MIMIC

The Medical Information Mart for Intensive Care

MinDong Sung, M.D.

DHLab, Yonsei University College of Medicine

2021-07-01

Introduction

MIMIC-III

- Provide critical care data for over 40,000 patients
  admitted to ICU at the Beth Israel Deaconess Medical Center (BIDMC).

- deidentified, and patient identifiers were removed

- MIMIC-III has been integral in driving large amounts of research in clinical informatics, epidemiology, and machine learning.

MIMIC-IV

- MIMIC-IV is intended to carry on the success of MIMIC-III
- Support a broad set of applications within healthcare.
- MIMIC-IV adopts a modular approach to data organization,
- Highlight data provenance
- Facilitate both individual and combined use of disparate data sources.

Methods

Acquisition

- Inclusion criteria
    - who were admitted to the BIDMC ED or any the ICU, 2008 - 2019.
- Extracted from the respective hospital databases
- A master patient list was created
- All source tables were filtered to only rows related to patients in the master patient list.

Preparation

- The data were reorganized to better facilitate retrospective data analysis.
    - the de-normalization of tables
    - removal of audit trails
    - reorganization into fewer tables
    - to simplify retrospective analysis of the database.
- Data cleaning steps were not performed
    - to ensure the data reflects a real-world clinical dataset.

Methods

De-identify

- Patient identifiers as stipulated by HIPAA were removed.
- Patient identifiers were replaced using a random cipher, resulting in deidentified integer identifiers for patients, hospitalizations, and ICU stays.
- Structured data were filtered using look up tables and allow lists.
- If necessary, a free-text deidentification algorithm was applied to remove PHI from free-text.
- Finally, date and times were shifted randomly into the future using an offset measured in days. A single date shift was assigned to each subject_id. As a result, the data for a single patient are internally consistent.

Backgrouds for hostpital data

How the patient data have been generated?

CASE

Think about patient flows in hospital

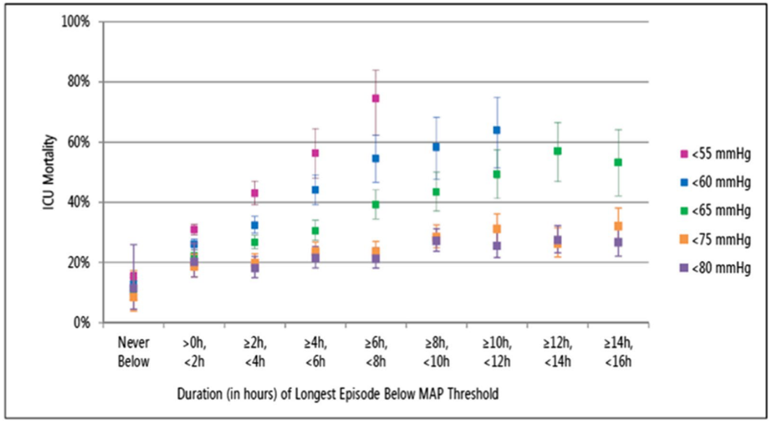
- Patient
- Admission
- Chartevents
- Provider orders(poe)/medication administration(emar)
- Labevents
- Chest X ray
- Diagnoses_icd
- Transfer
- chartevents
- datetimeevents
- inputevents
- outputevents
- procedureevents
- icustays

Papers from MIMIC data

Researches based on MIMIC data

Clinical
Model for clinicians
Algorithm

Mean arterial pressure and mortality in patients with distributive shock: a retrospective analysis of the MIMIC-III database (Annals of Intensive Care, 2018)



Number of Cases

MAP <55 mmHg	1731	2696	597	133	59	–	–	–	–
MAP <60 mmHg	874	2546	1114	374	154	84	61	–	–
MAP <65 mmHg	352	1682	1441	723	395	221	144	93	75
MAP <75 mmHg	71	412	725	756	651	543	368	311	258
MAP <80 mmHg	35	193	451	516	568	528	394	371	295

Fig. 1 ICU mortality by duration of longest episode with mean arterial pressure (MAP) < 80 mmHg (mauve), < 75 mmHg (orange), < 65 mmHg (green), 60 mmHg (blue) and 55 mmHg (pink) in all patients with distributive shock

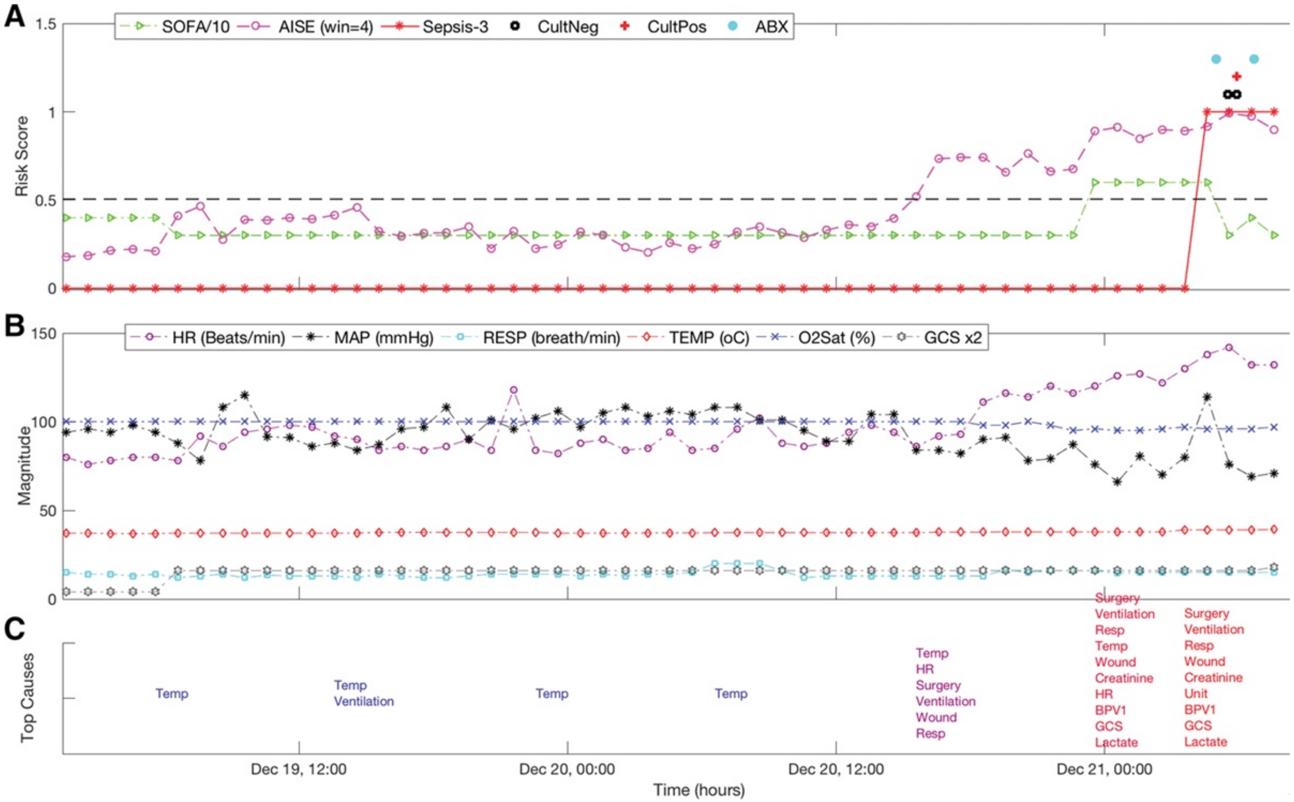
Vincent, JL., Nielsen, N.D., Shapiro, N.I. et al. Mean arterial pressure and mortality in patients with distributive shock: a retrospective analysis of the MIMIC-III database. Ann. Intensive Care 8, 107 (2018).

Transthoracic echocardiography and mortality in sepsis: analysis of the MIMIC-III database (ICM, 2018)

Method	OR	CI		*p* value
		2.5%	97.5%	
Doubly robust with unbalanced covariates	0.78	0.68	0.90	< 0.001
Doubly robust with all covariates	0.64	0.52	0.78	< 0.001
Propensity score IPW	0.84	0.78	0.92	< 0.001
Propensity score matching	0.78	0.66	0.92	< 0.001
Multivariate	0.64	0.53	0.78	< 0.001

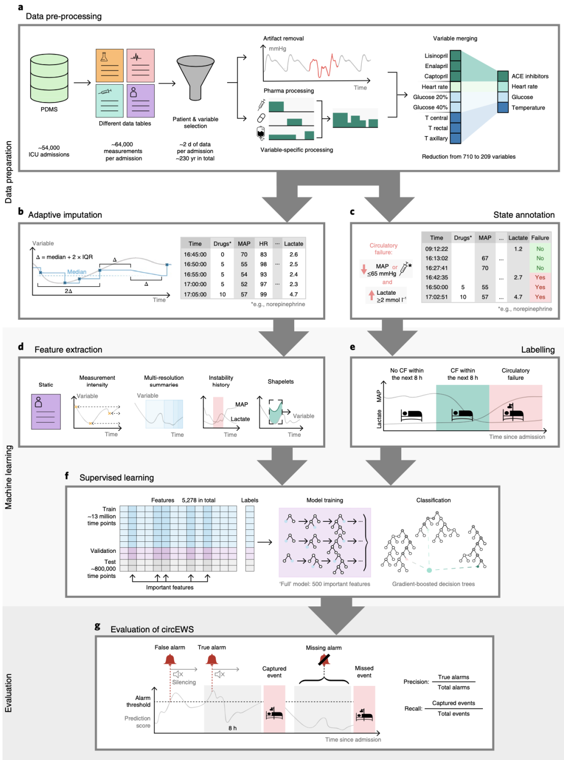
Feng, M., McSparron, J.I., Kien, D.T. et al. Transthoracic echocardiography and mortality in sepsis: analysis of the MIMIC-III database. Intensive Care Med 44, 884–892 (2018).

An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU (CCM 2018)



Nemati, Shamim et al., An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU, Critical Care Medicine: April 2018 - Volume 46 - Issue 4 - p 547-553

Early prediction of circulatory failure in the intensive care unit using machine learning (Nature Medicine 2020)



Hyland, Stephanie L., et al. "Early prediction of circulatory failure in the intensive care unit using machine learning." Nature medicine 26.3 (2020): 364-373.

Continuous blood pressure measurement from one-channel electrocardiogram signal using deep-learning techniques(Artificial Intelligence in Medicine, 2020)

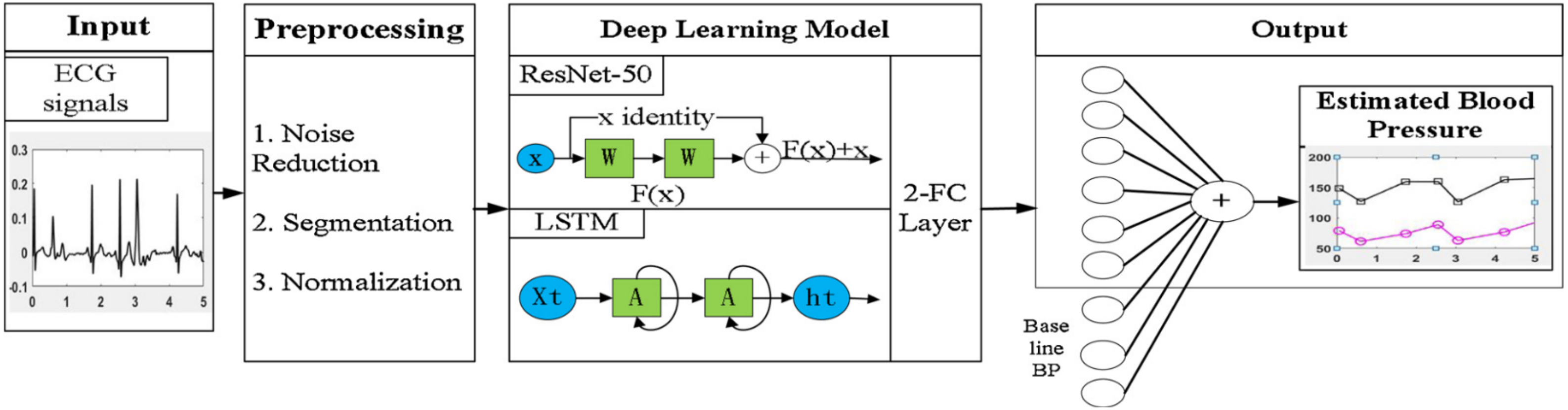


Fig. 1. Block diagram of the proposed model.

Fen Miao et al., Continuous blood pressure measurement from one-channel electrocardiogram signal using deep-learning techniques, Artificial Intelligence in Medicine, Volume 108, 2020

Clinical concept extraction using transformers (JAMIA 2020)

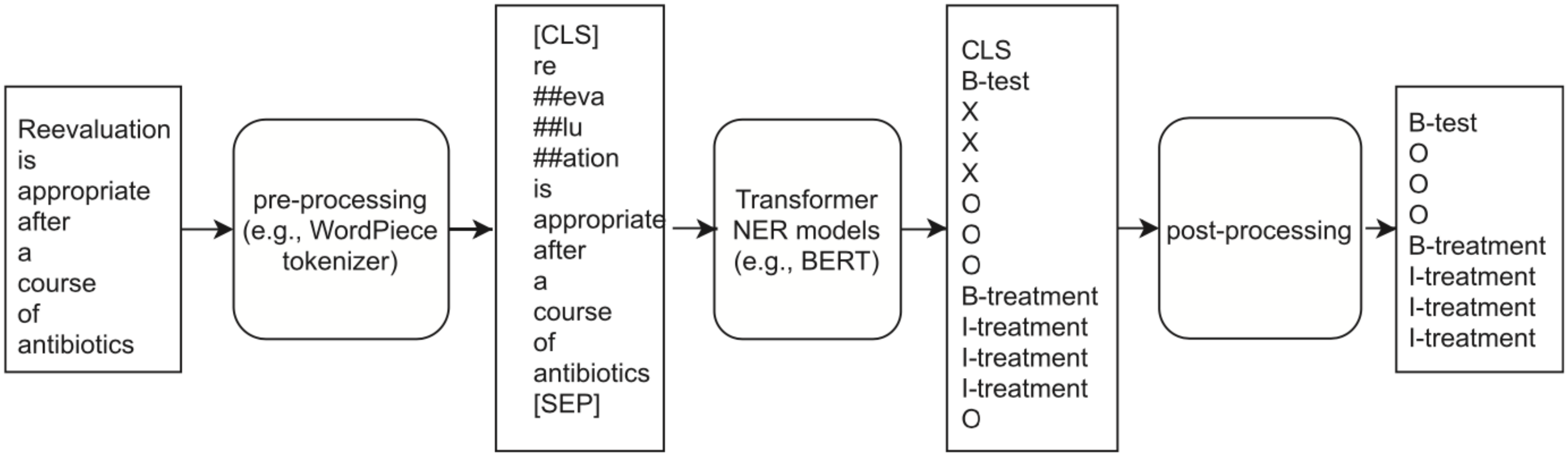


Figure 1. An overview of the workflow for a transformer-based NER pipeline using BERT as an example.

Xi Yang, Jiang Bian, William R Hogan, Yonghui Wu, Clinical concept extraction using transformers, Journal of the American Medical Informatics Association, Volume 27, Issue 12, December 2020, Pages 1935–1942,

Tables

What kind of data the MIMIC has?

Modular structure of MIMIC IV - Core

Patients

- Information that is consistent for the lifetime of a patient is stored in this table.

Admissions

- The admissions table gives information regarding a patient's admission to the hospital. Since each unique hospital visit for a patient is assigned a unique `hadm_id`, the admissions table can be considered as a definition table for `hadm_id`. Information available includes timing information for admission and discharge, demographic information, the source of the admission, and so on.

Transfers

- Physical locations for patients throughout their hospital stay.

Modular structure of MIMIC IV - Hosp

Laboratory measurements, microbiology cultures (labevents, microbiologyevents),

- Laboratory measurements sourced from patient derived specimens.
- Microbiology cultures.

Provider orders (poe, poe_detail),

- Orders made by providers relating to patient care.

Medication prescription/administration (prescriptions pharmacy; emar emar_detail)

- Prescribed medications.
- The Electronic Medicine Administration Record (eMAR); barcode scanning of medications at the time of administration.

Hospital billing information (diagnoses_icd, procedures_icd, hcpcsevents)

- Billed ICD-9/ICD-10 diagnoses for hospitalizations.
- Billed procedures for patients during their hospital stay.
- Billed events occurring during the hospitalization. Includes CPT codes.

Service(department information) related information (services).

- The hospital service(s) which cared for the patient during their hospitalization.

Modular structure of MIMIC IV - ICU

Charted information (chartevents).

- Charted items occurring during the ICU stay. Contains the majority of information documented in the ICU.

Intravenous and fluid inputs and patient output(inputevents, outputevents)

- Information documented regarding continuous infusions or intermittent administrations.
- Information regarding patient outputs including urine, drainage, and so on.

Procedures (procedureevents)

- Procedures documented during the ICU stay (e.g. ventilation), though not necessarily conducted within the ICU (e.g. x-ray imaging).

Information documented as a date or time (datetimeevents)

- Documented information which is in a date format (e.g. date of last dialysis).

Modular structure of MIMIC IV - code dictionary

d_icd_diagnoses

- Dimension table for diagnoses_icd; provides a description of ICD-9/ICD-10 billed diagnoses.

d_icd_procedures

- Dimension table for procedures_icd; provides a description of ICD-9/ICD-10 billed procedures.

d_labitems

- Dimension table for labevents; provides a description of all lab items.

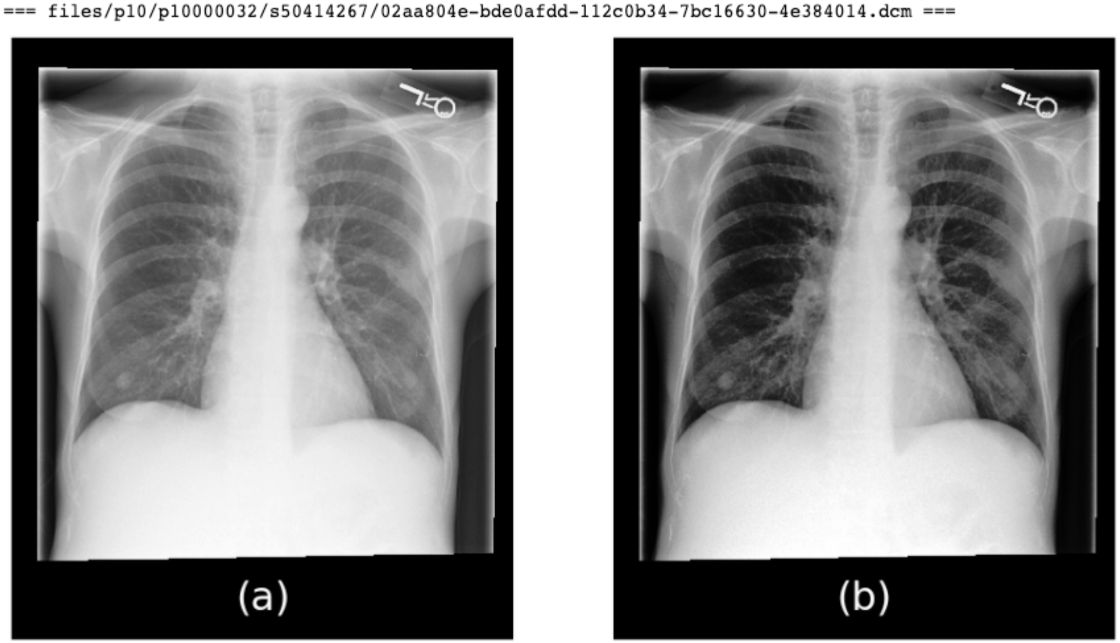
d_items

- Dimension table describing itemid. Defines concepts recorded in the events table in the ICU module.

Other data related to MIMIC

MIMIC-CXR / MIMIC-CXR-JPG (MIMIC IV)

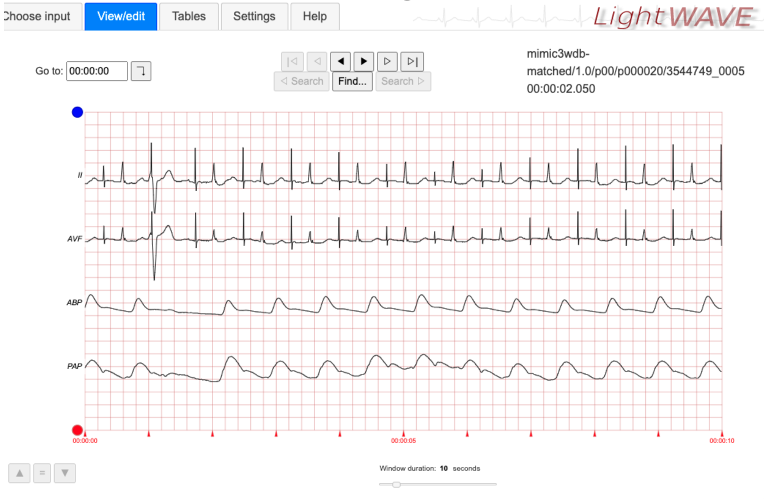
- Queried the BIDMC EHR for chest x-ray studies made in the emergency department between 2011 - 2016
- Extracted the set of patient identifiers associated with these studies.



1. Johnson AE, Pollard TJ, Berkowitz SJ, Greenbaum NR, Lungren MP, Deng CY, Mark RG, Horng S. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Scientific Data. 2019;6.
2. https://github.com/MIT-LCP/mimic-cxr

MIMIC-Waveform (MIMIC III)

- Waveforms almost always include one or more ECG signals, and often include continuous arterial blood pressure (ABP) waveforms, fingertip photoplethysmogram (PPG) signals, and respiration, with additional waveforms (up to 8 simultaneously) as available.
- Numerics typically include heart and respiration rates, SpO2, and systolic, mean, and diastolic blood pressure, together with others as available.
- Recording lengths also vary; most are a few days in duration, but some are shorter and others are several weeks long.



https://physionet.org/content/mimic3wdb/1.0/

How to access the data

Access

```sql
SHOW tables;
```

Tables_in_mimiciv
admissions
chartevents
d_hcpcs
d_icd_diagnoses
d_icd_procedures
d_items
d_labitems
datetimeevents
diagnoses_icd
drgcodes

```sql
SELECT * FROM d_icd_diagnoses LIMIT 20;
```

icd_code	icd_version	long_title
0010	9	Cholera due to vibrio cholerae
0011	9	Cholera due to vibrio cholerae el tor
0019	9	Cholera, unspecified
0020	9	Typhoid fever
0021	9	Paratyphoid fever A

```sql
SELECT * FROM diagnoses_icd LIMIT 20;
```

subject_id	hadm_id	seq_num	icd_code	icd_version
15734973	20475282	3	2825	9
15734973	20475282	2	V0251	9
15734973	20475282	5	V270	9
15734973	20475282	1	64891	9
15734973	20475282	4	66481	9

```sql
SELECT * FROM chartevents WHERE itemid IN (220050, 220051, 220052, 220045,
```

subject_id	hadm_id	stay_id	charttime	storetime	itemid	value	v
10003700	28623837	30600691	2165-04-24T05:30:00Z	2165-04-24T05:37:00Z	220210	14	
10003700	28623837	30600691	2165-04-24T05:30:00Z	2165-04-24T05:37:00Z	220045	65	
10003700	28623837	30600691	2165-04-24T06:00:00Z	2165-04-24T06:09:00Z	220210	14	
10003700	28623837	30600691	2165-04-24T06:00:00Z	2165-04-24T06:09:00Z	220045	56	
10003700	28623837	30600691	2165-04-24T06:09:00Z	2165-04-24T06:09:00Z	220210	14	
10003700	28623837	30600691	2165-04-24T06:09:00Z	2165-04-24T06:09:00Z	220045	55	
10003700	28623837	30600691	2165-04-24T07:00:00Z	2165-04-24T07:51:00Z	220210	14	
10003700	28623837	30600691	2165-04-24T07:00:00Z	2165-04-24T07:51:00Z	220045	57	
10003700	28623837	30600691	2165-04-24T08:00:00Z	2165-04-24T08:19:00Z	220045	56	
10003700	28623837	30600691	2165-04-24T08:00:00Z	2165-04-24T08:19:00Z	220210	14	

```sql
SELECT * FROM d_labitems LIMIT 10;
```

itemid	label	fluid	category	loinc_code
50801	Alveolar-arterial Gradient	Blood	Blood Gas	
50802	Base Excess	Blood	Blood Gas	
50803	Calculated Bicarbonate, Whole Blood	Blood	Blood Gas	
50804	Calculated Total CO2	Blood	Blood Gas	
50805	Carboxyhemoglobin	Blood	Blood Gas	

```sql
SELECT * FROM labevents LIMIT 10;
```

labevent_id	subject_id	hadm_id	specimen_id	itemid	charttime	storeti
5	10000019	25058216	74089533	51200	2129-05-21T21:10:00Z	2129-05-21T2
13	10000019	25058216	74089533	51250	2129-05-21T21:10:00Z	2129-05-21T2
44	10000032		52958335	50900	2180-03-23T11:51:00Z	2180-03-23T1
69	10000032		52958335	51000	2180-03-23T11:51:00Z	2180-03-23T1
93	10000032		81886159	51200	2180-03-23T11:51:00Z	2180-03-23T1

END