

# THE DATA MANAGEMENT GROUP

## P750

### Improving the quality of HLA typing data in Promise: a pilot study with the British Bone Marrow Registry

A. Griffiths<sup>1,\*</sup>, V. Rocha<sup>2</sup> on behalf of CTIWP

<sup>1</sup>British Bone Marrow Registry, NHS Blood & Transplant, Bristol,

<sup>2</sup>Department of Clinical Haematology, Oxford University Hospitals NHS Trust, Oxford, United Kingdom

**Introduction:** HLA matching is an important predictor of survival after allogeneic stem cell transplants [1]. However, EBMT data on HLA type and number of mismatches between patient and donor are often missing or unreliable in the unrelated donor transplant setting. This could be improved by obtaining data directly from donor registries.

**Material (or patients) and methods:** We identified 1039 allogeneic HSCTs carried out between April 2002 and December 2014 using matched unrelated donors from the British Bone Marrow Registry (BBMR), for which MED-A data were available in Promise. We obtained HLA-A, B and DRB1 typing for these donors and patients from BBMR records; for donors these are the results of confirmatory typing requests (CT) fulfilled by the BBMR, and for patients they are the data supplied by transplant centres when initiating a search for a BBMR donor.

We examined the two data sources for completeness, availability of DNA-based and high-resolution typing, and concordance of the EBMT donor typing with the validated results from BBMR records. Further work will be carried out to calculate match grades directly from the BBMR HLA typing, and compare these with EBMT data. We will also produce Kaplan-Meier curves to illustrate the effect of match grade on survival.

**Results:** In the EBMT database, 392 out of 1039 records (37.7%) had full data on donor HLA-A, B and DRB1 typing, 589 (56.7%) had no data, and the remaining 58 (5.6%) had partial data for some loci. Around 98% of existing HLA typing was DNA-based rather than serological (all loci), and around 75% of typing at the A and B loci and 96% at the DRB1 locus was at high resolution (allelic level). Results in terms of availability and resolution of HLA typing at these loci were very similar for patients.

From BBMR records, donor HLA-A, B and DRB1 typing was available in 1035 out of 1039 cases (99.6%). The proportions of donors with molecular typing and with high resolution typing were similar to the EBMT dataset. For patients, typing was available at all three loci in 899 out of 1039 cases (86.5%). Around 95% of A and B typing and more than 99% of DRB1 typing in patients was DNA-based; 60% of A and B typing and 92% of DRB1 typing was at high resolution.

Where data were available from both sources, the BBMR donor HLA typing was of a better quality than the EBMT data (higher resolution, or with splits resolved down to a single allele) in about half of all cases. For patients, the gain in information was smaller. Overall, more than 80% of EBMT records in this sample were improved by the addition of HLA typing data from BBMR records.

**Conclusion:** Current EBMT data on patient and donor HLA typing can be improved in quantity and quality by obtaining data from a donor registry. This should produce more accurate and precise estimates of the effect of matching on patient outcomes after haematopoietic stem cell transplants.

## References:

1. Passweg JR, Zhang MJ, Rocha V *et al*, "Donor characteristics affecting graft failure, graft-versus-host disease, and survival after unrelated donor transplantation with reduced-intensity conditioning for hematologic malignancies." *Biol Blood Marrow Transplant*. 2011;17(12):1869-73. PMID 21771571.

**Disclosure of Interest:** None declared.

## P751

### Linking MED-A-data to an external database: the GVHD-registry project

B. Lindner<sup>1,\*</sup>, G. Rumpold<sup>2</sup>, S. Gleich<sup>3</sup>, I. Hilgendorf<sup>4</sup>, J. Halter<sup>5</sup>, A. Lawitschka<sup>6</sup>, F. Ayuk<sup>7</sup>, H. Schoemans<sup>8</sup>, H. Greinix<sup>9</sup>, D. Wolff<sup>2</sup> on behalf of the German-Austrian-Swiss GVHD consortium

<sup>1</sup>Austrian Stem Cell Transplantation Registry (ASCTR), Medicine V, Medical University of Innsbruck, Innsbruck, <sup>2</sup>Evaluation Software Development (ESD), Rum, Austria, <sup>3</sup>Dept. of Internal Medicine III, University of Regensburg, Regensburg, <sup>4</sup>Dept. of Internal Medicine II, University of Jena, Jena, Germany, <sup>5</sup>Dept. of Hematology, University Hospital Basel, Basel, Switzerland, <sup>6</sup>St. Anna Children's Hospital, Medical University of Vienna, Vienna, Austria, <sup>7</sup>Dept. of Hematopoietic Stem Cell Transplantation, University Cancer Center Hamburg-Eppendorf, Hamburg, Germany, <sup>8</sup>Dept. of Hematology, University Hospital Leuven and KU Leuven, Leuven, Belgium, <sup>9</sup>Division of Hematology, Medical University of Graz, Graz, Austria

**Introduction:** Late acute and chronic graft-versus-host disease (GVHD) are the leading cause for late treatment related morbidity and mortality after allogeneic hematopoietic cell transplantation (HCT). To capture their incidence and outcome, the German-Austrian-Swiss GVHD Consortium has developed a specific GVHD-registry with concise description forms for evaluating patients with late acute and chronic GVHD according to the NIH criteria. Furthermore, the data of contributing centres can be linked to the MED-A-data of the EBMT database to prevent repetitive documentation and to allow assessment of the incidence of late acute and chronic GVHD in an unselected transplant population.

**Material (or patients) and methods:** Development and programming of the web-based GVHD-registry were performed by the Evaluation Software Development (ESD) company. The design of the database permits storing of detailed GVHD information at numerous time points including documentation of available histopathology and biomarker samples. Moreover, toxicity and comorbidity forms applying the newly developed post-transplant multimorbidity index (PTMI) are available. In order to import MED-A data into the GVHD-registry, identified variables in the EBMT database are linked as background transplant data to the GVHD registry. In cooperation with the EBMT Registry office, data export and import modalities have been established. As a standard procedure we use a pre-designed query from the EBMT Registry for the data export that contains all variables needed for the GVHD database. The import facilities have been designed in a way that all data of participating centres can be imported easily as often as a new export is done. In the import procedure new patient records are created for all patients not yet registered from an earlier import. In the GVHD registry GVHD information can only be added to patients who already exist in the EBMT database and whose data has been imported

in order to prevent duplications. In case of changes in any of the exported data fields, these will overwrite older values during the import procedure. If a patient in the GVHD registry database does not exist in an export/import procedure because of selecting e.g. a different year of transplant, this patient's data remain untouched.

**Results:** We are still in the final pilot phase before making the GVHD database available for all interested members of the GVHD consortium in January 2016, provided that they give their consent to the EBMT Registry office for data transfer to the GVHD-registry. In addition, consent to the GVHD registry rules is required, including participation in the scientific advisory board and joint evaluations.

**Conclusion:** This project is a promising example of combining existing EBMT database information with an external database to generate novel research data for a specific field of interest, in this case GVHD.

**Disclosure of Interest:** None declared.

## P752

### New database results-mapping software is an important source for clinical review

L. Little<sup>1,\*</sup>, R. Powles<sup>1</sup>

<sup>1</sup>Haemato-Oncology, Cancer Centre London, Parkside, London, United Kingdom

**Introduction:** The complexity of haematological malignancies, and individual treatment pathways means close monitoring and review are important for the delivery of safe, optimum care. We have developed a state of the art software package, in conjunction with Teleologic Ltd, that integrates real time haematology and biochemistry results electronically, linked with treatments (electronically prescribed on a Mosaik system), and support administered, into a haemato-oncology clinical database. The software allows clinicians quick access to panoramic graphical illustration of important results in relation to treatment, in the clinic setting.

**Material (or patients) and methods:** This web based system, accessed by patient ID, provides immediate access, in house and remotely, to graphical views of accurate patient data over a three month period, from a chosen date.

To produce the mapping graphs patient ID is entered to a tool bar on the computer monitor, and the date during any treatment inserted (or chosen from a drop-down calendar), providing graphs spanning a three month period from that date; changing the date will produce new graphical illustration in a rapid succession.

Fig 1 illustrates a multiple myeloma patient during sixth line treatment with Pomalidomide, prior to Autologous Stem Cell Transplantation. Treatment cycles are represented by the

vertical bar on the time line; pathology results including platelet levels, white cell and neutrophil counts are mapped on a log scale, with horizontal lines indicating the lower level of normal range, for easier visual recognition of trends. Above the graph blood and platelet support given is recorded, along a line indicating quantity. Growth factor support (peg-filgrastim and filgrastim) administered is represented at the base of the graph.

The database treatment screen has been developed to allow us to record other treatment related toxicity grades such as peripheral neuropathy, nausea, diarrhoea and so forth and linked to the treatment cycles. Results can be downloaded for clinical review and research.

**Results:** This poster will illustrate the multiple applications of the new software for clinical review, monitoring and treatment planning in day to day usage. For example, in risk managing haematological toxicities such as neutropenia and thrombocytopenia through treatment cycles; by recognising patterns of results retrospectively in the individual patient, in order to plan review and support. Biochemistry results are mapped on alternative graphs, through treatment, thereby, for example renal function can be monitored through graphs of creatinine, albumin and estimated Glomerular Filtration Rate (GFR) levels. Disease response to treatment, for example in Multiple Myeloma patients, can be viewed over any three month period by tracking Paraprotein and Free Light Chain ratios in graphical form, through treatment cycles.

**Conclusion:** Along with the traditional database facilities including audit, research and outcome analysis, the new mapping software improves the delivery of care to the individual patient through easily accessed panoramic views of accurate key data in graphical form, in the clinic setting. The process saves valuable clinic time and helps the clinical team in the delivery of optimum targeted and risk managed care to the haemato-oncology patient. The data generated is the basis of an inhouse research database.

**Disclosure of Interest:** None declared.

## P753

### Simultaneous Implementation of the Single European Code and ISBT 128

I. Slaper-Cortenbach<sup>1</sup>, Z. Szczepiorkowski<sup>2</sup>, P. Ashford<sup>3</sup>, P. Distler<sup>3,\*</sup>

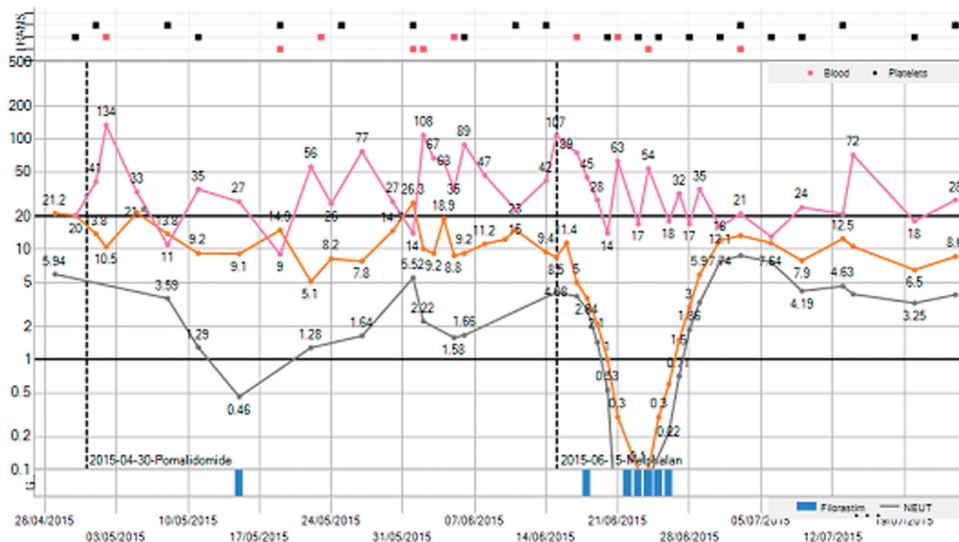
<sup>1</sup>University Medical Center Utrecht, Utrecht, Netherlands,

<sup>2</sup>Dartmouth-Hitchcock Medical Center, Lebanon, <sup>3</sup>ICCBBA, San Bernardino, United States

**Introduction:** Two initiatives have spurred the movement towards globally standardized coding for cells and tissues:

- In 2013, WHO developed a strategy for the governance of medical products of human origin (MPHO) that included

[P752]



global use of ISBT 128 to ensure unique identification, optimal traceability, and interoperability between countries and across all MPHO.

- AABB, FACT, and JACIE required the use of ISBT 128 terminology and an implementation plan for full ISBT 128 labelling.

In April 2015, the European Commission published Commission Directive (EU) 2015/565 that amends Directive 2006/86/EC specifying the implementation requirements for the Single European Code (SEC). Legislation requiring Tissue Establishments to apply the SEC to product labels takes effect 29 April 2017. Exceptions will exempt many cell therapy products from the SEC requirement but, depending on national legislation, some may still be required to carry an SEC. ISBT 128, managed by the non-profit ICCBBA, is the global standard for terminology, coding, and labelling of MPHO. It is compatible with the SEC and provides additional benefits.

**Material (or patients) and methods:** Experts in the fields of cells and tissues developed ISBT 128 international terminology using classes (high-level description, e.g., HPC, Apheresis) and attributes (additional details, e.g., Mobilized). Eurocet 128 developed a high-level terminology for tissues and cells (EU Tissue Code - EUTC). ISBT 128 classes have been mapped to their corresponding EUTC.

**Results:** More than 240 cell and tissue classes within ISBT 128 are mapped to 90 EUTCs allowing 3228 ISBT 128 cell and tissue product description codes to be used within the SEC system to describe products.

**Conclusion:** Not all cellular therapy products will be carrying the SEC, but if SEC must appear on the product label, this can be readily achieved using ISBT 128 codes in the SEC. Such an approach ensures that the consistency of ISBT 128 usage is maintained across all cellular therapy products whilst meeting the requirements of the Directive. This allows tissue establishments to:

- Provide descriptions of products useful in inventory management and biovigilance.
- Label with internationally standardized bar codes for more accurate records and improved traceability.
- Transfer products labelled with globally recognized bar codes beyond the EC countries.
- Comply with accreditation standards.
- Align with the global commitment of scientific and professional societies to use of ISBT 128.
- Align with the WHO call for globally consistent coding systems.

For those tissue establishments not already using ISBT 128, the SEC directive may require some changes to existing computer

systems. Simultaneous implementation of ISBT 128 with these changes may provide:

- Greater efficiency in that:
  - Many implementation steps (e.g., transition planning and validation) are the same for both systems.
  - Many processes (e.g., training and education of end users) can be done at the same time for both systems.
- ICCBBA provides mapping to EUTC so users do not need to do the mapping.
- Potential cost savings:
  - One computer upgrade instead of two.
- Efficiencies in implementation of the two systems together.

**Disclosure of Interest:** None declared.

**P754**

**Improving the sensitivity of supervised machine learning algorithms in predicting the odds of Day-100 TRM in acute and chronic leukemia patients post ALLO-SCT using different sampling techniques: An example of imbalance classification problem, 1042 cases**

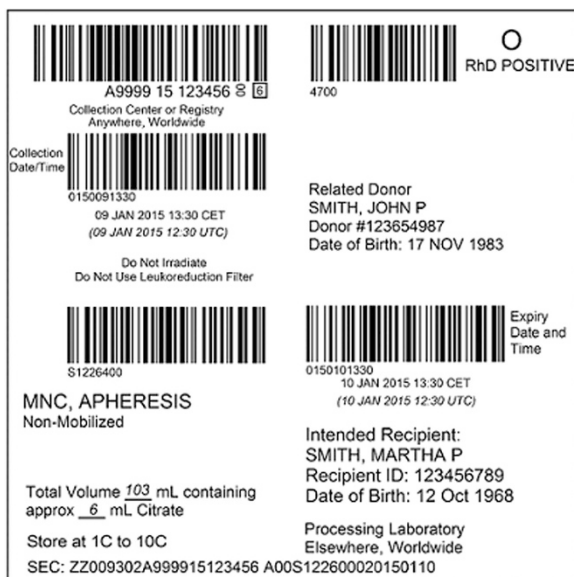
T. A. Elhassan<sup>1,\*</sup>, N. Chaudhri<sup>1</sup>, G. Aldawsari<sup>1</sup>, F. Alsharif<sup>1</sup>, H. Alzahrani<sup>1</sup>, S. Yousuf Mohamed<sup>1</sup>, W. Rasheed<sup>1</sup>, A. Hanbali<sup>1</sup>, S. Osman Ahmed<sup>1</sup>, M. Shaheen<sup>1</sup>, F. Alfraih<sup>1</sup>, R. Elfakih<sup>1</sup>, F. Almohareb<sup>1</sup>, M. Aljurf<sup>1</sup>  
<sup>1</sup>Oncology, KFSHRC, Riyadh, Saudi Arabia

**Introduction:** Day-100 Treatment Related Mortality (Day-100 TRM) is one of the important measures of allogeneic stem cell transplantation (ALLO-SCT) outcomes. However, predicting the odds of Day-100 TRM using machine learning (ML) classification algorithms is faced with some challenges due to the imbalance distribution of this event (TRM vs. No TRM) among the studying data set. Building a predictive model for this type of data results in a model with a low sensitivity towards the minority class yet high overall accuracy and specificity. This is because most the ML classification algorithms are built with an ultimate goal of improving the overall accuracy. Sampling the training data set to produce a balanced distribution of the class categories is one of the suggested solutions to overcome this problem.

**Material (or patients) and methods:** Different sampling techniques such as Random Oversampling (ROS), Random Undersampling (RUS) and Synthetic-Oversampling Technique (SMOTE) were utilized to improve model sensitivity towards the minority class (Day-100 TRM). A guided undersampling using Tomek link (TLink) was also utilized as a data cleaning method followed by the above mentioned sampling techniques. The different sampling techniques were compared using different ML algorithms such as Support Vector Machine (SVM), Artificial Neural Networks (ANN), Decision Tree (DT) using PART algorithms and Logistic Regression (LR). Model performance was evaluated using G-mean, F-statistics and Area Under the receiver operating characteristic Curve (AUC). Other performance measures such as overall accuracy, weighted accuracy, Negative Predictive Value (NPV) and precision were also calculated.

In this study 1,042 of patients with Acute Leukemias and Chronic Myeloid Leukemia underwent allogeneic transplantation between 1997 and 2013 were analyzed. Baseline characteristics such as age groups (< 20, 20-40 and >40), donor-sex-mismatch, disease stage (early, intermediate and advance), graft type (BM vs. PB), diagnosis (ALL, AML and CML) were used in building the predictive model. HLA mismatch and type of conditioning were not included because more than 95% of the patients were HLA-identical and received myeloablative conditioning. Analysis was performed using R studio.

**Results:** Sampling techniques have significantly improved the model ability of predicting the Day-100 TRM using all the studying machine learning algorithms. As expected, DT showed less sensitivity to the imbalance class distribution as compared to other ML algorithms. However, this is justifiable because DT is created using a splitting criterion that maximizes the information gain (entropy)/ Gini index which forces both classes to be addressed. LR showed inferior sensitivity



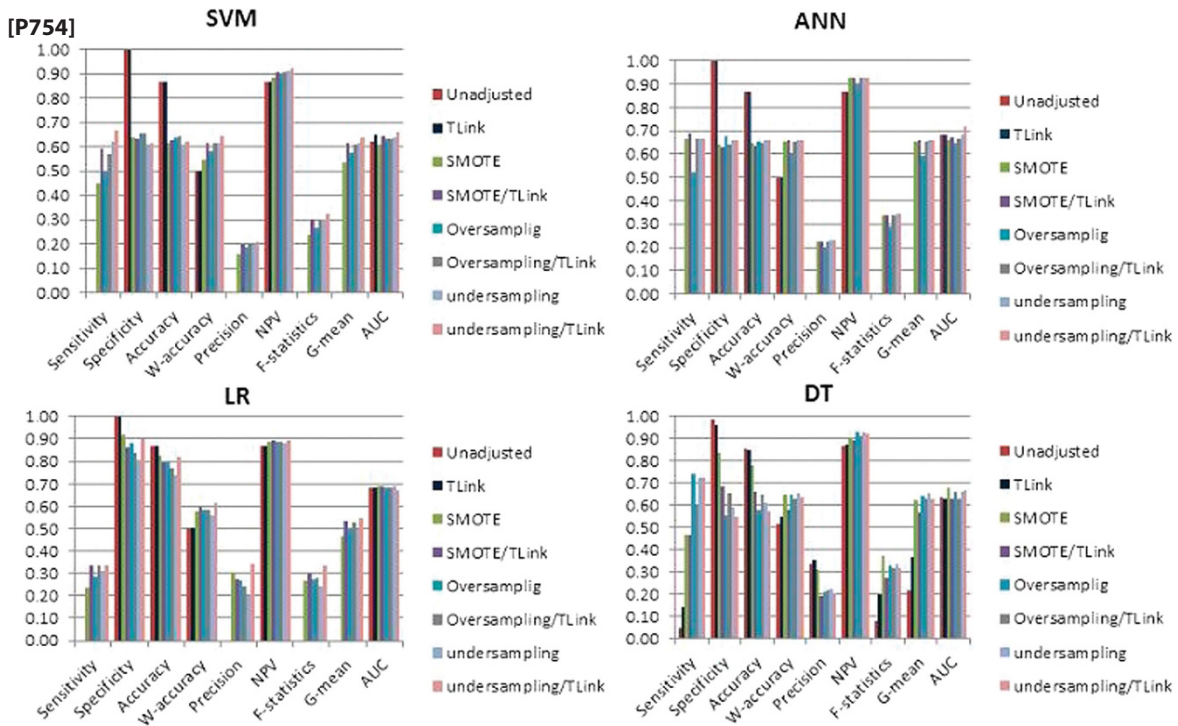


Figure 1: Performance comparison of different machine learning algorithms using different sampling techniques

compared to other ML algorithms. RUS/TLink showed the best performance using SVM, ANN and LR. On the other hand, RUS and ROS showed a comparable results using PART DT algorithm, Figure (1).

**Conclusion:** Sampling techniques have an important role in improving the predictive ability of the ML algorithms towards Day-100 TRM as a minority class. In fact, choosing the best sampling technique and the best ML algorithm that boosts the model performance measure is application-dependent where there is no absolute winner.

**Disclosure of Interest:** None declared.

### P755

#### Effectiveness of linkage disequilibrium for the identification of HLA null alleles

H.-P. Eberhard<sup>1</sup>, J. Mytilineos<sup>2</sup>, K. Fleischhauer<sup>3,\*</sup>, A. Schmidt<sup>4</sup>, C. Müller<sup>1</sup>

<sup>1</sup>ZKRD - German National Bone Marrow Donor Registry, <sup>2</sup>Institute of Clinical Transfusion Medicine and Immunogenetics, German Red Cross Blood Transfusion Service Baden Wuerttemberg-Hessen, Ulm, <sup>3</sup>Institute for Experimental Cellular Therapy, University Hospital Essen, Essen, <sup>4</sup>DKMS, Tübingen, Germany

**Introduction:** Most initial typing of donors and patients for HSCT is done today at some "high" resolution level accepting ambiguities outside the antigen recognition domain (ARD, i.e. exon 2+3 for class I and exon 2 for class II). The 150, 128, 89, 39 and 20 null alleles known for HLA-A, -B, -C, -DRB1 and -DQB1 respectively should be excluded as far as possible. The vast majority of these null alleles is defined by sequence variations within the ARS. Only 15 are caused by variations outside all exons and are rarely considered. However, the identification of null alleles defined in non-ARS exons is handled inconsistently and often one tries to reduce the number of futile tests using linkage disequilibrium information. In the context of a study on the frequency of high-resolution HLA-A~C~B~DRB1~DQB1-haplotypes in Germany we probe the recommendations of the current German Consensus for the selection of donors for allogeneic HSCT.

**Material (or patients) and methods:** Evidently, C\*04:09 N is the most frequent null allele with an allele frequency of over 1:20,000. It accounts for almost 60% of the 11 null alleles defined by non-ARS exon variations that were already observed in Germany (Table 1). The current German Consensus mentions only two additional ones (B51:11 N, A24:09 N), however A\*01:04 N and A\*02:83 N were found here to have a similar frequency and the cumulative frequency of the other 6 null alleles exceeds the frequency of A\*02:83 N.

**Results:** The consensus recommends the testing for C\*04:09 N in presence of B\*44:03. Our data support this recommendation since it reduces the number of tests by a factor of 6, missing in exchange only about 5% of the C\*04:09 N. The more complex strategies suggested for B51:11 N and A24:09 N deliver a comparable gain in efficiency, however they are missing half of the occurrences of the respective null allele.

Table 1 Null alleles with a frequency of more than 1 per million that require testing outside the ARS and observed in Germany

allele	allele frequency per million	fraction detected	detected	context to be tested according to German Recommendations
C04:09N	68.7	59.3%	55.8%	C*04:09N ~ B*44:03
B51:11N	12.0	10.4%	5.3%	A02:01 ~ B51:11N ~ C*15:02/13 ~ DRB1*04:02
A24:09N	10.1	8.7%	4.4%	A24:09N ~ B40 or A24:09N ~ B27
A01:04N	9.5	8.2%	0%	
A02:83N	6.8	5.8%	0%	
A02:15N	2.4	2.1%	0%	
C03:20N	2.3	2.0%	0%	
A24:11N	2.1	1.8%	0%	
	114.1	99.4%	65.5%	

**Conclusion:** Our numbers are based on haplotypes instead of complete genotypes. Nevertheless it is apparent that the current German Recommendations are missing about one third of the occurrences of null alleles and have to be reconsidered. Alternative strategies might improve the situation for some null alleles but their low frequencies make it difficult to underpin certain strategies with significant results.

**Disclosure of Interest:** None declared.