de bloedproducten is getransfundeerd aan patiënten met hart- en vaatziekten (25,1%) en nieuwvormingen (22,0%).

In Hoofdstuk 2 beschrijft een gedetailleerde studie naar de manier waarop de overleving van transfusieontvangers na transfusie moet worden geschat. In dit hoofdstuk wordt geïllustreerd dat de standaardmethode om overlevingskansen te schatten (de Kaplan-Meier methode) inderdaad kan worden toegepast om de overleving na (een willekeurige) transfusie te berekenen. Dit ondanks het feit dat patiënten meerdere transfusies kunnen ontvangen, wat impliceert dat er een afhankelijkheid tussen waargenomen follow-up tijden bestaat. Deze afhankelijkheid is echter niet van invloed op de geschatte overlevingskans na een transfusie omdat er geen afhankelijkheid is tussen de timing van de verschillende transfusies die een patiënt gekregen heeft. Er is echter wel een afhankelijkheid tussen transfusies gegeven aan een bepaalde patiënt in die zin dat de patiënt de eenheid van observatie is. Vandaar dat de variantie van de geschatte overleving wordt beïnvloed door het aantal waargenomen patiënten in plaats van door het aantal waargenomen transfusies. De met standaardmethoden berekende variantie van de overlevingskans is in deze toepassing daardoor verkeerd, en wijkt af naar beneden.

Het tweede deel van het proefschrift betreft de kosteneffectiviteit van verschillende bloedveiligheidsinterventies, de onzekerheid in de geschatte uitkomsten en de implicaties van deze onzekerheid voor de besluitvorming.

In hoofdstuk 3 wordt de kosteneffectiviteit van bacteriële kweek van bloedplaatjes (BCU) vergeleken met dat van pathogeen inactivatie (PRT). De kosten per 100.000 bloedplaatjes in Nederland wordt geschat op €2.520.794 voor de BCU en op €14.294.495 voor PRT. In vergelijking met de situatie zonder BCU en PRT worden de kosten per voor kwaliteit van leven gecorrigeerd levensjaar (QALY) geschat op €69.767 voor de BCU (95% CI: €13.960-€1.606.811) en €382.057 voor de PRT (95% CI: €110.731-€6.285.487). De verhouding van het verschil in kosten en het verschil in QALY's tussen de BCU en PRT (de relatieve kosteneffectiviteit) wordt geschat op €2.766.351 (95% CI: €846.639-€19.043.550). Er bestaat veel onzekerheid ten aanzien van de kans op sepsis als complicatie en (op het moment van publicatie van dit artikel) de overleving van ontvangers van bloedplaatjes. Daardoor bestaat er grote onzekerheid over de absolute kosteneffectiviteit van beide interventies. Deze onzekerheden maken de schatting van de kosteneffectiviteitsratio's van zowel BCU als PRT hoogst onzeker. Echter, ondanks deze grote onzekerheden, die eenzelfde effect hebben op zowel BCU als PRT, kan met zekerheid worden geconcludeerd dat BCU kosteneffectiever is dan PRT.

In hoofdstuk 4 is de kosteneffectiviteit bepaald van screening op hepatitis B met behulp van een nucleïnezuur-test (HBV NAT) naast de al bestaande serologische test. De analyse is uitgevoerd voor zowel het testen van individuele donaties (ID) als het gecombineerd testen van donaties van zes donoren (MP-6). De incrementele kosteneffectiviteitsratio (ICER) van het toevoegen van HBV MP-6-NAT of HBV ID-NAT in Nederland is respectievelijk €303.218 (95% CI: €233.001-€408.388) en €518.995 (95% CI: €399.359-€699.120) per QALY. Voor beide tests is de ICER is sterk afhankelijk van de leeftijd van de transfusieontvanger. De kosteneffectiviteit van de additionele HBV NAT test is beperkt door het relatief kleine aantal levensjaren dat wordt verloren bij een HBV-

transmissie. Ondanks de hogere effectiviteit is als gevolg van de hogere kosten HBV ID-NAT minder kosteneffectief dan MP-6-NAT.

In hoofdstuk 5 wordt een uitbreiding van het concept 'informatiewaarde' ('Value of Information',VOI) beschreven waarbij, in tegenstelling tot de conventionele benadering, onderscheid wordt gemaakt tussen de (klinische) uitkomst en waardering ervan. Met dit nieuwe 'toerekenbare VOI' concept ontstaat een extra perspectief op het VOI resultaat: namelijk die van de verwachte kosten en verwachte effecten. Dit biedt een beslisser de mogelijkheid om de waarde van additionele informatie te wegen op basis van de klinische uitkomsten naast het monetaire equivalent ervan. Dit concept opent de mogelijkheid tot het toepassen van het VOI concept als er sprake is van complexe beslissingscriteria, bijvoorbeeld wanneer de voorkeur wordt bepaald door een combinatie van criteria met betrekking tot kosteneffectiviteit en begrotingsbeslag. Aangezien de berekening van toerekenbare informatiewaarde niet meer vereist dan het reorganiseren van al beschikbare gegevens wordt deze uitbreiding aanbevolen voor alle informatiewaarde analyses.

Het derde en laatste deel van het proefschrift betreft de risico's van uit bloedplasma bereide geneesmiddelen.

In hoofdstuk 6 wordt een generiek probabilistisch model voor de analyse van de risico's van uit plasma bereide geneesmiddelen beschreven. Eerdere risicomodellen waren niet in staat om donorkenmerken adequaat te mee te nemen. Het nieuwe model maakt een analyse van de effectiviteit van verschillende strategieën ter beperking van contaminatierisico's mogelijk. Uit gevoeligheidsanalyses blijkt dat het contaminatierisico van een eindproduct lineair stijgt met de incidentie van infectieziekten in de donorpopulatie, en omgekeerd evenredig is met de opbrengst en de virusreductie-capaciteit van het productieproces. Het contaminatierisico is minder gevoelig voor veranderingen in het aantal gezamenlijk geteste donaties, de gemiddelde donatiefrequentie, en de vrijgaveperiode voor plasma. Er is slechts een beperkte afhankelijkheid van het type donatie (aferese of volbloed) en een verwaarloosbare afhankelijkheid van de grootte van de productiepool. Het probabilistische model biedt een enorme verbetering ten opzichte van de vroegere aanpak waar niet alleen donor kenmerken niet werden meegenomen, maar waarin ook het contaminatierisico werd bepaald op basis van een aaneenschakeling van 'worst case' aannames.

Hoofdstuk 7 beschrijft de ontwikkeling van een regressiemodel voor het evalueren van de robuustheid van de virusreductiecapaciteit van het productieproces van plasmamedicijnen voor variaties in procescondities. In het verleden werd deze bepaald met behulp van lineaire regressie op schattingen van virusreducties die waren bepaald op basis van seriële verdunningsassay-metingen van individuele experimenten. Het nieuwe model schat de regressiecoëfficiënten rechtstreeks op alle uitkomsten van de seriële verdunningsassays. Daarmee wordt een aantal conceptuele problemen met de eerdere aanpak opgelost en worden lagere *p*-waarden voor daadwerkelijk relevante covariaten gevonden. De nieuwe aanpak wordt aanbevolen voor de analyse van alle toekomstige robuustheidsstudies.

Hoofdstuk 8 heeft betrekking op de analyse van en besluitvorming rondom waargenomen infecties binnen de donorpopulatie. Het beschrijft statistische methodes en gereedschappen die zijn ontwikkeld en toegepast om verschillen of veranderingen in incidentie binnen de populatie van specifieke afnamecentra te detecteren. De analyses

laten zien dat er geen Nederlandse centra zijn met een significant verhoogde incidentie, noch dat er centra zijn met een significante toename in incidentie. Echter, op nationaal niveau is er wel een statistisch significante toename in de incidentie van hepatitis C-virus (*p*-waarde van 0,01) zichtbaar. De beschreven statistische tests zijn generiek en kunnen worden toegepast door iedere willekeurige organisatie voor bloedtransfusie of plasmafractionering.

In dit proefschrift is een heel scala aan modellen en hun toepassing beschreven: van donorepidemiologie via de verwerking van bloedplasma tot aan de analyse van de overleving van transfusieontvangers en de evaluatie van de kosten en effecten van veiligheidsinterventies. Al deze modellen zijn erop gericht om inzicht verhogen en zijn in staat om dit te realiseren, zij het tegen een prijs: ze persen de werkelijkheid in een rigide wiskundige structuur. Desondanks, mits toegespitst op de meest essentiële aspecten van een systeem, kunnen er valide conclusies mee worden afgeleid. Modellen bieden de mogelijkheid om nieuwe inzichten te ontwikkelen, alternatieven te onderzoeken en de resultaten bekijken vanuit een nieuw perspectief, en leveren daarmee een onvervangbare en onmiskenbare bijdrage aan de kwaliteit van de besluitvorming rondom bloedveiligheid.

DANKWOORD / ACKNOWLEDGEMENTS

Na heel wat jaren aan dit proefschrift te hebben gewerkt zijn er natuurlijk ook heel wat mensen om te bedanken. Dit kan zijn vanwege een inhoudelijke bijdrage, maar vooral ook doordat zij hebben bijgedragen aan de ontzettend fijne en inspirerende tijd die ik de afgelopen jaren bij het Julius heb gehad.

Op de eerste plaats wil ik bedanken mijn co-promotor Cees van der Poel, want zonder jou, Cees, zou dit werk er gewoonweg niet zijn geweest. Nadat ik na mijn eerste project bij het Julius had besloten dat onderzoek het voor mij toch niet was, kwam ik volkomen onverwacht en onvoorbereid met jou in een overleg. Er was een instantane klik, en die duurt voort tot op de dag van vandaag! Je enthousiasme voor de bloedtransfusie is aanstekelijk en je kennis van het onderwerp en de mensen die erin werkzaam zijn is enorm. Verder ben je een kei in het bewaken van een positieve werksfeer en het is dan ook een feest om met je samen te werken: heftige discussies over onderwerpen die soms helemaal niets met het onderzoek te maken hebben, onverwachte gedachtesprongen, veranderingen van perspectief, en vooral niet te vergeten de analogieën met de muziek. Ook ben ik altijd weer onder de indruk van je schrijfcapaciteit. Niet alleen rolt er in no-time een ijzersterke nieuwe notitie of onderzoeksaanvraag uit je pen, vooral je commentaar op mijn stukken is altijd weer ijzig scherp. Daarnaast ben je een bijzonder sensitief persoon: voordat je de drempel van onze kamer bent overgestapt en je hoed aan de kapstok hebt gehangen weet je al precies hoe ieders pet er die dag bij staat. Ook je politieke neus en goede contacten in het veld zijn meer dan eens cruciaal gebleken. Het zijn voor mij bijzonder mooie en leerzame jaren om op terug te kijken en ik hoop dan ook van harte dat je nog lang mee blijft draaien in onze groep!

En dan promotor Ben van Hout. Allesbehalve alledaags, iemand met gebruiksaanwijzing. Reden voor mij om je in het eerste jaar, voor zover mogelijk, te mijden. Ik heb je pas ontdekt toen ik met een biostatisticus niet uit een probleem kwam en daar vervolgens jouw mening over vroeg. Je hebt de echt unieke gave om het probleem van de ander eerst volledig te willen (en kunnen!) doorgronden voordat je ook maar begint om in de richting van mogelijke oplossingen te denken. Je bent in staat om razendsnel complexe zaken tot hun essentie terug te brengen en op basis daarvan haarscherpe conclusies te trekken. Dat laatste is nogal eens frustrerend, want je blijkt uiteindelijk inderdaad toch altijd weer gewoon gelijk te hebben… Er valt nog veel meer te zeggen over je onorthodoxe manier van werken en je geweldige gevoel voor humor. Ik hou het er maar bij dat je een geweldig pragmatisch denker bent met een bijzonder groot hart. Ik wil je hierbij dan ook bedanken voor alles wat ik van je heb mogen leren en voor de bijzonder open en vriendschappelijke manier waarop we hebben samengewerkt. Erg, heel erg jammer dat je zo ver weg zit en ik je zo weinig zie!

Ook wil ik graag bedanken promotor Gouke Bonsel. Beste Gouke, ik heb enorm genoten van jouw aanwezigheid in de bloedgroep, en daarbuiten. Je bent een bijzonder mens. De eerste keer dat je echt indruk op me maakte was toen we samen een uurtje achter de PC gingen zitten om een onderzoeksopzetje te dichten. De structuur en aanhangende tekst die je zo achteloos uit je mouw schudde was zo enorm sterk en produceerde je zo snel dat hij uit een andere dimensie leek te komen! Sinds die tijd is voor mij ook altijd de associatie met de alien gebleven... Eén van de dingen waar ik zeker aan heb moeten wennen is je onomwonden manier van kritiek leveren. Dat is in eerste instantie even

schrikken, maar geeft uiteindelijk zo veel helderheid! Daarnaast zijn er natuurlijk mooie herinneringen aan onze tripjes: Stockholm (onvergetelijke ervaring) en York (echt heel erg gelachen). Je bent echt een geboren verteller en weet een schat aan feitenkennis uit te dragen als ware het een thriller, waarbij het publiek aan je lippen gekluisterd is en alle gevoel voor tijd verliest. In het bijzonder wil ik je bedanken voor het feit dat op een aantal momenten waarop ik een steuntje in de rug nodig had jij altijd diegene was die, ondanks onze fysieke afstand, dit feilloos aanvoelde en er gevolg aan gaf. Ik hoop dan ook dat we in de toekomst zo links en rechts nog wat projecten samen kunnen doen.

Ook wil ik graag bedanken Wim Schaasberg. Beste Wim, al jaren draai je mee in onze groep. In die tijd heb je een ontzettend belangrijke rol vervuld als schakel naar het CBS. Maar veel belangrijker is jouw bijdrage geweest aan de continuïteit van deze club. Je bent daarin de berekenbare factor: altijd de stukken gelezen en becommentarieerd, altijd aanwezig op de vergaderingen en altijd een relativerende of komische kanttekening. Voor mij is dat een onmisbaar element geweest om alles weer in het juiste perspectief te kunnen zien en de motivatie op te kunnen brengen om toch gewoon weer lekker door te gaan. Bijzonder veel dank daarvoor.

Verder wil ik graag bedanken Barbara Borkent-Raven. Mijn eerste 'promovendus'. Bijna vier jaar geleden ben je begonnen aan je promotietraject. Toen nog als onzeker en een beetje bleu studentje dat verbaasd was over het feit dat ze bij ons was aangenomen. Het is machtig mooi om je in de afgelopen jaren te hebben zien uitgroeien tot een krachtige vrouw die weet wat ze wil en wat ze kan. Ik wil je bedanken voor je enorme inzet voor de PROTON studie, maar vooral ook voor de fijne tijd die we als kamergenoten en op buitenlandse trips hebben gehad. Bijzonder dankbaar ben ik ook voor de wijze waarop jij iedereen onder druk hebt gezet om mij vóór jou te laten promoveren. Petje af, het is je gelukt! When the going gets tough, the tough get going.

Eric Cator van de TU Delft wil ik bedanken voor een geweldig leuke en productieve samenwerking de afgelopen jaren. Beste Eric, jij bent min of meer toevallig bij een project gesleept door Roger Cooke en wat een geluk dat jij het was! Je bent met regelmaat voor mij volkomen onnavolgbaar, maar uiteindelijk komen we altijd wel weer uit waar we zijn moeten. Ik hou ontzettend van je directheid en je humor, en heb ontzettend veel van je geleerd de afgelopen jaren. Wat mij betreft gaan we ook in de toekomst weer nieuwe afstudeerders begeleiden.

En dan nog de co-auteurs van mijn eerste artikel: Luc Bonneux en Erik Buskens.
Luc, jou wil ik graag bedanken voor een geweldig leuke samenwerking, voor je heldere kijk, maar vooral ook voor je ondersteunende kritieken bij de zoveelste revisie van ons artikel. Jij hebt daarmee, zonder dat je het je waarschijnlijk bewust bent geweest, voor mij echt een verschil gemaakt. Bedankt!
Beste Erik, hoewel ons gezamenlijke project niet het succes is geworden wat we allebei hadden gehoopt ben ik blij dat ik onder jouw leiding de eerste schreden op het wetenschappelijke pad heb mogen zetten. Naast je gigantische gedrevenheid en je vermogen om altijd weer win-win situaties te vinden respecteer ik vooral dat je heel duidelijk je eigen pad kiest en daarin weet te volharden. Wij hebben elkaar tijdens ons

carotisproject aardig leren kennen, en ik ben blij dat ik weet dat ik altijd bij je aan kan kloppen.

Iemand die in mijn dankwoord niet mag ontbreken is Erik Koffijberg. Hoewel hij maar op één van mijn publicaties als co-auteur genoemd staat, is zijn bijdrage aan mijn werk heel groot. Erik, je bent als sparringpartner binnen het Julius voor mij absoluut onmisbaar. Je bent niet alleen snel en scherp, je bent ook nog eens bijzonder grappig! Zelden heb ik met zoveel plezier aan (de revisie van) een artikel gewerkt als bij ons value of information avontuur. Ik hoop dat er zich ook in de toekomst nog een paar van dat soort 'projectjes' aandienen. Je bent in de loop van de jaren uitgegroeid tot een echte vriend. Many, many thanks.

Ik heb voor dat ik bij het Julius kwam al bij een behoorlijk aantal bedrijven in de keuken mogen kijken als medewerker of als adviseur, maar de betrokkenheid en gedrevenheid die ik bij de medewerkers van Sanquin heb ervaren is in mijn ogen uniek. Ik wil dan ook graag iedereen van Sanquin met wie ik heb samengewerkt de afgelopen jaren daarvoor hartelijk bedanken. Een aantal mensen van Sanquin wil ik daarnaast nog graag even apart vermelden. Op de eerste plaats is dat Jan Over. Jan, alleen dankzij jouw nauwgezetheid en inhoudelijke kennis ben ik in staat geweest om een gedegen model voor plasmafractionering op te tuigen. Ik wil je danken voor je inzichten en vooral ook voor de fijne gesprekken tijdens onze carpoolritjes naar Amsterdam. Bij Sanquin jouw kamer binnenlopen voelt echt als thuiskomen! Dan nog Theo Cuijpers. Eén van die supergedreven mensen van Sanquin waarbij ik altijd aan kan kloppen met de meest onmogelijke vragen. Theo, veel dank voor je vele heldere en vooral ook snelle antwoorden. Tot slot Ernest Briët. Beste Ernest, hoewel wij slechts incidenteel contact hebben gehad ben ik altijd weer erg onder de indruk van de manier waarop en het gemak waarmee jij leiding geeft en neemt. Ik ben dan ook erg blij dat jij op het komende symposium dagvoorzitter wilt zijn!

Ik heb het altijd al erg leuk gevonden om studenten op individuele basis te begeleiden. Daarbij heb ik het geluk gehad om een aantal zeer getalenteerde studenten onder mijn hoede te krijgen. Ik wil Nan van Geloven bedanken voor haar bijzondere leergierigheid en geweldige inzet die heeft geleid tot een afstudeerprijs en een prachtige publicatie in dit proefschrift. Nan, mooi dat je uiteindelijk toch (zij het niet bij ons) in de statistiek terecht bent gekomen. Wie had dat gedacht!

Ook veel dank ben ik verschuldigd aan Ardine de Wit. Beste Ardine, je verstaat de kunst van het luisteren als geen ander. Ik kan altijd bij je terecht en je adviezen zijn per definitie helder, oprecht en adequaat. Een fijner afdelingshoofd had ik niet kunnen wensen.

Kamergenoten zijn altijd bijzonder omdat je simpelweg gedwongen bent om samen een relatie op te bouwen. Met jou, Kathelijn (Fischer), was dat een vanzelfsprekendheid, ondanks het feit dat we zo ontzettend verschillend zijn. Ik kan alles met je delen en krijg daarvoor terug je sprankelende energie, warmte, vriendschap en ijzersterke tips. Ik word er altijd weer blij van als je een dagje op de 7e bent!

I also would like to thank the following individuals with whom I have been – in some cases just figuratively – working together over the past years (in truly random order): Manon Kluijtmans, Theo Evers, Daniel Lewandowski, Mirjam Kretzschmar, Elise Destrée, Rianne Maillé, Leslie Beks, Sabrina Ramwadhdoebe, Gijs Hubben, Kirstin Mangundap, Marie-Josée Mangen, Marie-Emmanuelle Behr-Gross, Stephanie Thurston, Najida Begum, Dirk de Korte, Gosia Kaleta, Ton Los, Jo Wiersum, Roelck Cuperus, Henk van Stel, Welling Oei, Jan-Willem Smeenk, Roger Cooke, Charles Helsper, Jacoba Greving, Desanka Heidsieck, Paul Strengers, Martin Schipperus, Wendy Snijder, Rick Lopuhaä, Fokke Terpstra, Marieke Verschuuren, Lotte van Vlerken, Jeroen de Wit, Donal Cairns, Jean-Marc Spieser, Brian Custer, Frank van Engelenburg, Ed Slot, Muriel van Hemel, René van Hulst, Mayke Koek, Iris de Kruif.
…… en daarnaast nog alle Julianen natuurlijk!

Mark Wijnands wil ik nog even in het bijzonder vermelden omdat hij mij heeft geholpen bij de laatste stappen van het schrijven van dit proefschrift. Mark, ontzettend bedankt voor de verleende steun, maar misschien nog wel veel meer voor de geweldige muziek die we samen maken. En Giel Claessens, Ger Palmboom en Han van Eijk wil ik bedanken voor hun bijdrage aan het ontwerp voor de omslag. Han in het bijzonder voor het definitieve ontwerp: je hebt er echt iets geweldig moois van gemaakt!

Erik Gooijer en Reinier Bosman, dank dat jullie er altijd voor me zijn. Ook weer bij de verdediging van mijn proefschrift.

In het bijzonder wil ik nog mijn vrienden en (schoon)familie bedanken. Ik heb geboft met jullie allemaal! Mijn kinderen wil ik danken voor hun onvoorwaardelijke liefde. Zij hebben mijn leven met heel wat onverwachte dimensies verrijkt! En dan tot slot mijn vrouw Rosan. Al bijna 20 jaar door dik en dun, met pieken en dalen, alles wat het leven kleurt. Jouw liefde en steun voor mij en voor Tijn en Malou zijn overweldigend. Jij bent onze rots in de branding! Mijn wijze liefste vrouw, geweldig dat ik met jou mijn leven mag delen.

Utrecht,
8 augustus 2010

CURRICULUM VITAE & SCIENTIFIC PUBLICATIONS

Mart P. Janssen was born on February 21, 1964 in Nijmegen. He graduated from secondary school (VWO at the Canisius College Nijmegen) in 1982. Later that year he started his Mechanical Engineering training at the Technical University Eindhoven. After his graduation in 1988 he moved to Australia where he worked as a mechanical engineer for the ACT Electricity and Water supply (ACTEW, 1989-1990). After his return to the Netherlands he worked as a risk & reliability engineer for KEMA (1990-1993), on projects within the electricity supply and oil and gas industries. He moved to the UK Atomic Energy Authority, where he worked as a risk management consultant (1994-1996). Here he extended his expertise to transportation (road, rail and air), chemical and flood safety. Subsequently, he worked as senior risk-analysis specialist for the ministry of traffic and water (VWS, Bouwdienst Rijkswaterstaat, 1996-1999). During this time he was also part-time teacher at the department of Mechanical Engineering at The Hague University of Applied Sciences (Haagse Hogeschool, 1996-1997) and on the board of the Dutch Society for Reliability Technology (1995-1998, NVvB, now NVRB). After a sabbatical in 2000 he decided to change his focus to applied statistics and worked for VNU as senior marketing data analyst (Claritas, 2000-2001). In 2002 he started working at the Medical Technology Assessment (MTA) department of the Julius Center (UMC Utrecht) on a project concerning the cost-effectiveness of carotid artery stenting. Since 2004 he has been working on blood safety related technology- and risk assessments as a staff member of the Transfusion Technology Assessment group (TTA), a collaboration of the Julius Center and the Sanquin Blood Supply Foundation.

Borkent-Raven BA, Janssen MP, van der Poel CL, Schaasberg WP, Bonsel GJ, van Hout BA. *Survival after transfusion in the Netherlands*. Vox Sang 2010 Aug 18.

Borkent-Raven BA, Janssen MP, van der Poel CL. *Demographic changes and predicting blood supply and demand in the Netherlands*. Transfusion 2010 Jun 28.

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 Jul;99(1):54-64.

Janssen MP, Koffijberg H. *Enhancing Value of Information Analyses*. Value Health 2009 Sep;12(6):935-41.

van Geloven N, Cator EA, Lopuhaä HP, Janssen MP. *Regression analysis on serial dilution data from virus validation robustness studies*. Statistica Neerlandica 2009 Aug;63(3):245-57.

Janssen MP, Cator EA, van der Poel CL, Schaasberg WP, Bonsel GJ, van Hout BA. *Monitoring viral incidence rates: tools for the implementation of European Union regulations*. Vox Sang 2009 May;96(4):298-308.

Borkent-Raven BA, Janssen MP, van der Poel CL, de Wit GA, Bonsel GJ, van Hout BA. *Cost-effectiveness of additional hepatitis B virus nucleic acid testing of individual donations or minipools of six donations in the Netherlands*. Transfusion 2009 Feb;49(2):311-9.

Janssen MP, de Borst GJ, Mali WP, Kappelle LJ, Moll FL, Ackerstaff RG, Rothwell PM, Brown MM, van Sambeek MR, Buskens E. *Carotid stenting versus carotid endarterectomy: evidence basis and cost implications*. Eur J Vasc Endovasc Surg 2008 Sep;36(3):258-64.

Janssen MP, Over J, van der Poel CL, Cuijpers HTM, van Hout BA. *A probabilistic model for analyzing viral risks of plasma-derived medicinal products*. Transfusion 2008 Jan;48(1):153-62.

Janssen MP, van der Poel CL, Buskens E, Bonneux L, Bonsel GJ, van Hout BA. *Costs and benefits of bacterial culturing and pathogen reduction in the Netherlands*. Transfusion 2006 Jun;46(6):956-65.

Willems A, Janssen MP, Verstegen CP, Bedford T. *Expert quantification of uncertainties in a risk analysis for an infrastructure project*. Journal of Risk Research 2005 Jan;8(1):3-17.