

EBMT

European Group for Blood and Marrow Transplantation

THE EBMT REGISTRY DATABASE

For registration of

ALLOGENEIC or AUTOLOGOUS

Stem Cell
Transplantation



THE EBMT RELATIONAL DATABASE

Background

The old EBMT database was a two level relational database. Most of the fields were contained in the 1st level structure (made up of several parallel tables), indexed by transplant. The follow up records of each transplant were stored in a separate table, one level below, indexed by date. Despite its two level depth, the database was practically a flat file since the bulk of the information was contained in the main structure, which stored all the data pertaining to each transplant from the birth of the patient to 100 days after the transplant. It consisted of almost 1300 fields. This database was developed using software called "Project Manager".

This type of structure poses problems regarding the evolution of the database and its expansion to accommodate other treatments and indications. The main obstacles relate to:

- Unmanageable size: any single investigation (WBC levels, for example) requires as many fields as time points it needs to be recorded at, therefore any investigation at a new time point will require a new field for that time point. It is not unusual that the staging for a certain disease may require 20 or 30 separate items to be filled in, so if staging is to be done at a new time point (before DLI, for example) that requires the addition of 20 to 30 new variables, or maybe more if the requirement is made for several diseases.
- Duplication of patient information. The database tables were indexed by transplant, not patient, so in each transplant record we have the same patient information until the occurrence of the first transplant for that patient
- Duplication of follow up information. Since certain information for the same patient –typically whether the patient is alive or dead- may be relevant for all the transplant records, it may get replicated for as many records per patient as there are transplants.

The EBMT needed to be able to increase the size of the database whenever new treatments, investigations or indications for treatment became available. It is important to develop a system, which allows new questions to be posed with a minimum overhead. The current structure tries to provide a reasonable solution to these needs.

Current database

This document should be read in conjunction with the Power Point file "Presentation on the EBMT database structure" which can be found on <http://www.ebmt.org/4Registry/registry4.html>.

LEVELS

The number of levels is five. All tables contain "navigation" items which are used by the designer to support data entry but which have no reflection on the clinical items or registry data quality items. All tables also have database administration items to enable proper functioning inside ProMISe.

LEVEL 1

This is the CENTRE table. It contains essential information for each transplant centre. Each transplant centre is one record and the Centre Identification Code (CIC) is the primary index (PI). This table is not present in the database as such, but linked from the membership database.

LEVEL 2

PATIENT (AA_) (152 FIELDS)

- patient data, which will not change throughout the lifetime of the patient, such as date of birth, initials, date of death, etc. Each patient is one record. the combination of items ID (centre identification) and IDAA (patient identification in that centre) uniquely identify each record and constitute the PI.
- administrative variables to be used for monitoring data quality, or to permit viewing flexibility in circumstances such that a patient may need to be shared among various centres; this may happen, for example, when a patient receives a second transplant in a different centre from that which transplanted him/her before, or when a patient is being followed up away from the transplant centre, or when a harvest centre wants recognition for the harvest.
- summary of the registration: dates of transplants, last survival status, last date seen
- a reference to the STUDY table where information regarding studies affecting this patient is kept.

LEVEL 3

The bulk of the database will be at this level. It consists of six tables. All these tables share the PATIENT table as their parent table. The PI of four tables is the ID, IDAA combined with the date the data was collected. This date may differ from table to table, since a record is created in a table only if data needs to be stored in that table.

The other two tables contain study information and the HLA of the patient. These tables are indexed with an integer.

STUDY (BA_) (46 fields)

This table is only used if the patient has been entered into a prospective trial. Contains:

- information pertaining to treatment allocation, and number in the study, etc. The PI is defined with a number.
- Fields to aid WP study co-ordinators to keep tabs on data queries, mailing, completeness of the data, etc.

DIAGNOSIS (BB_) (120 fields)

The table contains all the information pertaining to all diagnoses. It does not contain any staging information, which may be repeated at several time points for the same patient (ie: lymph nodes involved, haptoglobins in serum, etc.). The date in the PI is the date the diagnosis was made. Most patients will only have one record in this table. However, patients, who developed a secondary disease after the transplant, or have two diseases simultaneously for which transplant can be an indication, or

whose transplant diagnosis was secondary to a prior diagnosis may have as many records as diagnoses registered. The diagnoses are differentiated as to whether or not they are indication for transplant.

TREATMENT (BC_)(214 fields)

The table contains all information pertaining to all treatments received by the patient. This includes lines of treatment given to the patient before the first transplant ever, all the transplants, any treatment between transplants, treatment for relapse or progression, pre-emptive treatment, etc. All the treatments, regardless of the time point or reason for them, share the same database fields. For example, the same field will be used for chemotherapy given as 1st line treatment, for mobilisation, for conditioning, for relapse, for a second mobilisation, for a second conditioning, etc.

Treatment at each time point is contained in a different record. Therefore, one single patient may have several treatment records. However, there will only be one record for each treatment even if the treatment is made up of several therapies such as different chemotherapies, plus radiotherapy, plus serotherapy, all given at slightly different dates. The date used for the PI will be the very first date that particular treatment started. This is to avoid a needless proliferation of records within a very short period of time.

It contains:

- Indication for the treatment
- Type of treatment
- Characteristics associated to the treatment (sequential number, etc.)
- Response or status after this particular treatment

This table is parent to level 4 tables in which a more detailed description of the treatment (chemotherapy, monoclonal antibodies, progenitor cells, donor characteristics) can be registered.

ASSESSMENT(1) (BE_)(341 fields)

The table contains all investigations performed at different patient assessments. This includes haematology, biochemistry, staging, organs involved, bone metastasis, etc. The number of records per patient is indefinite, although they will usually coincide with the standard time points of diagnosis, 1st line treatment, transplant, response, progression, etc. This is the largest table, both in terms of records and of items. The date used for the PI is the date of the investigation.

- Type of investigations done (cytogenetics, molecular, scans, etc.)
- Biochemical and haematological values
- Disease involvement
- Clinical abnormalities
- Staging including the current disease status in terms of CR, PR, etc.
- Viral profile
- Patient physical state
- Incidence of complications, including GvHD
- Engraftment and chimaerism status
- Transfusion status

This table is parent to level 4 tables in which a more detailed description of some of the assessments (complications, immunophenotyping, cytogenetics, infections, involvement) can be registered.

HLA OF THE PATIENT (BH_)(7 fields)

It records the HLA of the patient. It could form part of the ASSESSMENT(1) table but a decision was made to keep it separate in the event that very detailed HLA items could be added if necessary. The last PI is a code which indicates the combination of the loci and the type of technique used.

LEVEL 4

This level contains mostly a series of smaller tables, children to one of the tables in Level 3. It allows detailed description of investigations or treatments mentioned in the parent tables. The PI is made of the ID and IDAA, the date of the parent table, and the type of information being stored (chemotherapy regimens for treatment, CD's for immunophenotyping, etc.) Each type is stored in a separate record. The addition of new types does not require the addition of new fields, but simply the addition of a new label to the label set which describes it.

TREAT COMPL (CA_)(7 fields)

This table is the child of the ASSESSMENT(1) table. It stores complications which are considered a consequence of the treatment. When complications are indicated as present in the ASSESSMENT(1) table, these complications can be described in the TREAT COMPL table. There is no limit to the number of complications that can be added at any time point. The last PI is the complication itself.

IMMUNOPHENOTYPE (CB)(4 fields)

This table is the child of the ASSESSMENT(1) table. When immunophenotyping is done, this is indicated in the ASSESSMENT(1) table. A full description can then be entered in this table. There is no limit to the number of CD's tested that can be added at any time point. Information on whether a CD was or was not tested, or whether it was positive or not can also be added. The last PI is the phenotype itself.

CYTOGENETICS (CC)(13 fields)

This table is the child of the ASSESSMENT(1) table. When cytogenetic investigations are done, this is indicated in the ASSESSMENT(1) table. A full description can then be entered in this table. There is no limit to the number of abnormalities that can be added at any time point. Information on whether an abnormality was or was not present can also be added. The last PI is the abnormality itself.

INFECTIONS (CG)(16 fields)

This table is the child of the ASSESSMENT(1) table. It stores infections associated to the treatment. When infections are indicated as present in the ASSESSMENT(1) table, these infections can be described in the INFECTIONS table. There is no limit to the number of infections that can be added at any time point. It allows entry of type of infection, pathogen, method used for isolation, length of episode, number of days hospitalised for the episode, etc. The last PI is a number with no intrinsic significance which serves only for the purpose of rendering the record unique.

INVOLVEMENT (CK)(7 fields)

This table is the child of the ASSESSMENT(1) table. It stores organs and type of involvement. When organ involvement is indicated as present in the ASSESSMENT(1) table, these organs can be described in the INVOLVEMENT table. There is no limit to the number of involvements that can be added at any time point. It allows entry of organ involved, type of involvement, presence or absence, whether primary or metastatic, etc. The last PI is the organ involvement itself.

MOLECULAR (CL)(3 fields)

This table is the child of the ASSESSMENT(1) table. When molecular investigations are done, this is indicated in the ASSESSMENT(1) table. A full description can then be entered in this table. There is no limit to the number of markers that can be added at any time point. Information on whether a marker was or was not present can also be added. The last PI is the marker itself.

DIS COMPL (CN)(4 fields)

This table is the child of the ASSESSMENT(1) table. It stores complications which are considered a consequence of the diagnosis. Currently it is only in use for Autoimmune diseases. When complications are indicated as present in the ASSESSMENT(1) table, these complications can be

described in the DIS COMPL table. There is no limit to the number of complications that can be added at any time point. The last PI is the complication itself.

CIRCULATING AB(CO)(4 fields)

This table is the child of the ASSESSMENT(1) table. It stores the presence of antibodies in the patient. Currently it is only in use for Autoimmune diseases. When antibodies are indicated as present in the ASSESSMENT(1) table, these antibodies can be described in the CIRCULATING AB table. There is no limit to the number of antibodies that can be added at any time point. The last PI is the antibody itself.

QUESTIONNAIRE (CP)(2 fields)

This table is the child of the ASSESSMENT(1) table. It stores the answers to quality of life and disease symptoms questionnaires. Currently it is only in use for Autoimmune diseases. When questionnaires are indicated to have been used in the ASSESSMENT(1) table, the answer to each question can be described in the QUESTIONNAIRE table. There is no limit to the number of questions that can be added at any time point. The last PI is the question itself.

DRUGS (CHEMO) (CD)(13 fields)

This table is the child of the TREATMENT table. It allows for entry of details related to any drug treatment such as chemotherapy regimen, dose, brand, duration of treatment, route of administration, number of cycles and the date at which each particular therapy –within the same treatment- started and ended. The last PI is the drug/regimen itself.

ANTIBODY TREATMENT (CE) (9 fields)

This table is the child of the TREATMENT table. It allows for entry of details related to monoclonal antibody treatment such as type, radiolabelling, duration of treatment and the date at which each particular therapy –within the same treatment- started and ended. The last PI is the antibody itself.

STEM CELL COUNTS (CF) (5 fields)

This table is the child of the TREATMENT table. It allows for entry of all cell counts relating to the cells actually infused during transplant. The type of cell, counts and the timing of the measurement are stored here. The last PI is a code indicating the combination of source of the stem cells and the timing of the measurement.

DONOR (CH)(65 fields)

This table is the child of the TREATMENT table. It enters all the donor information such as viral profile, sex, age, measurements related to the blood (used only for cord blood), cytokine administration, etc. Since this table is the child of a transplant record in the TREATMENT table, it can have as many records as necessary thus allowing for multiple donors. This table is parent to another table in which HLA information for the donor is stored. The last PI is a number with no intrinsic significance which serves only for the purpose of rendering the record unique.

LEVEL 5

There is only one table at this level, which is used to record the HLA of the donor, and is a child of the table where the donor is registered.

HLADONOR (DA)(7 fields)

This table is the child of the DONOR table. It records the HLA of the donor at the time of transplant. The last PI is a code which indicates the combination of the loci and the type of technique used.