

HBYS ve LBYS'ni etkin kullanımı: Veri madenciliği



Muhittin A. SERDAR
Acıbadem Üniversitesi
Clinlab Laboratuvarı

Laboratuvarın bilişim teknolojisi ile tanışması



809-873
Hunayn ibn Ishaq



1743-1794
Antoine L. Lavoisier



1825-50



1900-50



1960

1980.....
Dr. Masahide Sasaki



1990.....

2000....

Bilişim Teknolojileri;

- **verileri saklamak,**
 - **işlemek ve**
 - **iletmek**
- İçin kullanılan teknolojilerdir.

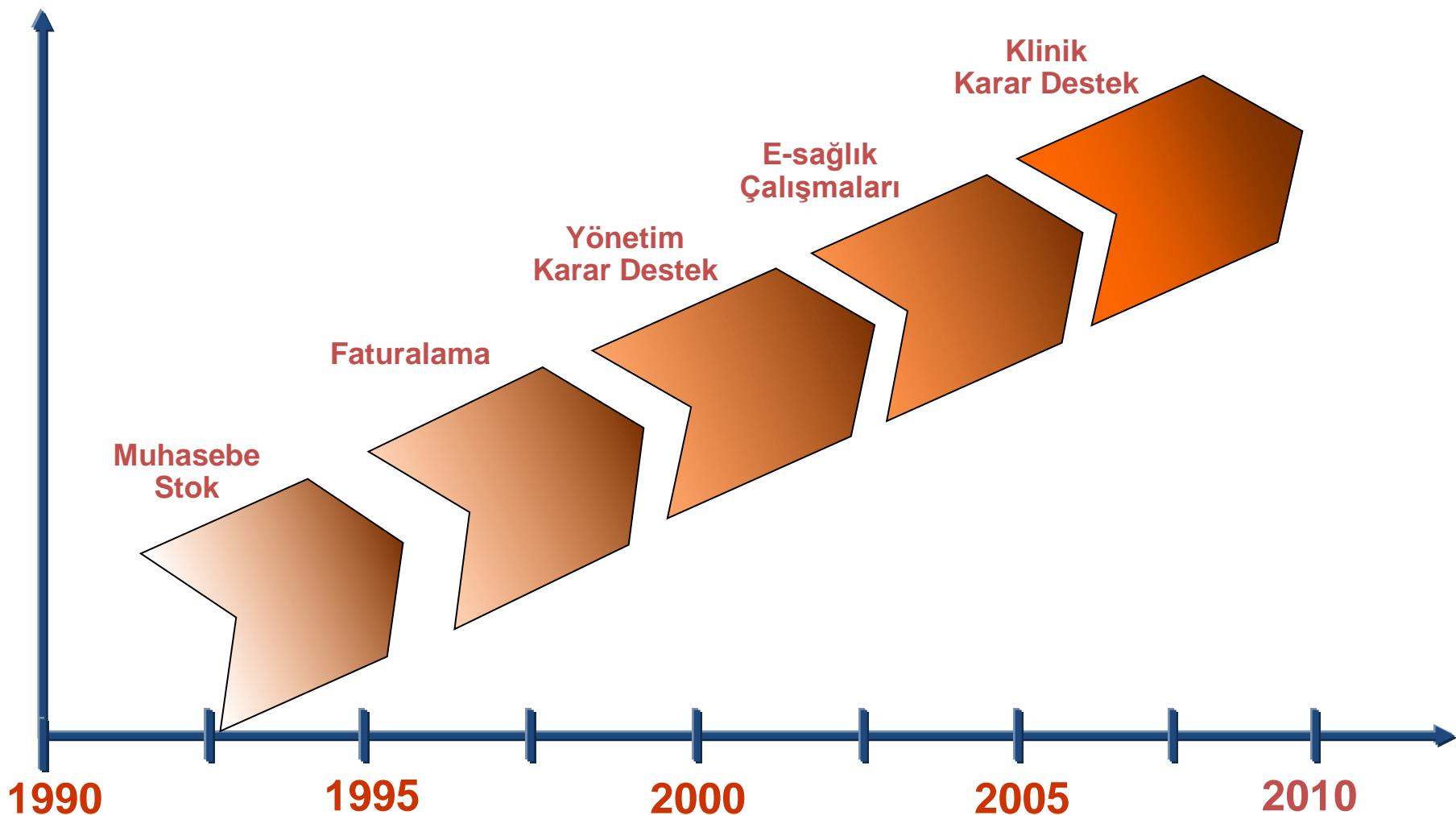


Bilişim Teknolojilerinin içeriği

- **bilgisayar donanımı ve yazılımı,**
 - **bilgisayar ağları,**
 - **iletişim teknolojileri ve araçları**
 - **yetişmiş insan gücü,**
 - **prosedürler,**
 - **internet, intranet ve**
- gibi çok sayıda bileşene sahiptir**



Türkiye'de Sağlık Bilgi Sistemlerinin Evrimi



Tıp Bilişimi (Medical Informatic) Nedir?

Bilgi teknolojilerini kullanarak yapılan
tıbbi çalışma (tanı, tedavi),
eğitim,
iletişim,
veri, bilgi toplama ve bilgi işleme
bilgiyi yönetme,
tıbbi karar verme ve
bilimsel çözümleme
yöntemlerini içeren bir bilim dalıdır.



**Tıp Bilişimi, Tıp, bilişim teknolojileri ve farklı disiplinlerinin
kesiştiği bir yerde bulunmaktadır.**

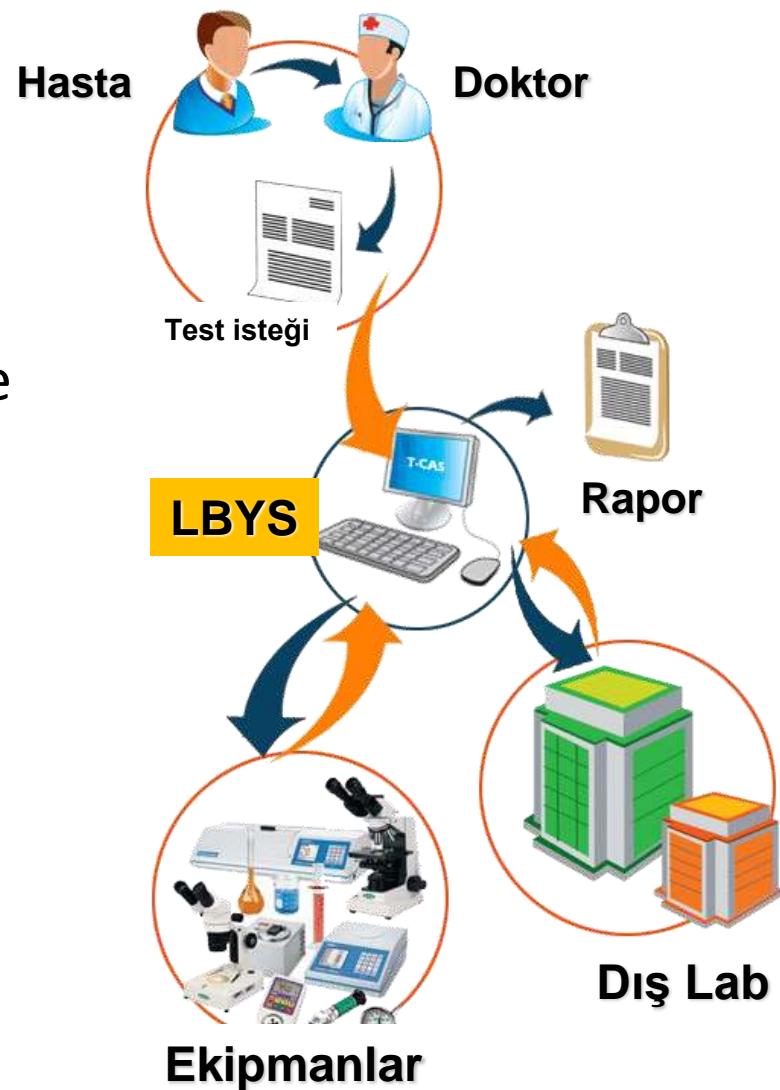
Laboratuvarların bilişim teknolojileri ile en önemli ilişkisi: Laboratuar Bilgi Yönetim Sistemleri (LBYS)

Bilişim dilinde

Tıbbi laboratuar süreçleri sonucunda ortaya çıkan verilerin oluşturulması, işlenmesi ve depolanmasını sağlayan yazılım ve donanımlar bütünüdür.

Klinik Laboratuar boyutunda

Laboratuara örneğin girişinden, hastalara sonuçların ulaşmasına kadar olan çevrimi **yöneten bir bilgi sistemidir.**



Laboratuar Bilgi Yönetim Sistemleri (LBYS)

**Laboratuar
İş akışı ve yönetimi**

**Kalite Kontrol ve
Güvencesi**

**Faturalandırma
Maliyet analizi**

Yönetimsel kontrol

Arşivleme

**Laboratuar cihazlarıyla
bağlantı kurma**

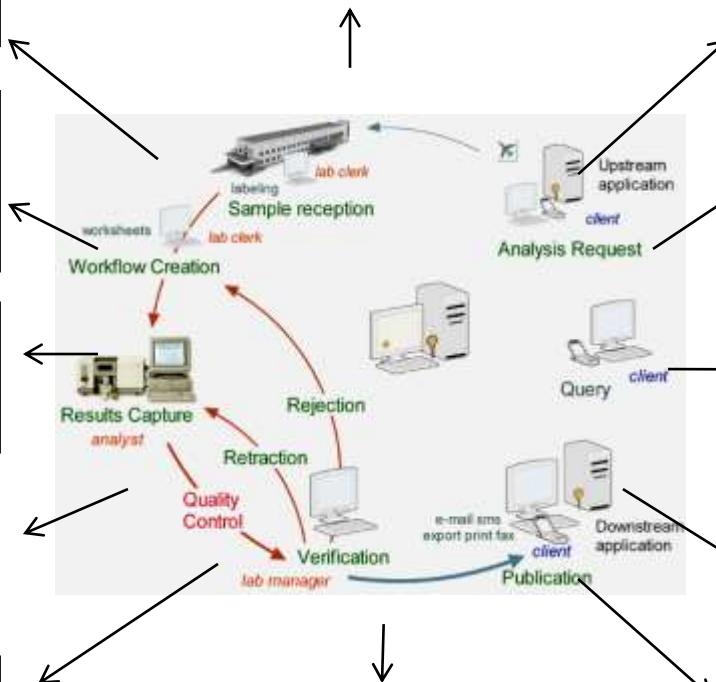
**Numune (Örnek)
İzleme**

Veri girişi

Veri raporlama

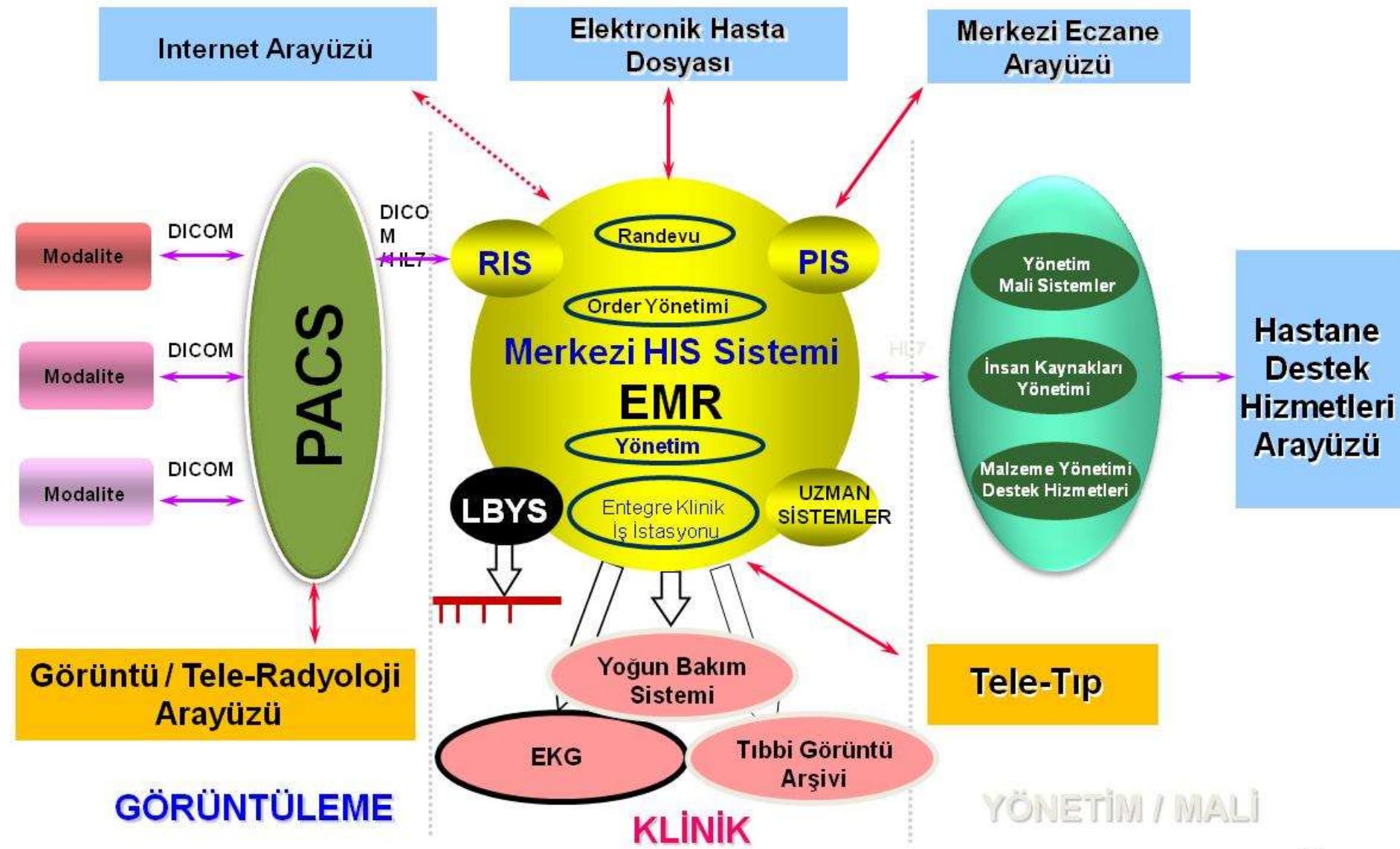
**Hata kaynakları
tespiti**

**Karar verme,
Araştırma !
Kod oluşturma
Depo ve kontrolü**

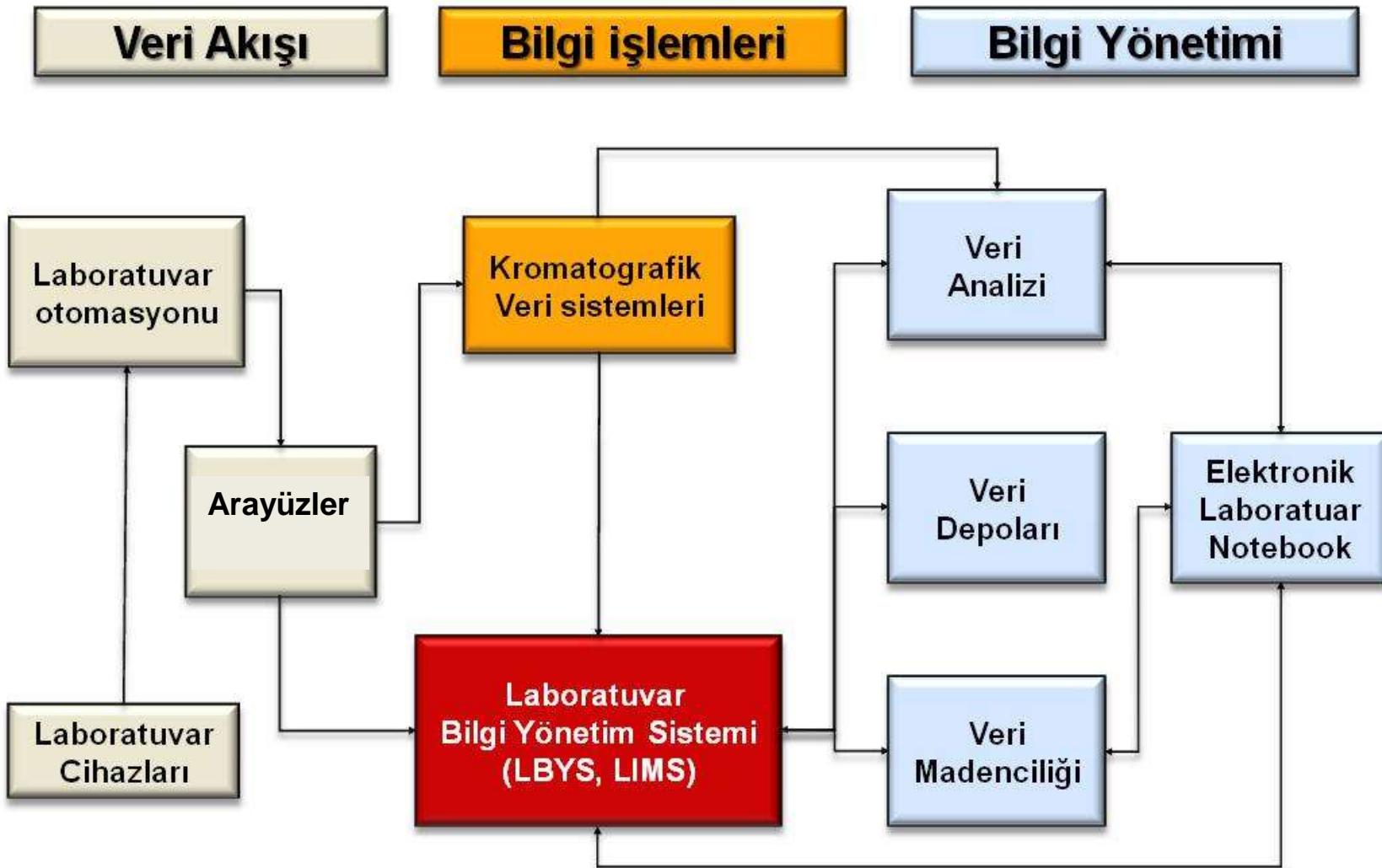


RAPOR

Hastane Bilgi Yönetim Sistemi (HBYS)



Laboratuvarlarda Bilgi Akışı



Niçin veriler toplanır ?

- Hasta ve toplum sağlığı için
- Değerlendirme yapmak için
- Kaliteyi artırmak için
- Maliyeti azaltmak için
- Karar verme sürecinde kullanmak
(veri madenciliği, araştırma vs)
- Denetlemek (veya cezalandırmak) için



Karar verme süreci

- Bir karar verici için verilen kararın doğruluğu,
 - yeteneklerine,
 - deneyimine,
 - bilgi birikimine,
 - veri kümesinin yeterliliğine bağlıdır
- Diğer bir deyişle kararın başarısında,
 - verilerin doğru depolanması,
 - doğru sınıflanması,
 - doğru ayıklanması,
 - doğru işlenmesi ve
 - doğru yorumlanması çok önemlidir.

Veri madenciliği

- Veri kendi başına degersizdir.
- Bir amaca yönelik işlenmiş veriye bilgi denir.
- **Veri madenciliği:**
 - Büyük miktarda veri içinden
 - tahmin yapmamızı sağlayacak bağıntı ve kuralların,
 - bilgisayar programları kullanılarak aranmasıdır.

*VM, beklenilen değil beklenilmedik ilişkileride önemser

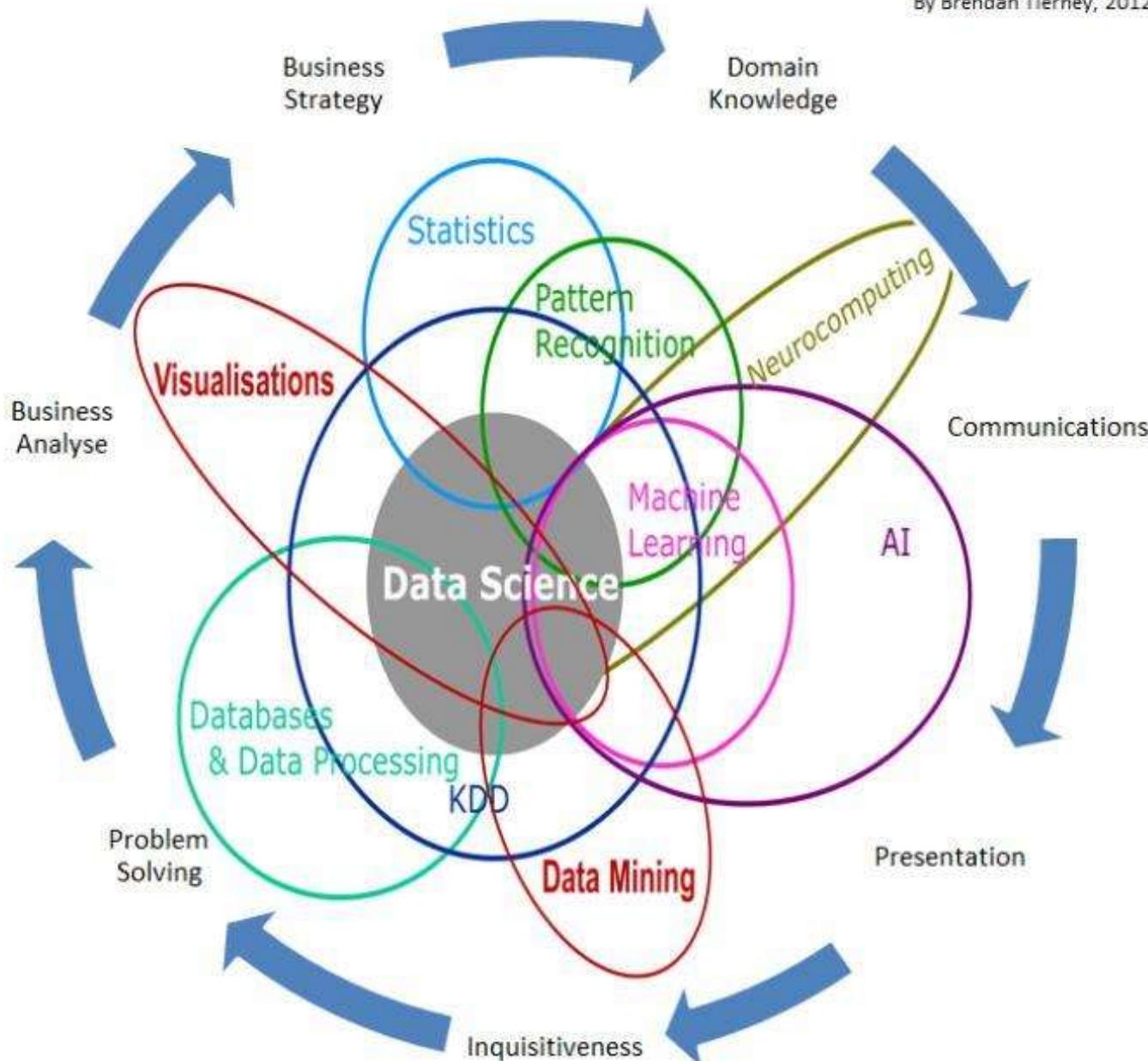


Veri madenciliği ortamı



Data Science Is Multidisciplinary

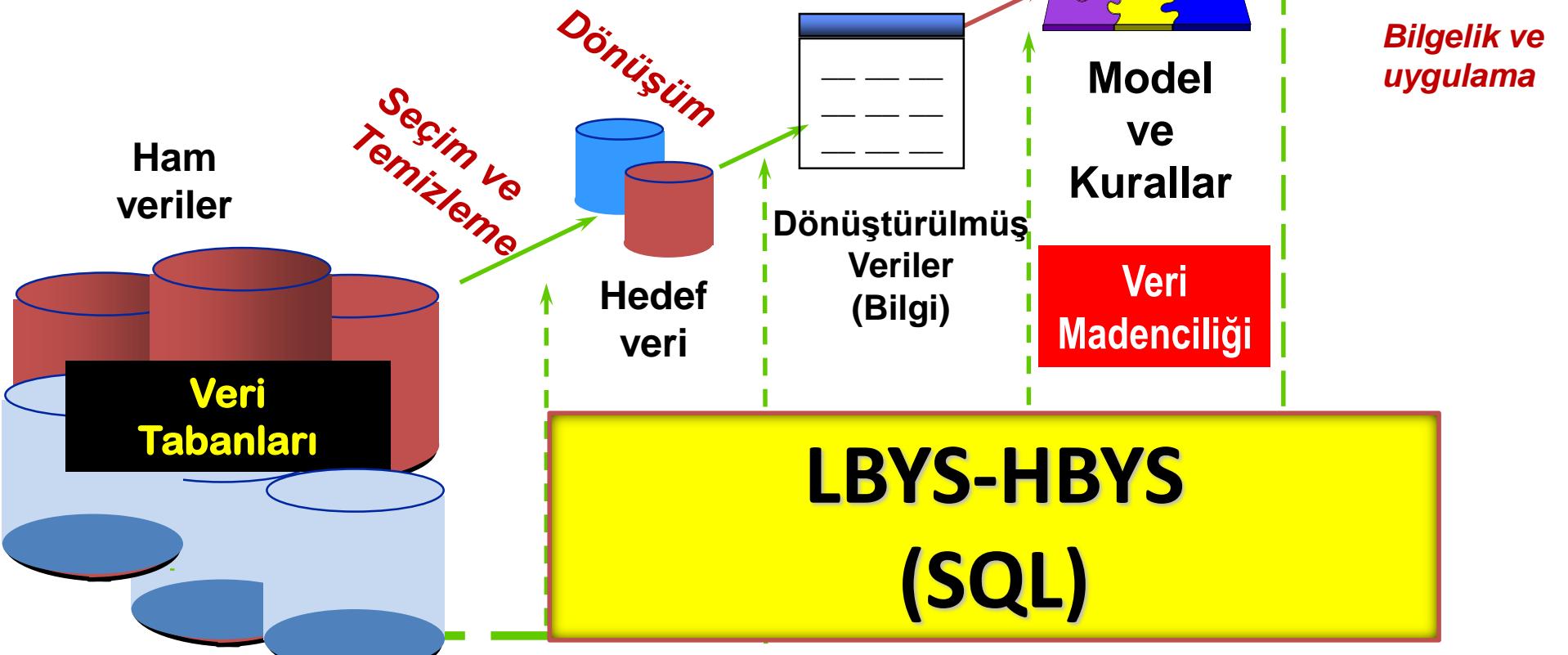
By Brendan Tierney, 2012



Veri madenciliği

verileri veri ambarından alır, bunları derler, düzenler ve yorumlar.

- ❖ **Veri Temizleme:** Gürültülü ve tutarsız verileri çıkarmak
- ❖ **Veri Bütünleştirme:** Birçok data kaynağını birlestirebilmek
- ❖ **Veri Seçme:** Yapılacak olan analiz ile ilgili olan verileri belirlemek
- ❖ **Veri Dönüşümü:** Verinin veri madenciliği yöntemine göre hale dönüşümünü gerçekleştirmek
- ❖ **Veri Madenciliği:** Verilerdeki örüntülerin belirlenmesi için veri madenciliği yöntemlerinin uygulanması
- ❖ **Örüntü Değerlendirme:** Bazı ölçütlerde göre elde edilmiş ilginç örüntüler bulmak ve değerlendirmek
- ❖ **Bilgi Sunumu:** Elde edilen bilgilerin kullanıcılara sunumunu



"Big Data" in Laboratory Medicine

Moderators: Nicole V. Tolan^{1*} and M. Laura Parnas²

Experts: Linnea M. Baudhuin,³ Mark A. Cervinski,⁴ Albert S. Chan,⁵ Daniel T. Holmes,⁶ Gary Horowitz,⁷ Eric W. Klee,⁸ Rajiv B. Kumar,⁹ and Stephen R. Master¹⁰

- Orta büyülükte bir laboratuvar yılda 3-4 milyon hasta sonucu üretir.
- Laboratuvar için büyük veri **çok veri** değil aynı zamanda **hızı, karakteri** çeşitliliği fazla olan veri
- **Genetik veri algoritmaları** biraz daha farklılık gösteriyor. Sadece total gen dizisi değil bunun klinike ilişkilendirilmiş olması gereklidir
- **Hasta bakımına katkı temel odak** (örneğin vankomisin tedavisinin TAT etkiliyor mu?)
- Kalite değerlendirilmesinde önemli (TAT, hata oranı kritik değerlerin oranı, MA gibi)
- **LBYS** yapılandırılması kritik önemli
- Genellikle **R** yazılımı üzerinde duruyorlar
- **Çok veri hasta sonuçlarını iyileştirecek anlamına gelmez.** Temel gürültü çoktur ve bundan verinin temizlenip düzgün algoritmalar kurmak kolay değildir.



Review

Clinical chemistry in higher dimensions: Machine-learning and enhanced prediction from routine clinical chemistry data

Alice Richardson ^a, Ben M. Signor ^b, Brett A. Lidbury ^b, Tony Badrick ^{b,c,*}

^a National Centre for Epidemiology & Population Health, Australian National University, 62 Mills Rd, Acton, ACT 2601, Australia

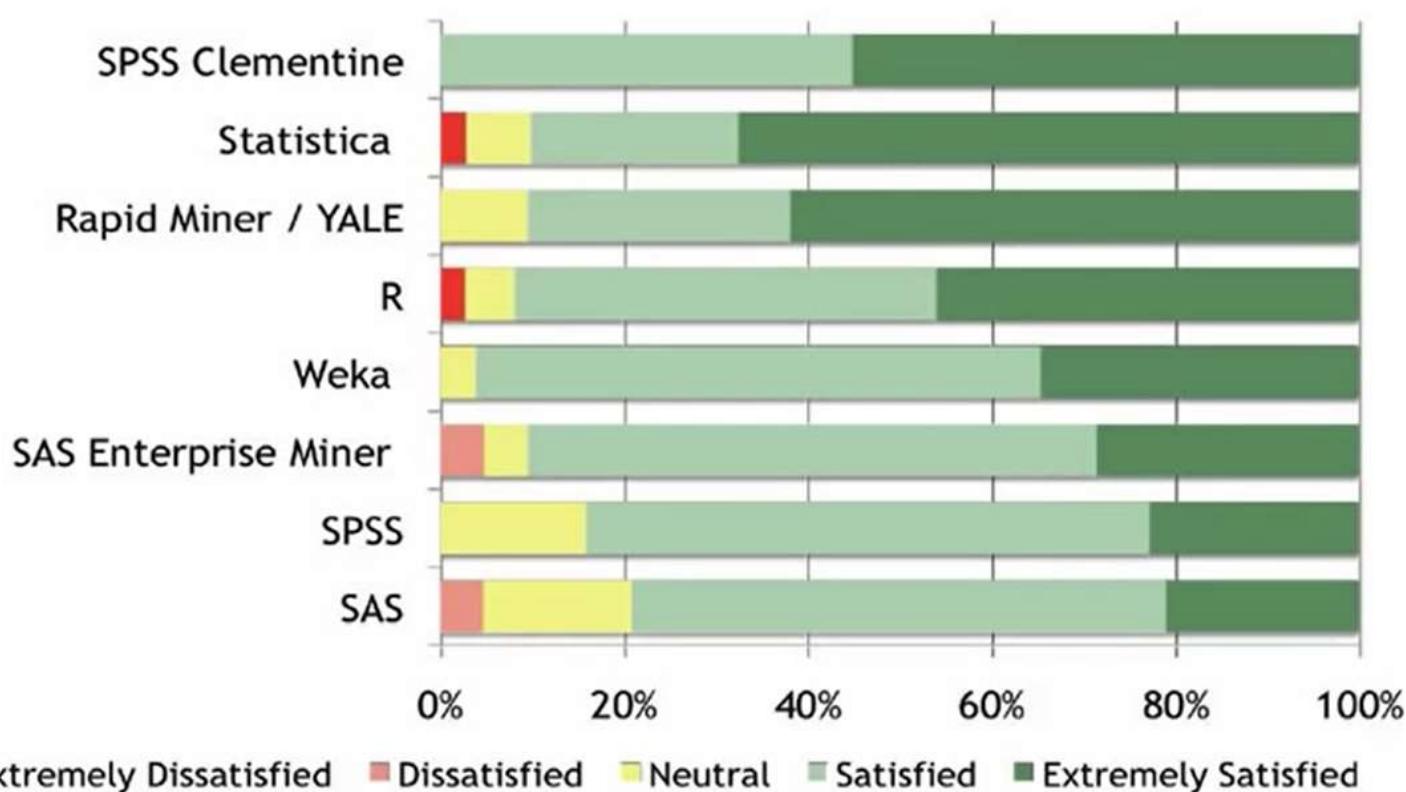
^b Pattern Recognition and Pathology, John Curtin School of Medical Research, Australian National University, 62 Mills Rd, Acton, ACT 2601, Australia

^c RCPAQAP, Suite 201/8 Herbert Street, St Leonards, NSW 2065, Australia

The seven myths
Myth 1. Big data is universally big
Myth 2. Big data means never having to say what your research question is
Myth 3. Big data means never having to say what your model is
Myth 4. Big Data means never having to consider sampling theory, a standard error, or a p-value
Myth 5. Big data means more valuable information
Myth 6. Big data means observational data can be used to measure causal relationships
Myth 7. Classical statistical methods are inadequate to deal with big data

Veri madenciliği yazılımları

SATISFACTION



Veri madenciliği yazılımları

What is your primary analytics platform?

Poll Results (single answer required):



What Analytics, Data mining, Big Data software you used in the past 12 months for a real project (not just evaluation) (708 users)		
Software	% users in 2012	% users in 2011
R (245)	30.7%	23.3%
Excel (238)	29.8%	21.8%
Rapid-I RapidMiner (213)	26.7%	27.7%
KNIME (174)	21.8%	12.1%
Weka / Pentaho (118)	14.8%	11.8%
StatSoft Statistica (112)	14.0%	8.5%
SAS (101)	12.7%	13.6%
Rapid-I RapidAnalytics (83)	10.4%	Not asked in 2011
MATLAB (80)	10.0%	7.2%
IBM SPSS Statistics (62)	7.8%	7.2%
IBM SPSS Modeler (54)	6.8%	8.3%
SAS Enterprise Miner (46)	5.8%	7.1%

XLSTAT Your data analysis solution
Version 2016.1
New color themes
Enhanced Gage R&R for attributes and much more

Order an Upgrade now!

Modeling data Machine learning Correlation/ Association tests Parametric Non tests Machine learning Correlation/ Association tests Parametric Non tests

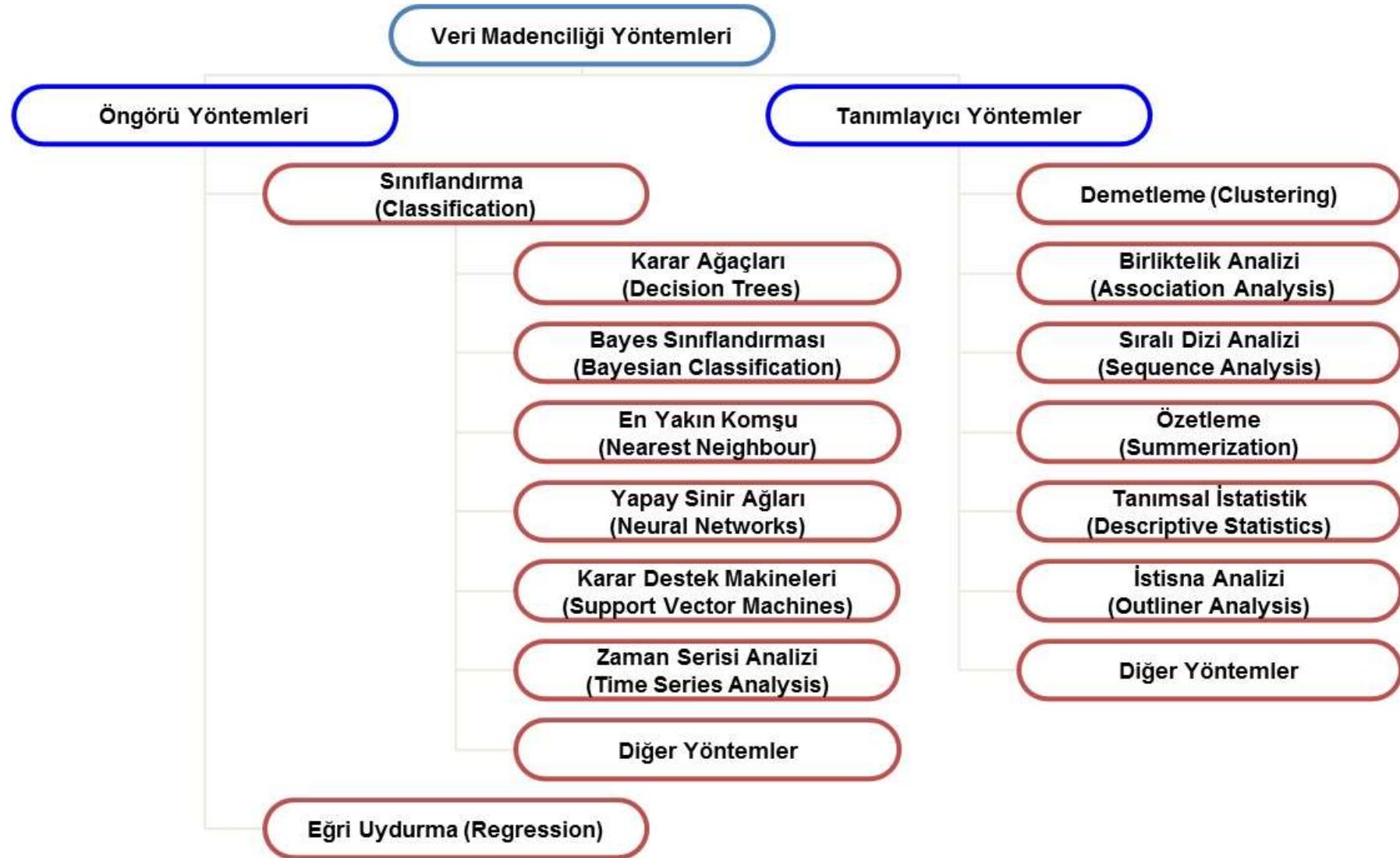
- Distribution fitting
- Linear regression
- ANOVA
- ANCOVA
- Repeated measures ANOVA
- Mixed Models
- MANOVA
- Logistic regression
- Log-linear regression
- k-means clustering
- Gaussian Mixture Models
- Classification and regression trees
- K Nearest Neighbors
- Naive Bayes classifier
- Association rules

Sensory data Conjoint Time series Monte Carlo Power Statistical Process Control (SPC) Design Of Experiments (DoE) Survival Method Dose effect OMICs data Multiblock data PLS Path Modeling

Quantitative data: Workbook = mg pth vtd.xls / Sheet = cucuk / Range = cucuk!\$F:\$K / 182 rows and 6 columns

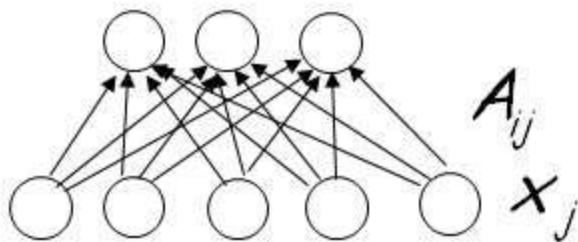
		Statistic	Minimum	Maximum	Range	1st Quartile	Median	3rd Quartile	Mean	Standard deviation (n)
1.33	Vitamin D.	4.0000	91.0000	87.0000	14.7250	20.6500	28.6000	23.2511	13.7978	
1.41	Kalsiyum	6.3000	11.5000	5.2000	9.1000	9.6000	10.0000	9.5736	0.7247	
1.49	Fosfor. İnc	1.8500	7.9200	6.0700	4.3700	4.8200	5.4000	4.8674	0.8835	
1.51	Alkalen Fo	47.0000	884.0000	837.0000	159.0000	212.0000	256.7500	222.4725	111.3416	
1.59	Parathorm	2.3000	79.6700	77.3700	21.0925	32.1500	46.2450	35.0585	16.8073	
1.62	Magnezyu	1.1100	2.8400	1.7300	1.9300	2.0750	2.2000	2.0643	0.2469	
1.63										
1.64										
1.66										
1.68										
1.71										
1.71										
1.72										

Veri Madenciliği Yöntemleri

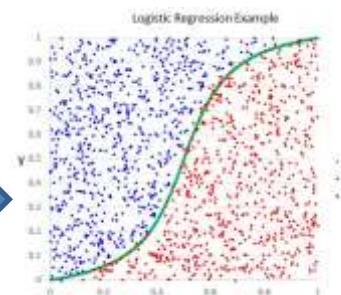
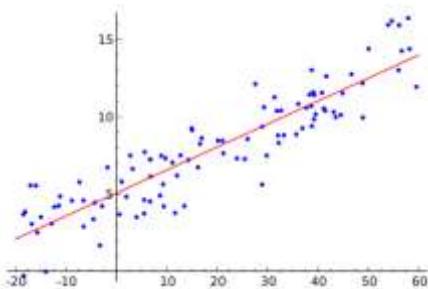


Regresyon

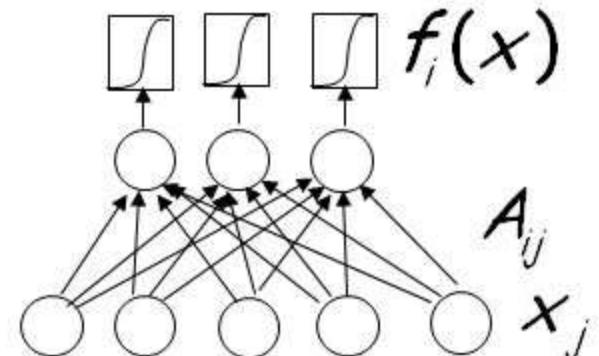
linear regression



$$Y_n = AX_n + b$$

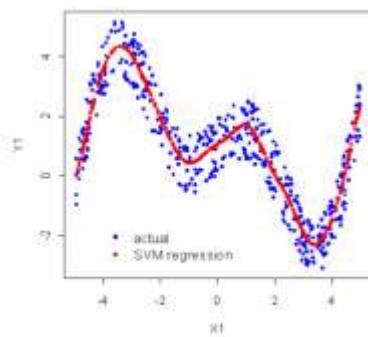


logistic regression



$$P(Y_n = 1 | X_n) = f(AX_n + b)$$

$$f(X) = \frac{1}{1 + \exp[-(AX + b)]}$$

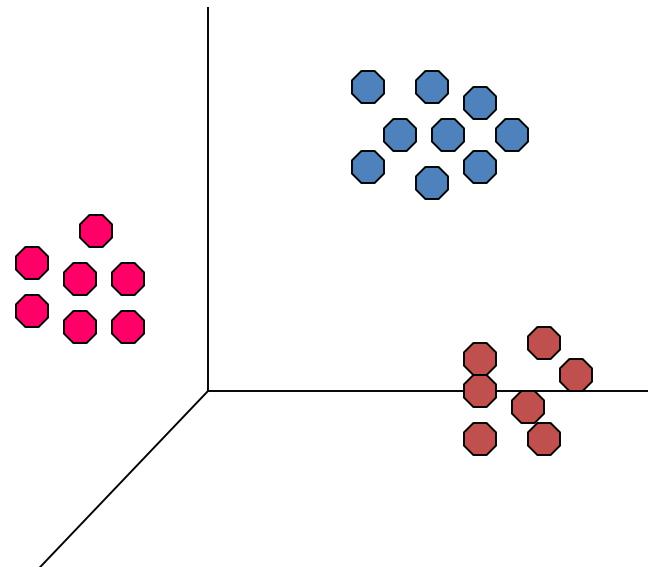


Kümeleme analizleri

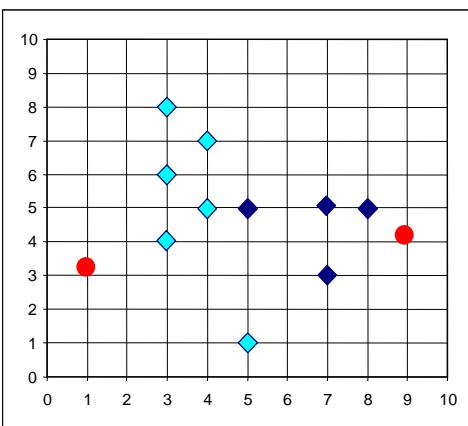
- ✗ 3-Boyutlu uzayda öklid uzaklığuna dayanan kümeleme.

Küme içi uzaklıklar
minimize edilir.

Kümeler arası uzaklıklar
maksimize edilir.



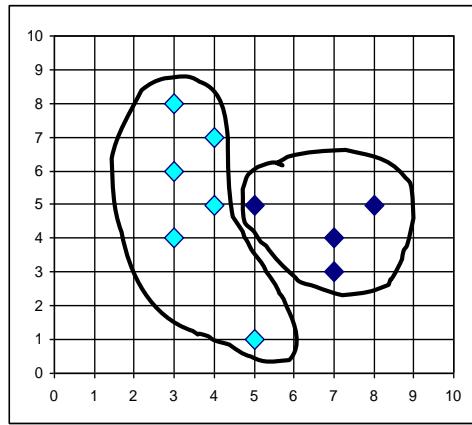
K-Ortalama(Means) Kümeleme Yöntemi



K=2

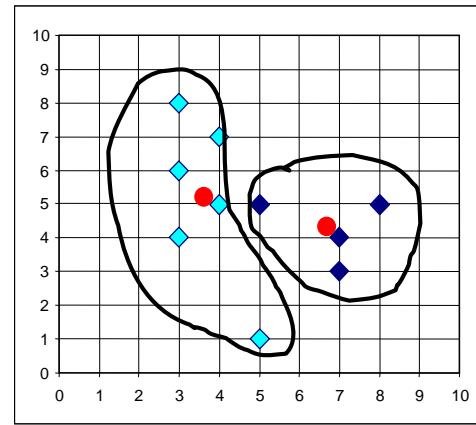
Arbitrarily choose K
object as initial
cluster center

Assign
each
objects
to most
similar
center



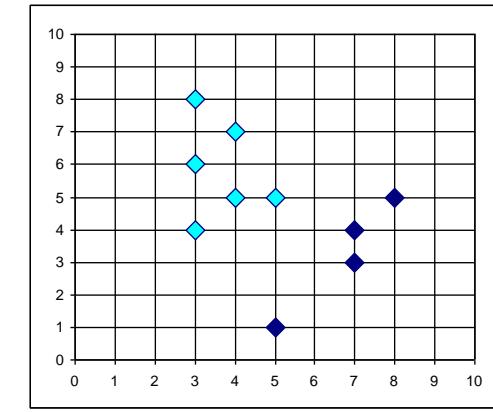
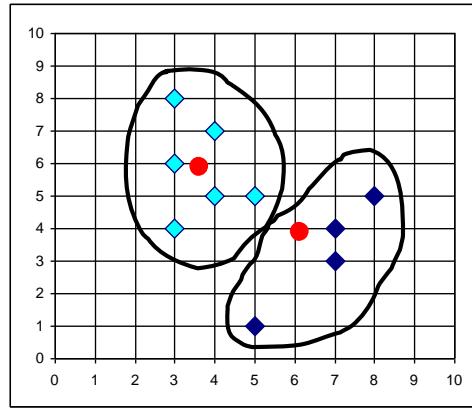
reassign

Update
the
cluster
means



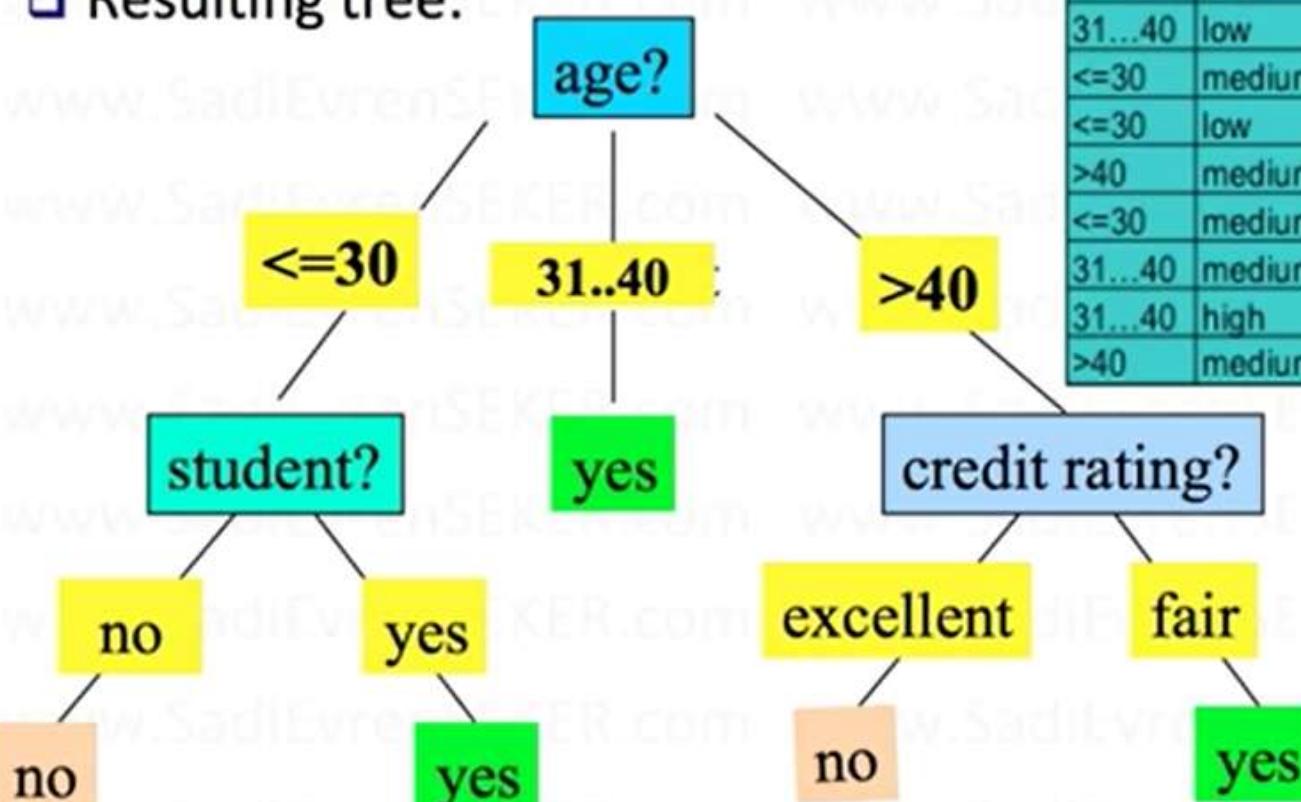
reassign

Update
the
cluster
means



Decision Tree Induction: An Example

- Training data set: Buys_computer
- The data set follows an example of Quinlan's ID3 (Playing Tennis)
- Resulting tree:



age	income	student	credit rating	buys comp
≤ 30	high	no	fair	no
≤ 30	high	no	excellent	no
31..40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31..40	low	yes	excellent	yes
≤ 30	medium	no	fair	no
≤ 30	low	yes	fair	yes
>40	medium	yes	fair	yes
≤ 30	medium	yes	excellent	yes
31..40	medium	no	excellent	yes
31..40	high	yes	fair	yes
>40	medium	no	excellent	no

Identification of Asthma Phenotypes Using Cluster Analysis in the Severe Asthma Research Program

Wendy C. Moore^{1,2}, Deborah A. Meyers^{1,2}, Sally E. Wenzel², W. Gerald Teague², Huashi Li¹, Xingnan Li¹, Ralph D'Agostino, Jr.³, Mario Castro², Douglas Curran-Everett², Anne M. Fitzpatrick², Benjamin Gaston², Nizar N. Jarjour², Ronald Sorkness², William J. Calhoun², Kian Fan Chung², Suzy A. A. Comhair², Raed A. Dwe², Elliot Israel², Stephen P. Peters^{1,2}, William W. Busse², Serpil C. Erzurum², and Eugene R. Bleecker^{1,2}, for the National Heart, Lung, and Blood Institute's Severe Asthma Research Program^{2*}

¹Wake Forest University School of Medicine, Center for Human Genomics; ²The Severe Asthma Research Program, Bethesda, Maryland; and ³Wake Forest University School of Medicine, Public Health Sciences, Winston-Salem, North Carolina

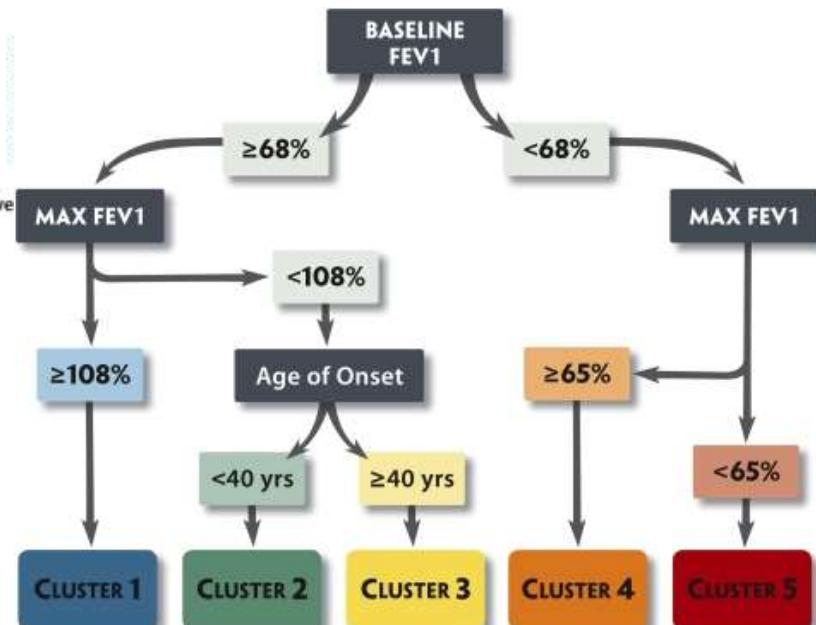


TABLE 1. DEMOGRAPHICS AND CLINICAL CHARACTERISTICS OF SUBJECTS

	Total Cohort	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	P Value*
Number of subjects	726	110	321	59	120	116	
Age at enrollment, years	37 (14) [†]	27 (8)	33 (12)	50 (8)	38 (13)	49 (11)	<0.0001
Sex, % female	66	80	67	71	53	63	0.0006
Race, % White/AA/other	64/28/8	62/29/9	63/30/7	73/22/5	62/33/5	68/20/12	0.17
Body mass index (BMI)	29 (8)	27 (5)	28 (8)	33 (9)	31 (9)	31 (7)	<0.0001
% with BMI >30	37	24	31	58	44	51	<0.0001

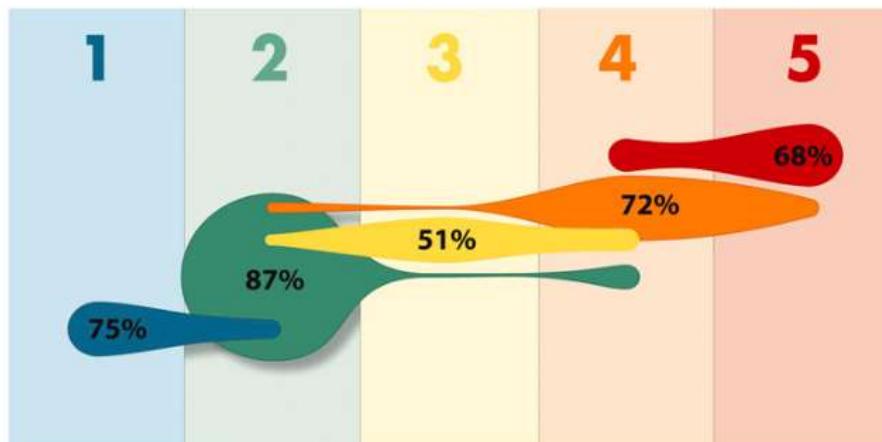


Figure 2. Tree performance. Using the algorithm generated by the tree analysis, 80% of subjects are assigned to the correct cluster of asthma severity. Colors are maintained from the tree diagram (blue = mild atopic asthma; green = mild to moderate atopic asthma; yellow = late-onset nonatopic asthma; orange = severe atopic asthma; red = severe asthma with fixed airflow). Individual figure size is proportional to the frequency of a specific cluster. The percentage of subjects from that cluster that are correctly assigned is indicated numerically within the shape.

Hierarchical clustering

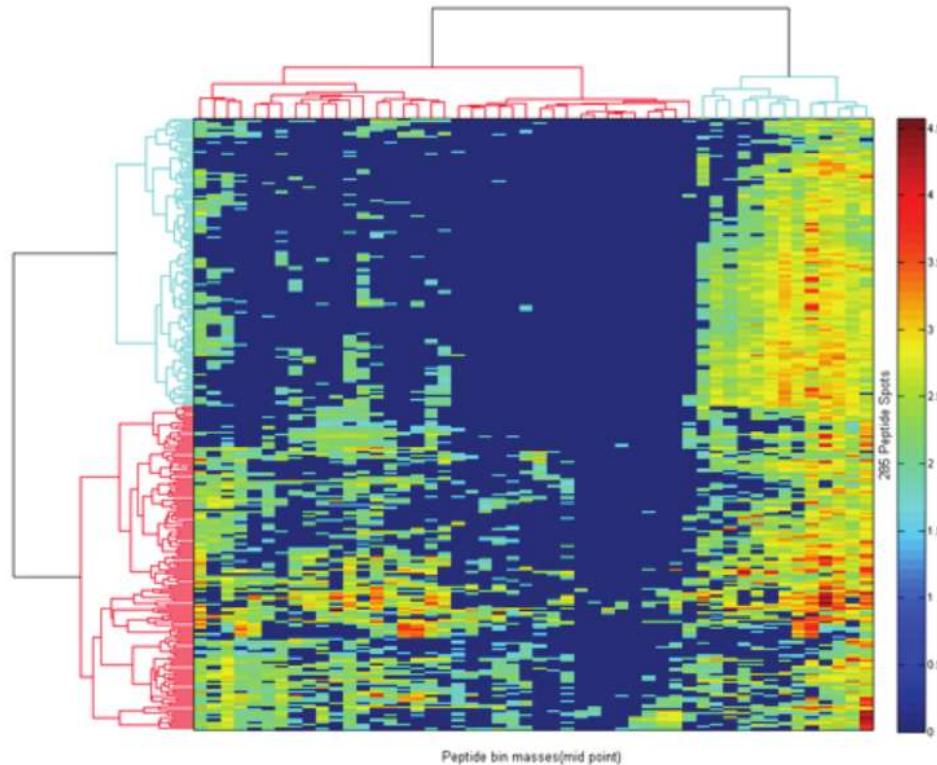
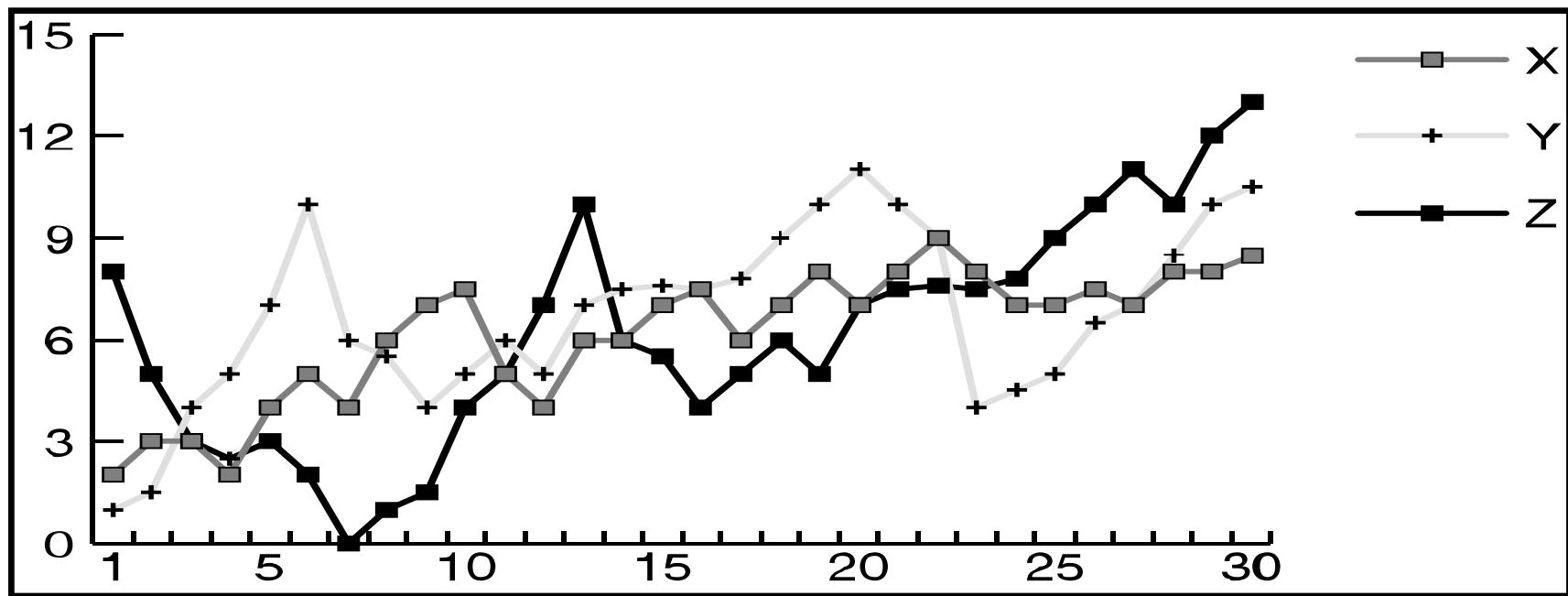


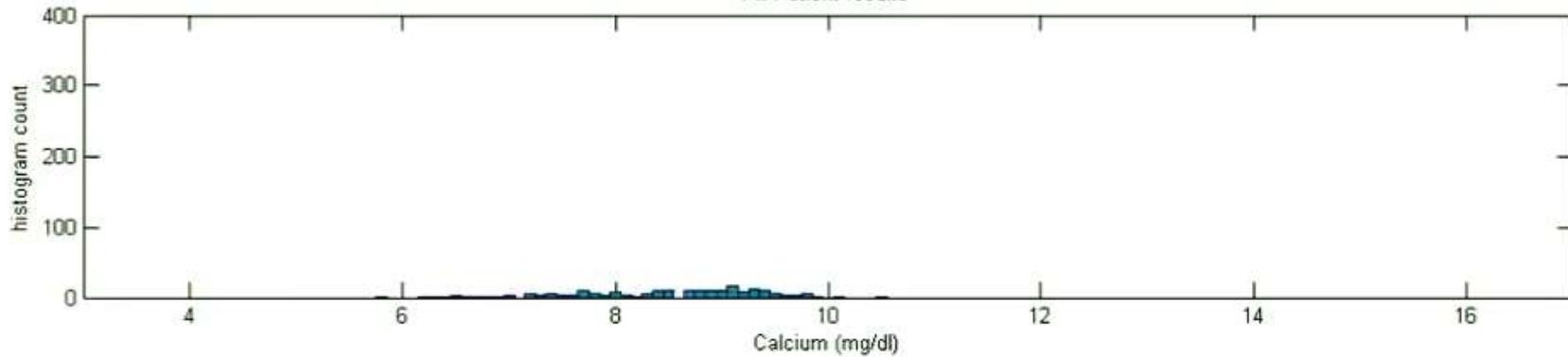
Fig. 3. Heatmap visualization of a hierarchical clustering result. Two well separated protein clusters (horizontal dendrogram) and two well separated bin clusters (vertical dendrogram) can be observed at the top-level. Figure adapted from (Ventoura et al., 2007).

Zaman Serisi Analizi

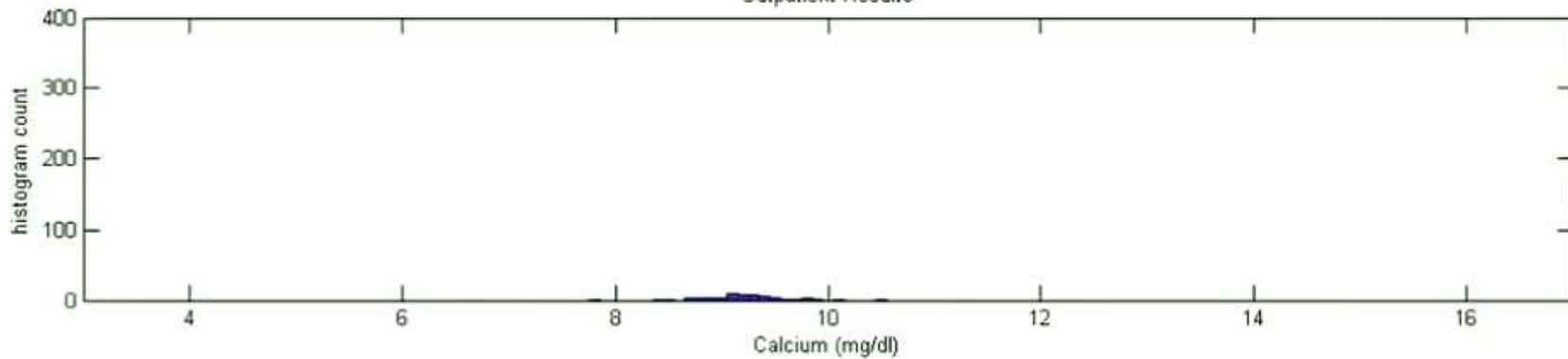
- Örnek: MA, EWMA



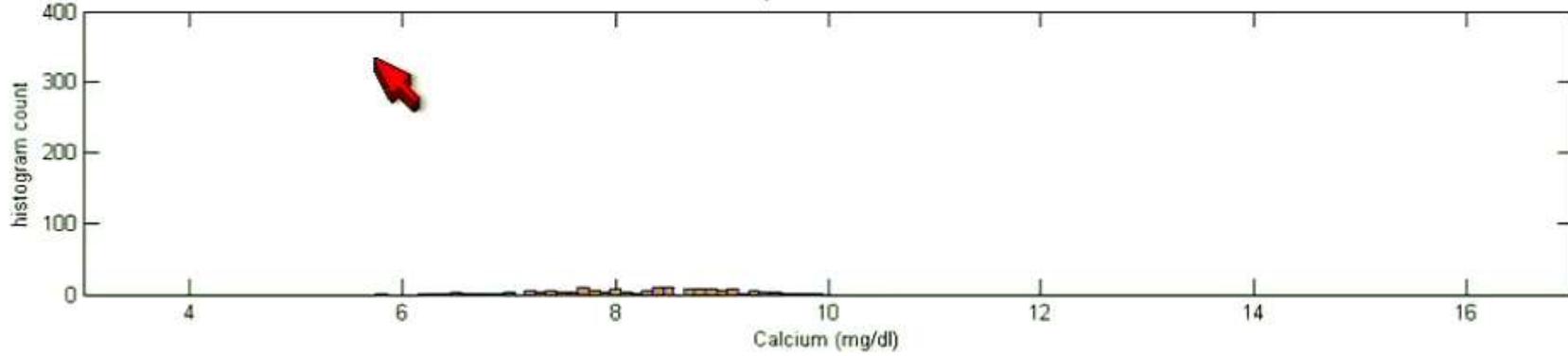
Hours (starting from midnight): 00:00
All Patient results



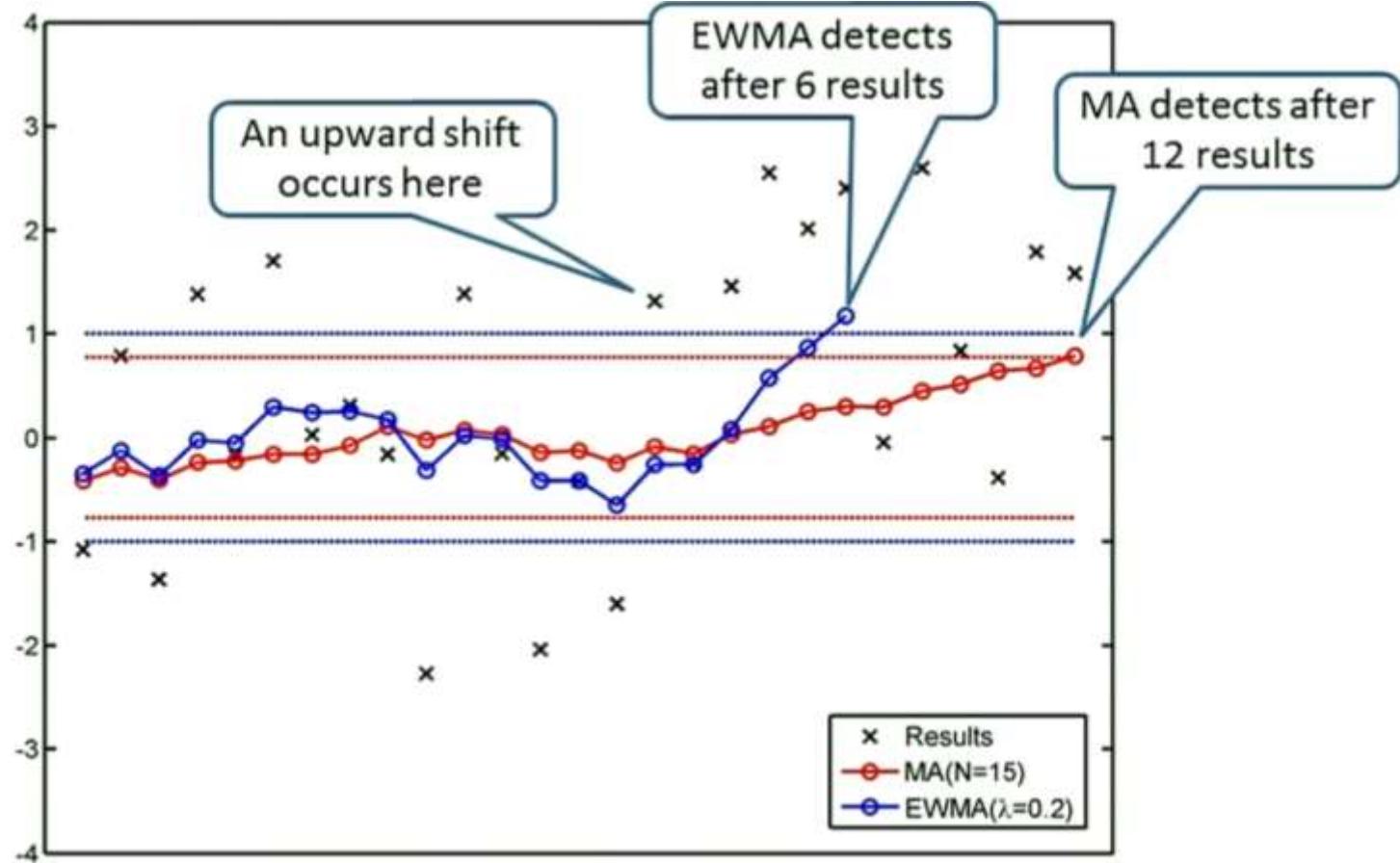
Outpatient Results



Inpatient Results

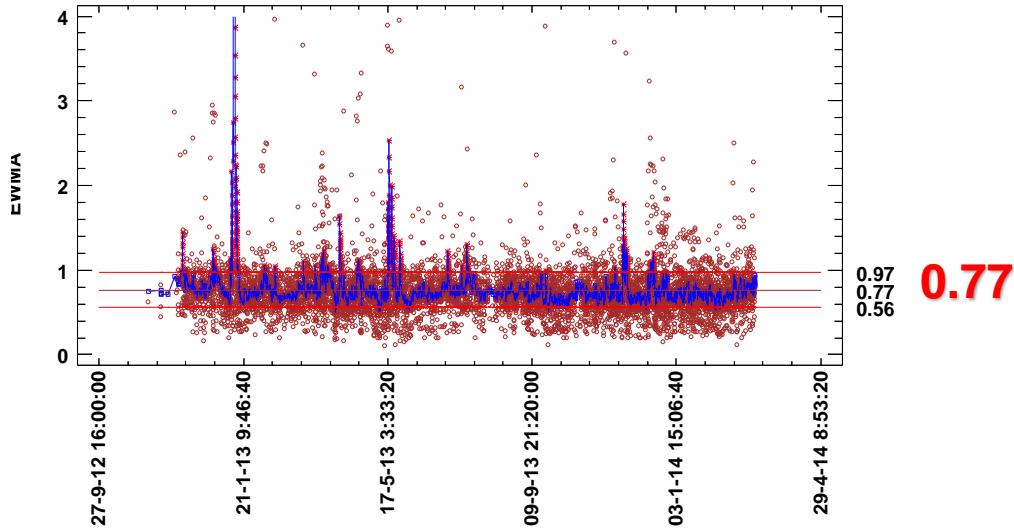


Hareketli ortalama-Moving Avarage (MA) Exponantially weighted moving avarage (EWMA)

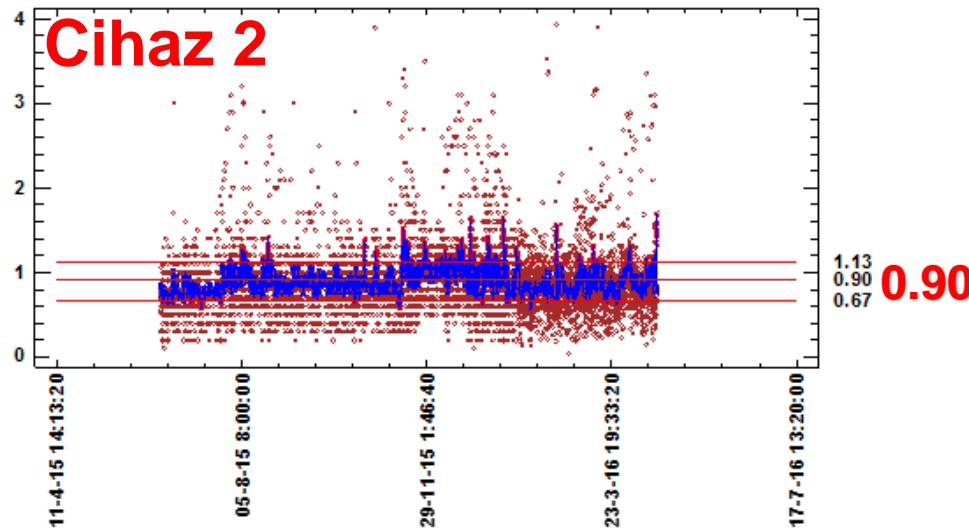


Günlük EWMA Uygulamalarımız

Cihaz 1

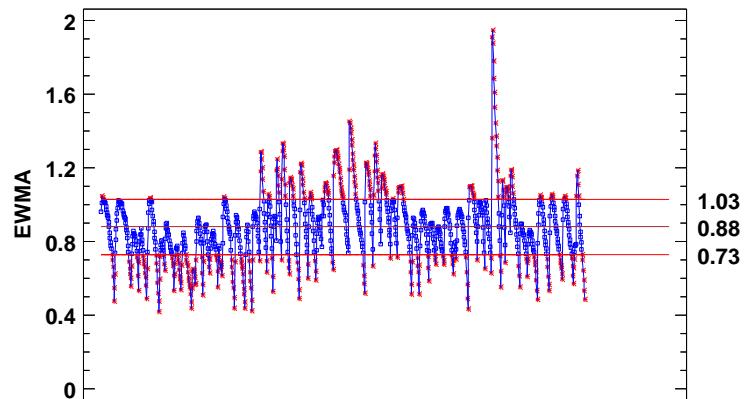


Cihaz 2

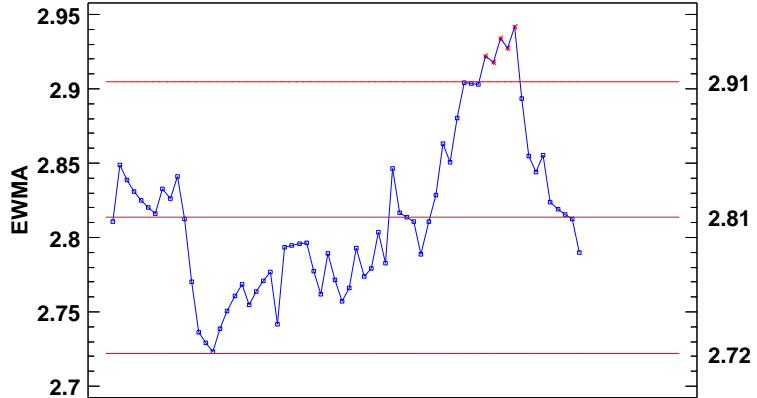


Cihaz 1

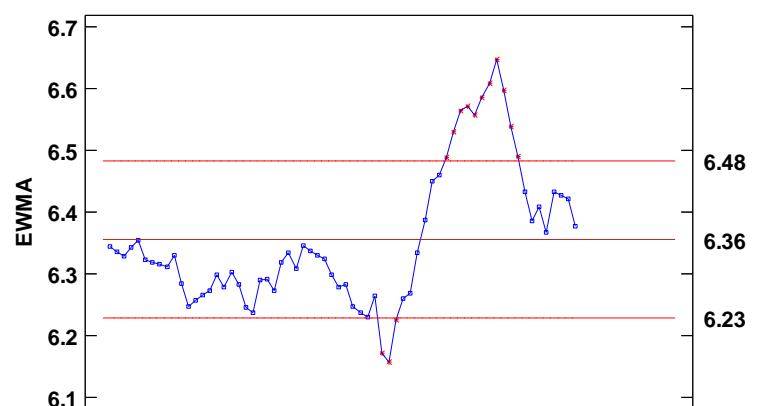
DÜZYEY 1 KONTROL HASTA SONUÇLARI



DÜZYEY 1 KONTROL

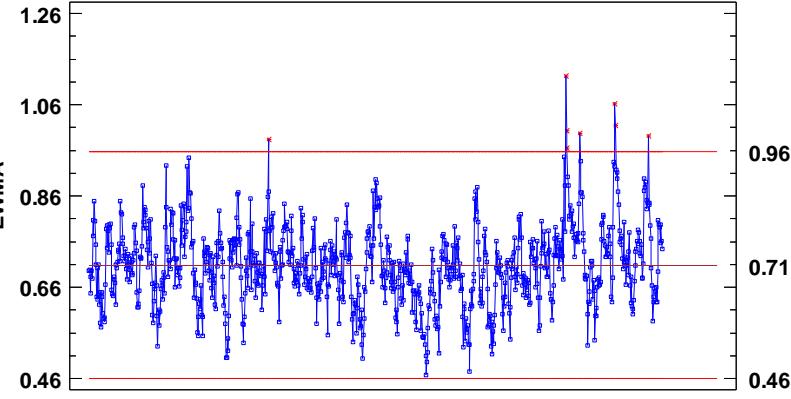


DÜZYEY 2 KONTROL

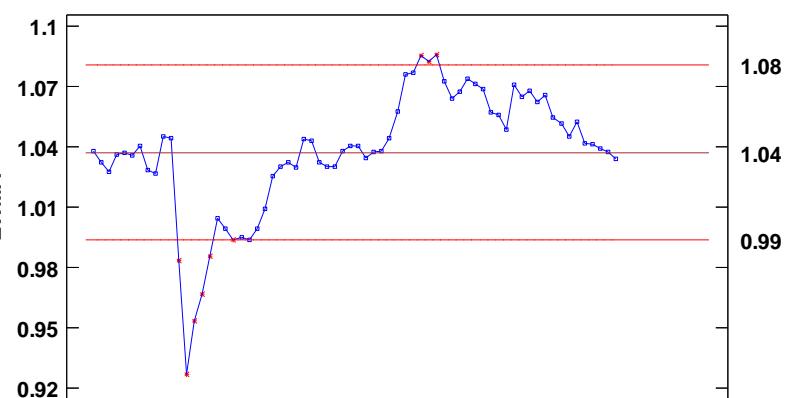


Cihaz 2

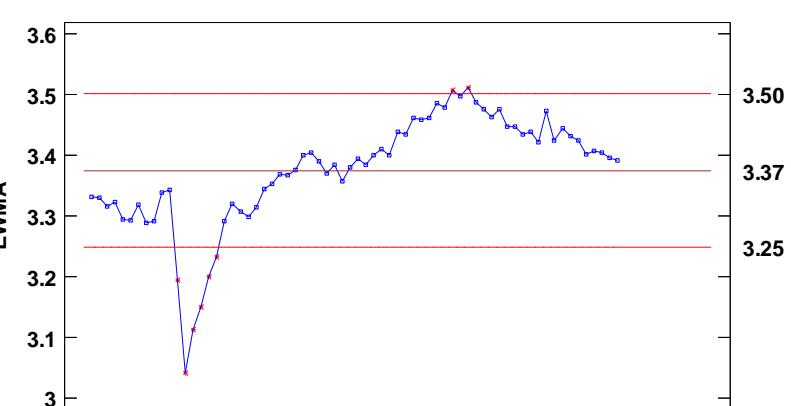
EWMA



EWMA

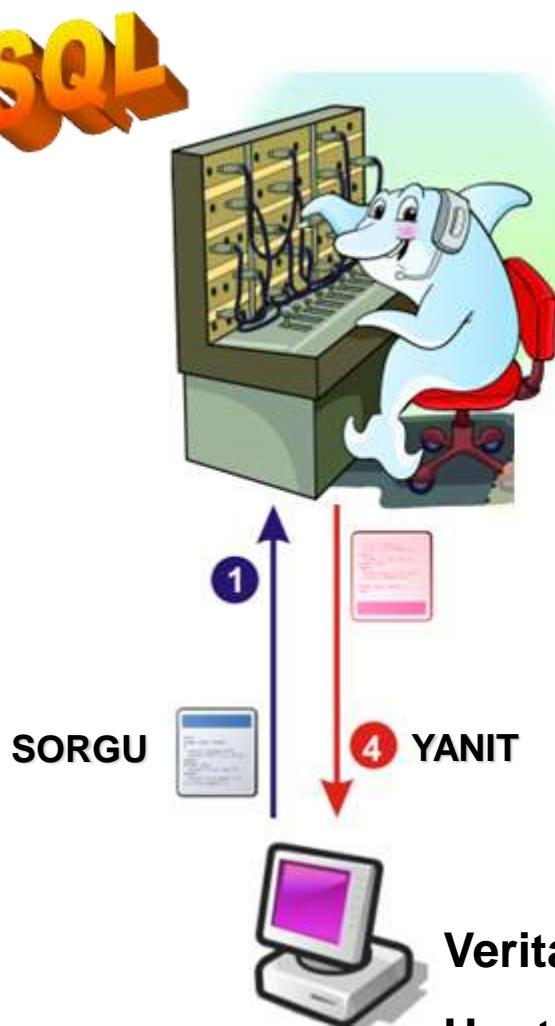


EWMA



Veri tabanları ve SQL

SQL



SQL

SELECT * FROM
WHERE
DELETE
UPDATE
INSERT
GROUP BY
JOIN
HAVING
COUNT
SUM
AVG
VS

Veritabanı bir yada birden fazla tablodan oluşur.

Her tablonun bir adı vardır.

Bu tablolardaki kayıtlar dizilerde saklanır.

SQL çalışmalarında veriler kimiksizleştirilmek zorundadır

SQL- export format

 [Export Wizard]

Step 1 of 5

The wizard allows you to specify details of how to export your data. Which export format would you like?

Table Type

- Paradox file (*.db)
- QuattroPro file (*.wq1)
- DBase file (*.dbf)
- SQL script file (*.sql)
- Text file (*.txt)
- XML file (*.xml)
- HTML file (*.htm; *.html)
- MS Access database (*.mdb)
- Excel spreadsheet (*.xls)
- MS Windows Clipboard
- Excel file (*.xls)
- Rich Text format (*.rtf)
- Word file (*.doc)
- SPSS format (*.sav)
- SYLK (Symbolic Link) (*.slk)
- Adobe Acrobat Document (*.pdf)
- DIF (Data Interchange Format) (*.dif)
- LDAP DataInterchangeFormat (*.ldif)
- Lotus 1-2-3 file (*.wk1)

Save

<< < Back Next > >> Cancel

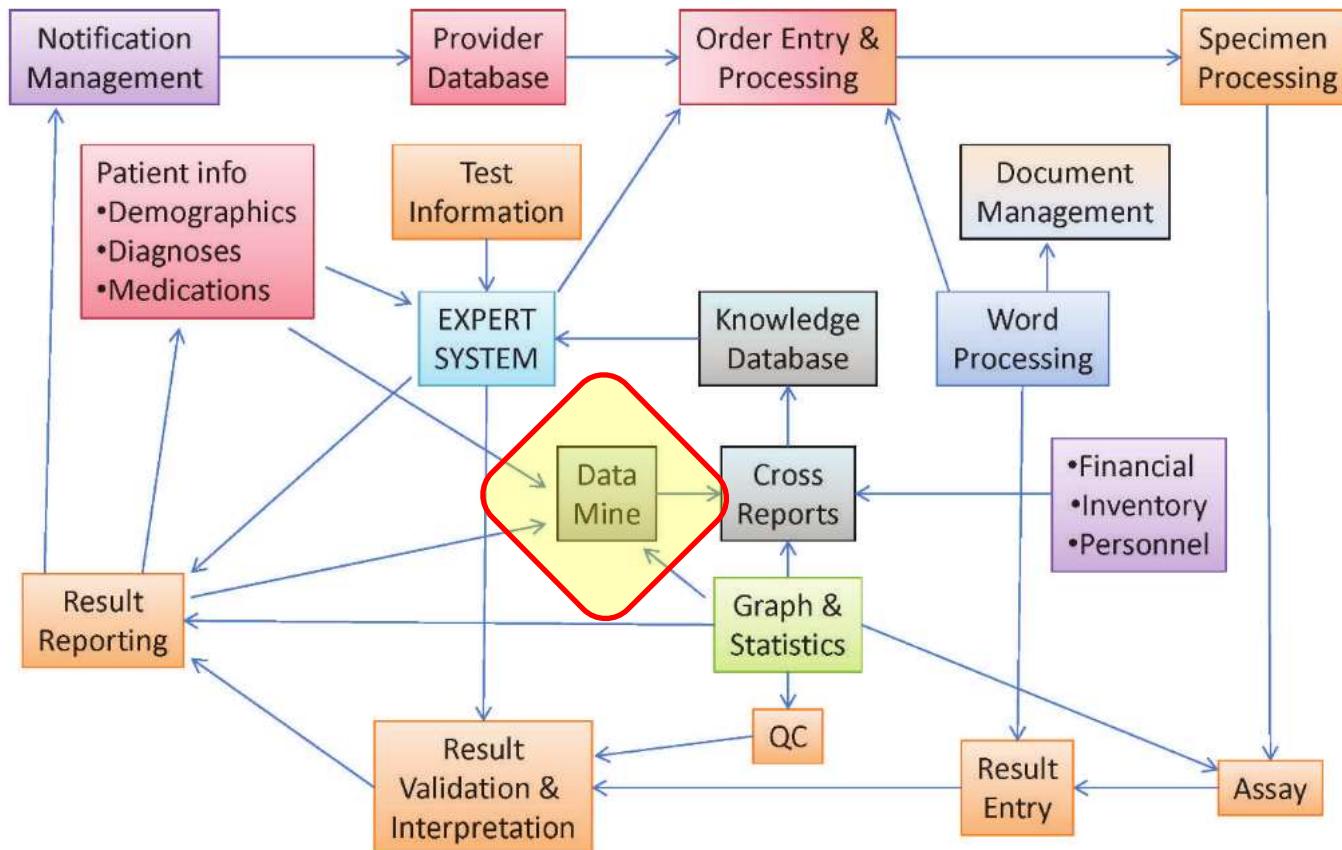
[Örnek](#) [Mikrobiyoloji](#) [Raporlar](#)[İstatistik](#)[Sistem](#)[Yönetici](#)[Diğer](#)

Hoşgeldiniz

[Tek Test İzle](#)[Test tekrar oranları](#)[Kısmi İşlem Yapılmış Örnekler](#)[Test Sayıları](#)[Dağılımlar](#)[Örnek Dağılım](#)[Onay süreleri](#)[Test Sonuç Süreleri](#)[Ret sebeplerine göre dağılım](#)[Bölümlere göre hasta sayıları](#)[Test gruplarına göre örnek sayıları](#)[Test gruplarına göre örnek türü dağılımı](#)[Durumlara göre hasta sayıları](#)[En Çok Gelenler](#)[Özel Sorgu](#)[Genel Hizmet Dökümü](#)[Örnek Kabulden Geçmeyen Hastalar](#)

The Ideal Laboratory Information System

Jorge L. Sepulveda, MD, PhD; Donald S. Young, MD, PhD



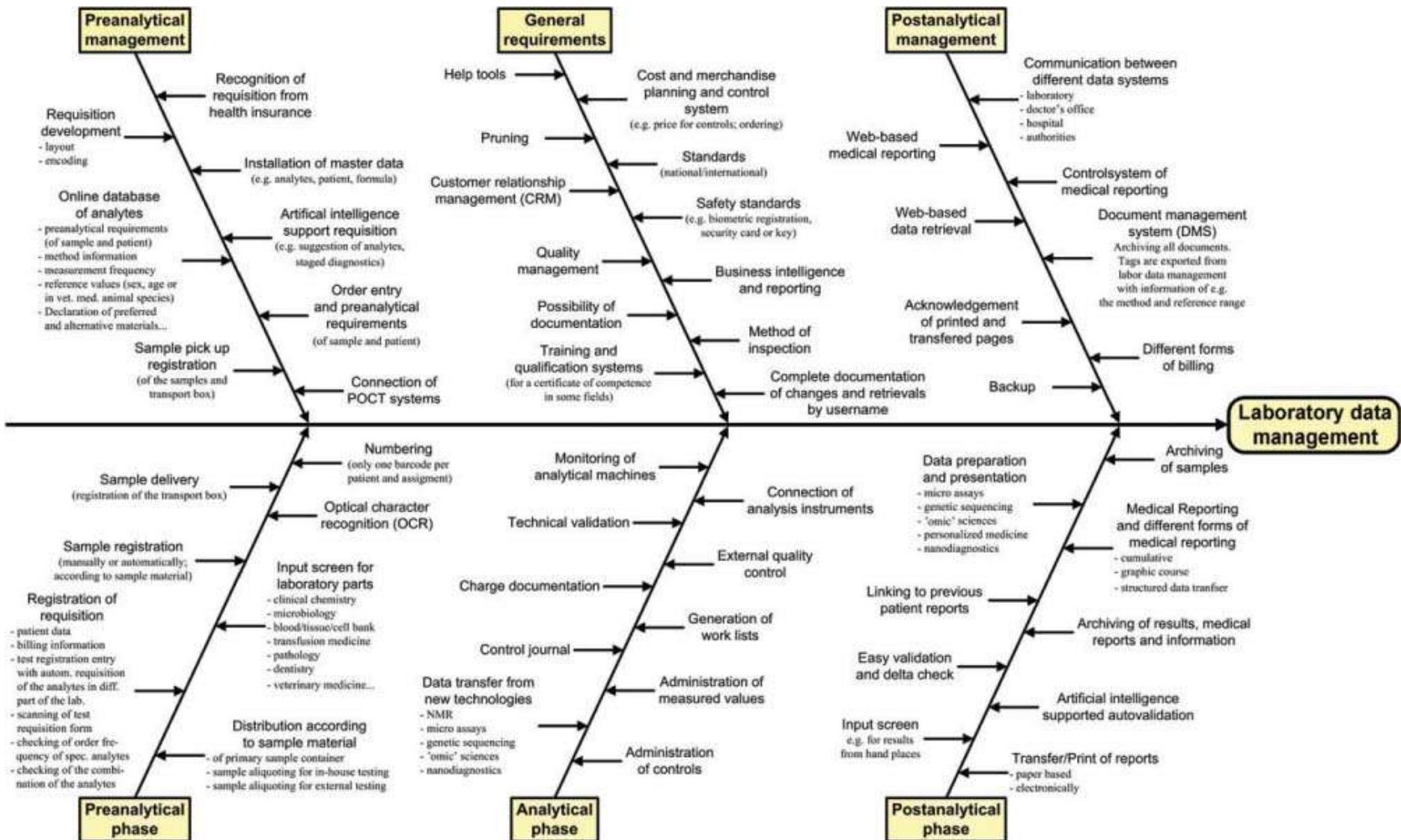
Modules contributing to the ideal laboratory information system.

The Ideal Laboratory Information System

Jorge L. Sepulveda, MD, PhD; Donald S. Young, MD, PhD

Advanced data warehouse and mining capabilities should be available in an advanced LIS. Examples of useful queries and reports include the following:

1. Search functions for combinations of laboratory results and clinical information, such as diagnoses, medications, and treating specialty, producing reports with user customizable display of queried and nonqueried fields such as patient demographics, specimen accession data, and location.
2. Laboratory testing turnaround time reports with the ability to consolidate or split the various components, such as order to collection, collection to receipt in laboratory, receipt to testing, and testing to report, and the capability to group by accession areas, individual test or test groups, hours or shifts, employee, patient location, clinics, providers, etc.
3. Surveillance data online reporting to public health agencies in their required format
4. Nosocomial infection tracking and antibiograms reporting the frequency distribution of microbial susceptibility to antimicrobial agents.
5. Laboratory utilization reports by provider, provider group, specialty, clinics, wards, patient types, diagnoses and diagnostic groups, ICD-9/10 codes, etc, to include test type, volumes, and costs per case.
6. Patient outcome analysis using laboratory data-mining capabilities and clinical data extracted from the HIS. Examples of useful parameters to correlate with laboratory testing include mortality, morbidity, hospital length of stay, and cost of care, grouped by diagnostic groups.





Özel Sorğu

Yeni Sorğu F3 Bul F2 Excel Yaz | Sorğu Kaydet Sorğu Seç F5 Kapat

Bul

Sonuç

Tarih Aralığı

Baş: 21.09.2000 00:00

Son: 21.09.2016 23:59

Cinsiyet

- Hepsi
- Yalnız Erkekler
- Yalnız Bayanlar

Rutin/Acıl

- Hepsi
- Yalnız Acıl Örnekler
- Yalnız Rutin Örnekler

Durum

- Hepsi
- Yatan
- Ayaktan

Sorgu Adı:

Birimler

Yaşlar:

Test Grupları

- Biyokimya
- Hematoloji
- Mikrobiyoloji



Testler

 Tüm sonuçlar Tüm testler aynı anda olmalı (and) Test Seç

Alanin Aminotransferaz (ALT)

Kolesterol

Örnek Türleri

Kurumlar

Tanılar

Seçili birimler için listeleme şartı

- Birimin çalıştığı testleri getir
- Birime ait hastaların testlerini getir

Sorguda listlenecek alanlar

- Kimlik No
- Cinsiyet
- Doğum Tarihi
- Yaş
- ÖrnekNo
- Tanı
- Sonuç Tarihi
- Sonuç Açıklama
- Örnek Türü
- Doktor
- Tarih
- Bölüm

Microsoft Excel - hasta_glikoz_misol_ugit_2635 hasta super.xls [Compatibility Mode] - Microsoft Excel + Analyse-it®

File Home Insert Page Layout Formulas Data Analyse-it View Add-ins ABBYY FineReader YI Power Pivot Team XLSTAT Tell me what you want to do...

Cut Copy Format Painter Paste

Styles

Normal Bad Good Neutral Calculation Check Cell Explanatory T... Input

Insert Delete Format Clear Sort & Find & Filter Select

G13 A B C D E F G H I J K L M N O P Q R S T

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Çalıştır	Yaz	Tarih	Glikoz: 0 Dakika	Glikoz: 60 Dakika	Glikoz: 120 Dakika	Hemoglobin A1c	İnsülin: 0 dakika	İnsülin: 60 dakika	İnsülin: 120 dakika	K	L	M	N	O	P	Q	R	S	T
2	E	45	30.11.2011 08:10	105	166	100	5.9	1847	153.1	37.15										
3	E	40	05.03.2012 09:28	103	192	82	5.6	1317	97.9	44.13										
4	E	55	11.04.2012 08:49	110	184	122	5.8	1491	69.64	30.23										
5	E	11	23.04.2012 09:15	94	119	133	5.6	1267	40.09	70.71										
6	E	53	27.04.2012 10:02	97	142	179	6	22.3	103.4	102.7										
7	K	43	09.06.2012 11:18	95	126	105	5.7	7.7	97.57	66.05										
8	K	46	22.06.2012 09:18	88	170	134	5.5	6.84	136.1	138.8										
9	K	48	06.07.2012 09:41	102	142	100	5.8	11.79	78.98	37.75										
10	K	38	24.07.2012 10:00	94	158	123	5.5	10.75	77.54	105.6										
11	K	37	10.08.2012 09:50	93	94	65	5.3	9.12	19.28	18.31										
12	K	31	22.09.2012 09:39	89	110	90	5	13.09	34.58	34.99										
13	K	29	25.09.2012 09:11	89	147	81	5.3	1.59	52.12	21.05										
14	K	16	14.03.2013 10:14	86	122	86	5.1	1428	88.21	17.18										
15	K	49	08.03.2013 11:29	88	125	116	5.2	8.37	47.22	26.02										
16	E	27	09.03.2013 10:07	109	137	141	5.4	28.99	132.9	198										
17	E	45	30.05.2013 10:06	91	115	88	5.4	10.07	67.3	23.23										
18	K	24	05.07.2013 08:27	86	138	157	5.3	9.27	23.48	99.55										
19	K	21	06.07.2013 09:25	99	117	63	5.7	7.78	36.14	18.19										
20	E	29	14.08.2013 09:19	92	138	109	5.5	10.9	86.12	31.86										
21	K	38	03.09.2013 10:41	100	188	80	5.3	8.48	86.78	38.69										
22	K	31	09.09.2013 09:22	97	174	95	5	37.68	42.21	43.89										
23	E	36	17.09.2013 09:53	94	111	119	5.6	10.7	59.7	66.74										
24	E	46	21.11.2013 10:03	99	191	85	5.8	12.19	138.9	39.08										
25	K	44	26.11.2013 08:51	101	195	147	5.5	15	115.5	134										
26	K	36	10.12.2013 10:24	83	114	130	5	3.27	38.13	36.57										
27	E	25	31.01.2014 09:37	98	112	93	5.5	7.3	80.57	30.18										
28	K	54	07.02.2014 08:39	100	206	168	5.7	9.37	160.7	201.9										
				147	222	240	5.5	20.78	334.7	240.4										

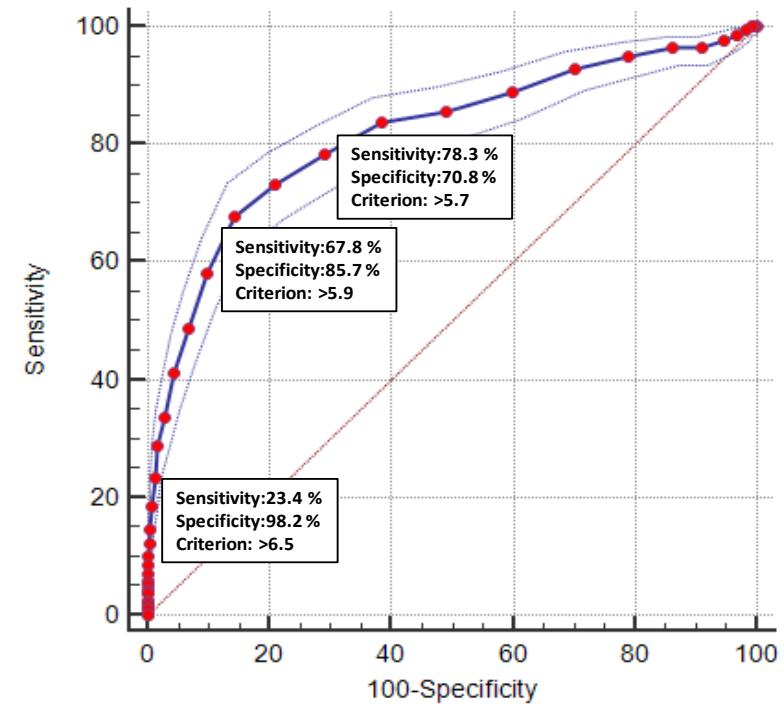
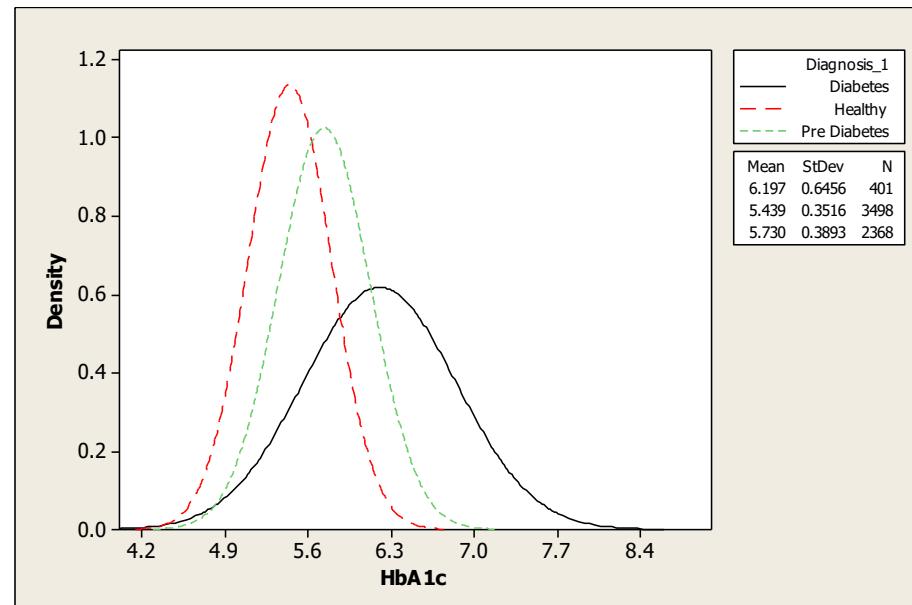
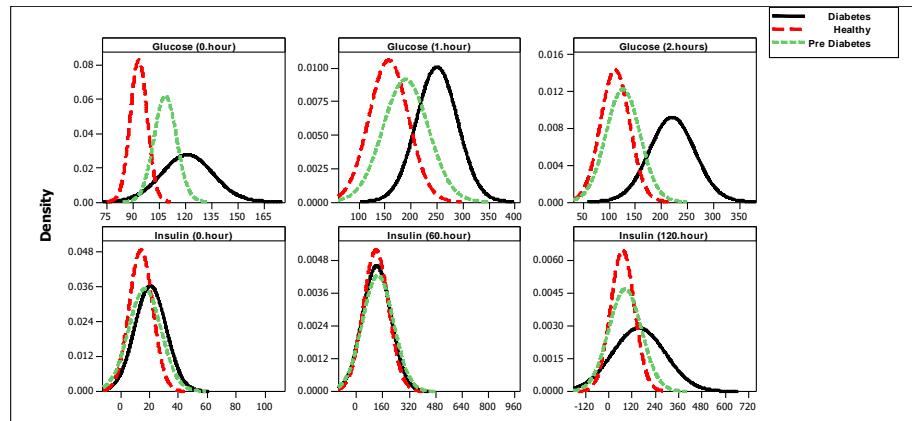
Sheet1

Ready



Desktop 2:36 AM 21-Sep-16

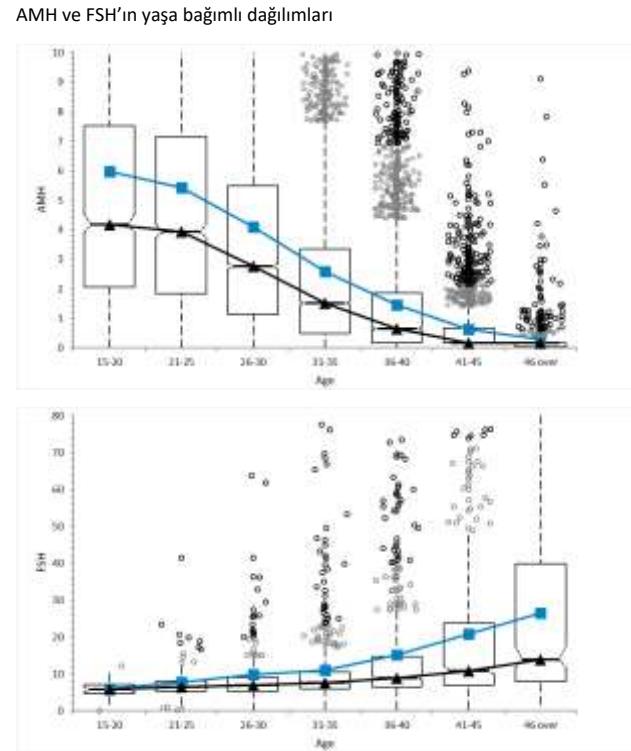
Diagnostic efficiency of glycosylated hemoglobin test in comparison to oral glucose tolerance test: A data mining study



ANTİ MÜLLERİAN HORMONIN YAŞ, LH, FSH VE ESTRADIOL İLİŞKİSİNİN BELİRLENMESİNDE ÇOKLU LINEER REGRESYON MODELİNİN UYGULANMASI:

Bir veri madenciliği çalışması

- Giriş;** AMH reproduktif yaşılanmanın iyi bir göstergesi olarak kadının menopoz zamanını tahmin etmek için yaygın olarak kullanılmaktadır. Bu bu very madenciliği çalışmasında retrospektif çalışmada over rezervinin değerlendirilmesinde AMH'nın yaş ve diğer rezerv belirteci hormon (L,FSH,Estradiol) seviyeleriyle ilişkisini belirlemesi amaşanmıştır
- Gereç ve yöntem;** Çalışma retrospektif olarak Acıbadem Labmed laboratuvarına başvuran 25294 bayan birey dahil edilmüktür. AMH'in yaşa (18-45) bağlı değişimleri ile LH, FSH ve Estradiol arasında çoklu regresyon analizi ile incelenmiştir. AMH Beckman Gen II ELISA kit ile diğer hormon ölçümleri ise electrochemiluminescence teknik ile Roche Elecsys Cobas analizörü üle yapılmıştır.
- Bulgular;** AMH, yaş, LH, FSH ve estradiol ile anlamlı korelasyon gösterirken progesterone ile ilişkisinin olmadığı delerlendirilmiştir. Yapılan çoklu lineer regresyon analizinde $AMH = 11.018 - (0.220 \times \text{yaş}) - (0.066 \times FSH) + (0.044 \times LH) - (0.004 \times \text{Estradiol})$ denklemi elde edilmiş ve modelin $R^2 = 0.627$ olarak saptanmıştır.
- Sonuç;** Overin fonksiyonel ya da biyolojik yaşıni tahmin etmede pek çok metod geliştirilmiştir. Yaş ile birlikte FSH, LH ve E2 ölçümlerinin over rezervini gösteren indirekt belirteçlerdir. Bunlara arasında belirlediğimiz regresyon modeli ile AMH un diğer hormonlar ile ilişkisi matematiksel olarak yapılmış ve çok daha iyi tahminlerinin geliştirilebilceği değerlendirilmiştir.



$$AMH = 11.018 - (0.220 \times \text{yaş}) - (0.066 \times FSH) + (0.044 \times LH) - (0.004 \times \text{Estradiol})$$

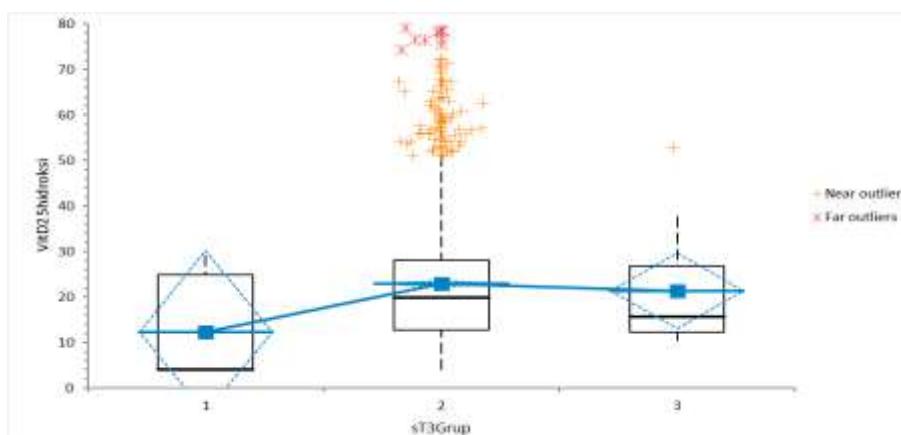
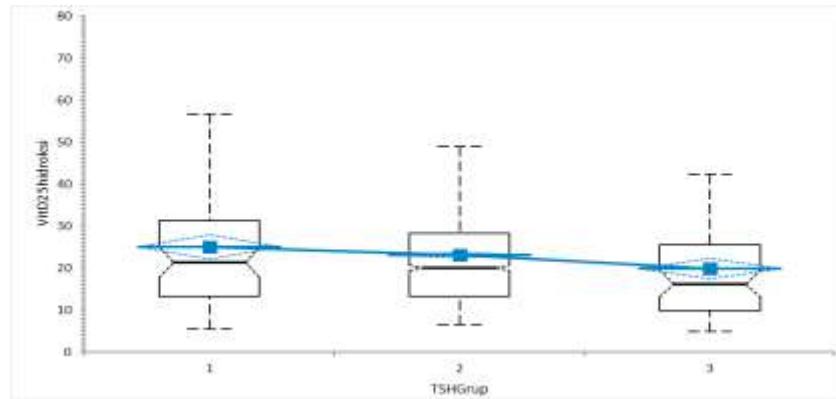
TİROID HORMONLARI VE ANTİKORLARININ YAŞ İLE DEĞİŞİMLERİ VE 25-OH VİTAMİN D DÜZEYİ İLE İLİŞKİLERİ

Amaç: Bu çalışmada retrospektif olarak hasta laboratuvar sonuçlarından, tiroid fonksiyonunun belirteçleri olan tiroid sitümülen hormon (TSH), serbest tiroksin (sT4), serbest triyodotironin (sT3), anti tiroid peroksidaz (anti-TPO) ve anti tiroglobulin (anti-Tg) düzeylerinin yaşa ve cinsiyete bağlı değişimleri ve 25 hidroksi vitamin D (25OHD) düzeyleri ile ilişkisinin incelenmesi amaçlanmıştır.

Gereç ve Yöntem: Acıbadem Labmed laboratuvarlarında tiroid fonksiyon testleri yapılmış, yaşıları 1-91 arasında değişen, **2574 kişi** (434 erkek, kadın 2140 kadın) çalışmaya dahil edilmiştir. TSH, sT4, sT3, anti TPO, anti Tg ölçümüleri elektrokemiluminesan immunoassay (ECLIA) ile Elecsys analizatöründe yapılmıştır. 25OHD ölçümü sıvı kromatografisi kütle spektrometresi ile yapılmıştır.

Bulgular: Serum TSH 1. ve 7. dekatta anlamlı yüksek olduğu saptanmıştır. sT3 düzeylerinin yaşa bağımlı şekilde istatistik olarak azaldığı saptanmıştır. Ancak sT4 düzeylerinin yaşa bağımlı değiştiği gözlenmemiştir. TSH düzeyleri referans limitleri içinde olan vakalarda ($0.27\text{-}4.84 \text{ mIU/L}$) 25OHD düzeyleri $23.0 \pm 15.6 \text{ ng/mL}$ iken TSH seviyesi yüksek olan vakalarda 25OHD seviyelerinin ($19.9 \pm 16.5 \text{ ng/mL}$) anlamlı düşük olduğu tespit edilmiştir ($p<0.001$).

Sonuç: Yaşa bağımlı olarak sT3'ün azalmasının nedeni sT4'ten sT3'e periferal dönüşümün azalması ile ilişkili olduğu değerlendirilmiştir. sT3 seviyelerine özellikle ileri yaşlarda dikkat edilmesi uygundur. İlave olarak, 25OHD düzeylerinin düşüklüğünde vakaların hipotroidi yönünden değerlendirilmesi gerekmektedir.



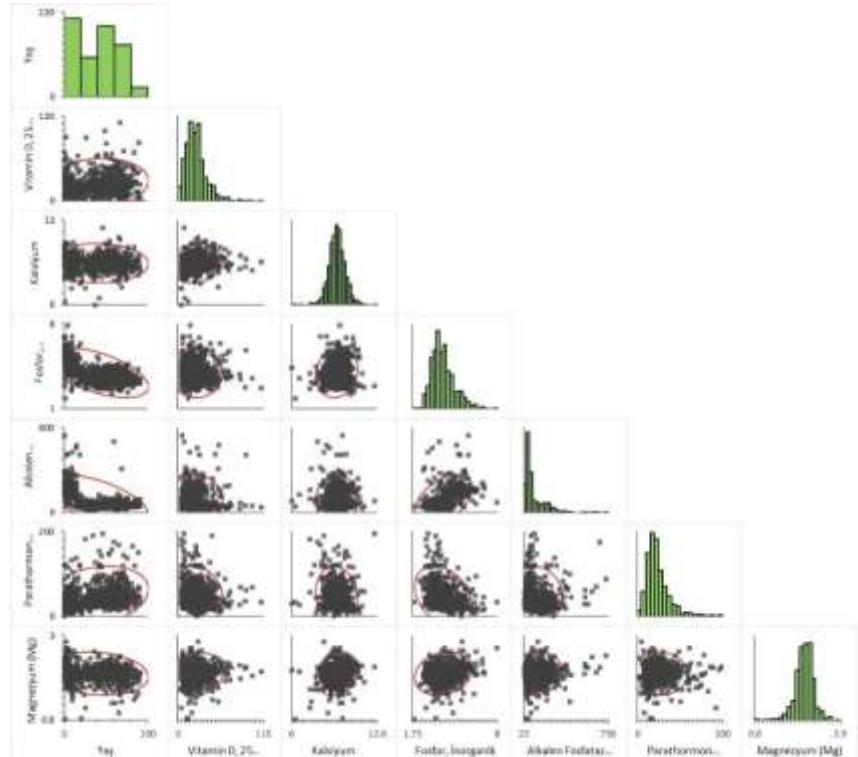
Kemik metabolizması belirteçleri ile Magnezyun düzeylerinin ilişkisi: Bir veri madenciliği çalışması.

Amaç: Magnezyum (Mg) elementi kemik metabolizması açısından kilit ama az bilinen bir moleküldür. Mg eksikliğinde, paratiroid hormonu (PTH) ve 25OH Vitamin D seviyeleri düşmektedir ve buna bağlı osteoporoz gelişimi oluşabilmektedir. Bu çalışmada Mg seviyesinin osteoporoz ile ilgili parametreler arasındaki ilişki geriye dönük incelenmiştir.

Yöntem: Acıbadem Labmed laboratuvarlarında aynı anda serum intakt PTH, VitD, ALP, Ca, P ve Mg analizleri yapılmış olan ver gerekli dışlama ve temizleme işlemleri yapıldıktan sonra geriye kalan **640** kişinin verileri anonimize edilerek çalışma grubu oluşturulmuştur. Popülasyonun, 183 çocuk (6.4 ± 5 yıl) ve 457 kişi erişkin (52.2 ± 16.9 yıl) olarak ayrılmıştır; analizler total popülasyon ve belirtilen yaş gruplarında ayrı ayrı yapılmıştır. Çalışmada Microsoft Excel Analyze-it ve SPSS PASW v18.0.0, XLSTAT programından yararlanılarak gerçekleştirilmiştir.

Bulgular: Popülasyonun tümünde, yaşa bağımlı olarak Mg, Ca ve P seviyeleri düşüş göstermektedir. Özellikle serum Mg ve VitD-25OH seviyeleri arasında özellikle çocukların olmak üzere pozitif anlamlı korelasyon bulunmuştur ($r=0.213$, $p<0.01$). İlave olarak tüm popülasyon, Mg değeri düşük olan bireylerde VitD-25OH seviyelerinin de anlamlı düşük olduğu tespit edilmiştir ($p<0.05$). İlave olarak çocukların Mg düşüklüğünde, Ca, VitD ve P seviyeleri gözlenmektedir ($p < 0.001$). Erişkinlerde ise, Mg düşüklüğünde yalnızca Ca seviyelerinin de düşüş olduğu görülmüştür ($p < 0.05$). PTH ile Mg arasında anlamlı ilişki gözlenmemiştir.

Sonuç: Bu veriler Mg seviyesinin kemik metabolizmasının değerlendirilmesinde ve özellikle çocukların VitD eksikliğinde değerlendirilmesinde dikkatle irdelemesi gereklidir. Ayrıca düşük Mg seviyelerine, düşük Ca ve P metabolizma değerlendirilmesinde önemlidir. Vitamin D eksikliğinin, tedavi planlaması ve takibi sırasında mutlaka Mg, Ca, P ve iPTH değerlendirilmesinde



A linear regression model for estimating D-dimer levels from hsCRP, WBC, neutrophil and procalcitonin: A data mining study

D-dimer (a fibrin degradation product) test is used to aid in the diagnosis of intravascular coagulation. The aim of this study is to investigate the correlation between D-dimer levels and other inflammatory markers including procalcitonin.

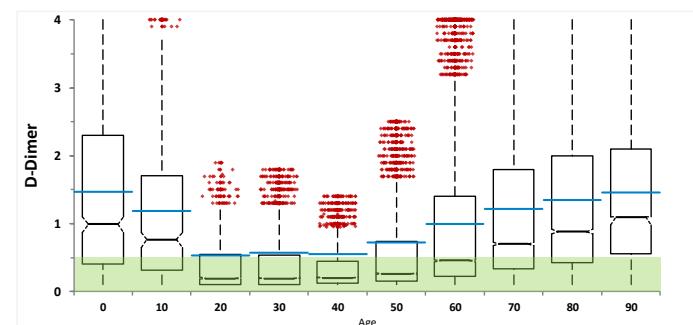
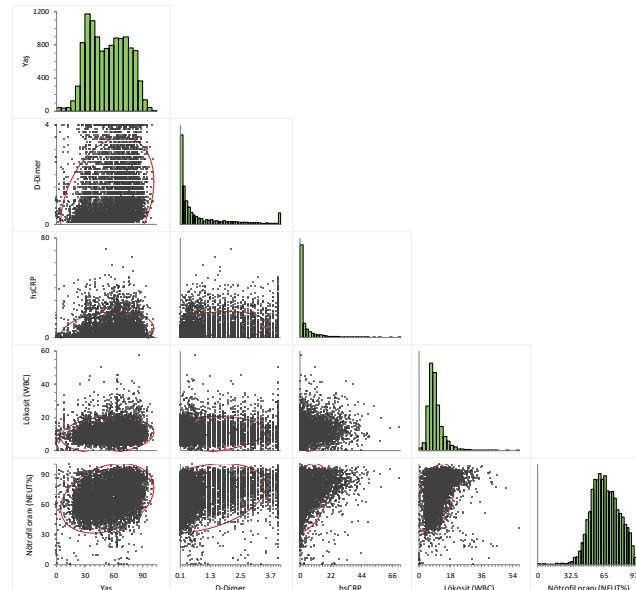
Anonymized data on D-dimer, fibrinogen, hsCRP, WBC, Neutrophil% (NEUT%) and procalcitonin levels from 50,107 patients (mean age \pm SD, 53.37 ± 20.6) were used for the correlation (Excel Analyze-it v4.60.4) and linear regression (PASW Statistics 18 v18.0) analysis between the measured parameters.

There was a significant ($p<0.05$) age-dependent increase in D-dimer levels between different age groups. Patients with the highest D-dimer levels were also found to have an increased frequency of hsCRP levels. D-dimer levels showed a significant correlation with hsCRP, WBC, NEUT%, and procalcitonin ($p=0.344$) describing the positive association between these parameters were built. The resulting equation is as follows:

$$\text{D-Dimer} = (\text{hsCRP} * 0.054) + (0.011 * \text{Age}) + (0.006 * \text{WBC}) + (0.04 * \text{NEUT\%}) - 0.929.$$

Correlation analysis between procalcitonin and D-dimer levels gave Pearson's correlation coefficient of 0.159.

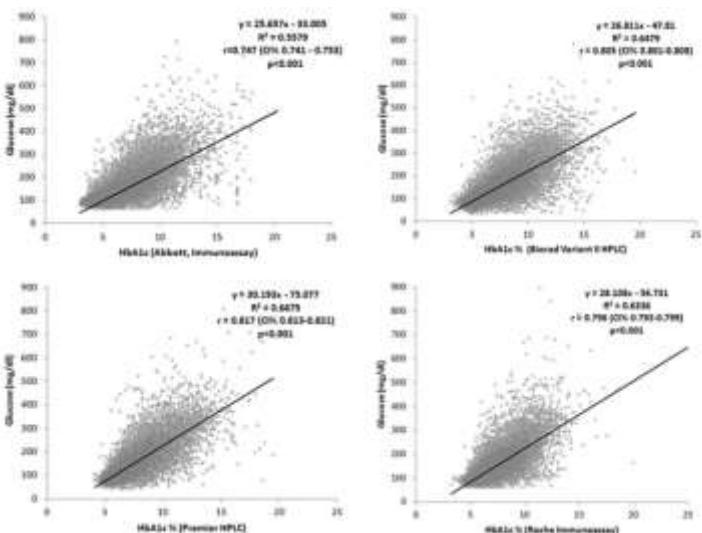
Our results suggest that the age-dependent variations should be taken into account while interpreting D-dimer test results. In addition, NEUT% ratio was found to be the most important parameter for estimating D-dimer levels. Our equation can be used when the D-dimer test is not available or for control purposes only.



A Simple and Easy Process for the Determination of Estimated Plasma Glucose Level in Patients Presenting to Hospital: An Example of Multicentric Data Mining

Journal of Laboratory Automation
1–5
© 2016 Society for Laboratory Automation and Screening
DOI: 10.1177/211068215624
jla.sagepub.com
SAGE

Muhittin A. Serdar¹, Macit Koldaş², Mustafa Serteser¹, Okhan Akın³,
Cigdem Sonmez⁴, Ozlem Gülbahar⁵, Filiz Akbiyik⁶, and İbrahim Ünsal¹



Abstract

The aim of the present study was to determine the relation between the simultaneous fasting plasma glucose level and HbA1c in a large population of patients presenting to the hospital, based on various measurement methods available for HbA1c. HbA1c levels of 162,210 patients presenting to various hospitals and laboratories were measured based on seven different systems, and at the same time, eAG levels were calculated based on HbA1c levels. The correlation coefficients (r) between serum plasma glucose and HbA1c levels were found to be 0.809, 0.774, 0.779, 0.817, 0.704, 0.796, and 0.747 in Bio-Rad Variant II, Tosoh G8, ADAMS A1c, Trinity Boronate Affinity, Chromsystems HPLC, Roche Tina-quant, and Abbott Architect, respectively. The concordance correlation coefficients between the eAG levels as calculated with the formulas provided in the text and the eAG levels as calculated according to NGSP directions (where $eAG = (28.7 \times HbA1c) - 46.7$) were found to be between 0.9339 and 0.9866. Despite the progress made for the standardization of HbA1c measurements, the relation between serum glucose and HbA1c still demonstrated certain discrepancies pertaining to the differences in measurement methodologies. As a conclusion, each laboratory could determine different eAG levels depending on the data originated by their individual analyzer.

Comparison of four immunoassay analyzers for relationship between thyroid stimulating hormone (TSH) and free thyroxine (FT4)

Muhittin A. Serdar¹,

Emre İspir³,

Taner Ozgurtas⁴,

Ozlem Gulbahar²,

Zahid Ciraci²,

Hatice Pasaoglu²,

Ismail Kurt⁴

ABSTRACT

Objective: There is an inverse log/linear relationship between TSH and FT4 due to the negative feedback of these hormones on the pituitary. The objective of our study was to compare this relationship of TSH and FT4 between four different immunoassay analyzers.

Methods: In our study, four data sets obtained from the database of four different hospital laboratories each using only a single method of measurement was used to evaluate this relationship between TSH and FT4. These data sets with their assay methods include: 21,102 test results measured by Modular E170 Analyzer (Roche Diagnostics, Germany), 20,241 test results measured by Access Dxi 800 Unicel (Beckman Coulter, USA), 22,444 test results measured by Architect i2000sr (Abbott Laboratories, Abbott Park, Illinois, U.S.A) and 20,200 test results measured by ADVIA Centaur XP (Siemens Diagnostics, Tarrytown, NY). Inverse logarithmic relationship were determined from each data sets and compared between analyzers.

Results: The correlation coefficients were -0.439 [95% CI, (-0.450)- (-0.428)], -0.488 [95% CI, (-0.498)- (-0.478)], -0.353 [95% CI, (-0.364)- (-0.342)], -0.430 [95% CI, (-0.441)- (-0.419)] for Dxi 800 Unicel, Modular E 170, Architect i2000sr and ADVIA Centaur XP, respectively. In our study, all immunoassay analyzers showed poor correlation in a concentration range of TSH between 1.0 to 10.0 uIU/mL and the inverse log/linear relationship was not observed. Inverse relationship between TSH and FT4 is the result of negative thyroid pituitary hypothalamic feedback mechanism and this relationship is important especially in the diagnosis and treatment of hypo and hyperthyroidism. In data sets obtained from four different immunoassay analyzers, poor and statistically different correlation was observed between analyzers at TSH values ranged from 1.0 to 10.0 uIU/mL compared with FT4.

Conclusion: These variations between analyzers may affect the clinical decisions especially in the evaluation of subclinical hypothyroidism, clinicians and laboratory specialists should be aware of these situation.

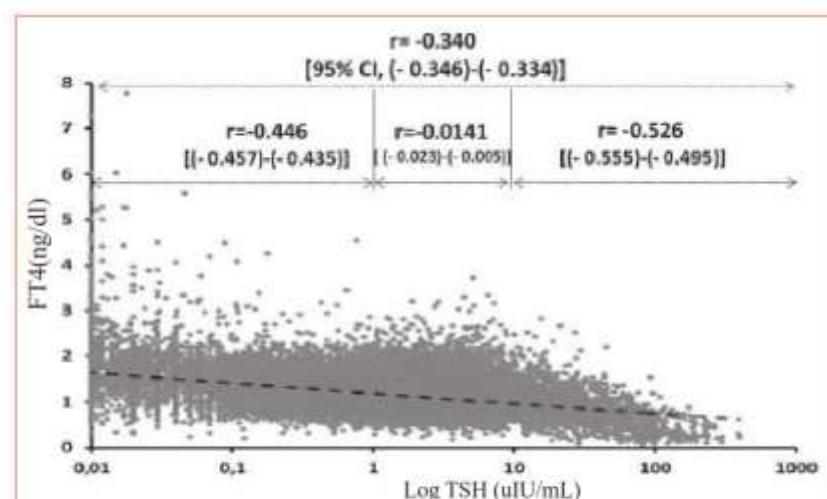


Figure 2. The correlation coefficients with their 95% confidence intervals at different (0-1, 1-10 and over 10 uIU/ml) and the overall TSH concentrations were depicted for combined four data sets. Conversion factor for FT4 (ng/dL → pmol/L): 12.87.

Analysis of changes in parathyroid hormone and vitamin D levels with respect to age,

gender and season: A Data Mining Study*

Muhittin A. Serdar^{1,2}, Başar Batu Can¹, Meltem Kilercik¹, Zeynep A. Durer¹, Fehime Benli

Aksungar¹, Mustafa Serteser¹, Abdurrahman Coskun¹, Aysel Ozpinar¹, and Ibrahim Unsal¹

- **Abstract**
- **Background and Objectives:** 25(OH)D and parathyroid hormone (PTH) are important regulators of calcium homeostasis. The aim of this study was to retrospectively determine the cut-off for sufficient 25(OH)D in a four season region and the influence of age, seasons, and gender on serum 25(OH)D and PTH levels.
- **Design:** Laboratory results of 9890 female and 2723 male individuals aging 38.8 ± 22.1 years who had simultaneous measurements of 25(OH)D₃ and PTH were retrospectively analyzed by statistical softwares. Serum 25(OH)D₃ and PTH levels were measured by mass spectrometry method and by electrochemiluminescence immunoassay, respectively.
- **Results:** Mean serum 25(OH)D levels showed a sinusoidal fluctuation throughout the year and were significantly ($p < 0.01$) higher in summer and autumn. On the other hand, PTH levels were significantly higher ($p < 0.01$) in women and showed an opposite response to seasonal effects relative to 25(OH)D. Lowest levels of 25(OH)D were detected in people aged between 20 and 40 years old whereas PTH hormone levels were gradually increasing in response to aging. The significant

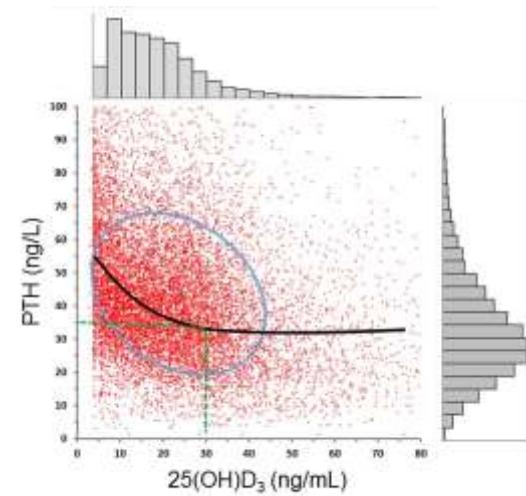
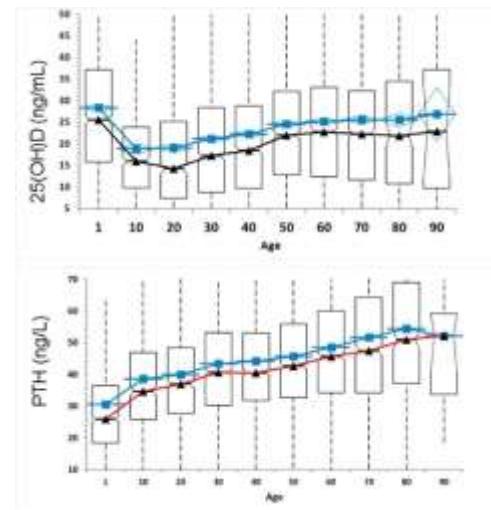
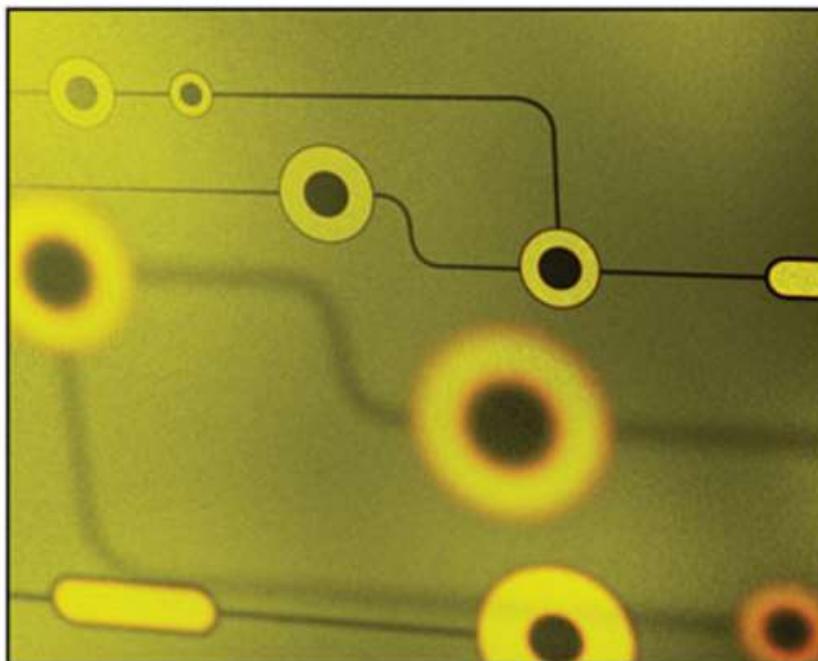


Figure 5: Relation between PTH and 25(OH)D levels.

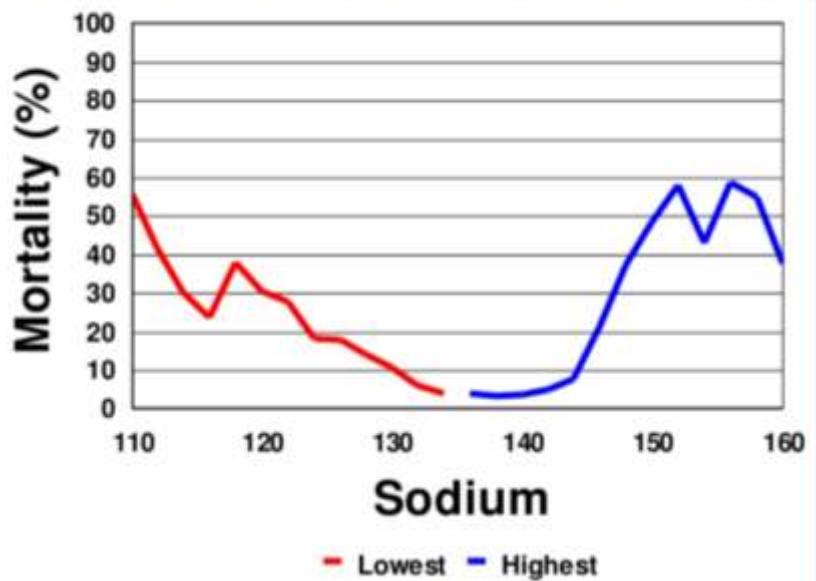
PREMIER REFERENCE SOURCE

DATA MINING AND MEDICAL KNOWLEDGE MANAGEMENT

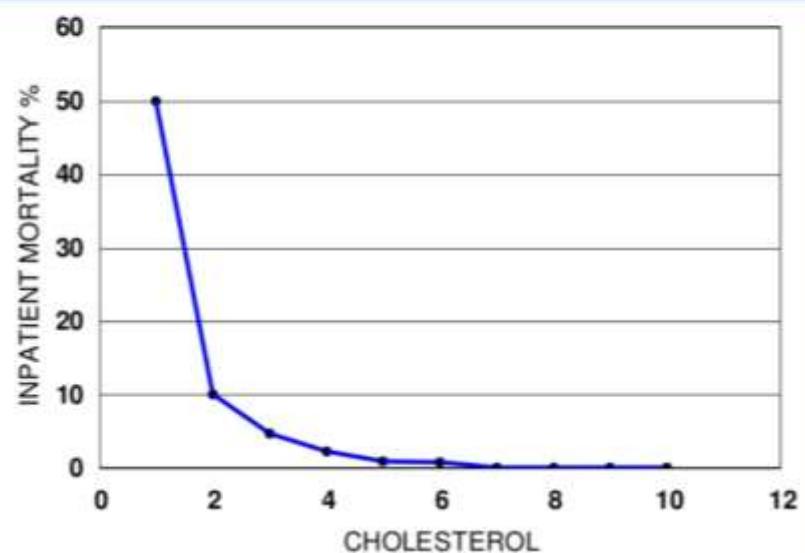
CASES AND APPLICATIONS



Sodium & Inpatient Mortality



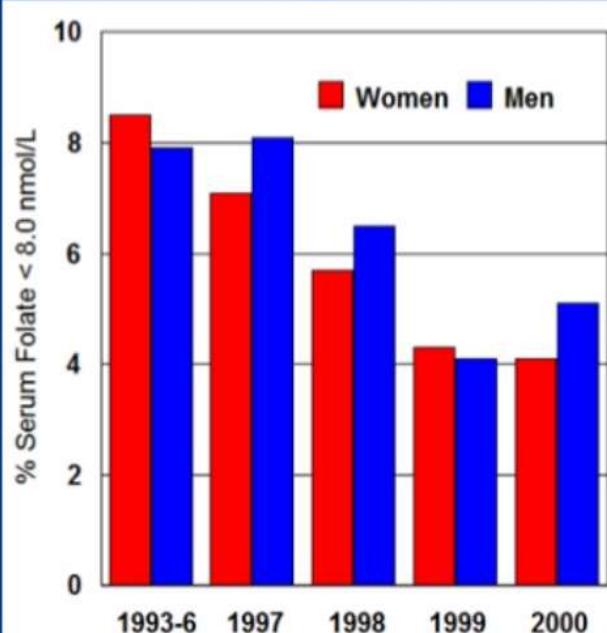
Cholesterol and Inpatient Mortality



Folate deficiency trends

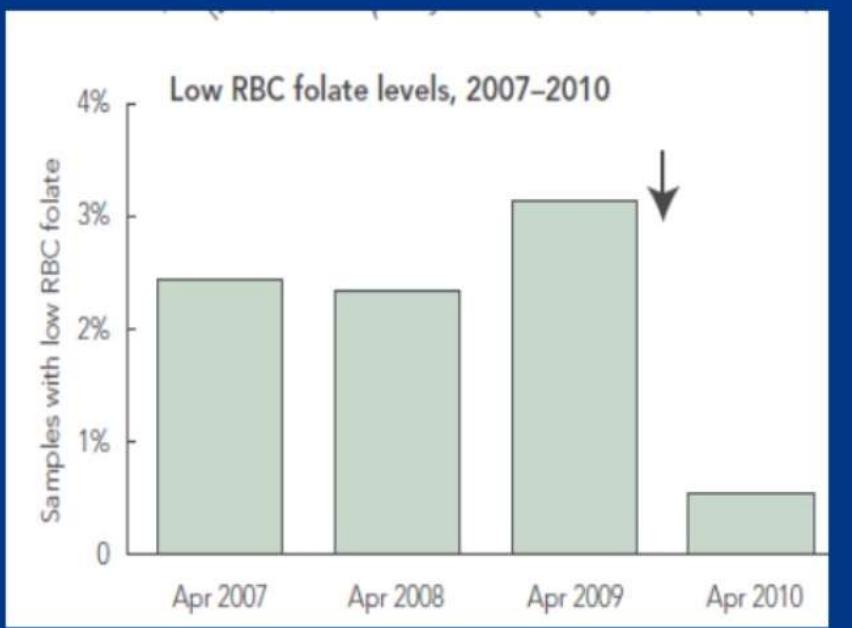
Changes in serum folate concentrations following voluntary food fortification in Australia

Mets & Sikaris, et al. MJA 2002; 176
(Since 1995)



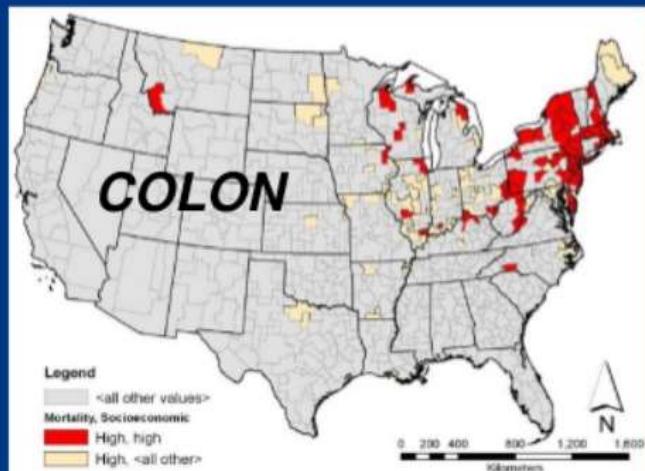
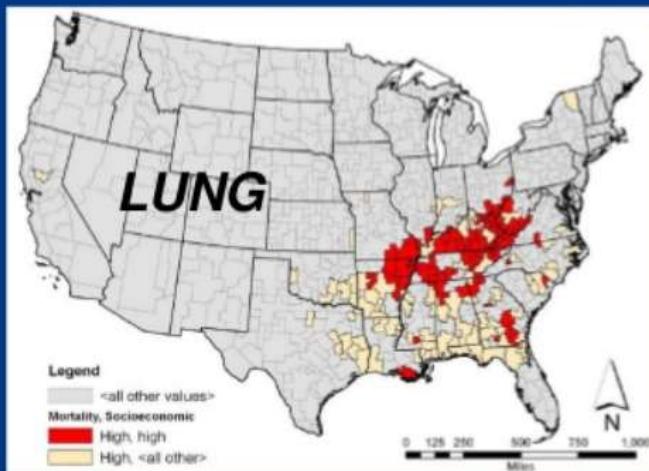
The impact of mandatory fortification of flour with folic acid on the blood folate levels of an Australian population

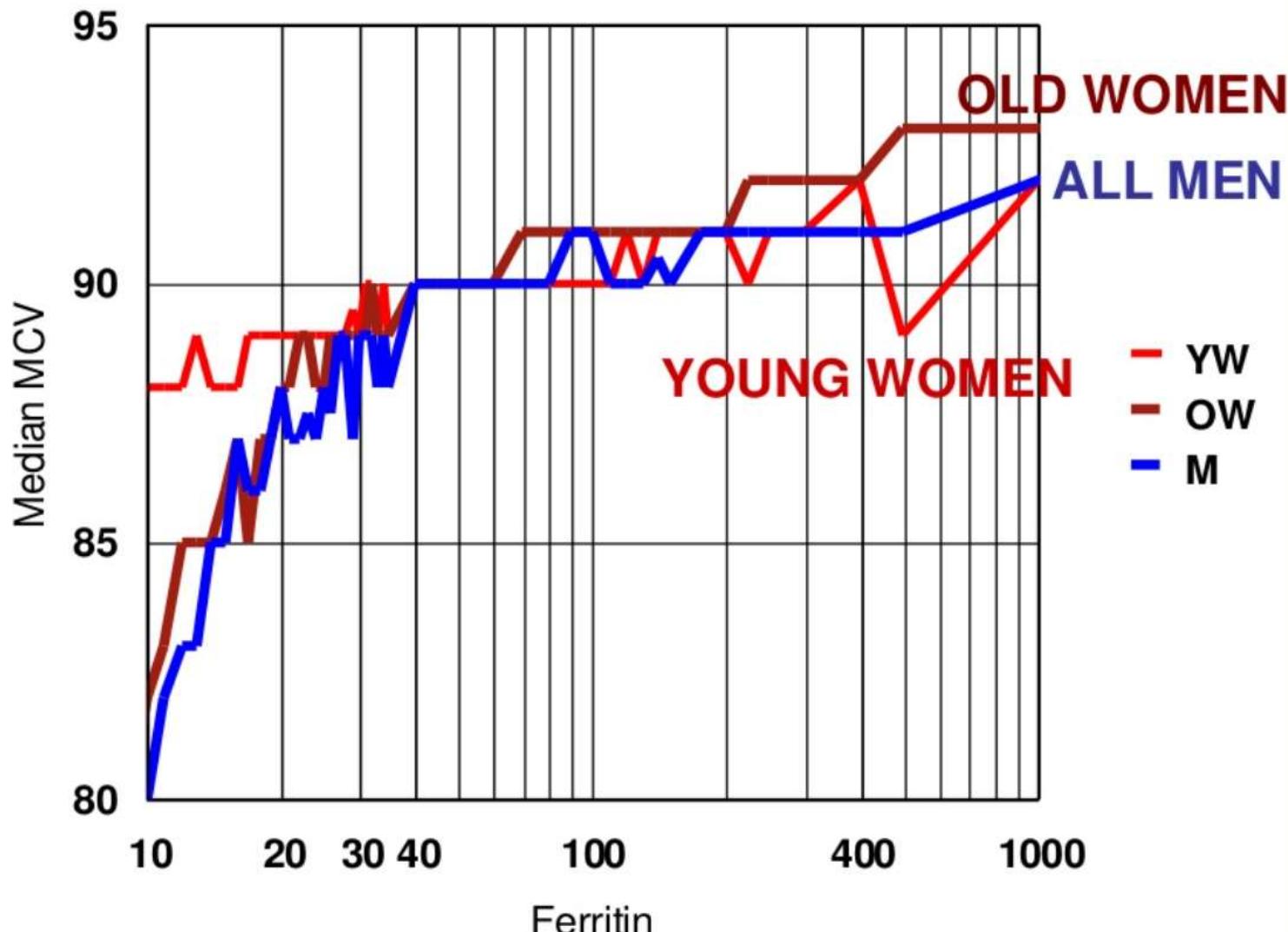
Brown et al. MJA 2011; 194 (2): 65-67.
(Since Sept 2009)

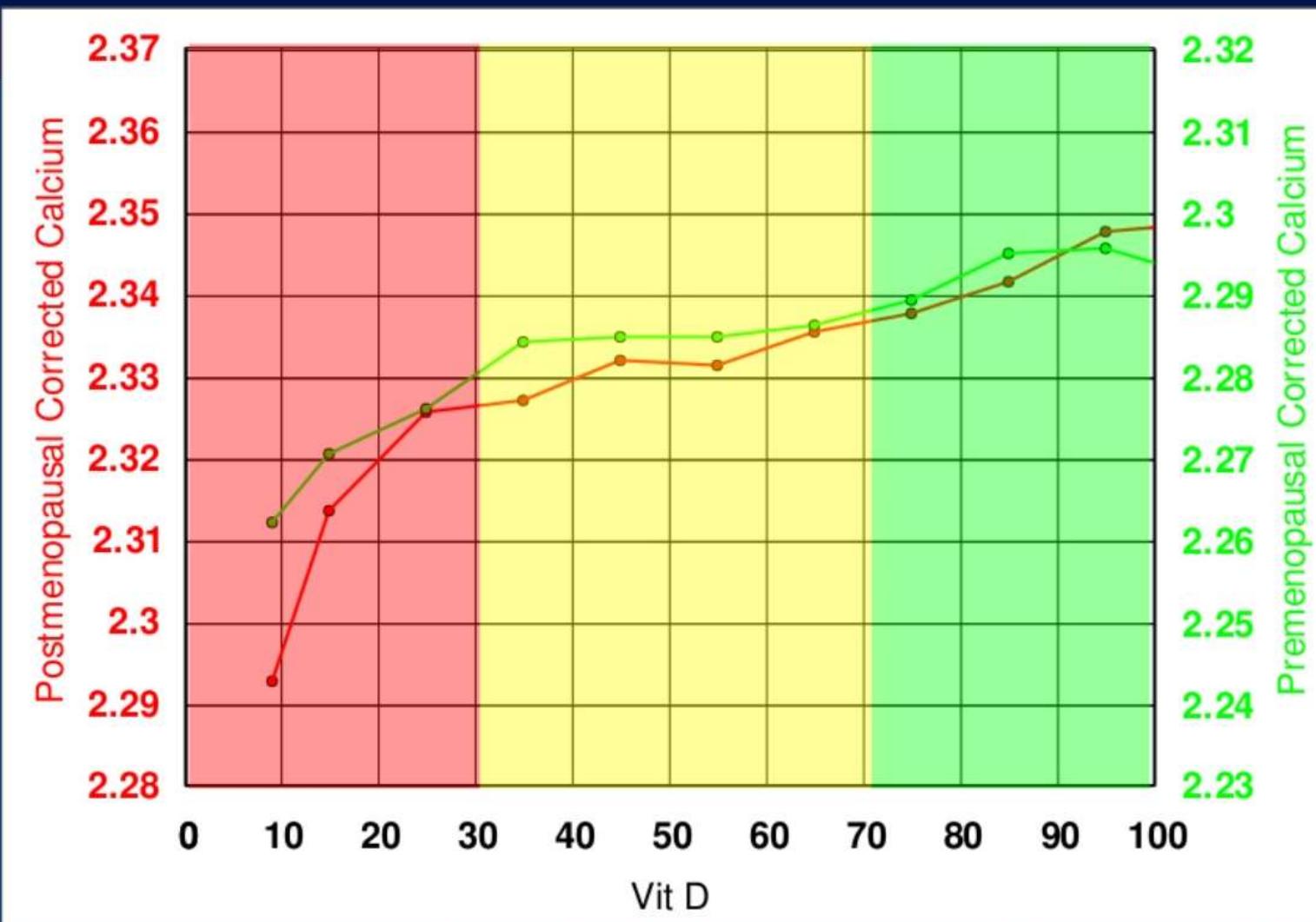


Socioeconomic inequality of cancer mortality in the United States: a spatial data mining approach

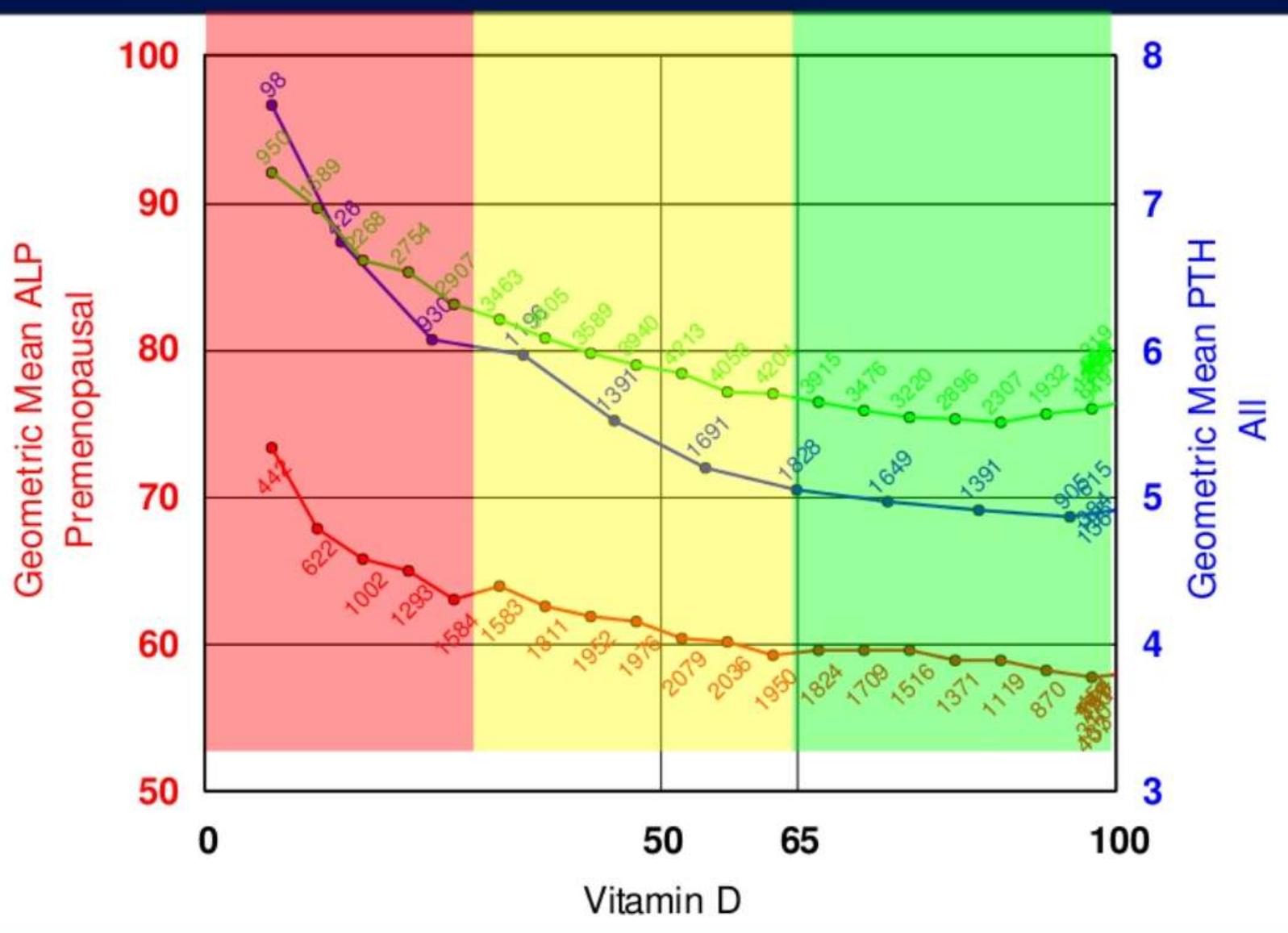
Srinivas Vinnakota* and Nina SN Lam







Lu ZX, Dahanayaka K, Lambrianou J, Ratnakee S, Sikaris KA, "How much Vitamin D is sufficient? An evidence based approach." Clin Biochem Rev 2007; 28:S29



Lu ZX, Dahanayaka K, Lambrianou J, Ratnakee S, Sikaris KA, "How much Vitamin D is sufficient? An evidence based approach." Clin Biochem Rev 2007; 28:S29

Sonic Reference Intervals 2009



Deriving Sonic Network Reference Intervals For Pregnancy

K Sikaris¹, Z Lu¹, D Kanowski², L Price², R Flatman³, G Caldwell³, N Taylor³, T Yen³, S Sacks⁴, M Metz⁵, R Hanlon⁶, J Andriolo⁷

References 1. Australian Pathology, Collingwood, Victoria; 2. Sullivan Neurology Pathology, Taringa, Queensland; 3. Angkor Eye & Ear Pathology, Macquarie Park, New South Wales; 4. ClinPath Laboratories, West Perth, Western Australia; 5. ClinPath Laboratories, Kent Town, South Australia; 6. Capital Pathology, Hobart, ACT; 7. North West Pathology, Burnie, Tasmania.

Introduction

As part of the Bionic Reference Interval Project, reference intervals for different stages of pregnancy need to be defined and validated. This would allow the network to measure from site specified intervals based on either manufacturer suggested limits, literature or locally defined reference limits to a common set of validated reference intervals.

Mythic 8

Following a survey and literature review of existing reference intervals, I&E ($n=7,870$) and RT ($n=9,970$) data was extracted from the Apollo Laboratory Information System at Melbourne Pathology and Sullivan Pathology for Bhattacharya analysis. As gestational age was also recorded, partitioning into trimesters specific intervals could be investigated. Both laboratories use Roche Modular platforms for routine chemistry.

Result

Table . Reference Intervals for fT4 s and fT3 s in pregnancy by trimester (Trim).

Conclusion

All reference intervals for pregnancy were lower than in a nonpregnant young woman with the exception of ALP which rose dramatically at the start of the third trimester.



Deriving Sonic Network Reference Intervals For Children

K Sikaris¹, Z Lu¹, D Kanowski², L Price², R Flatman², G Caldwell², N Taylor³, T Yen³, S Sacks⁴, M Metz⁵, R Hanlon⁶, J Andriolo⁷.

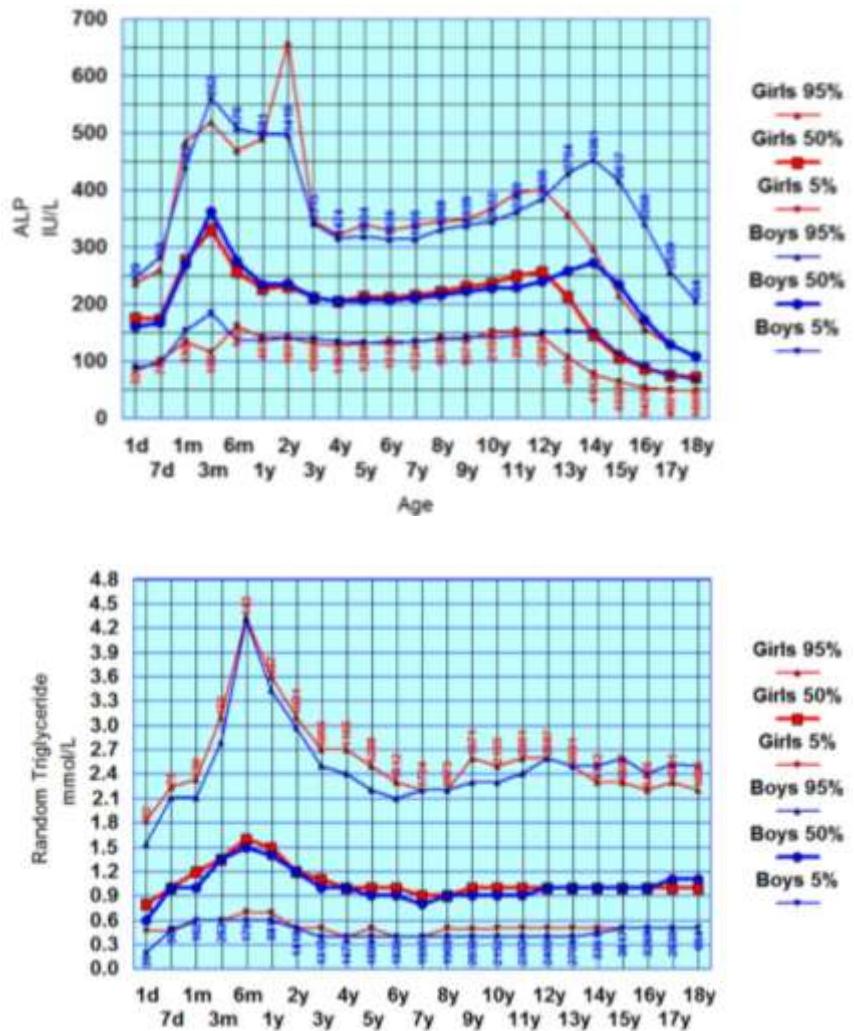
Wesbourne Pathology, Collingwood, Victoria, Sullivan Worldwide Pathology, Tiaro, Queensland, Angler Rocker Edge Pathology, Macquarie Park, New South Wales, TotalPath Laboratories, West Perth, Western Australia, Path Laboratories, Kenilworth, South Australia, Hospital Pathology, Dandenong, Victoria, North West Pathology, Burnie, Tasmania

Introduction

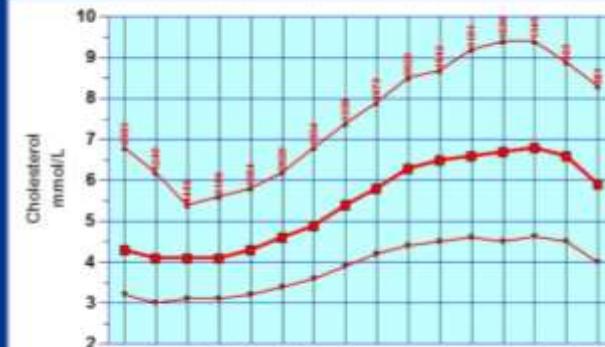
part of the World Table, UAE and LFT reference intervals in childhood.

zhongguo cheng

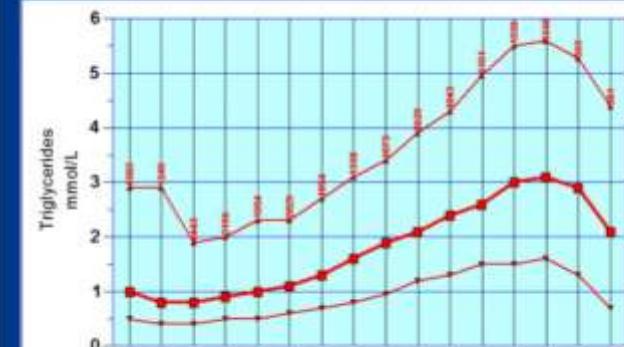
The results imply that paediatric reference intervals change at varying stages of childhood depending on the physiological changes affecting each athlete. Defining the age for making a reference interval transition in childhood is just as important as determining the limits of the reference interval for the age interval selected.



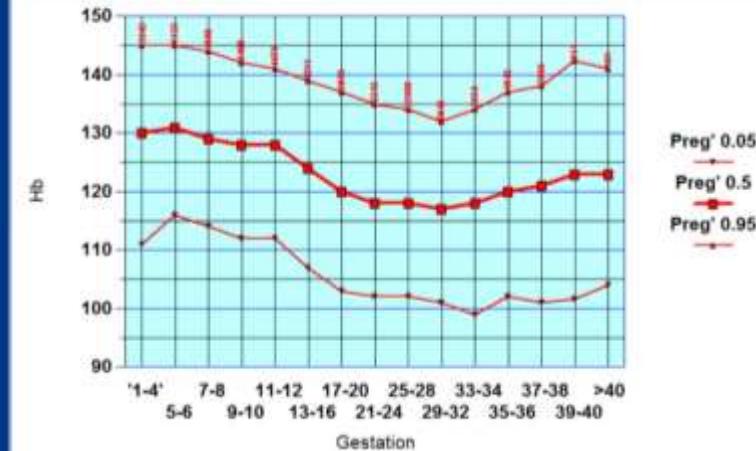
Cholesterol in Pregnancy



Triglycerides in Pregnancy



Hb in Pregnancy



Uncovering the Patterns in Pathology Ordering by Australian General Practitioners: a Data Mining Perspective

Zoe Yan Zhuang, Leonid Churilov Ken Sikaris

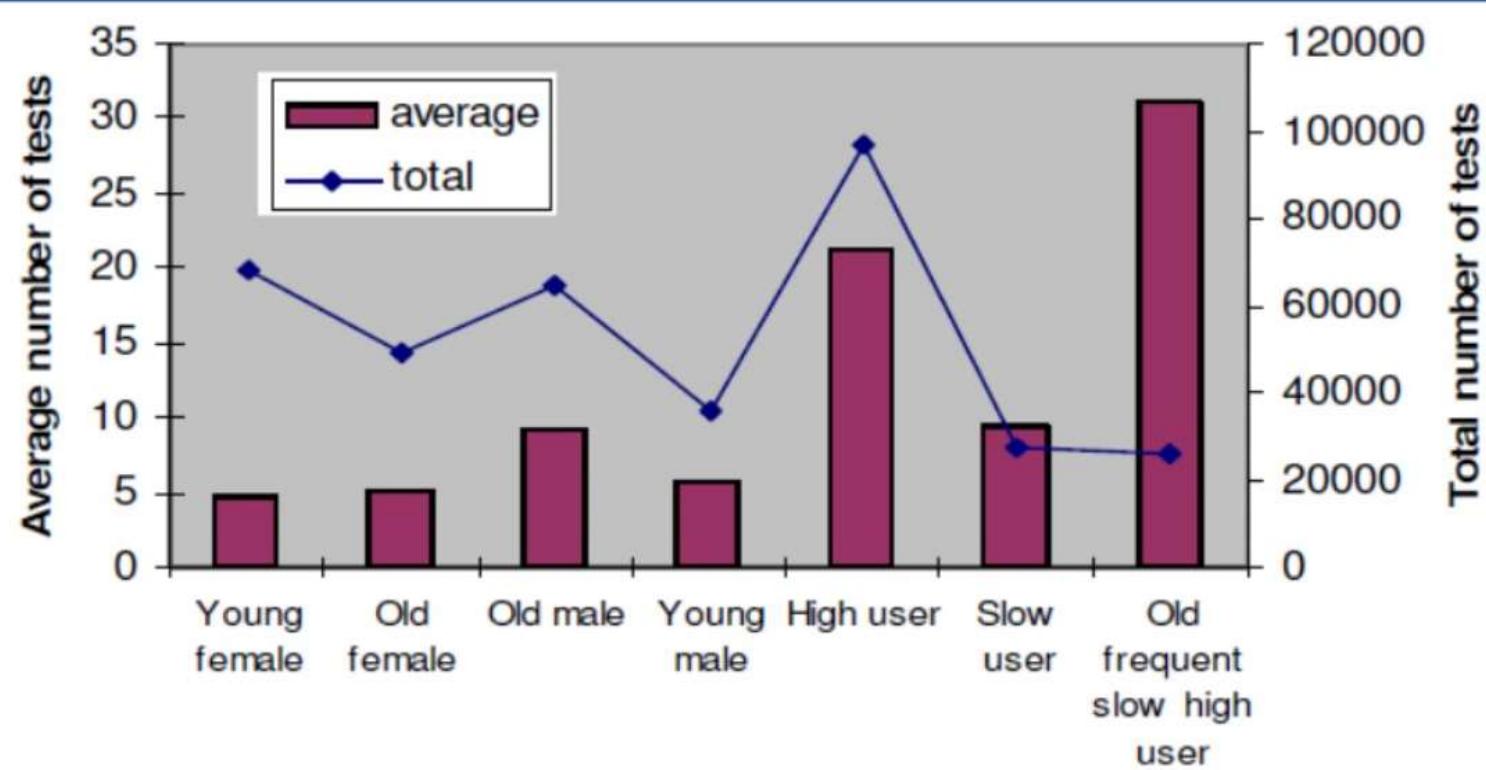


Figure 2. Average and total number of tests for each cluster

A framework for an intelligent decision support system: A case in pathology test ordering

Zoe Y. Zhuang, Carla L. Wilkin ^{*}, Andrzej Ceglowski ¹

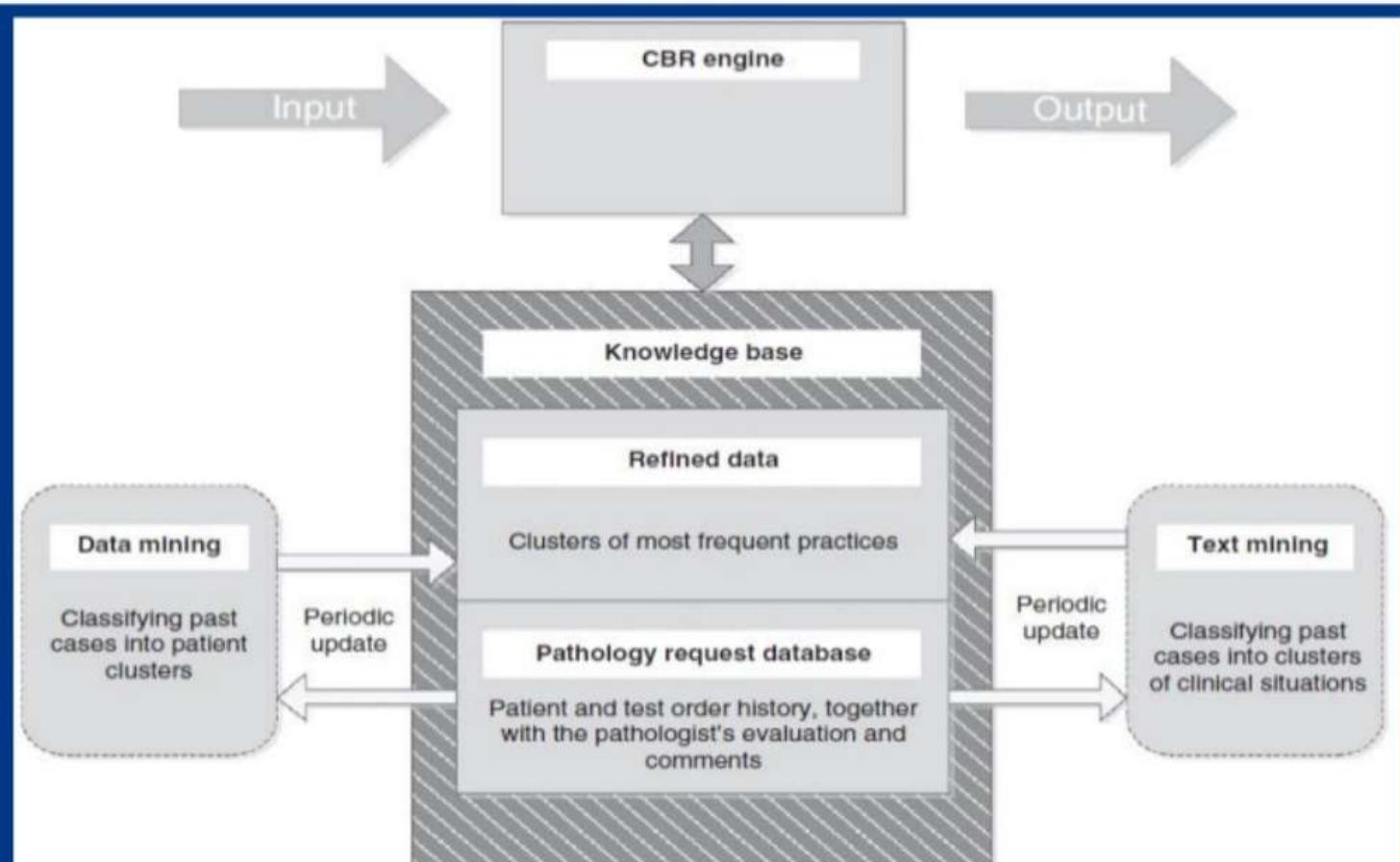


Fig. 3. Simplified framework for intelligent decision support.

Mutlaka kuralları olmalı



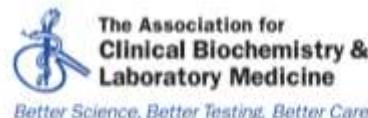
Data mining brings a lot of benefits to businesses, society, governments as well as the individual. However, **privacy, security, and misuse of information** are the big problems if they are not addressed and resolved properly

Model oluşturma laboratuvar, klinik bilgi, araştırma-istatistik ve bilişim bilgileri ister



Better Science, Better Testing, Better Care

Annals of Clinical Biochemistry
2015, Vol. 53(2) 251–258
© The Author(s) 2014
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0004563214541969
acb.sagepub.com



Better Science, Better Testing, Better Care

Serum total bilirubin concentrations are inversely associated with total white blood cell counts in an adult population

Wan-Ni Tsai¹, Ya-Yu Wang^{2,3,4}, Jiin-Tsae Liang⁵, Shih-Yi Lin^{4,6}, Wayne Huey-Herng Sheu^{4,6,7} and Wen-Dau Chang²

Abstract

Background: Several studies have shown that serum bilirubin has a protective effect against cardiovascular disease and that inflammation plays an important role in its pathogenesis. This cross-sectional study investigated whether there was an association between blood total white blood cell count, a simple indicator of inflammation, and serum total bilirubin concentration in a cohort of an adult population in Taiwan.

Methods: A total of 2458 apparently healthy adults (1246 men and 1212 women) who attended health examination at a medical centre in central Taiwan were enrolled. We collected anthropometric measurements, fasting blood test results, lifestyle habits and medical history.

Results: Total white blood cell counts decreased progressively when the concentrations of total bilirubin increased as demonstrated in the total bilirubin quartiles. Both total bilirubin concentrations and total white blood cell counts showed significant correlations with the components of metabolic syndrome, including triglyceride and high-density lipoprotein cholesterol concentrations. Multivariate linear regression analysis revealed that, for both genders, total bilirubin showed an independent negative correlation with total white blood cell count after adjusting for conventional cardiovascular risk factors.

Conclusions: Higher serum total bilirubin concentrations within the reference range were associated with lower blood total white blood cell counts, regardless of other classical cardiovascular risk factors.

Response to article: serum total bilirubin concentrations are inversely associated with total white blood cell counts in an adult population

The article by Tsai et al. in Annals¹ describes a significant inverse association between total serum bilirubin and total white (blood) cell count (WCC) in a cohort of healthy Taiwanese adults. This relationship was detected in cross-sectional data for cases with total serum bilirubin within the reference interval and total WCC less than $15 \times 10^9/L$. We have investigated similar relationships in two distinct cohorts of Australian patients who had routine pathology conducted via Sullivan Nicolaides Pathology (Taringa, Queensland), including assessments of inflammatory markers.

Sonuç

- Laboratuvar uzamanı görev tanımı ve eğitim içeriğinin değişiyor
- Bilişim teknolojileri ve istatistik alt yapılarını geliştirmeliyiz. Büyük verileri anlar işler hale getmek zorundayız (*)
- LBYS ve HBYS etkin kullanımı çok önemli
- Veri madenciliği çalışmaları basitten oldukça kompleks bir yelpazede yapılabilir. En önemli sorun **verilerinin işlenebilir hale gelmesi** için temizlenmesi sürecidir
- Diğer bütün branşlar (klinisyen, bilişim, istatistik uzmanı gibi) **ortak çalışma** kültürünü artırmak zorundayız

