User Guide for the MIMIC II V3.0 Database

Laboratory for Computational Physiology
Institute for Medical Engineering & Science,
Massachusetts Institute of Technology

User Guide version1.0

Last Modified: December18th, 2014

Table of contents:

1.0 Clinical overview of data in version 3.0 of MIMIC II database

The new version 3.0 (Dec. 2014) of MIMIC II contains around 48,000 patients, including over 15,000 newly added adult patients (neonates patients will be added at a later release). The total count of hospital admissions is now nearly 58,000 with over 60,000 ICU stays. Version 2.6 contains ICU clinical data from 2001 to 2008; version 3.0 extended the data set to October of 2012.

The comparison of patients/admissions/icustays between v2.6 and v3.0 is listed below.

Table 1. Comparison of patient volume between MIMIC2V26 and MIMIC2V30

Table Name (Primary Key column)	Count (MIMIC2V26)	Count (MIMIC2V30)	Difference	New ID data in MIMIC2V30 starting point
D_Patients (subject_id)	32,536	48,018	15,482	33,000
Admssions (hadm_id)	36,095	57,955	21,860	37,000
ICUStayEvents (icustay_id)	40,426	60,618	20,192	48,000

2.0 Clinical database

2.1 Patients (D_Patients table)

The source table for new patients added to D_PATIENTS comes from the PATIENTS table in the MetaVision (for ICU Adult patients) database; the date range is 7/12/2007-9/25/2012.

The structure of the D_Patients table in version 3.0 did not change much, but we did add a new column - the first 3 digits of zip codes. This information should provide new information for geographic based studies. Table 2 lists all columns in the D_Patients table.

In addition to data from source databases and data sets, this table also incorporates DOD (Date of Death) information from social security up to Feb. 2013.

Another change to this table is that we shifted the ages of all patients over age 90 to an older age (over 200) for de-identification purposes. In the old version (2.6), only patients who are alive and over 90 were shifted. This change is to minimize confusion for users who are interested in age related studies.

Table 2. Description of the columns in the D_PATIENTS table

Column name	Data type	New Column	Remarks
SUBJECT_ID	NUMBER(7)	N	Primary key
SEX	VARCHAR2(1)	N	
DOB	DATE	N	
DOD	DATE	N	
HOSPITAL_EXPIRE_FLG	VARCHAR2(1)	N	
ZIPCODE	VARCHAR2(5)	Y	

2.2 Care Giver (D_Caregivers table)

The Caregiver IDs (CGID) are stored in table D_Caregivers, which contains about 11,000 rows in version 2.6 and about 15,000 rows in version 3.0. However, we discovered that in version 2.6, one caregiver could be assigned with multiple CGIDs. In version 3.0, we added one new column called CG_UNIQUEID, which is the unique ID for caregivers. For the multiple CGIDs for one Care Giver, we chose one ID as the active one and assigned status 'A' (Active) in the CGID_STATUS column, all other CGIDs were assigned status 'M' (Merged), indicating a 'merged' status of the CGID. Only the CGID with CGID_status='A' was used in other tables like MEDEVENTS, CHARTEVENTS etc.

Another change to this table in version 3.0 is that we added a new column - 'DESCRIPTION', which gives more detailed information about the Care Giver. For example, a 'MD' can be 'Attending' or 'Resident/Fellow/PA/NP'.

Table 3. Description of the columns in the D_CAREGIVERS table

Column name	Data type	New Column	Remarks
CGID	NUMBER	N	Care giver ID (the old ones from v2.6 is not unique)
CG_UNIQUEID	NUMBER	Y	Unique ID of a care giver
LABEL	VARCHAR2(6)	N	Title of Care Giver
DESCRIPTION	NVARCHAR2(30)	Y	More detailed description of care giver
CG_STATUS	VARCHAR2(1)	Y	'A' indicates an active CGID, 'M' indicates a merged CGID.

2.3 Care Unit (D_Careunits table)

This table did not change much between version 2.6 and version 3.0, the only difference is that CUID 54 used to be 'CSRU', it's now called 'CVICU', so the label of this care unit in version 3.0 includes both names.

2.4 ADT data (Admissions, Censusevents, ICUStayevents and ICUStay_days tables)

Admissions

The source data for the ADMISSIONS table comes from the BIDMC admission/discharge/transfer data set, which included hospital admissions and discharges for all patients from 01/02/2001 to 10/31/2012.

The ADMISSIONS table in MIMIC2V26 has a date range of 4/3/2001 to 9/16/2008. As mentioned earlier, the date range for new patient admissions from MetaVision database is 7/12/2007 - 9/25/2012. Therefore, some ICU patients in the MetaVision adult patients table have been included in the MIMIC2V26 database.

The data mapping and merging process between MIMIC II v2.6 and the MetaVision Database started by filtering out new hospital admissions first; among the new hospital admissions, there are old patients and new patients. For old patients, we needed to map to the existing SUBJECT_ID; for new patients, we generated new SUBJECT_IDs.

Some changes to the ADMISSIONS table in v3.0 are as follows:

1) For new admissions, the ADMIT_DT and DISCH_DT in the v3.0 ADMISSIONS table contain date and time information; data coming from MIMIC II v2.6 only contains date information.
2) Three new columns were added to this table – 'ADM_DIAGNOSIS', 'FIRST_SERVICE_UNIT', 'LAST_SERVICE_UNIT'

Table 4. Description of the columns in the ADMISSIONS table

Column name	Data type	New Column	Remarks
HADM_ID	NUMBER	N	Primary Key
SUBJECT_ID	NUMBER	N	Foreign Key – referring to D_PATIENTS table
ADMIT_DT	TIMESTAMP(6) WITH TIME ZONE	N	
DISCH_DT	TIMESTAMP(6) WITH TIME ZONE	N	
ADM_DIAGNOSIS	CLOB	Y	

FIRST_SERVICE	VARCHAR2(4)	Y	
LAST_SERVICE	VARCHAR2(4)	Y	

Censusevents

The Censusevents table tracks the changes of beds or care units and transfer of patients. For MIMIC2V30, this table is generated based on the PATIENTTRACKING table in the MetaVision database.

ICUStayevents

The ICUSTAYEVENTS table is generated from the CENSUSEVENTS table. An ICUSTAY_ID is generated to mark any new ICUSTAY events for patients. In many cases, patients can be in and out of the ICU care units multiple times during one hospital admission. For these cases, we set the following rule regarding ICUSTAY_IDs:

For patients transferred out of ICU units but re-admitted to the same ICU care unit within 24 hours, it's considered as one ICUSTAY event with the same ICUSTAY_ID. However, if the patient was re-admitted back to the same ICU care unit after 24 hours, it's considered as a new ICUSTAY event, and is assigned a new ICUSTAY_ID.

ICUStay_days tables

The ICUSTAY_DAYS table is generated from the ICUSTAYEVENTS table; it basically listed the ICUSTAYEVENTS by days and marked the first and last day of ICUSTAY.

2.5 Patient data

2.5.1 Demographics (Demographic_Details table)

This table contains patient demographic information such as ethnicity, religion, marital status as well as admission source, admission type and insurance information. For version 3.0, this table contains ITEMID as well as descriptions, making word-based searches much easier.

2.5.2 Items (D_Items table)

In version2.6, the ITEMID and LABELs (descriptions) for different events tables (such as chartevents, medevents, ioevents, labevents) are stored in separate tables; the ranges of the ITEMIDs in these tables are listed in Table 2.

Table 5. The ITEMS tables in version 2.6

Table name	Range of ITEMID
D_Chartitems	1 - 20009
D_Meditems	1 – 405
D_IOitems	-1 - 6807
D_Labitems	50001 - 50735
D_Codeditems	60001 - 101885
D_Demographicitems	200001 - 200088

In version 3.0, we combined all D_...ITEMS tables from v2.6 and the MetaVision PARAMETERS (ITEMS) to form the new D_ITEMS table.

When we attempted to merge the Metavision data with MIMIC2V26 data, one of the big challenges was to properly map and merge the D_...ITEMS table. Since Metavision data does not separate ITEMIDs into the same categories (Med, IO, Coded, Lab, Chart etc.) and their names can be totally different, it's nearly impossible to map them directly to the six different D_...ITEMS tables and the ITEMIDs in MIMIC2V26. To solve the problem, we took the approach of combining all D_...ITEMS tables in v2.6 and the PARAMETERS Table in MetaVision, thus forming the new D_ITEMS table.

Please note that, in version 2.6 and earlier versions of MIMIC2 database, one specific ITEM can have more than one ITEMID. This is still the case in version 3.0; users will have to add the MetaVision ITEMID to the list of ITEMIDs by running a name (or word)-based search in the D_ITEMS table.

Since we needed to merge all the D_...ITEMS tables into one table in version 3.0, we had to make sure different types of ITEMIDs (from different D_...ITEMS tables in version 2.6) are within their own numeric range (no overlapping). Therefore, two types of ITEMIDs (MED and IO) from version 2.6 had to be shifted to a higher range of integers to avoid overlap with ITEMIDs from D_Chartitems (see Table 2).

The 'ORIGIN' column added to the new D_ITEMS table indicates the source table of the data (ITEMIDs). The following table lists the ORIGINs and ITEMID ranges in version 3.0.

Table 6. The D_ITEMS table in version 3.0 and ranges of ITEMIDs

v3.0 Table name	ORIGIN	Range of ITEMID	Difference	Source table in v2.6 /database
D_ITEMS	CHART	1 - 20009	None	D_Chartitems (v2.6)
	MED	30001 – 30405	+ 30000	D_Meditems (v.2.6)
	IO	40000 – 46808	+ 40001	D_IOitems (v2.6)
	LAB	50800 – 51554	Re-generated ITEMIDs	D_ Labitems (v2.6)
	CODED	60001 - 101885	None	D_Codeditems (v2.6)
	DEMOGRAPHIC	200001 - 200088	none	D_Demographicitems(v2.6)
	METAVISION	220003 - 228647	+220000	MetaVision DB

Since the D_ITEMS table in v3.0 combined all six of the D_...ITEMS tables in v2.6 and we kept all related columns, the D_ITEMS table does have more columns than any of its source tables. The following table lists the column names, data types and source tables of all columns in MIMIC2V30.D_ITEMS table.

Table 7. Description of the columns in the D_ITEMS table

Column name	Data type	New column	Source tables/Database
ITEMID	NUMBER(7)	N	D_Chartitems, D_Meditems, D_IOitems, D_Codeditems, D_Labitems, D_Demographicitems and METAVISION Database
LABEL	VARCHAR2(100)	N	D_Chartitems, D_Meditems, D_IOitems, D_Codeditems, D_Demographicitems and METAVISION Database
ABBREVIATION	VARCHAR2(50)	Y	METAVISION Database
ORIGIN	VARCHAR2(12)	Y	
CODE	VARCHAR2(10)	N	D_Codeditems
CATEGORY	VARCHAR2(50)	N	D_Chartitems, D_IOitems, D_Codeditems, D_Labitems, D_Demographicitems and METAVISION
UNITID	NUMBER(5)	Y	METAVISION Database
UNITNAME	VARCHAR2(50)	Y	METAVISION Database
TYPE	VARCHAR2(40)	N	D_Codeditems and METAVISION
DESCRIPTION	VARCHAR2(150)	N	D_Chartitems, D_Codeditems
LOWNORMALVALUE	FLOAT(126)	Y	METAVISION Database
HIGHNORMALVALUE	FLOAT(126)	Y	METAVISION Database
ALLERGYACTION	NUMBER(3)	Y	METAVISION Database
LOINC_CODE	Varchar2(7)	N	Lab data from BIDMC
LOINC_DESCRIPTION	Varchar2(100)	N	Lab data from BIDMC
OLD_LABITEMID	NUMBER(7)	Y	D_Labitems
OLD_TEST_NAME	VARCHAR2(50)	Y	D_Labitems
OLD_LOINC_CODE	VARCHAR2(7)	Y	D_Labitems

2.5.3 Diagnosis (ICD9, DRGEVENTS)

The ICD9 table is an important table for users to check patient diagnoses. The structure of this table did not change in the new version.

But for the first time, the admission diagnosis (adm_diagnosis) information is now available in the ADMISSIONS table in v3.0. (This diagnosis is assigned by the admitting office at the moment of

admission based on input from the admitting physician. It may be modified after the patient is treated in the hospital.)

Another table that contains diagnosis related data is the DRGEVENTS table; we included descriptive columns for the ITEMID in the new version. The following table lists all columns of DRGEVENTS in v3.0.

Table 8. Description of the columns in the DRGEVENTS table

Column name	Data type	New Column	Remarks
SUBJECT_ID	NUMBER(7)	N	Foreign Key – referring to D_PATIENTS table
ITEMID	NUMBER(7)	N	Foreign Key – referring to ADMISSIONS table
TYPE	VARCHAR2(12)	N	
CODE	VARCHAR2(10)	N	
DESCRIPTION	VARCHAR2(100)	Y	
COST_WEIGHT	NUMBER	Y	

2.5.4 Medications (MEDEVENTS, ORDERENTRY, POE_MED_ORDER)

In version 3.0, a new table – ORDERENTRY, which contains all medical treatment order information, is added to medication related tables. MEDEVENTS, ADDTIVES and IOEVENTS now all contain ORDERID, which is a foreign key referring to the ORDERENTRY table. Since the new source database, MetaVision, has totally different schema design and table structures, we had to add some new columns to the medication related tables while trying to preserve all the data and columns from the old version.

One big difference you may notice is that all new medication records have 'START' and 'END' times; users can easily calculate the total amount of medication using these two columns. Because of this, the A_MEDDURATIONS table is no longer needed. But users can still use that table in v2.6 for any checking and calculations for the old data.

The following table lists the columns of MEDEVENTS; new columns are noted.

Table 9. Description of the columns in the MEDEVENTS table

Column name	Data Type	New Column	Remarks
SUBJECT_ID	NUMBER(7)	N	Foreign key, referring to D_PATIENTS
ICUSTAY_ID	NUMBER(7)	N	Foreign key, referring to ICUSTAYEVENTS
ORDERID	NUMBER(7)	Y	Foreign key, referring to ORDERENTRY
ITEMID	NUMBER(7)	N	Foreign key, referring to D_ITEMS

LABEL	VARCHAR2(100)	Y	Included for easy word-based searches
SOLITEMID	NUMBER(7)	N	
SOLITEMLABEL	VARCHAR2(100)	Y	Included for easy word-based searches
CHARTTIME	TIMESTAMP(6) WITH TIME ZONE	N	Used for old data, null for new data
ELEMID	NUMBER(7)	N	
REALTIME	TIMESTAMP(6) WITH TIME ZONE	N	Used for old data, null for new data
STARTTIME	TIMESTAMP(6) WITH TIME ZONE	Y	Used for new data, null for old data
ENDTIME	TIMESTAMP(6) WITH TIME ZONE	Y	Used for new data, null for old data
VALUE	NUMBER	N	Called 'DOSE' in v2.6
UOM	VARCHAR2(100)	Y	Called 'DOSEUOM' in v2.6
SOLITEMVALUE	NUMBER	N	Called 'SOLVOLUME' in v2.6
SOLITEMUOM	VARCHAR2(100)	N	Called 'SOLUNITS' in v2.6
CGID	NUMBER	N	
CUID	NUMBER	N	
STOPPED	VARCHAR2(20)	N	

Here is a list of all columns in the newly added table – ORDERENTRY.

Table 10. Description of the columns in the ORDERENTRY table

Column name	Data Type	Remarks
ORDERID	NUMBER(10)	Primary key
LINKORDERID	NUMBER(10)	
SUBJECT_ID	NUMBER	Foreign key, referring to D_PATIENTS
ICUSTAY_ID	NUMBER(7)	Foreign key, referring to ICUSTAYEVENTS
CGID	NUMBER	Foreign key, referring to D_CAREGIVERS
ISSUEDATE	TIMESTAMP(6) WITH TIME ZONE	
ORDERCATEGORY	NVARCHAR2(84)	
PATIENTWEIGHT	FLOAT(126)	
ISOPENBAG	NUMBER(1)	
CANCELREASON	NUMBER(5)	
COMMENTS	NVARCHAR2(1000)	
LOCATIONNAME	NVARCHAR2(153)	
ROUTE	NVARCHAR2(50)	
DURATION	NUMBER	

DURATIONUOM	NVARCHAR2(70)	
TOTALVOLUME	FLOAT(126)	
TOTALVOLUMEUOM	NVARCHAR2(70)	
CONTINUEINNEXTDEPT	NUMBER	

Another table that's related to medication is POE_MED_ORDER, the source data for this table comes from BIDMC data sets, not from the MetaVision database. This table documents medications that were *ordered* not administered, and includes orders well beyond the ICU stay.

In version 2.6, this table was divided into two tables, POE_MED and POE_ORDER. However, in the new data set we have already combined the data from these two tables, so we just created one table.

Table 11. **Description of the columns in the POE_MED_ORDER table**

Column name	Data Type	Remarks
SUBJECT_ID	NUMBER(7)	Foreign key, referring to D_PATIENTS
HADM_ID	NUMBER(7)	Foreign key, referring to ADMISSIONS
ICUSTAY_ID	NUMBER(7)	Foreign key, referring to ICUSTAYEVENTS
START_DT	TIMESTAMP(6) WITH TIME ZONE	
STOP_DT	TIMESTAMP(6) WITH TIME ZONE	
DRUG_TYPE	VARCHAR2(80)	
DRUG	VARCHAR2(80)	
DRUG_NAME_POE	VARCHAR2(80)	
DRUG_NAME_GENERIC	VARCHAR2(50)	
FORMULARY_DRUG_CD	VARCHAR2(90)	
GSN	VARCHAR2(180)	
NDC	VARCHAR2(90)	
PROD_STRENGTH	VARCHAR2(90)	
DOSE_VAL_RX	VARCHAR2(90)	
DOSE_UNIT_RX	VARCHAR2(90)	
FORM_VAL_DISP	VARCHAR2(90)	

| FORM_UNIT_DISP | VARCHAR2(90) | |
| ROUTE | VARCHAR2(60) | |

2.5.5 Charts (CHARTEVENTS table)

The CHARTEVENTS table is the largest table in the database and contains all bedside patient records. In version 2.6, the CHARTEVENTS table has about 196 million rows of data; in version 3.0, the row count of the CHARTEVENTS table is 234 million.

One change we made to the CHARTEVENTS table in v3.0 is that we renamed 'CHARTTIME' column to 'TIME' and 'REALTIME' column to 'VALIDATIONTIME'. The new names were introduced to avoid confusion often associated with the 'CHARTTIME' and 'REALTIME' columns in v2.6. The new names in v3.0 should be easier to understand: 'TIME' refers to the actual time the measurements were taken, and 'VALIDATIONTIME' refers to the time when the values of the measurements were recorded.

For the new data coming from MetaVision, the event time is recorded as 'TIME' and followed by the 'VALIDATION' time; we adapted the naming conventions of the new data source. Another new column from the MetaVision database is called 'COMMENTS', which contains units of measure and the normal range of a measurement. This column can be very useful to users.

Please note that there is no 'VALUE2' for the new data coming from MetaVision. One good example is the recording of blood pressure:

In MIMIC2V26, the itemid=51 (LABEL= 'Arterial BP') included 'VALUE1' and 'VALUE2' for systolic and diastolic blood pressures. For new patients in MIMIC2V30, there are two ITEMIDs for blood pressure measurements: for systolic blood pressure, the ITEMID=220179, and for diastolic blood pressure, ITEMID=220180.

Table 12. Description of the columns in the CHARTEVENTS table

Column name	Data Type	New Column	Remarks
SUBJECT_ID	NUMBER(7)	N	Foreign key, referring to D_PATIENTS
HADM_ID	NUMBER(7)	Y	Foreign key, referring to ADMISSIONS
ICUSTAY_ID	NUMBER(7)	N	Foreign key, referring to ICUSTAYEVENTS
ITEMID	NUMBER(7)	N	Foreign key, referring to D_ITEMS
LABEL	VARCHAR2(100)	Y	Included for easy word-based searches
TIME	TIMESTAMP(6) WITH TIME ZONE	N	Called 'CHARTTIME' in v2.6

ELEMID	NUMBER(7)	N	
VALIDATIONTIME	TIMESTAMP(6) WITH TIME ZONE	N	Called 'REALTIME' in v2.6
CGID	NUMBER(7)	N	Foreign key, referring to D_CAREGIVERS
CUID	NUMBER(7)	N	Foreign key, referring to D_UNITS
VALUE1	VARCHAR2(110)	N	Can be numeric or alphabetic values
VALUE1NUM	NUMBER	N	Numeric values only
VALUE1UOM	VARCHAR2(120)	N	Unit of measure for vlaue1
COMMENTS	NCLOB	Y	New data from MetaVision, contains units and normal range of the measurements.
VALUE2	VARCHAR2(110)	N	For old data only, Can be numeric or alphabetic values
VALUE2NUM	NUMBER	N	For old data only, Numeric values only
VALUE2UOM	VARCHAR2(20)	N	For old data only, Unit of Measure for vlaue2
RESULTSTATUS	VARCHAR2(20)	N	For old data (v2.6) only
STOPPED	VARCHAR2(20)	N	For old data (v2.6) only
WARNING	NUMBER(1)	Y	For new data
ERROR	NUMBER(1)	Y	For new data

2.5.6 Fluids (IOEvents, Additives, Totalbalevents)

In version 2.6, patient input/output (IO) data is recorded in the IOEVENTS, D_IOITEMS, DELIVERIES, TOTALBALEVENTS and additives tables.

In version 3.0, we retired the DELIVERIES table since the data contained in this table (RATE, RATEUOM etc.) are included in the IOEVENTS table. As mentioned earlier, the D_IOITEMS table has been merged into the D_ITEMS table.

Please note that the ITEMIDs for IO or ADDITIVES ITEMS have been shifted up by 40001 to avoid overlapping with ITEMIDs for the CHART ITEMS (refer to table 3).

As with the MEDEVENTS table, the IOEVENTS and ADDITIVES tables in v3.0 have some new columns like 'ORDERID', a foreign key referring to the ORDERENTRY table. Users can get more detailed medical order-related information from the ORDERENTRY table.

As in the MEDEVENTS table, the new IO data includes 'STARTTIME' and 'ENDTIME', making calculations of total INPUT/OUTPUT volumes much easier.

Another change to these tables is that the 'LABEL' of ITEMID is included in the table; users can do word-based searches right on the events table, no need to join with the D_ITEMS table first.

Table 13. Description of the Columns in the ADDITIVES table

Column name	Data Type	New Column	Remarks
SUBJECT_ID	NUMBER(7)	N	Foreign key, referring to D_PATIENTS
ICUSTAY_ID	NUMBER(7)	N	Foreign key, referring to ICUSTAYEVENTS
ORDERID	NUMBER(10)	Y	Foreign key, referring to ORDERENTRY
ITEMID	NUMBER(7)	N	Foreign key, referring to D_ITEMS
LABEL	VARCHAR2(100)	Y	Included for easy word-based searches
IOITEMID	NUMBER(7)	N	
IOITEMLABEL	VARCHAR2(100)	Y	Included for easy word-based searches
CHARTTIME	TIMESTAMP(6) WITH TIME ZONE	N	Used for old data, null for new data
STARTTIME	TIMESTAMP(6) WITH TIME ZONE	Y	Used for new data, null for old data
ENDTIME	TIMESTAMP(6) WITH TIME ZONE	Y	Used for new data, null for old data
ELEMID	NUMBER(7)	N	
CGID	NUMBER	N	
CUID	NUMBER	N	
VALUE	NUMBER	N	Called 'AMOUNT' in v2.6
UOM	NVARCHAR2(70)	Y	Called 'DOSEUNITS' in v2.6
IOITEMVALUE	NUMBER	N	Called 'SOLVOLUME' in v2.6
IOITEMUOM	NVARCHAR2(101)	N	Called 'SOLUNITS' in v2.6

Table 14. Description of the columns in the IOEVENTS table

Column name	Data Type	New Column	Remarks
SUBJECT_ID	NUMBER(7)	N	Foreign key, referring to D_PATIENTS
ICUSTAY_ID	NUMBER(7)	N	Foreign key, referring to ICUSTAYEVENTS
ORDERID	NUMBER(7)	Y	Foreign key, referring to ORDERENTRY

ITEMID	NUMBER(7)	N	Foreign key, referring to D_ITEMS
LABEL	NVARCHAR2(366)	Y	Included for easy word-based searches
CHARTTIME	TIMESTAMP(6) WITH TIME ZONE	N	Used for old data, null for new data
ELEMID	NUMBER(7)	N	
ALTID	NUMBER(7)	N	
REALTIME	TIMESTAMP(6) WITH TIME ZONE	N	Used for old data, null for new data
STARTTIME	TIMESTAMP(6) WITH TIME ZONE	Y	Used for new data, null for old data
ENDTIME	TIMESTAMP(6) WITH TIME ZONE	Y	Used for new data, null for old data
CGID	NUMBER	N	
CUID	NUMBER	N	
VALUE	NUMBER	N	Called 'VOLUME' in v2.6
UOM	NVARCHAR2(101)	N	Called 'VOLUMEUOM' in v2.6
UNITSHUNG	NUMBER(5)	N	Called 'SOLVOLUME' in v2.6
UNITSHUNGUOM	VARCHAR2(20)	N	Called 'SOLUNITS' in v2.6
NEWBOTTLE	NUMBER	N	
STOPPED	VARCHAR2(20)	N	
ESTIMATE	VARCHAR2(20)	N	

Table 15. Description of the columns in the TOTALBALEVENTS table

Column name	Data Type	New Column	Remarks
SUBJECT_ID	NUMBER(7)	N	Foreign key, referring to D_PATIENTS
ICUSTAY_ID	NUMBER(7)	N	Foreign key, referring to ICUSTAYEVENTS
CHARTTIME	TIMESTAMP(6) WITH TIME ZONE	N	
ELEMID	NUMBER	N	
REALTIME	TIMESTAMP(6) WITH TIME ZONE	N	
CGID	NUMBER	N	Foreign key, referring to D_CAREGIVERS
CUID	NUMBER	N	Foreign key, referring to D_CAREUNITS
ITEMID	NUMBER	N	Foreign key, referring to D_ITEMS
LABEL	VARCHAR2(100)	Y	
VOLUME	VARCHAR2(100)	N	Called 'PERVOLUME' in v2.6
CUMITEMID	NUMBER	Y	

CUMLABEL	VARCHAR2(40)	Y	
CUMVOLUME	VARCHAR2(100)	N	
UOM	NVARCHAR2(10)	Y	
ACCUMPERIOD	VARCHAR2(100)	N	
APPROX	VARCHAR2(100)	N	
RESET	NUMBER	N	
STOPPED	VARCHAR2(20)	N	

2.5.7 Notes (Noteevents table)

In version 3.0, we added three new types of notes in the NOTEEVENTS table: ECG reports, Echo reports, and the physician notes from MetaVision.   Here is a list of NOTES categories in v3.0 and v2.6 (Extracted from the NOTEEVENTS table, exact upper/lower case match).

Table 16. Comparison of Notes *categories* in NOTEEVENTS table

Categories of Notes in v3.0	*Categories of Notes in v2.6*
ECG_REPORT	*(Not available in v2.6)*
ECHO_REPORT	*(Not available in v2.6)*
DISCHARGE_SUMMARY	DISCHARGE_SUMMARY
NURSING/OTHER	Nursing/Other
PROVIDER_NOTE	*(Not available in v2.6)*
RADIOLOGY_REPORT	RADIOLOGY_REPORT

The Nursing/Other category contains the nursing and respiratory therapist notes collected (2008 or earlier) from the CareVue system. Provider notes collected from the MetaVision are under the PROVIDER_NOTE category and include notes by physicians, nurses, therapists and others.   The table structure of NOTEEVENTS did not change between v2.6 and v3.0. However, the CHARTTIME of DISCHARGE_SUMMARY notes used 'DISCHARGE_TIME' in v3.0 while the 'ADMISSION_TIME' was used in v2.6.

Note: Some entries in the NOTEEVENTS table contain empty text due to the null entry in the original text. These entries will be removed in the final release to avoid confusion.

2.5.8 Procedures (Procedureevents table)

In v2.6, Procedureevents table is a relatively simple table, contains only 5 columns: SUBJECT_ID, HADM_ID, ITEMID, SEQUENCE_NUM, PROC_DT.

In v3.0, six new columns are added to this table as shown in the following table. The new procedures all have ORDERIDs, related order information can be found in ORDERENTRY table. We also included 'LABEL'

for each ITEMID, make it easier for word-based searches. Also, each new procedure has a 'START' and 'END' time.

In v2.6, checking for INTUBATION/EXTUBATION event and time was not an easy task. In the new version, it's very easy to find these events in the Procedureevents table.

Table 17. Description of the columns in the PROCEDUREEVENTS table

Column name	Data Type	New Column	Remarks
SUBJECT_ID	NUMBER(7)	N	Foreign key, referring to D_PATIENTS
HADM_ID	NUMBER(7)	N	Foreign key, referring to ADMISSIONS
ORDERID	NUMBER(7)	Y	Foreign key, referring to ORDERENTRY
ORDERCATEGORYNAME	VARCHAR2(30)	Y	
ITEMID	NUMBER(7)	N	Foreign key, referring to D_ITEMS
LABEL	VARCHAR2(100)	Y	Included for easy word-based searches
PROC_DT	DATE	N	
SEQUENCE_NUM	NUMBER	N	
STARTTIME	DATE	Y	Used for new data, null for old data
ENDTIME	DATE	Y	Used for new data, null for old data
CGID	NUMBER	Y	

2.5.9 Laboratory and microbiology tests

In v3.0, the Labevents table was re-generated from the raw lab tests data set from BIDMC. As there have been additions of new lab tests over the years or name changes of lab tests, and the majority of lab tests don't have LOINC codes, it was very hard to map to the Lab ITEMIDs in v2.6. Therefore, the ITEMIDs for Labevents are re-generated based on current lab tests. However, whenever possible, the old lab ITEMID was mapped and listed in the D_ITEMS table (ORIGIN='LAB') (see table 4). Also, we tried our best to map to the current LOINC codes as well.

Like other facts (…events) tables in v3.0, we included related ITEMID descriptive columns like 'TESTNAME', 'FLUID', 'CATEGORY' and 'LOINC_CODE', much easier for word-based searches.

Please note that, we collect all labs for a patient independent of when they were done. Some labs are from subsequent clinic visits, for example.

Table 18. Description of the columns in the LABEVENTS table

Column name	Data Type	New Column	Remarks
SUBJECT_ID	NUMBER	N	Foreign key, referring to D_PATIENTS
HADM_ID	NUMBER	N	Foreign key, referring to ADMISSIONS
ITEMID	NUMBER(38)	N	Foreign key, referring to D_ITEMS
CHARTTIME	TIMESTAMP(6) WITH TIME ZONE	N	
TEST_NAME	VARCHAR2(50)	Y	Included for easy word-based searches
VALUE	VARCHAR2(200)	N	
VALUEUOM	VARCHAR2(15)	N	
FLAG	VARCHAR2(10)	Y	
FLUID	VARCHAR2(40)	Y	
CATEGORY	VARCHAR2(20)	Y	
LOINC_CODE	VARCHAR2(15)	Y	

The Microbiologyevents table in v3.0 also included names, descriptions of the ITEMIDs. The following table lists all the columns in this table in v3.0, all new columns are noted.

Table 19. Description of the columns in the V3.0 MICROBIOLOGYEVENTS table

Column name	Data Type	New Column	Remarks
SUBJECT_ID	NUMBER	N	Foreign key, referring to D_PATIENTS
HADM_ID	NUMBER	N	Foreign key, referring to ADMISSIONS
TIME	DATE	N	'CHARTTIME' in v2.6
SPEC_TYPE_CD	VARCHAR2(10)	Y	Included for easy word-based searches
SPEC_ITEMID	NUMBER(7)	N	Foreign key, referring to D_ITEMS
SPEC_TYPE_DESC	VARCHAR2(100)	Y	Included for easy word-based searches
ORG_CD	VARCHAR2(4)	Y	Included for easy word-based searches
ORG_ITEMID	NUMBER(7)	N	
ORG_NAME	VARCHAR2(100)	Y	Included for easy word-based searches
ISOLATE_NUM	VARCHAR2(10)	N	
AB_CD	VARCHAR2(2)	Y	Included for easy word-

			based searches
AB_ITEMID	NUMBER(7)	N	
AB_NAME	VARCHAR2(100)	Y	Included for easy word-based searches
DILUTION_AMOUNT	VARCHAR2(10)	N	
DILUTION_COMPARISON	VARCHAR2(10)	N	
INTERPRETATION	VARCHAR2(1)	N	

3.0 Summary

MIMIC2 II version 3.0 is quite different from previous versions since this is the first version requiring the mapping and combination of ICU data from two totally different source databases (CareVue and MetaVision). As we didn't have good documentation for the MetaVision database and the database structure was totally different, it was very hard to find all the matching data from this large database. The mapping process was mainly based on searching and studying the data/tables in MetaVision.

This new version mapped and combined data from MIMIC2V26 (45 tables), MetaVision (506 tables), additional data sets from BIDMC (23 tables) and Social Security data sets (2 tables).

Since the source databases were so different, even with our best effort to make the new database schema similar to the MIMIC II v2.6 tables and structures, there had to be some changes. The majority of the new changes added great value to the new database, but there might be some changes that users will have to re-learn and get used to. We are hoping to document the changes as thoroughly as we can to make this learning curve easier.

Also, since the data mapping and merging process was a challenging process and our first try, there might be issues or bugs. We have gone through some data integrity checks but we would appreciate if you can report to us any potential issues or bugs you find in the process of testing/using this new version of MIMIC II database as well.

One more note is that since almost all tables have grown much bigger in this version, you may find your search process slower. We are still in the process of tuning the database and tables to make your searches faster. Your feedback on this aspect would be much appreciated as well! Email your comments and observations to mimic-support@physionet.org.