

Bioinformatics and Biomedical Informatics

Hồ Tú Bảo

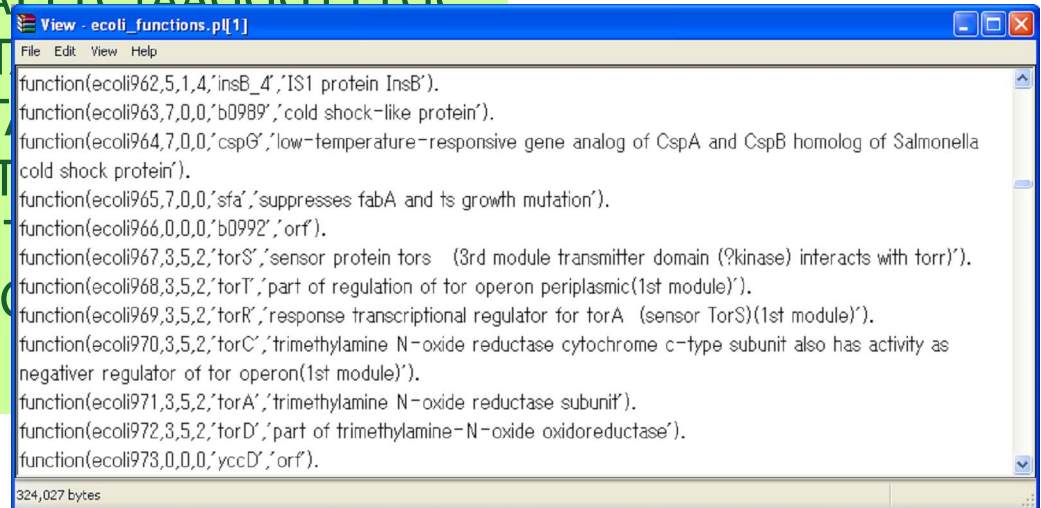
Viện Nghiên cứu Cao cấp về Toán



How biological data look like?

A portion of the DNA sequence, consisting of 1.6 million characters, is given as follows (about 350 characters, 4570 times smaller):

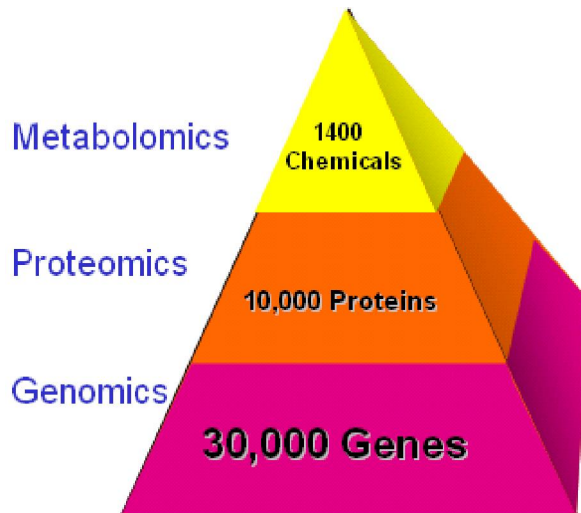
```
...TACATTAGTTATTACATTGAGAACTTTATAATTAAGGATTCATGTAA
ATTTCTTATTTGTTTATTTAGAGGTTTTAAATTTAATTTCTAAGGGTTTGC
TGGTTTCATTGTTAGAATATTTAACTTAATCAAATT
ATTAGGATTAATTAGGTAAGTAAATAAAATTTCTCT
TTTTAAATTTAAGGAGATAAAAATACTACTCTGTTT
GATTTAAATACTAAAGGGTTTATATATATGAAGTAG
ATGGTATAGAAAGCTTAAATATTAAGAGTGATGAAC
```



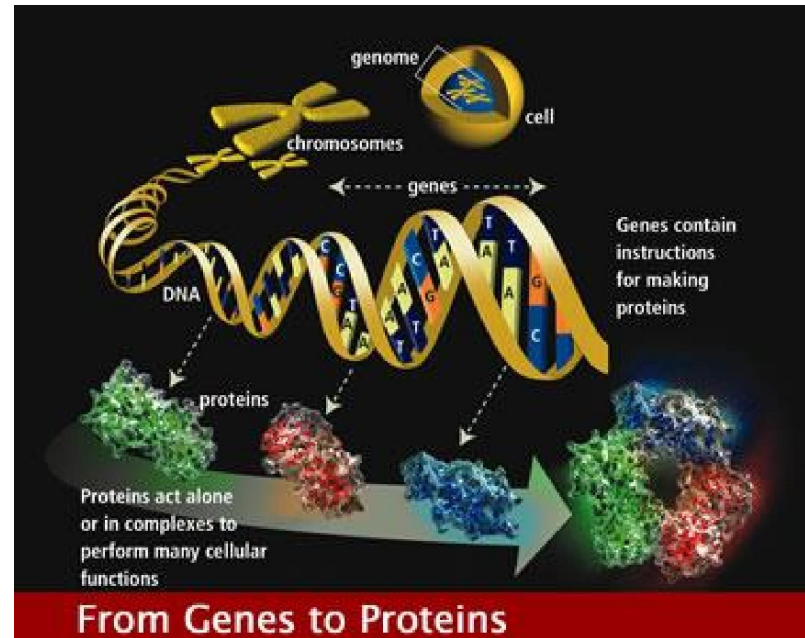
```
function(ecoli962,5,1,4,'insB_4','IS1 protein InsB').
function(ecoli963,7,0,0,'b0989','cold shock-like protein').
function(ecoli964,7,0,0,'cspG','low-temperature-responsive gene analog of CspA and CspB homolog of Salmonella cold shock protein').
function(ecoli965,7,0,0,'sfa','suppresses fabA and ts growth mutation').
function(ecoli966,0,0,0,'b0992','orf').
function(ecoli967,3,5,2,'torS','sensor protein torS (3rd module transmitter domain (?kinase) interacts with tor)').
function(ecoli968,3,5,2,'torT','part of regulation of tor operon periplasmic(1st module)').
function(ecoli969,3,5,2,'torK','response transcriptional regulator for torA (sensor TorS)(1st module)').
function(ecoli970,3,5,2,'torC','trimethylamine N-oxide reductase cytochrome c-type subunit also has activity as negativer regulator of tor operon(1st module)').
function(ecoli971,3,5,2,'torA','trimethylamine N-oxide reductase subunit').
function(ecoli972,3,5,2,'torD','part of trimethylamine-N-oxide oxidoreductase').
function(ecoli973,0,0,0,'yccD','orf').
```

Many other kinds of biological data

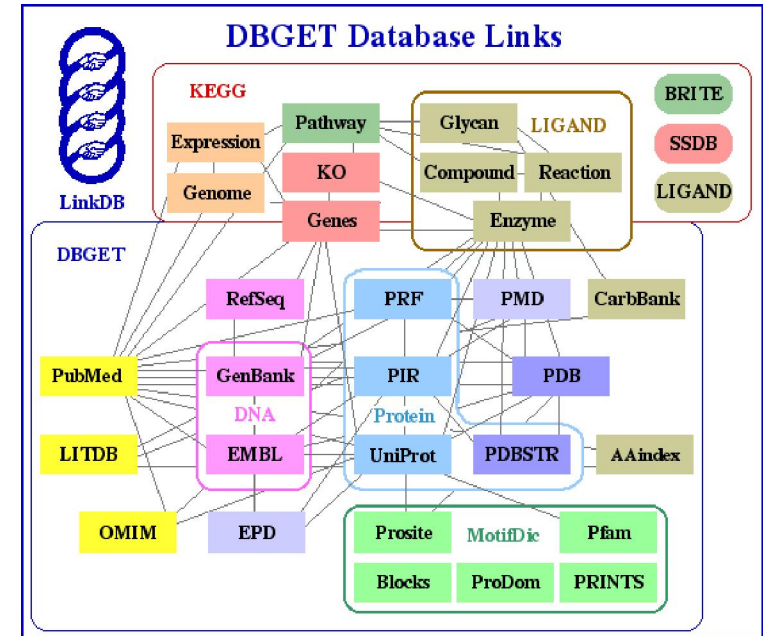
Problems in bioinformatics



Gene – Genome - Genomics
Protein – Proteome - Proteomics
Metabolite – Metabolome - Metabolomics



Central Dogma of Molecular Biology



High-throughput biology

Problems in bioinformatics

Sequence analysis

- Sequence alignment
- DNA sequence analysis
- Statistical sequence matching

Genomics

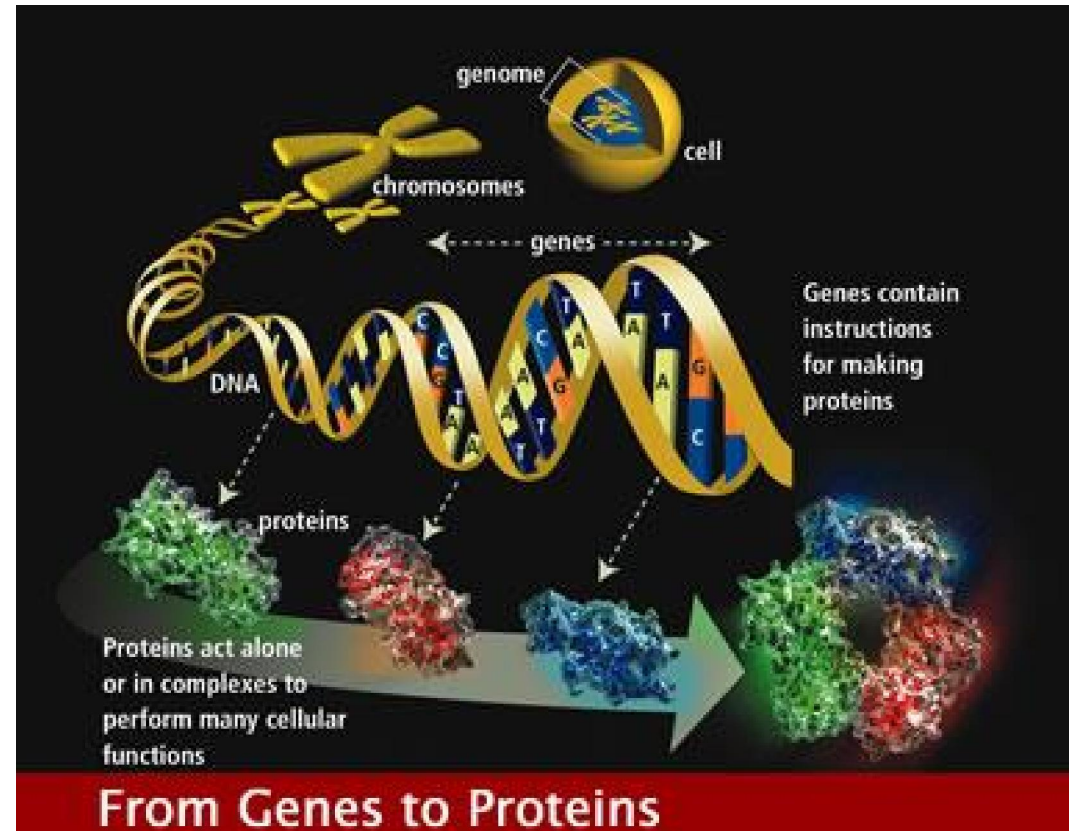
- Gene finding & prediction
- Functional genomics
- Structural genomics
- Gene expression analysis

Proteomics

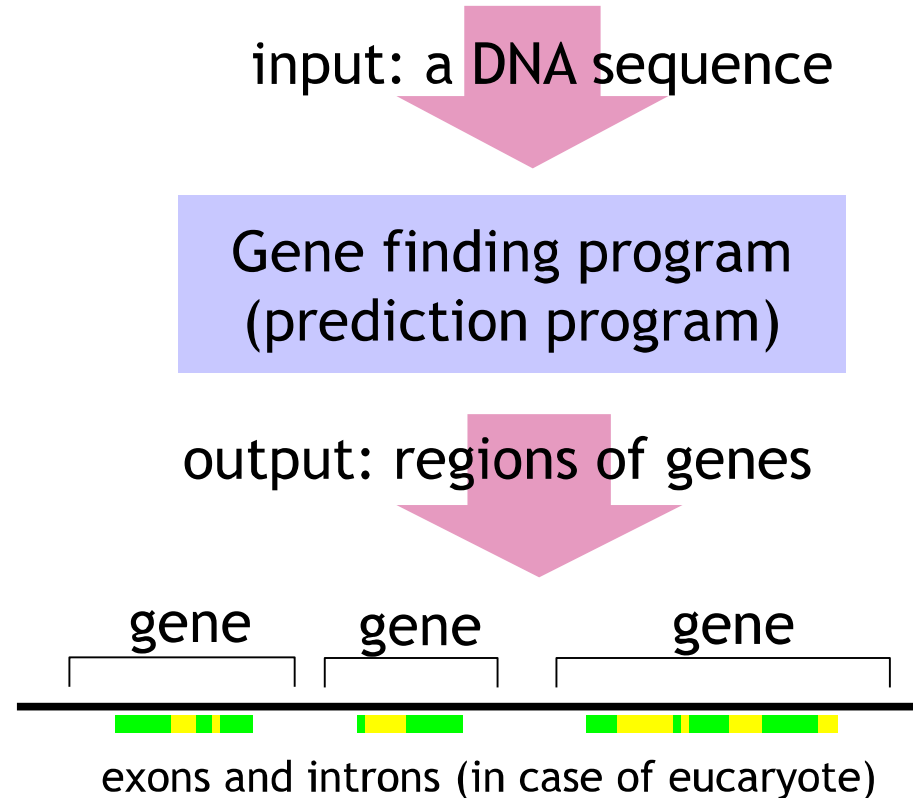
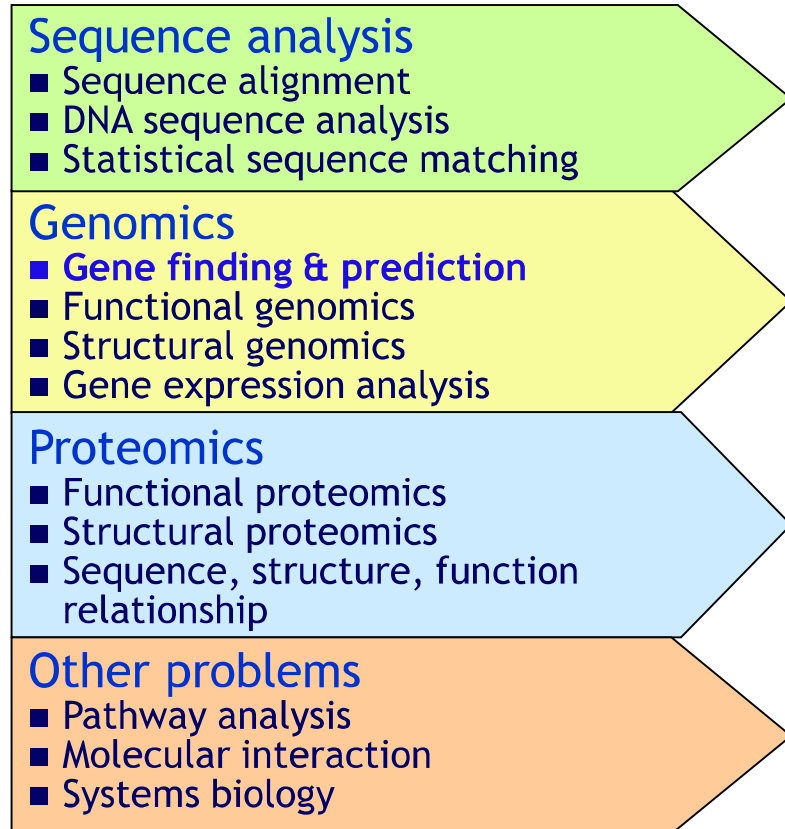
- Functional proteomics
- Structural proteomics
- Sequence, structure, function relationship

Other problems

- Pathway analysis
- Molecular interaction
- Systems biology



Problems in bioinformatics



Problems in bioinformatics

Sequence analysis

- Sequence alignment
- DNA sequence analysis
- Statistical sequence matching

Genomics

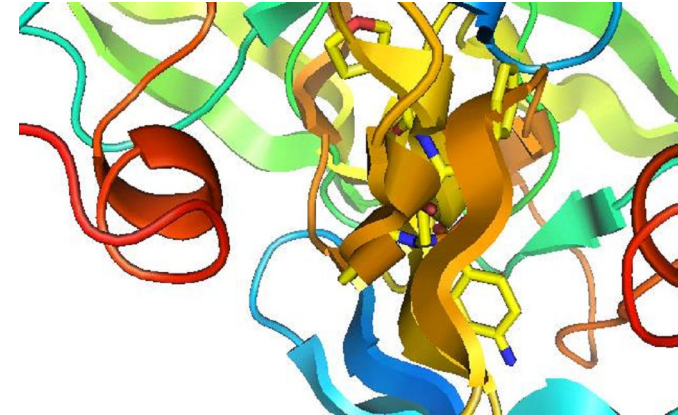
- Gene finding & prediction
- Functional genomics
- Structural genomics
- Gene expression analysis

Proteomics

- Functional proteomics
- Structural proteomics
- Sequence, structure, function relationship

Other problems

- Pathway analysis
- Molecular interaction
- Systems biology



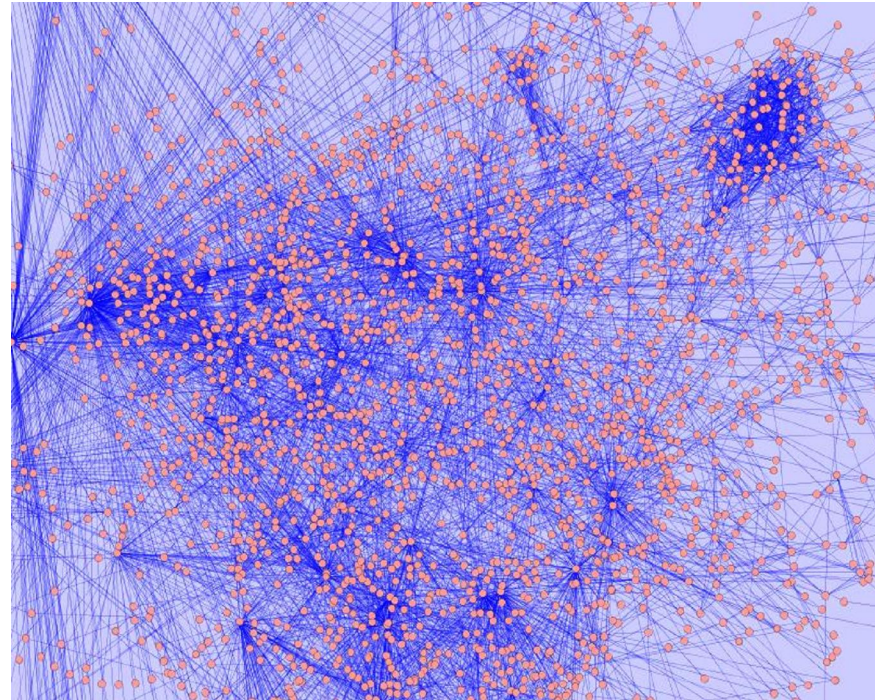
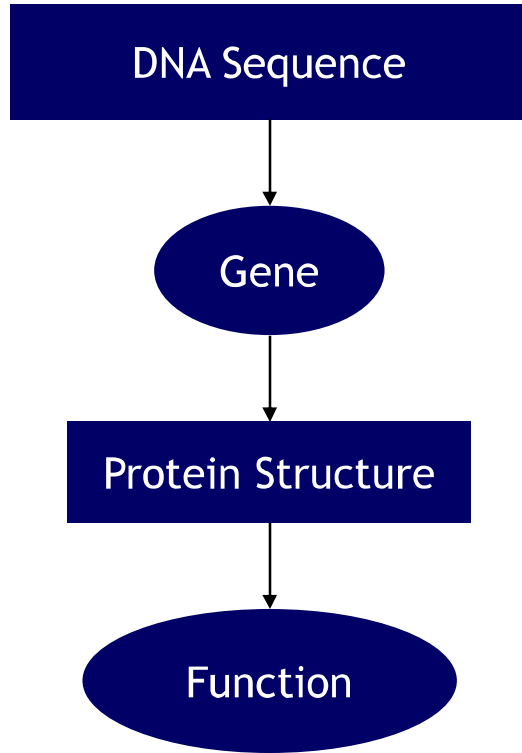
PPI: Given a sequence or structure

- in which part it interacts?
- with which protein it interacts?
- how tightly it interacts?

Our approaches: Inductive logic programming, Bayesian networks.

Nguyen, T.P., Ho, T.B. (2011). Detecting Disease Genes Based on Semi-Supervised Learning and Protein-Protein Interaction Networks, *Artificial Intelligence in Medicine*, Vol. 54, 63-71

Interactions and functions



7000 Yeast interactions among 3000 proteins

DNA → RNA → protein

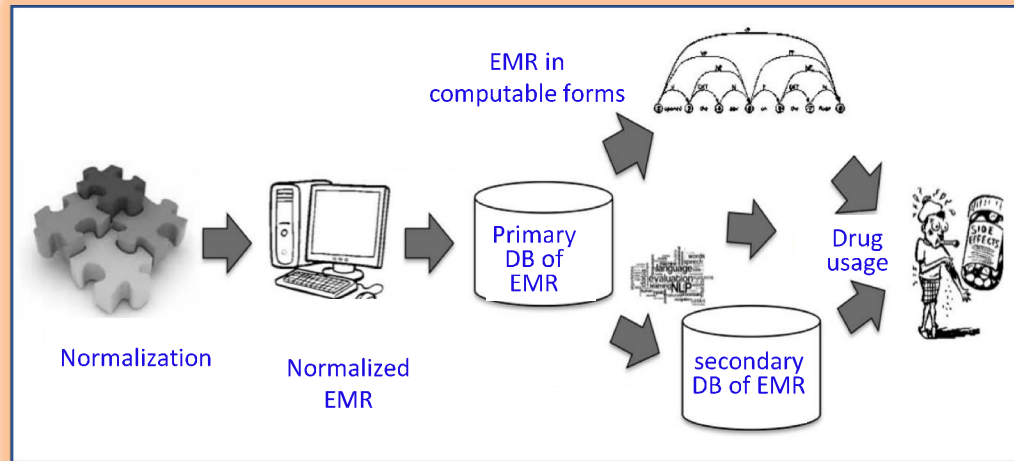
**Sequence → Structure
→ Function**

**Interaction → Network
→ Function**

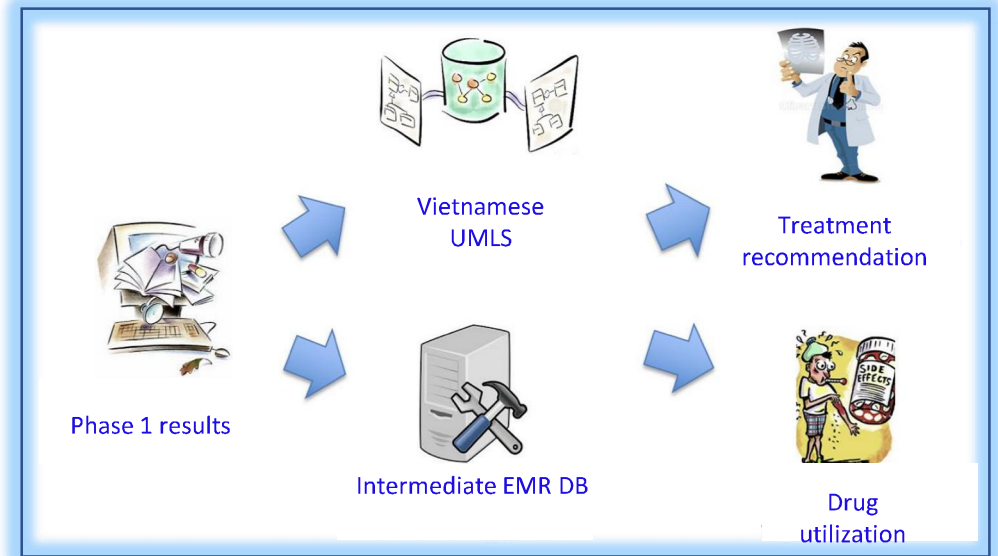
Some work in bioinformatics & biomedical informatics

- Pham, T.H., Clemente, J., Satou, K., Ho, T.B. (2005). Computational Discovery of **Transcriptional Regulatory Rules**, *Bioinformatics*, Oxford University Press, Vol. 21, Supp. 2, 101-107
- Tran, D.H., Satou, K., Ho, T.B. (2008). Finding **microRNA regulatory modules** in human genome using rule induction, *Journal BMC Bioinformatics*, Vol. 9, No. Supp 11, 1-10, MedCentral.
- Nguyen, T.P., Ho, T.B. (2011). Detecting Disease Genes Based on Semi-Supervised Learning and **Protein-Protein Interaction** Networks, *Artificial Intelligence in Medicine*, Vol. 54, 63-71
- Bui, N.T., Ho, T.B., Kanda, T. (2015). A semi-supervised tensor regression model for **siRNA efficacy** prediction, *BMC Bioinformatics*, 16:80, 2015
- Ho, T.B., Le, L., Dang, T.T., Siriwon, T. (2016). Data-driven Approach to Detect and Predict **Adverse Drug Reactions**, *Current Pharmaceutical Design Journal*, Vol. 22, No. 23, 3498-3526
- Hoang, K.H., Ho, T.B. (2019). Learning and Recommending **Treatments from Electronic Medical Records**, *Knowledge-Based Systems*, Vol. 181.
- Dang, T.T., Nguyen, T.H., Ho, T.B. (2020). **Causality** Assessment of **Adverse Drug Reaction**: Controlling Confounding Induced by Polypharmacy, *Current Pharmaceutical Design*, Vol. 25(10)
- Nuttapong S., Anon P., T.B, Ho, Ekawwit N. (2021). Improving Sentiment Analysis on **Clinical Narratives** by Exploiting UMLS Semantic Types, *Information Sciences*, Vol. 527, 356-368.

Core technology development for EMR in Vietnamese



Phase I (2015-2017)



Phase 2 (2019-2021)

Project members and collaborators



Pr. K. Takabayashi



Pr. Takahiro Suzuki



Pr. Tatsuo Kanda



Pr. Đàm Hiếu Chí



Pr. Đồng T.B.Thủy



Dr. Nguyễn V. Bính



Pr. Nguyễn Đ Công



Pr. Cao Hoàng Trụ



Dr. Võ T.N. Châu



Dr. Nguyễn T.M. Huyền



Pr. Lê Lý



Mr. Nguyễn N Hợ



Mr. Huỳnh T. Anh



Pr. Hồ Tú Bảo



PhD Studdents:

Hoàng K. Hưng



G. Moharasan



S. Nuttapon



P. Ouankhamchan



M. Matsuo



Phan S. Thành

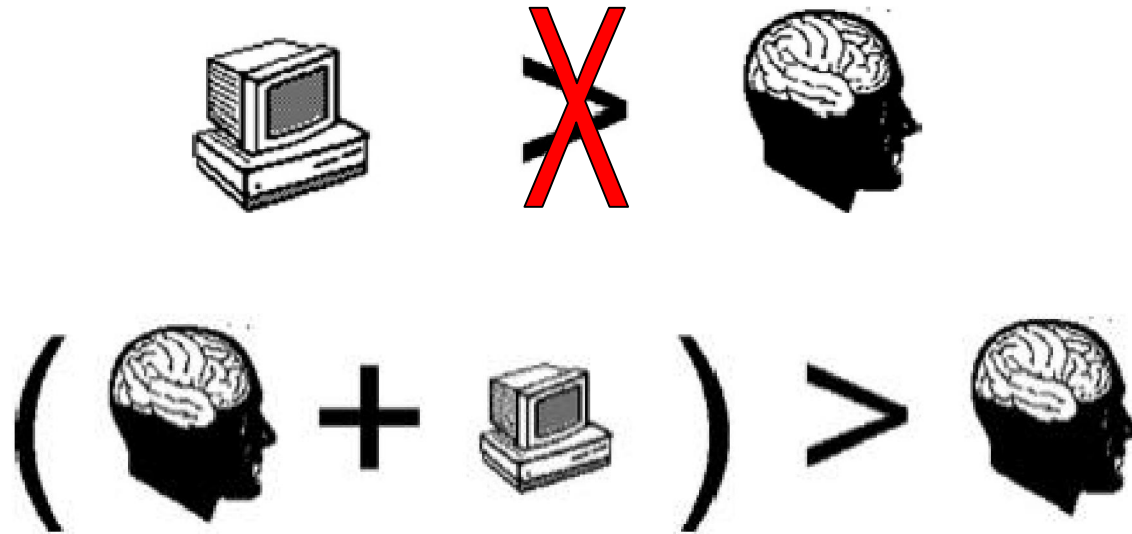


S. Taewijit



Đặng T. Thái

“Fundamental theorem” in biomedical informatics

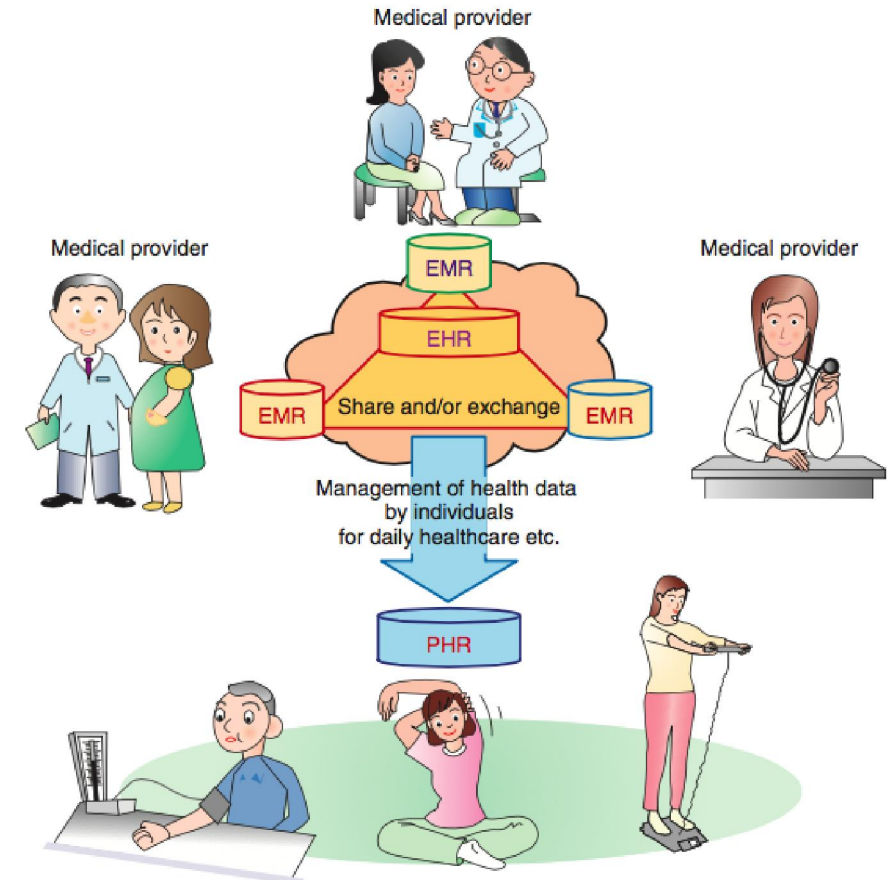


“Fundamental theorem”

Charles P. Friedman. J Am Med Inform Assoc. 2009;16:169 –170

EMR, EHR, and PHR?

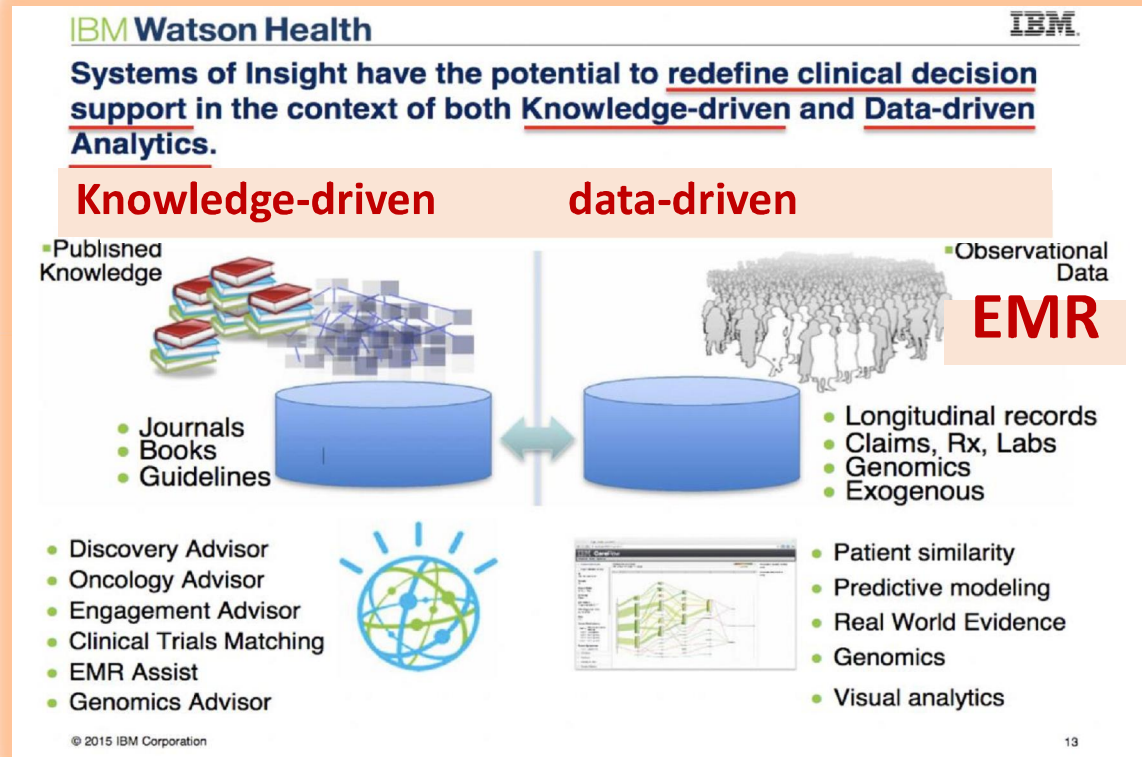
- **EMR - electronic medical records** – created, gathered, managed, and consulted by authorized clinicians and staff *within one health care organization*.
- **EHR - electronic health records** - conforms to nationally recognized interoperability standards and that can be created, managed, and consulted by authorized clinicians and staff *across more than one health care organization*.
- **PHR - personal health records** - conforms to nationally recognized interoperability standards and that can be drawn from multiple sources while being *managed, shared, and controlled by the individual*.



Yasuo Ishigure, Trends, Standardization, and Interoperability of Healthcare Information, NTT Technical Review 2017

Opportunities for medicine and biomedical research

Knowledge and data integration



Genomic medicine



Two typical problems



Diagnosis – Chẩn bệnh



Treatment – Trị bệnh

Two kinds of data in EMRs

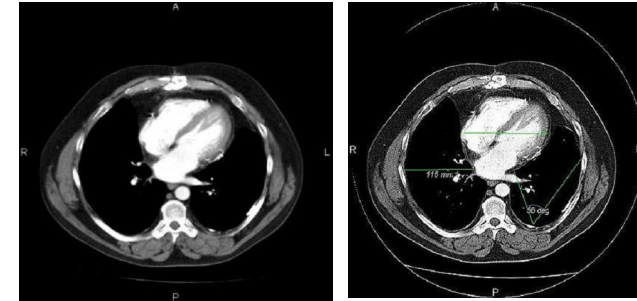
984,20123,1216,0,3354-02-05 05:40:00 EST,3354-02-05 06:01:00
EST,4270,"N",54,"Nursing/Other", "MICU nursing admission note 7AM", "
MICU nursing admission note 7AM
Pt is 68 yo male adm [**Hospital1 2**] EW [**2-4**] s/p fall 2 weeks ago while
in [**State 552**] where he landed on his left side, having left sided pain. CXR x
2 at hospital, no fx, pt sent home. Took motrin for pain steadily last 2 weeks. ^
SOB, anorexia last 2-3 days. Flew to [**Location (un) 175**] for medical care. In
EW, + EKG changes, + troponin/MB. ARF, cr 3.4, K 5.4. Given Kayexelate, D50, IV
insulin, CaGluc. Heparin gtt started for EKG changes, ?PE. No CT d/t ARF. VQ
scan showed low prob PE. Also FSBS 300s, covered by SQ insulin.
Vanco/levoquin for ? UTI. Desatt on RA, 100% NRB with SATs 100%. CXR no
rib Fx. Bicarb gtt for acidosis, gap 26. Hemodynamically stable, BP decreased
80s x 2 while sleeping, increased when awake. A&O x 3. Tx MICU for further
management. ARF probably d/t motrin use.

Neuro - A&O x 3. C/O left sided pain when turning, otherwise comfortable.
MAE.

Resp - Weaned O2 NC 6L, SATs 94%. Lungs clear, diminished at bases. No SOB.

CV - BP 103-118/54-59. NSR 70s-80s, no ectopy. Heparin gtt 1450U/hr. PTT
>150, shut off @ 4:30. Restarted 6:00 @ 1200U/hr. K 5.7->6.5. EKG unchanged.
2amps CaGluc, insulin 10U IV, 30gm Kayexelate given. Pt has had no stool from
any kayexelate given. 4:30 lytes will not reflect

Heterogeneous
and longitudinal



X-ray, CT scan, MRI, ... in PACS

MCHC	327.0	g/L	280 - 360	280 - 360	06/10/2016 14:5
MCV	81.2	fL	83.0 - 98.0	83.0 - 98.0	06/10/2016 14:5
MPV	9.6	fL	6.0 - 13.0	6.0 - 13.0	06/10/2016 14:5
MDA	1.5	GPI	0.2-0.8	0.2-0.8	06/10/2016 14:5
MD%	21.9	%	5 - 8	5 - 8	06/10/2016 14:5
P-LCR	22.2	%			06/10/2016 14:5
PDW	11.3	fL	6.0 - 10.0	6.0 - 10.0	06/10/2016 14:5
RBC(Hồng cầu)	4.67	/mm ³	4.0 - 5.9	4.0 - 5.9	06/10/2016 14:5
RDW	40.1	%	8.0 - 12.0	8.0 - 12.0	06/10/2016 14:5
THR(Tiểu cầu)	238	/mm ³	150 - 450	150 - 450	06/10/2016 14:5
WBC(Bạch cầu)	6.9	/mm ³	4.0 - 10.0	4.0 - 10.0	06/10/2016 14:5
Tổng phân tích nước tiểu (B)					
pH	7.0		4.8-7.4	4.8-7.4	06/10/2016 14:5
BIL (Bilirubin)	Âm tính	umol/L	<3.4	<3.4	06/10/2016 14:5
BLO (Hồng cầu)	Vết	/μ	<5	<5	06/10/2016 14:5
GLU (Glucose nước tiểu)	Âm tính	mmol/L	3.7 - 6.2	3.7 - 6.2	06/10/2016 14:5
KET (Ketone)	Âm tính	mmol/L	<5	<5	06/10/2016 14:5
LEU (Bạch cầu)	-	/μ	<10	<10	06/10/2016 14:5

Lab examination (blood, cardiogram...)

CLINICAL DATA (clinical text)

PARA-CLINICAL DATA (numbers)

Learning and recommending treatments

ICUSTAY

26,2538-10-29,4320,"N",1,1,"Y","Y",2538-10-26 03:18:00 EST,2538-10-29
16:25:00EST,58.95198,"adult",5107,"N","CCU","CCU","CCU","CCU",185.42,100.4,100.4,
100.4,16,5,16,5,1,5,

ICD DISGNOSIS AGE

SUBJECT_ID,HADM_ID,SEQUENCE,CODE,DESCRIPTION
25,5726,1,"410.71","SUBENDOCARDIAL INFARCTION INITIAL EPISODE OF CARE"
25,5726,2,"250.11","DIABETES MELLITUS WITH KETOACIDOSIS TYPE I NOT STA"
25,5726,3,"414.01","CORONARY ATHEROSCLEROSIS OF NATIVE CORONARY ARTERY"
25,5726,4,"401.9","UNSPECIFIED ESSENTIAL HYPERTENSION"

DEMOGRAPHIC EVENTS DATA

SUBJECT_ID,HADM_ID,MARITAL_STATUS_ITEMID,MARITAL_STATUS_DESCR,ETHNICITY_ITEMID,ETHNICITY_DESCR,OVERALL_PAYOR_GROUP_ITEMID,OVERALL_PAYOR_GROUP_DESCR,RELIGION_ITEMID,RELIGION_DESCR,ADMISSION_TYPE_ITEMID,ADMISSION_TYPE_DESCR,ADMISSION_SOURCE_ITEMID,ADMISSION_SOURCE_DESCR
25,5726,200050,"MARRIED",200083,"WHITE",200067,"PRIVATE",200081,"UNOBTAINABLE",200028,"EMERGENCY",200029,"EMERGENCY ROOM ADMIT"

MEDEVENTS DATA

SUBJECT_ID,ICUSTAY_ID,ITEMID,CHARTTIME,ELEMID,REALTIME,CGID,CUID,VOLUME,DOSE,DOSEUOM,SOLUTIONID,SOLVOLUME,LUNITS,ROUTE,STOPPED
25,28,45,2538-10-26 04:30:00 EST,1,2538-10-26 04:57:00 EST,2691,1,0,8,"Uhr",18,100,"ml","IV Drip",
25,28,142,2538-10-26 04:30:00 EST,1,2538-10-26 05:00:00 EST,2691,1,0,2,"mcgkgmin",13,100,"ml","IV Drip",
25,28,45,2538-10-26 04:45:00 EST,1,2538-10-26 04:57:00 EST,2691,1,0,10,"Uhr",18,100,"ml","IV Drip",
25,28,142,2538-10-26 04:45:00 EST,1,2538-10-26 05:00:00 EST,2691,1,0,2,"mcgkgmin",13,100,"ml","IV Drip",
25,28,45,2538-10-26 05:00:00 EST,1,2538-10-26 05:23:00 EST,2049,1,0,10,"Uhr",18,100,"ml","IV Drip",
25,28,142,2538-10-26 05:00:00 EST,1,2538-10-26 05:23:00 EST,2049,1,0,2,"mcgkgmin",13,100,"ml","IV Drip",
25,28,45,2538-10-26 05:15:00 EST,1,2538-10-26 06:07:00 EST,2691,1,0,10,"Uhr",18,100,"ml","IV Drip",
25,28,142,2538-10-26 05:15:00 EST,1,2538-10-26 06:07:00 EST,2691,1,0,2,"mcgkgmin",13,100,"ml","IV Drip",
25,28,45,2538-10-26 05:30:00 EST,1,2538-10-26 06:07:00 EST,2691,1,0,10,"Uhr",18,100,"ml","IV Drip",
25,28,142,2538-10-26 05:30:00 EST,1,2538-10-26 06:07:00 EST,2691,1,0,2,"mcgkgmin",13,100,"ml","IV Drip",

MEDURATIONS DATA

SUBJECT_ID,ICUSTAY_ID,ITEMID,ELEMID,STARTTIME,STARTREALTIME,ENDTIME,CUID,DURATION
25,28,45,1,2538-10-26 04:30:00 EST,2538-10-26 04:57:00 EST,2538-10-29 16:25:00 EST,1,5035
25,28,142,1,2538-10-26 04:30:00 EST,2538-10-26 05:00:00 EST,2538-10-29 16:25:00 EST,1,5035
25,28,45,1,2538-10-26 04:45:00 EST,2538-10-26 04:57:00 EST,2538-10-29 16:25:00 EST,1,5020
25,28,142,1,2538-10-26 04:45:00 EST,2538-10-26 05:00:00 EST,2538-10-29 16:25:00 EST,1,5020
25,28,45,1,2538-10-26 05:00:00 EST,2538-10-26 05:23:00 EST,2538-10-29 16:25:00 EST,1,5005
25,28,142,1,2538-10-26 05:00:00 EST,2538-10-26 05:23:00 EST,2538-10-29 16:25:00 EST,1,5005

POE-MED DATA

POE_ID,DRUG_TYPE,DRUG_NAME,DRUG_NAME_GENERIC,PROD_STRENGTH,FORM_RX,DOSE_VAL_RX,DOSE_UNIT_RX,FORM_VAL_DISP,FORM_UNIT_DISP,DOSE_VAL_DISP,DOSE_UNIT_DISP,DOSE_RANGE_OVERRIDE
1930588,"BASE","D5W","","250mL Bag","","250","ml","250","ml",,,
1930589,"BASE","NS","","500mL Bag","","500","ml","500","ml",,,
1936709,"BASE","SW","","100mL Bottle","","100","ml","100","ml",,,
1929791,"MAIN","Aspirin","Aspirin",325mg Tab,"325","mg","1","TAB",,,
1929796,"MAIN","Potassium Chloride","Potassium Chloride",20mEq Packet,"20","mEq","1","PKT",,,
1929797,"MAIN","Atorvastatin","Atorvastatin",40mg Tab,"40","mg","2","TAB",,,
1929819,"MAIN","Potassium Chloride","Potassium Chloride",20mEq Packet,"20","mEq","2","PKT",,,
1930558,"MAIN","Potassium Chloride","Potassium Chloride",20mEq Packet,"20","mEq","2","PKT",,,
1930691,"MAIN","Pantoprazole","Pantoprazole",40mg Tab,"40","mg","1","TAB",,,
1931503,"MAIN","Calcium Gluconate","Calcium Gluconate",1g/10mL Vial,"1","g","2","VIAL",,,
1931745,"MAIN","Zolpidem Tartrate","Zolpidem Tartrate",5mg Tab,"5","mg","1-2","TAB",,,
1931746,"MAIN","Acetaminophen","Acetaminophen",325mg Tab,"325-650","mg","1-2","TAB",,,

POR-ORDER DATA

POE_ID,SUBJECT_ID,HADM_ID,ICUSTAY_ID,START_DT,STOP_DT,ENTER_DT,MEDICATION,PROCEDURE_TYPE,STATUS,ROUTE,FREQUENCY,DISPENSE_SCHEID,IV_FLUID,IV_RATE,INFUSION_TYPE,SLIDING_SCALE,DOSES_PER_24HRS,DURATION,DURATION_INTVL,EXPIRATION_VAL,EXPIRATION_UNIT,EXPIRATION_DT,LABEL_INSTR,ADDITIONAL_INSTR,MD_ADD_INSTR,NURSE_ADD_INSTR
1929790,25,5726,28,2538-10-26 05:00:00 EST,2538-10-27 03:00:00 EST,2538-10-26 04:00:00 EST,"Insulin","IV Piggyback","Inactive (Due to a change order)","IV Drip","INFUSION",,,,,,"Ongoing","Enter on Label",,,,"Fingersticks every hour IV Drip Rate: 8 UNIT/HR","Specify blood glucose goal",
1929795,25,5726,28,2538-10-26 05:00:00 EST,2538-10-26 04:00:00 EST,2538-10-26 04:00:00 EST,"Potassium Chloride","IV Piggyback","discontinued","IV","ONCE","5",,,,,1,1,"Doses",,,,"Enter on Label",,,,"CARDIAC MONITORING AND CENTRAL LINE ARE REQUIRED WHEN SELECTING CONCENTRATED PRODUCT (20 mEq/50 mL). 20 mEq/50 mL preparations are given via central line only. Fluid restricted patients may receive 40 mEq in 500 mL NS or D5W. No more than 60 mEq placed in one liter of fluid per BIDMC policy.",,"Cardiac monitoring and central lines are required for rates > 10 mEq/hr."

NOTEVENTS DATA

"25,5726,28,0,2538-10-26 07:51:00 EST,2538-10-26 08:33:00 EST,1807,"N",1,"Nursing/Other","NURSING PROGRESS NOTE",
NURSING PROGRESS NOTE
58 Y/O MALE ADMITTED FROM [**Hospital1 2**] ER (TRANSFERRED FROM [**Hospital6 110**]). HE INITIALLY PRESENTED TO [**Hospital6 110**] WITH C/O N/V, DIZZINESS. HE IS S/P INSULIN PUMP INSERTION IN [**2538-5-6**]. HIS PUMP FAILED ON SATURDAY AND BEGAN FEELING POORLY. HE WAS ADMITTED WITH A BLOOD GLUCOSE > 575. HE ALSO HAD ST CHANGES ON EKG. HE WAS TREATED WITH IV LOPRESSOR, INTEGRILLIN, IV NS, INSULIN. HE REFUSED ASA STATING IT MAKES HIS STOMACH UPSET. ADMITTED TO CCU FOR R/O MI PROTOCOL.

This is a 58 yr old male Pt who presented to [**Hospital6 **] with C/O N/V & dizziness- He had an insulin pump inserted in 6/04 & on Saturday [**10-25**] it failed- blood sugar was > 500- Also, his EKG showed new ST depressions (no C/O CP & cardiac enzymes negative)- Pt was transferred to [**Hospital1 2**] EW on integrilin & insulin gts for further care- Pt was admitted to CCU- R radial A line was placed- Pt developed a sinus arrhythmia HR 40-70's with hypotension (SBP 60-70's)- atropine given for ? bradycardia induced hypotension. IV fluids wide open & dopamine gtt started- EKG SA HR 50-70's with return of ST-T waves changes in lateral leads- PA line inserted into R II- RA 8-PAP 42/22-PCWP 15-16- decision was made to cath Pt due to persistent hypotension Cardiac cath revealed moderately severe single vessel CAD (LCx large vessel proximal 60-70%) normal LV systolic function- no intervention done- ? elective stent LCx when stable- CO high with low SVR- ? sepsis

CV-R/MI. HR 70-80NSR, BP by R radial Aline 110-140/60-70. ASA, plavix (loaded w/ 300mg this am) cont., lopressor 12.5mg bid added. No c.o CP, weakness, dizz. PA line- CVP 8-10, PA 28-38/16-18, CO [**7-15**], SVR 500. Has received ~10liters of IVF over 48hr, u/o 3000 over same time. R femoral Aline d/c @ 12n w/o complication by Card fellow, site C&D w/ transparent dsq, no hematoma, no oozing. Pulses dpl-1+, baseline. Endo/Fluids- IDDM on insulin gtt @2-3u/hr w/ small and improving po intake.FS 92-152. IVF D5.45NS @ 100cc/h (dec'd from 150/hr this am). U/O 80-120/hr clear urine. + 2500 for day.

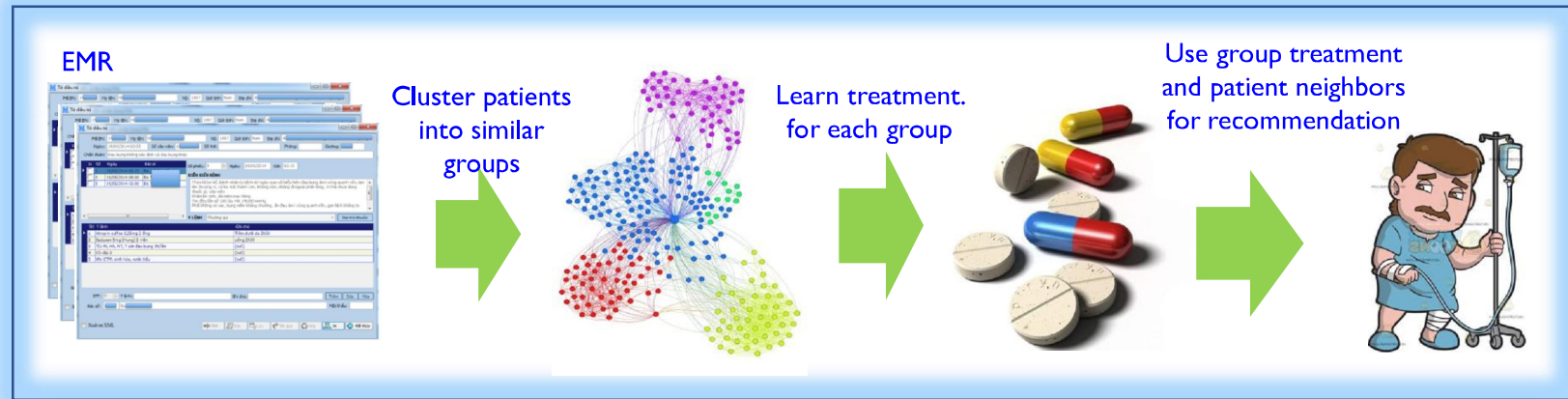
CCU Nursing Progress Note-7a-7p 58 y/o male admitted [**10-25**] w/ N/V/dizz, IDDM w/ failed pump, FS 576 to [**Hospital6 **], EKG changes. Placed on insulin gtt, IVF and tx to [**Hospital1 2**]. Over w/e,hypotensive- Dopa and Levo; PA line placed w/ High CO, low SVR; cathed, RCA 70% stenosis, RI MI; DKA. Much improved overnight and today. Anion gap now closed. Heparin. R fem Aline d/c. Cont INS gtt, IVF, antith. Plan for Stent of RCA [**10-28**]. NPO p MN. Neuro- A&O x3, MAE, much less irritable w/ cardiac explanation/education by MD/RN CCU team. Able to assist w/ position change. To be OOB this evening when PA line DIC.

CHARTDURATION DATA

SUBJECT_ID,ICUSTAY_ID,ITEMID,ELEMID,STARTTIME,STARTREALTIME,ENDTIME,CUID,DURATION
25,28,781,0,2538-10-26 03:59:00 EST,2538-10-26 04:30:00 EST,2538-10-29 16:25:00 EST,1,5066
25,28,1536,0,2538-10-26 03:59:00 EST,2538-10-26 04:30:00 EST,2538-10-29 16:25:00 EST,1,5066
25,28,1535,0,2538-10-26 03:59:00 EST,2538-10-26 04:30:00 EST,2538-10-29 16:25:00 EST,1,5066
25,28,1534,0,2538-10-26 03:59:00 EST,2538-10-26 09:29:00 EST,2538-10-29 16:25:00 EST,1,5066
25,28,1532,0,2538-10-26 03:59:00 EST,2538-10-26 09:29:00 EST,2538-10-29 16:25:00 EST,1,5066
25,28,1529,0,2538-10-26 03:59:00 EST,2538-10-26 04:30:00 EST,2538-10-29 16:25:00 EST,1,5066
25,28,1525,0,2538-10-26 03:59:00 EST,2538-10-26 04:30:00 EST,2538-10-29 16:25:00 EST,1,5066
25,28,1523,0,2538-10-26 03:59:00 EST,2538-10-26 04:30:00 EST,2538-10-29 16:25:00 EST,1,5066
25,28,1522,0,2538-10-26 03:59:00 EST,2538-10-26 09:29:00 EST,2538-10-29 16:25:00 EST,1,5066
25,28,1162,0,2538-10-26 03:59:00 EST,2538-10-26 04:30:00 EST,2538-10-29 16:25:00 EST,1,5066

MIMIC is an EMR database given by Harvard Medical School to the research community

Learning and recommending treatments

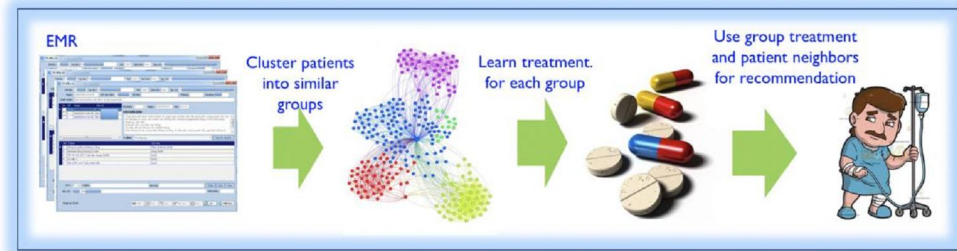


1 Convert EMRs into computable forms → Patient clustering

2 Learn treatment periods → Prescription tree construction

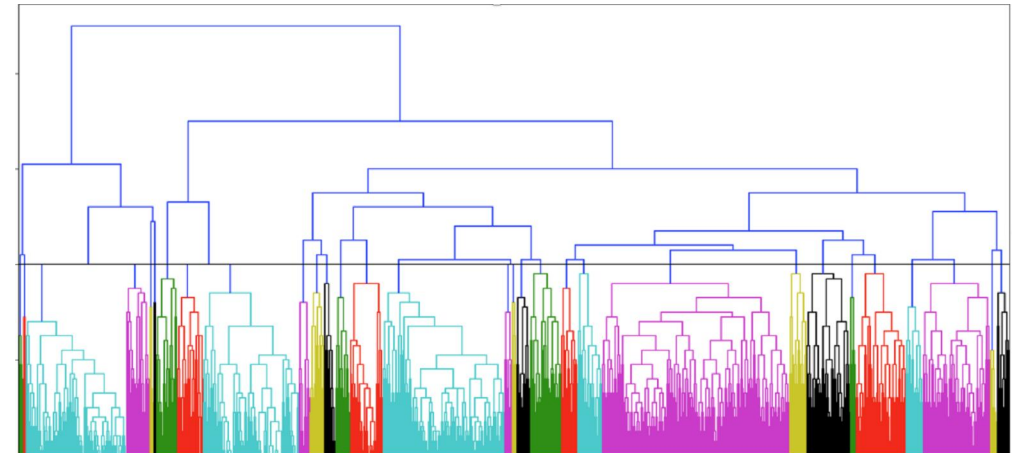
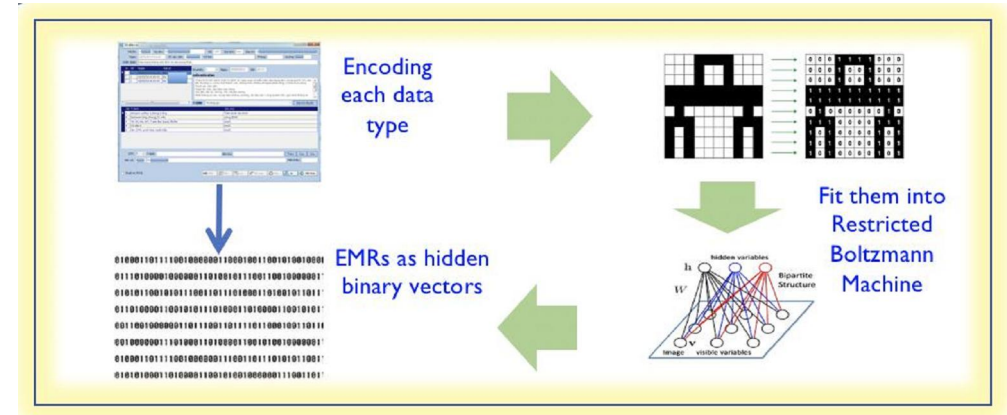
3 Recommend treatment with and without weights

Learning and recommending treatments

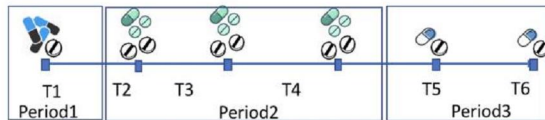
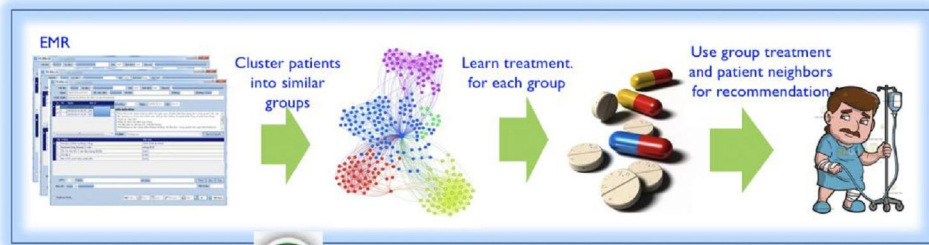


1

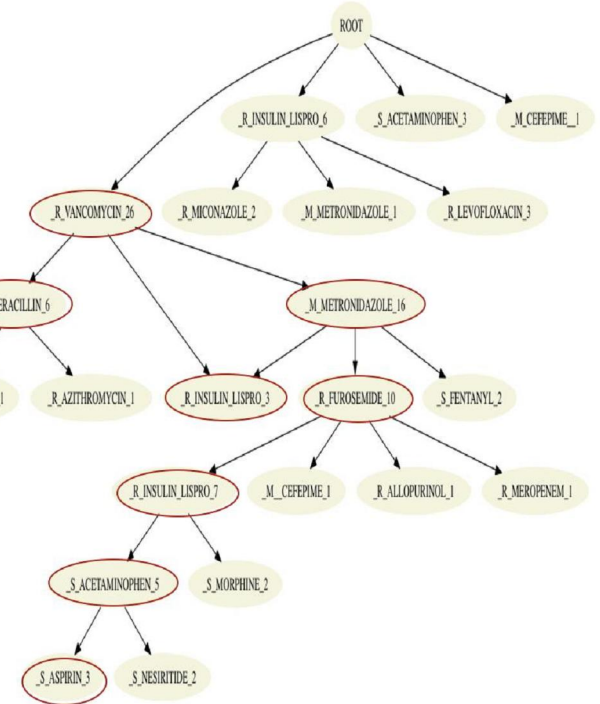
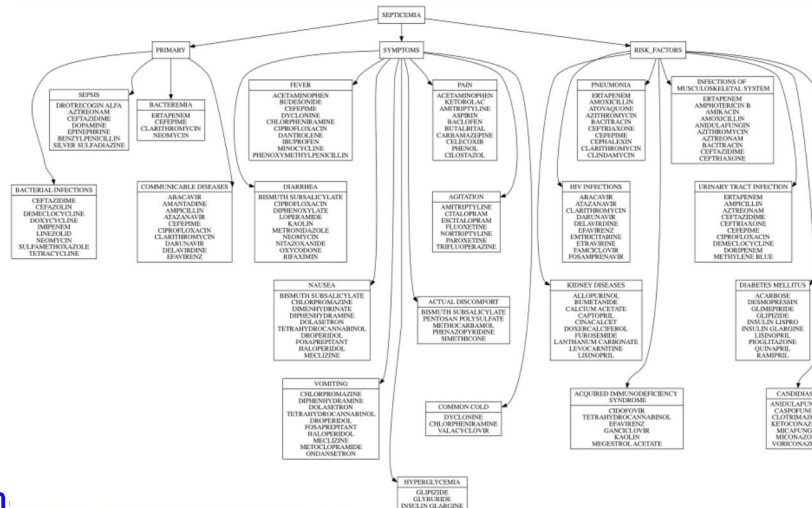
Converting EMRs into computable forms → Patient clustering



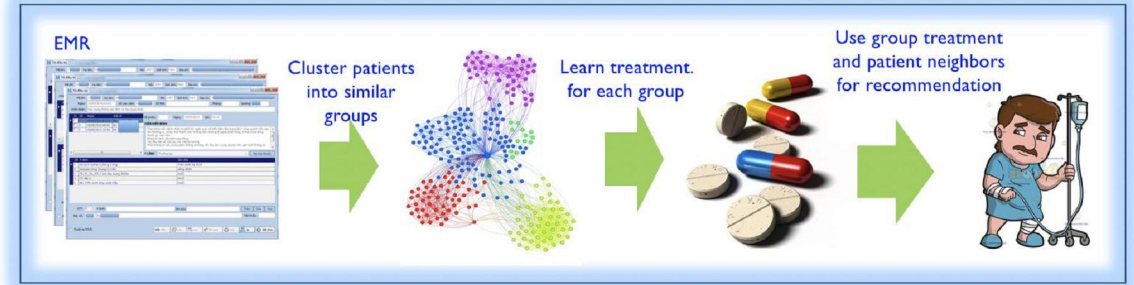
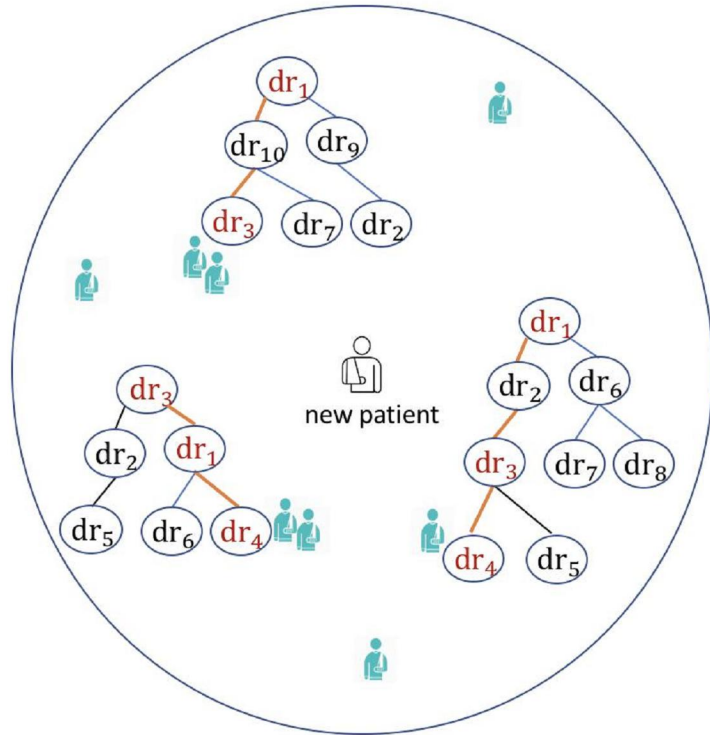
Learning and recommending treatments



Algorithm to learn prescription tree



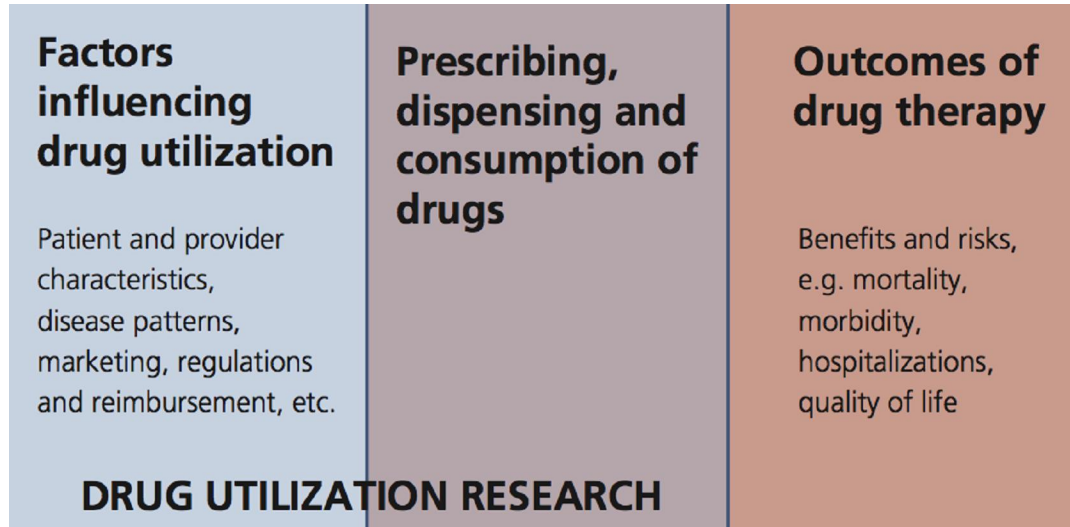
Learning and recommending treatments



3 Recommend treatment with and without weights

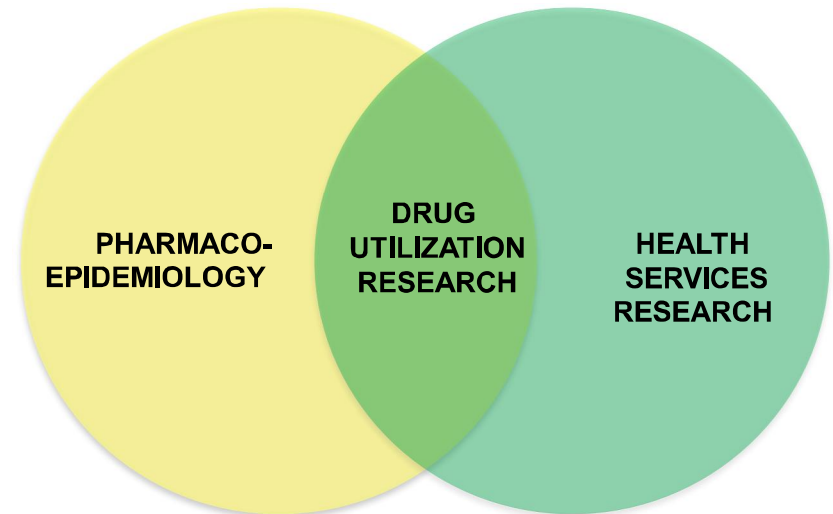
RESPIRATORY COHORT									
Framework	Precision			Recall			F score		
	@3	@5	@10	@3	@5	@10	@3	@5	@10
TRoS	41.33	33.06	21.62	42.58	56.52	71.07	41.94	41.71	33.15
WTRoS	42.15	33.59	21.68	44.02	57.44	71.19	43.06	42.38	33.23
ICF + SGD	42.13	33.4	21.75	43.83	56.68	71.16	42.95	42.03	33.31
ICF + IALS	20.1	23.19	17.96	19.33	39.0	59.89	19.7	29.06	27.62
ICF + ADA	42.15	33.44	21.72	43.48	56.78	70.9	42.8	42.08	33.25

Drug utilization research

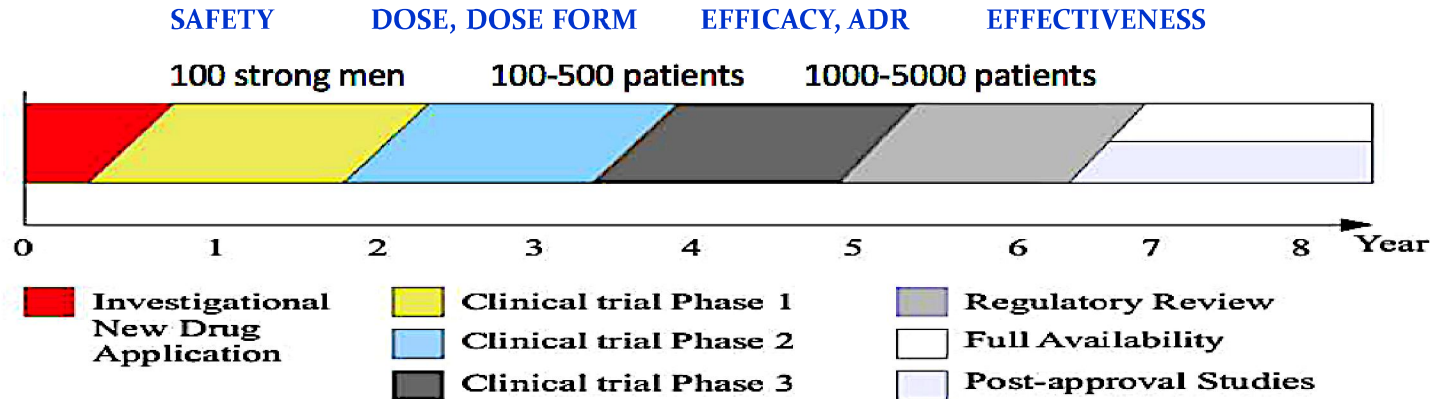


The principal aim of drug utilization research is to assess whether drug therapy is rational or not.

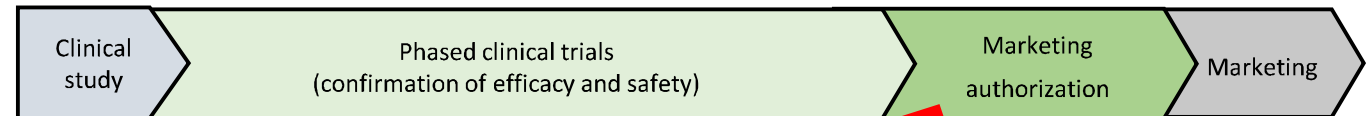
WHO. Introduction to Drug Utilization Research, 2003.



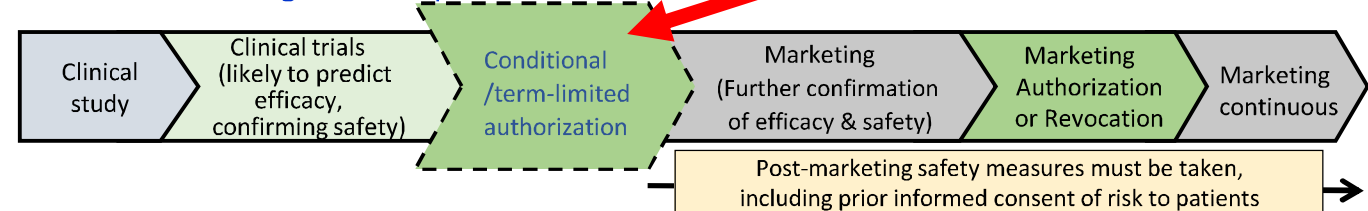
Post-market surveillance of drug



Traditional approval process



New scheme for regenerative product



Efficacy vs. effectiveness

- **Efficacy** is the capacity to produce an effect. In medicine, it is the ability of an intervention or drug to produce a desired effect. It is tested by **explanatory clinical trials** under **ideal and controlled circumstances**.
- **Effectiveness** is the capability of producing a desired result. In medicine effectiveness relates to **how well a treatment works in practice**. It is tested by **pragmatic clinical trials**.



Ian Ford, Pragmatic trials, *The New England Journal of Medicine*, 454-463, 2016.

Two problems under investigation

- Construction of contingency tables from EMRs for evaluating the effectiveness of drug utilization
- Detection and prediction of adverse drug reaction when using multiple drugs

	Treatment 1	Treatment 2
Effective level 1	n11	n12
Effective level 2	n21	n22
Effective level 3	n31	n32

Hypertension drug effectiveness

	Coversyl I 5mg	Amidile G (Amlodipin 5mg)
Less effective (0, 0.01)	2	14
Effective (0.01, 0.55)	9	50
Strong Effective (0.55, 1)	2	1

	Less effective	Effective	Strong effective	Sample size
Coversyl 5mg	2 (2.67)	9 (9.83)	2 (0.5)	13
Amidile G(Amlodipin 5mg)	14 (13.33)	50(49.17)	1(2.5)	65
Total	16	59	3	78

The values of $(O - E)^2/E$

	Less effective	Effective	Strong effective
Coversyl 5mg	0.168	0.07	4.5
Amidile G(Amlodipin 5mg)	0.034	0.014	0.9
Total	$\chi^2 = 5.686$ $d.f \text{ of } \chi^2 = (3 - 1)(2 - 1) = 2$		

- Compare Coversyl 5mg vs Amidile G(Amlodipin 5mg) using EMRs of hypertension patients who took only those drugs for the study.

$$H_0: p_{Coversyl_lessEff} = p_{AmildileG_lessEff};$$

$$p_{Coversyl_Eff} = p_{AmildileG_Eff}$$

$$p_{Coversyl_strongEff} = p_{AmildileG_strongEff}$$

- With d.f. = 2, the tabulet upper 5% point of χ^2 is 5.99, the null hypothesis is not rejected.
- No difference in the patients treated by Coversyl 5mg and Amidile G (Amlodipin 5mg).

Ho T.B., Hoang K.H., Dang T.T., Drug utilization research with pragmatic clinical trials using electronic medical records (in preparation)

Acute bronchitis drug effectiveness

	Acetyl cystein - (ESOMEZ 200mg)	Acetyl cystein (Adomuc 200mg)
Less effective (0, 0.01)	7	7
Effective (0.01, 0.55)	20	26
Strong Effective (0.55, 1)	1	2

	Less effective	Effective	Strongly effective	Sample size
ESOMEZ 200mg	7 (6.222)	20 (20.444)	1 (1.333)	28
Adomuc 200mg	7 (7.778)	26(25.556)	2(1.667)	35
Total	14	46	3	63

The values of $(O - E)^2/E$

	Less effective	Effective	Strong effective
ESOMEZ 200mg	0.097	0.01	0.083
Adomuc 200mg	0.078	0.008	0.0665
Total	$\chi^2 = 0.3425$ $d.f \text{ of } \chi^2 = (3 - 1)(2 - 1) = 2$		

- Compare ESOMEZ 200mg and Adomuc 200mg using EMRs of hypertension patients who took only those drugs for the study.

$$H_0: p_{\text{Esomez_lessEff}} = p_{\text{Adomuc_lessEff}}$$

$$p_{\text{Esomez_Eff}} = p_{\text{Adomuc_Eff}}$$

$$p_{\text{Esomez_strongEff}} = p_{\text{Adomuc_strongEff}}$$

- With d.f. = 2, the tabulet upper 5% point of χ^2 is 5.99, the null hypothesis is not rejected.
- No difference in the patients treated by ESOMEZ 200mg and Adomuc 200mg.

Digital transformation in healthcare



Working with hospitals

Seminar at
Military Medical
Department with
leaders of 17
Northern
hospitals and
seminar at
hospital 108

