Ana Szarfman, MD, PhD, Medical Officer; Jonathan G. Levine, PhD, Statistician; and Joseph M. Tonning, MD, MPH, Medical Officer; OPaSS Immediate Office, CDER, FDA; Szarfman@cder.fda.gov

Use of Advanced Computer Methods to Simplify the Analysis of Complex Clinical Drug Safety Data

Harvard School of Public Health, Department of Biostatistics, Schering-Plough Workshop on Statistical Issues in Drug Safety Monitoring, Boston, Massachusetts, June 2-3, 2005 http://www.biostat.harvard.edu/events/schering-plough/ and 26th Spring Symposium New Jersey Chapter of ASA. Statistical Data Mining, June 7, 2005

Disclaimer

- **Views expressed in this presentation are those of the speaker**

Overall Summary

- **Communication Roadblocks**
  - Too Many Data Standards
- **Contrast between Clinical Practice and Pharmacoepidemiology**
  - Safety Questions and Answers
- **Practical tools for speeding the analysis of complex safety data**
  - *Drugs under review:* Application of emerging standards
  - *Drugs already marketed:* Data mining using MGPS and HBLR
- **Take home messages**

Understanding Drug Safety Systematically

- **Access to multiple collections of medical data to cross-reference potential signals *for replication***
  - Medical records such as Health Maintenance Organization records, military databases; clinical trial data; animal pharmacology data, etc.
- **Full communication between experts *for improved interpretation by tapping know-how***
  - Need real-time access of the same data by different experts for analysis and re-analysis w/o starting from scratch

Communication Roadblocks: Too Many Individual *Data* Standards

- **Safety data are complex**
  - High dimensional, high variations between subjects, and missing information
- **Safety data are disorganized: Uniform data standards are just emerging**
  - Data quality problems are not limited to spontaneous reports (e.g., same concomitant medication having multiple drug name variations)
- **Data and tools have to be constantly re-configured**
  - This creates **"difficult to audit," "difficult to share," and "difficult to learn"** paradigms

Szarfman A., Talarico L., Levine J.G. In: Comprehensive Toxicology. 4:363-379,1997

Complex data, missing data example

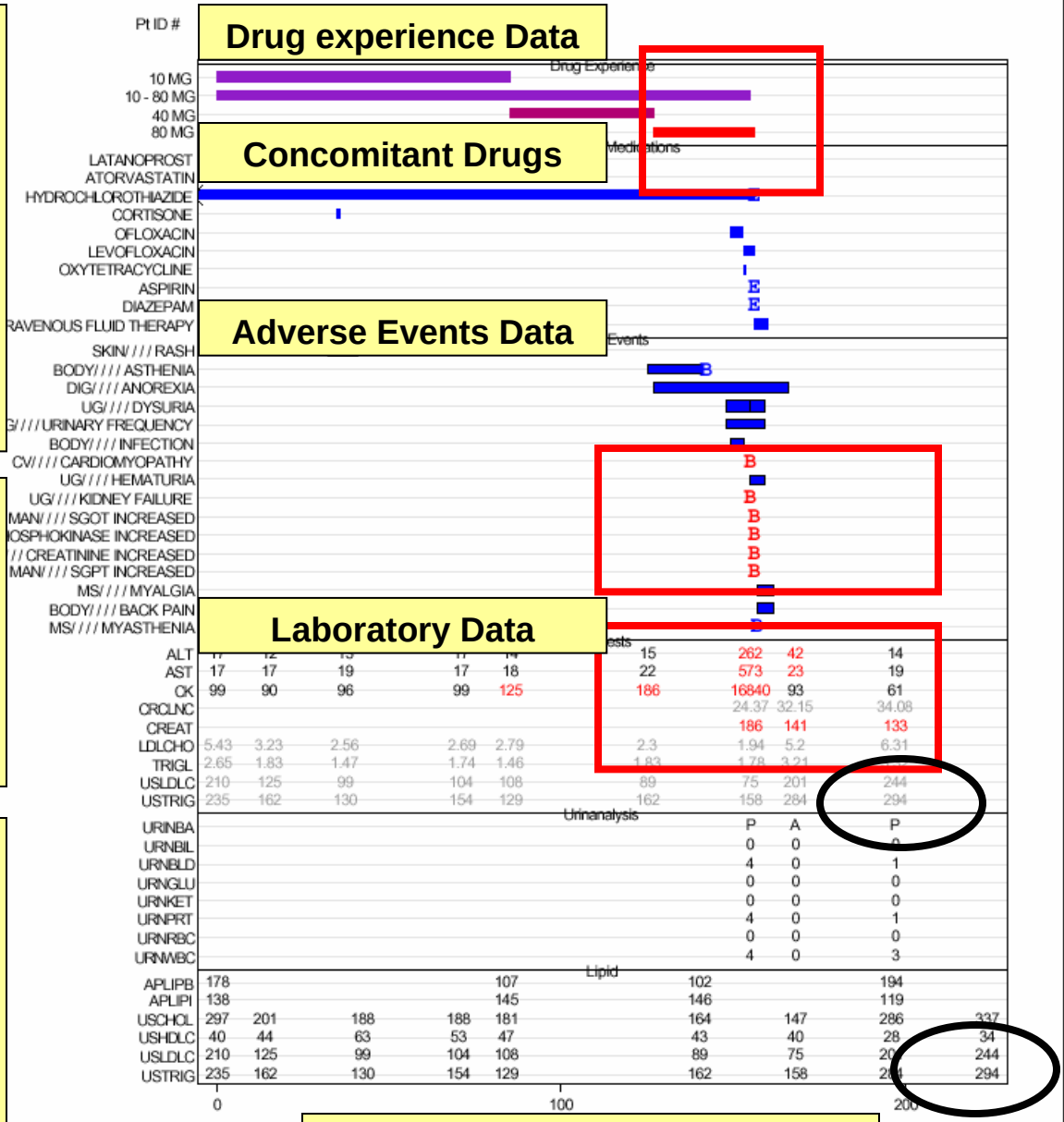
B: Begin no End

E: End no Begin

Individual Patient Profile

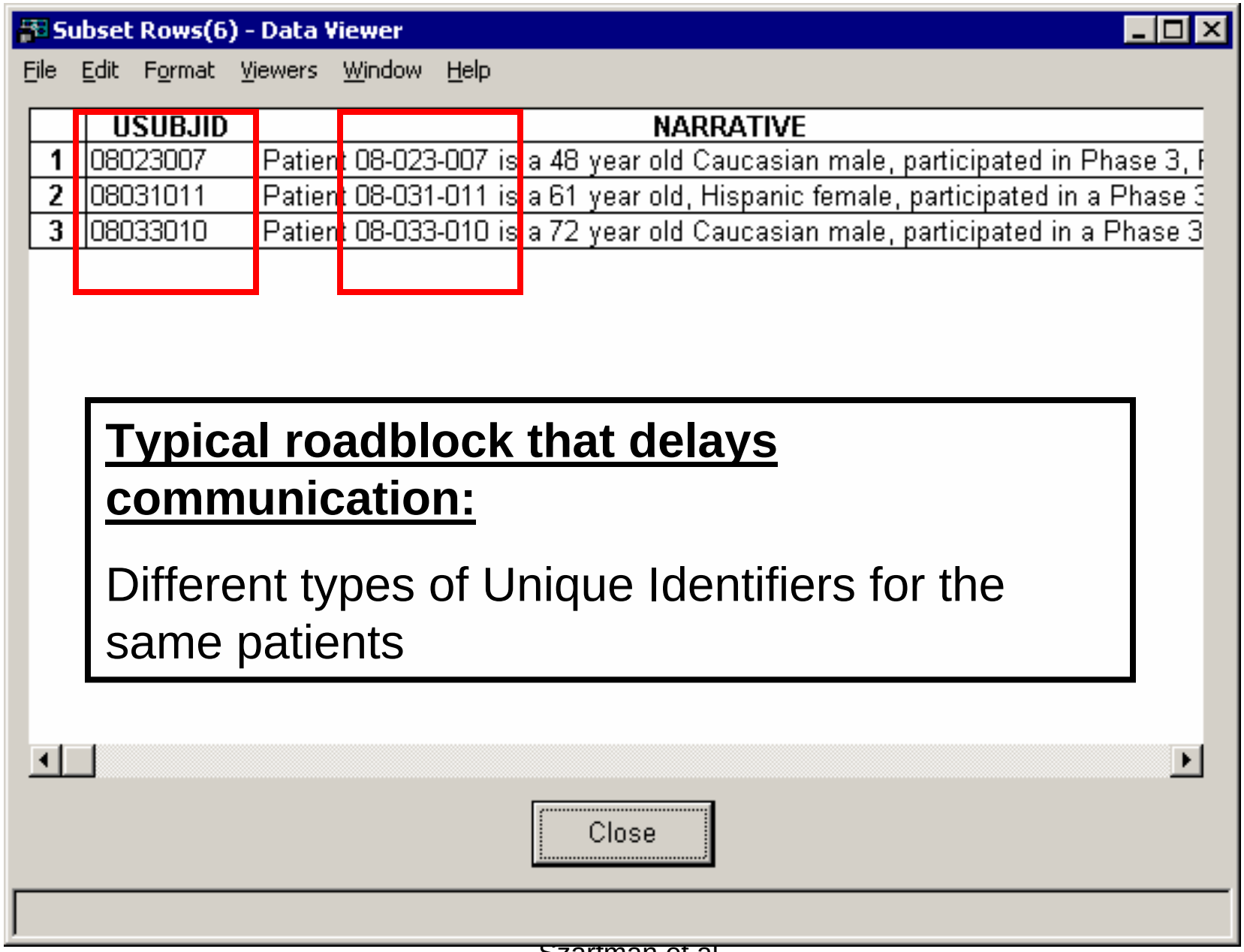
Linkage of several data tables using the same timeline

Other Patients had the same profile with the high dose for the test drug.
This was not seen with the other drugs and with the other doses studied



Narrative

diabetes: N; smoker: N; arteriosclerosis: Y; CO myalgia, k female
nt (during the Week 16 visit), the subject reported a lack of energy that was attributed at the time to recent intensive
travel. The subject was titrated fro ct noted loss of appetite on Day 128 and increasing fatigue on Day 143, but she did not report these
symptoms to the site until approxi d the investigator's office. She was asked to come in and was seen for a suspected urinary tract infection. On Day
156, the subject had severe acute renal failure (COSTART: kidney failure) and severe myopathy (COSTART: myopathy). On Day 157, she was hospitalized with elevated creatine
phosphokinase (COSTART: creatine phosphokinase increased), elevated creatinine (Cr) (COSTART: creatinine increased), elevated SGOT (AST) (COSTART: SGOT increased), and


Subset Rows(6) - Data Viewer
File Edit Format Viewers Window Help
USUBJID
NARRATIVE
1 08023007 Patient 08-023-007 is a 48 year old Caucasian male, participated in Phase 3,
2 08031011 Patient 08-031-011 is a 61 year old, Hispanic female, participated in a Phase 3
3 08033010 Patient 08-033-010 is a 72 year old Caucasian male, participated in a Phase 3
Typical roadblock that delays communication:
Different types of Unique Identifiers for the same patients
Close

Too Many Individual *Data* Standards In Place

- **Hamper communication between**
  - **Statisticians,** clinicians, epidemiologists, clinical safety reviewers, animal pharmacologists, clinical pharmacologists, chemists, and other experts
  - **Consumers, clinicians, regulators, industry, legislators**
- **Hamper advancing the art *and the science* of drug safety monitoring and risk assessment**

Too Many Personal Standards In Place to Understand Safety

- **Safety data that is convincing to one person is unconvincing to another**
- **We do not know systematically what is a precursor of what**
  - e.g., is schizophrenia or drug treatment the precursor of tardive dyskinesia?
  - No systematic objective markers of drug toxicity in place
- **Labeling is no panacea**
  - Messages given need to be understood – We need follow-up of new data
- **Any given reviewer can only look at so much data in one day by using manual methods**

Contrast Between Clinical Practice and Pharmacoepidemiology

- **Paradigm in clinical practice:**
  - **More answers than questions**
    - In everyday clinical practice, practical diagnostic tests provide answers to questions never asked
    - Clinicians frequently are surprised by the results
    - Formal and informal gold standards and positive and negative controls are in place and are evolving

However …

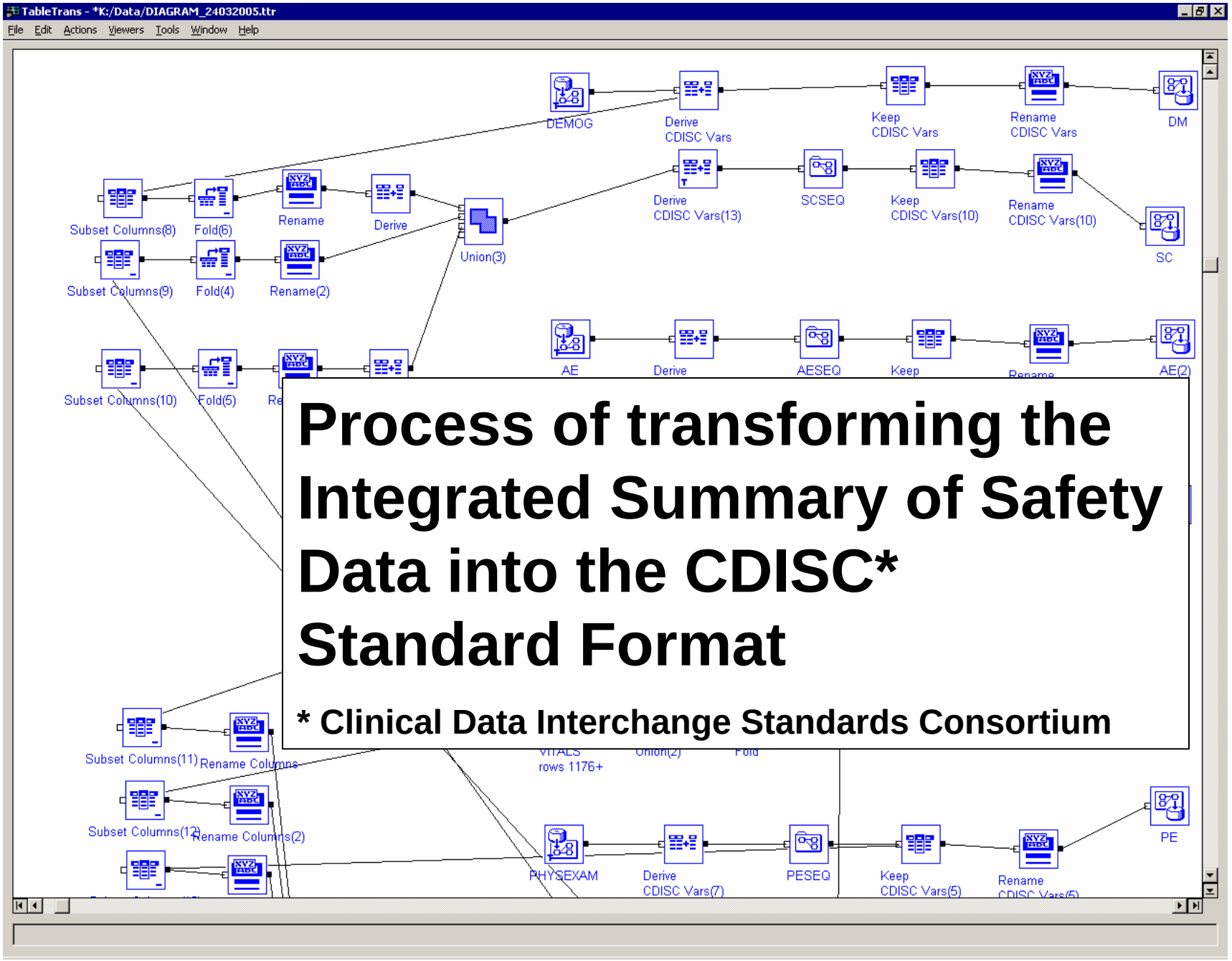
- **Paradigm in traditional statistics and epidemiology**
  - Harm, has to be pre-specified – but experts do not know how to pre-specify harm, patients at risk, and records to study in a systematic way
- **In drug safety we also need practical tools capable of giving us answers to questions that we do not know how to pre-specify**
  - Practical tools are easier to implement when uniform data standards are in place

Practical Tools For Speeding the Analysis of Complex Safety Data

- ***Drugs under review for approval***
- ***Drugs already on the market***

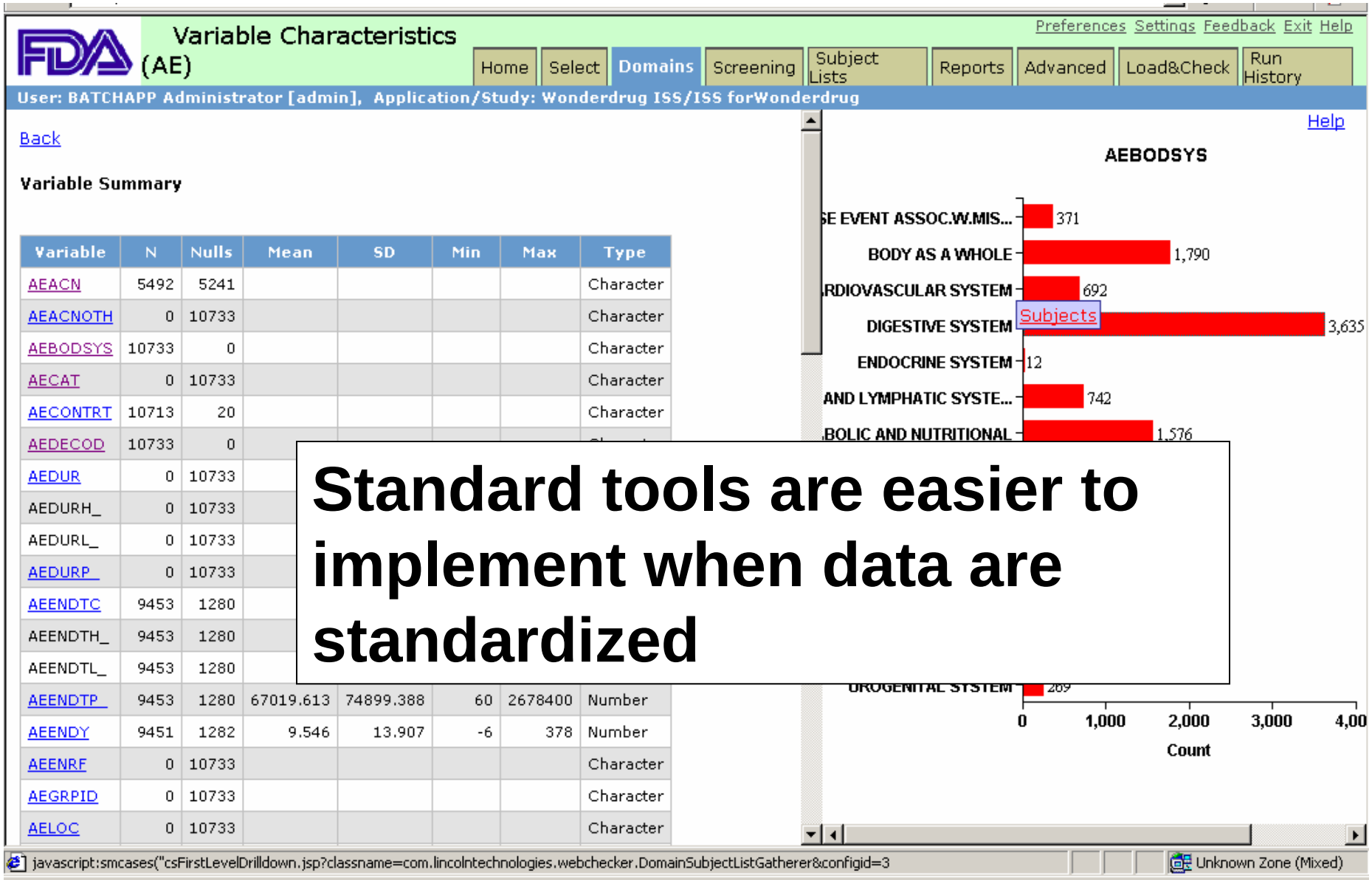
***Example for drugs under review for approval*:**

Application of emerging data standards and data visualization tools to speed the review of the integrated clinical trial data of t*wo NDAs*


TableTrans - *K:/Data/DIAGRAM_24032005.ttr
File Edit Actions Viewers Tools Window Help
DEMOG
Derive
CDISC Vars
Keep
CDISC Vars
Rename
CDISC Vars
DM
Subset Columns(8)
Fold(6)
Rename
Derive
Union(3)
Derive
CDISC Vars(13)
SCSEQ
Keep
CDISC Vars(10)
Rename
CDISC Vars(10)
SC
Subset Columns(9)
Fold(4)
Rename(2)
AE
Derive
AESEQ
Keep
AE(2)
Subset Columns(10)
Fold(5)
Process of transforming the Integrated Summary of Safety Data into the CDISC* Standard Format
* Clinical Data Interchange Standards Consortium
Subset Columns(11)
Rename Columns
VITALS
rows 1176+
Union(2)
Fold
Subset Columns(12)
Rename Columns(2)
PHYSEXAM
Derive
CDISC Vars(7)
PESEQ
Keep
CDISC Vars(5)
Rename
PE

Use of CDISC's standardized data definitions and variable names

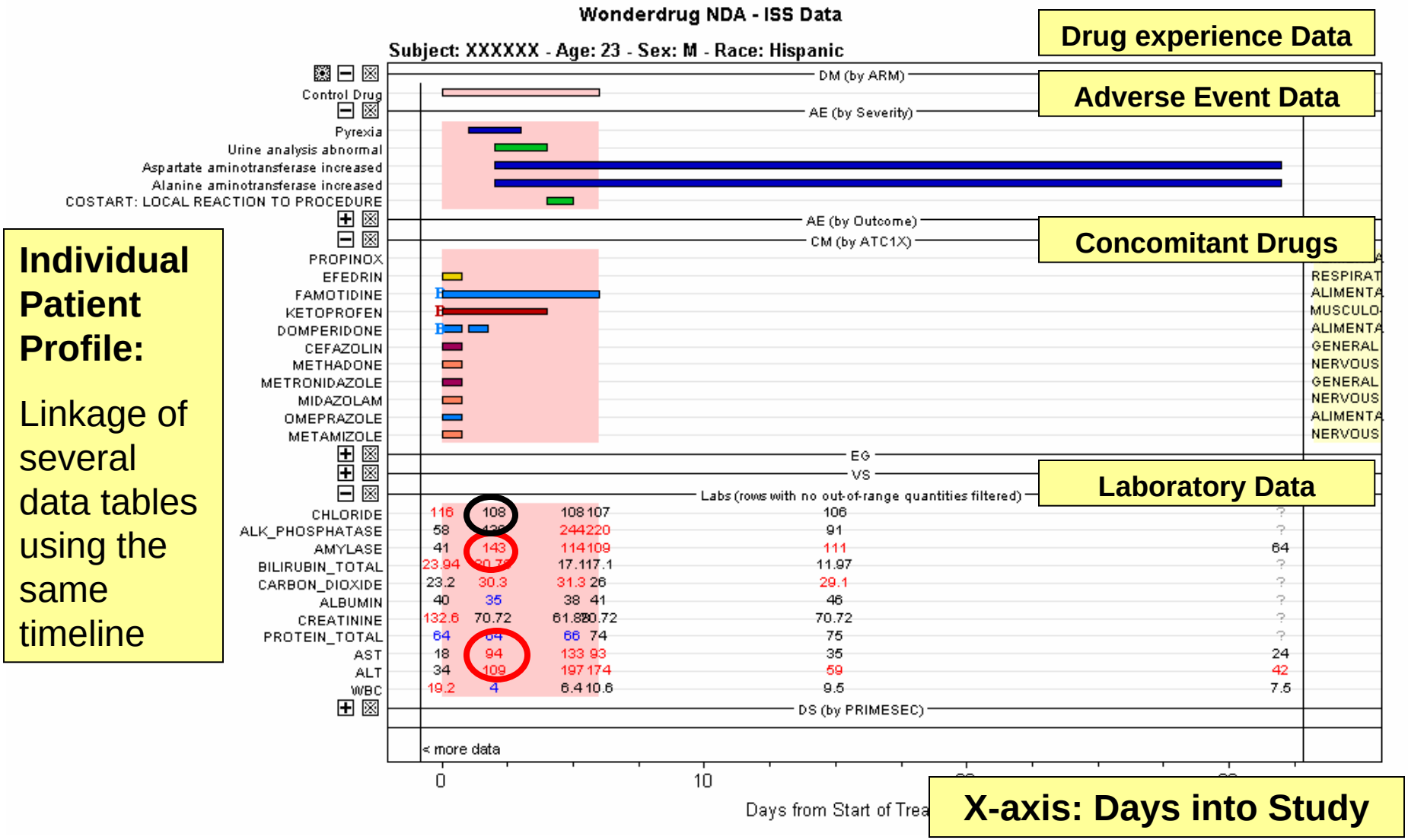
- **Simplifies the application of standardized web-based analytical tools**
- **The use of the same data and tools**
  - Reduces the time by reviewers on data management and analysis
  - Enhances communication between reviewers of the two NDAs and their supervisors
  - Improves data interpretation, and the quality, speed, and transparency of review process


Variable Characteristics
(AE)
Preferences Settings Feedback Exit Help
Home Select Domains Screening Subject Lists Reports Advanced Load&Check Run History
User: BATCHAPP Administrator [admin], Application/Study: Wonderdrug ISS/ISS forWonderdrug
Back
Variable Summary
Variable N Nulls Mean SD Min Max Type
AEACN 5492 5241 Character
AEACNOTH 0 10733 Character
AEBODSYS 10733 0 Character
AECAT 0 10733 Character
AECONTRT 10713 20 Character
AEDECOD 10733 0
AEDUR 0 10733
AEDURH_ 0 10733
AEDURL_ 0 10733
AEDURP_ 0 10733
AEENDTC 9453 1280
AEENDTH_ 9453 1280
AEENDTL_ 9453 1280
AEENDTP_ 9453 1280 67019.613 74899.388 60 2678400 Number
AEENDY 9451 1282 9.546 13.907 -6 378 Number
AEENRF 0 10733 Character
AEGRPID 0 10733 Character
AELOC 0 10733 Character
Help
AEBODSYS
SE EVENT ASSOC.W.MIS... 371
BODY AS A WHOLE 1,790
RDIOVASCULAR SYSTEM 692
DIGESTIVE SYSTEM 3,635
Subjects
ENDOCRINE SYSTEM 12
AND LYMPHATIC SYSTE... 742
BOLIC AND NUTRITIONAL 1,576
UROGENITAL SYSTEM 269
0 1,000 2,000 3,000 4,00
Count
Standard tools are easier to implement when data are standardized
javascript:smcases("csFirstLevelDrilldown.jsp?classname=com.lincolntechnologies.webchecker.DomainSubjectListGatherer&configid=3
Unknown Zone (Mixed)

Assessing Potential Liver Injury by Analyzing Increases in Serum Alanine Aminotransferase (ALT) and Total Serum Bilirubin (TBILI) IN ONE STEP

Individual Patient Profile:

Linkage of several data tables using the same timeline



Examples for drugs already on the market:

Data Mining Voluntary Reports of Adverse Drug Events

The two Methods We Use:

1. MGPS (Multi-Item Gamma Poisson Shrinker)

2. HBLR (Hierarchical Bayesian Logistic Regression)

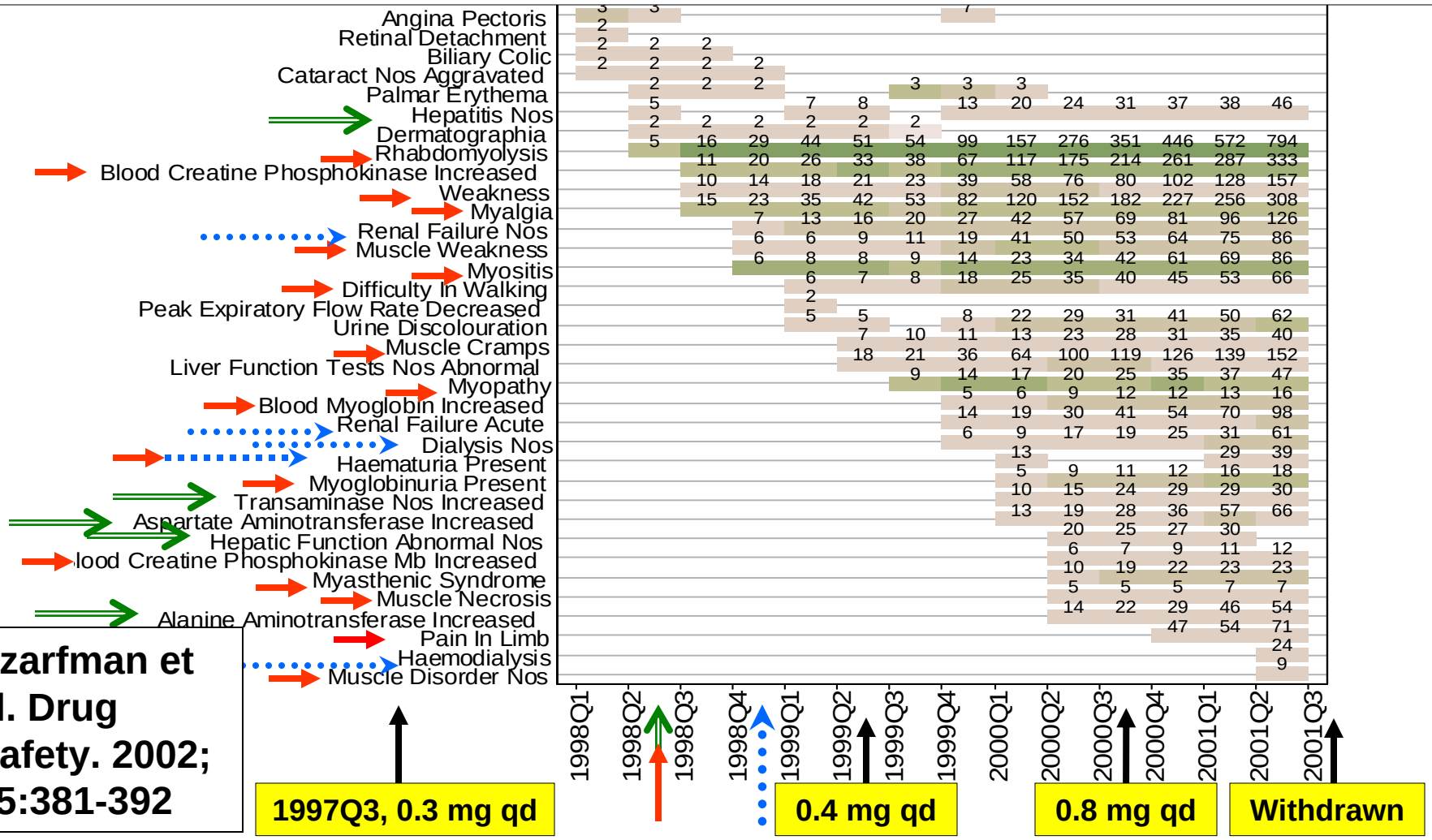
The AERS (Adverse Event Reporting System) Database

- **Most of the safety issues leading to drug withdrawal from the market were discovered using the AERS database**
  - David Graham, Advisory Committee Meeting Presentation on May 18, 2005
- **Contains over 2.5 million voluntary reports on marketed drugs since 1968**
- **Data are pre-processed and cleaned prior to data mining**
  - Some duplicate reports remain – Multiple manufacturers may report the same case

1. MGPS (Multi-Item Gamma Poisson Shrinker) data mining program

- **Algorithm for detecting signals of higher-than-expected drug-event combinations**
- **Through statistical modeling, identifies signals in the large AERS database**
- **Handles complex stratification (>900 categories)**
- **Developed by Dr. William DuMouchel**
  - Work initiated by Dr. Ana Szarfman
- **Currently, a CRADA is in place to develop data mining tools to assess the safety of drugs**
  - Drs. Paul Seligman and Miles Braun are principal investigators

Cerivastatin DM Story — A notable example of detection of early MGPS signals!



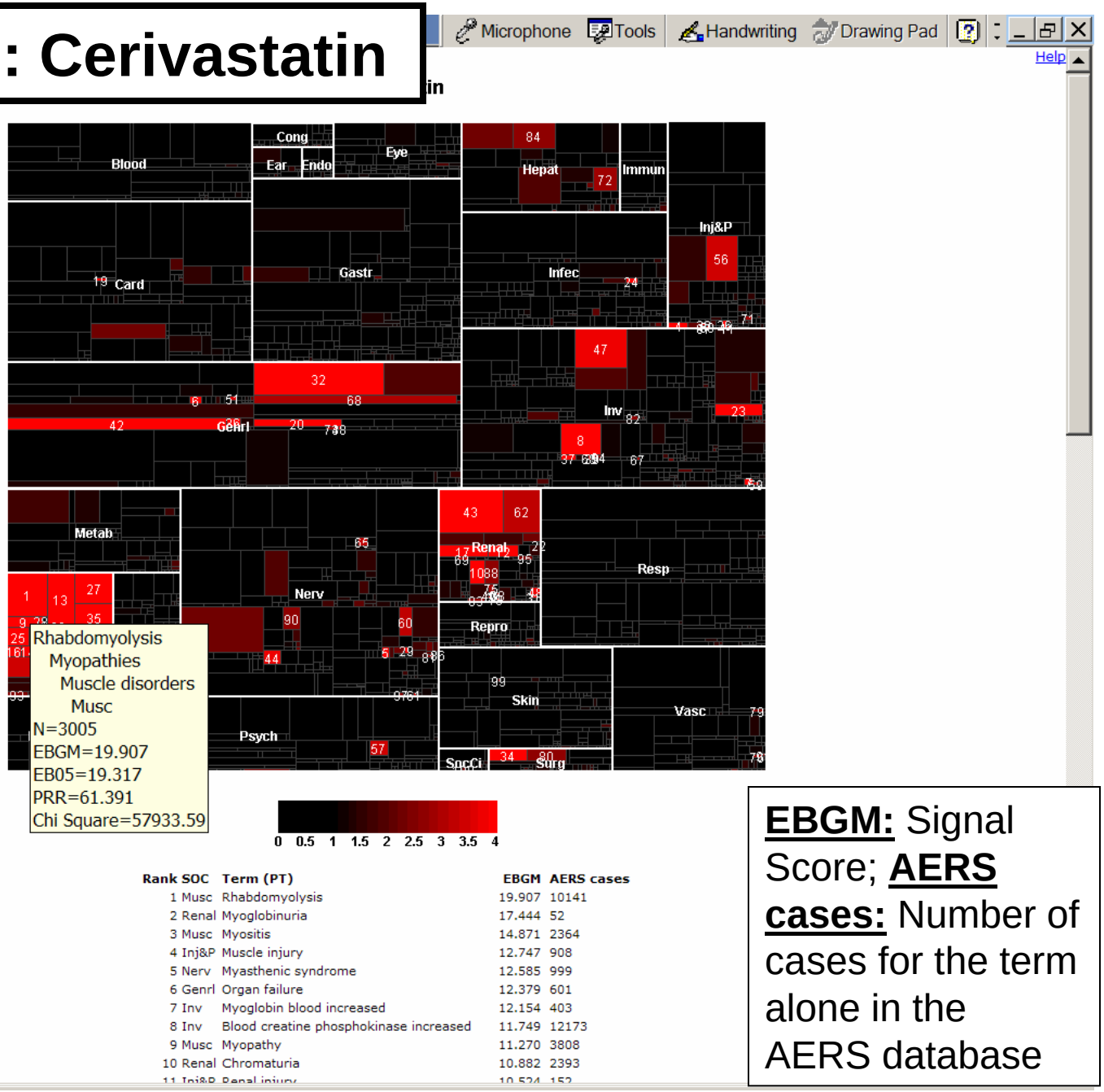
Szarfman et al. Drug Safety. 2002; 25:381-392

Answers: Mild liver event codes; severe muscle events including rhabdomyolysis; and renal failure signals beginning in 1998 (darker shading corresponds to stronger signals)

Cerivastatin — A notable example of detection of early MGPS signals!

- **Rhabdomyolysis and “CPK increase” in 1998Q2**
- **Renal failure in 1998Q4**
- **Dosage increases occurred twice: From 0.3 mg qd at approval, to 0.4 mg qd in 1999Q2, to 0.8 mg qd in 2000Q3**
- **Bayer sold marketing rights in Europe**
- **Withdrawn from the market in 2001Q3**
- **It cannot be argued that these problems were known before these actions took place**

Sector Map: Cerivastatin



Rank	SOC	Term (PT)	EBGM	AERS cases
1	Musc	Rhabdomyolysis	19.907	10141
2	Renal	Myoglobinuria	17.444	52
3	Musc	Myositis	14.871	2364
4	Inj&P	Muscle injury	12.747	908
5	Nerv	Myasthenic syndrome	12.585	999
6	Genrl	Organ failure	12.379	601
7	Inv	Myoglobin blood increased	12.154	403
8	Inv	Blood creatine phosphokinase increased	11.749	12173
9	Musc	Myopathy	11.270	3808
10	Renal	Chromaturia	10.882	2393

Answer: **The strong renal signals were unexpected**

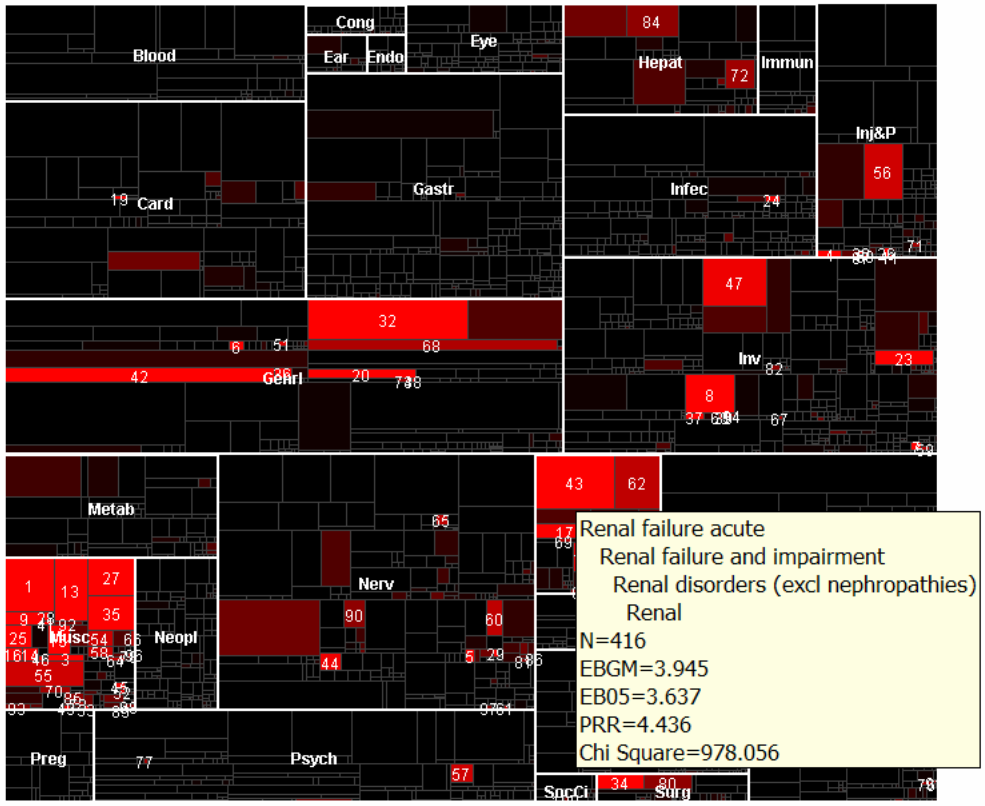
EBGM: Signal Score; **AERS cases:** Number of cases for the term alone in the AERS database

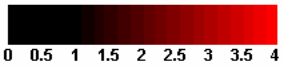
Sector Map:
Assessing Safety Profiles Using a Dictionary of Medical Terms

- **The color, size, position in space, grouping, and ranking of tiles provide a "big picture" overview of the adverse event profile of a drug**

- **Color:**
  - **Red corresponds to stronger signals**
- **Size:**
  - **A large tile (with a white border) defines each SOC (System Organ Class) in the MedDRA dictionary**
  - **The box size for each PT term (preferred adverse event term) is based on the number of serious cases of the term alone in the AERS database**
  - **The size of each PT is stable over displays of different drugs**
- **Position in space:**
  - **SOCs and PTs are always represented in the same area of the sector map**
  - **The position of each SOC and PT is stable over displays of different drugs**
- **Grouping:**
  - **PTs are grouped by high level term (HLT), high level group (HLGT), and SOC**
- **Ranking:**
  - **PTs are ranked in descending order of EBGM values**

Sector Map: Cerivastatin

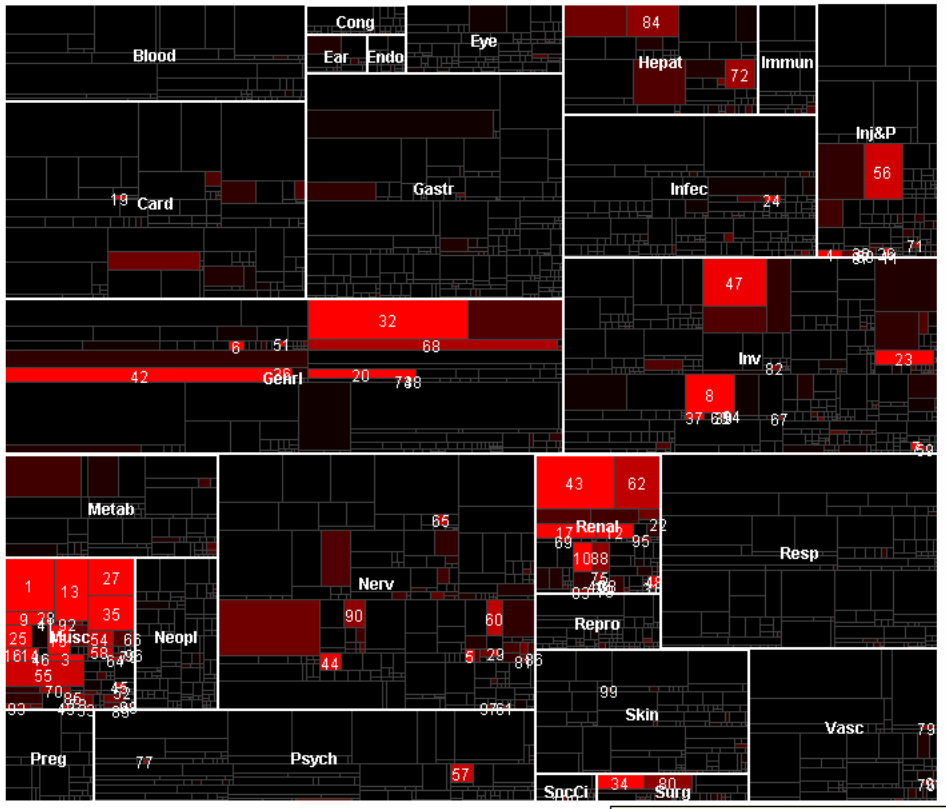




Rank	SOC	Term (PT)	EBGM	AERS cases
1	Musc	Rhabdomyolysis	19.907	10141
2	Renal	Myoglobinuria	17.444	52
3	Musc	Myositis	14.871	2364
4	Inj&P	Muscle injury	12.747	908
5	Nerv	Myasthenic syndrome	12.585	999
6	Genrl	Organ failure	12.379	601
7	Inv	Myoglobin blood increased	12.154	403
8	Inv	Blood creatine phosphokinase increased	11.749	12173
9	Musc	Myopathy	11.270	3808
10	Renal	Chromaturia	10.882	2393

EBGM: Signal Score; **AERS cases:** Number of cases for the term alone in the AERS database

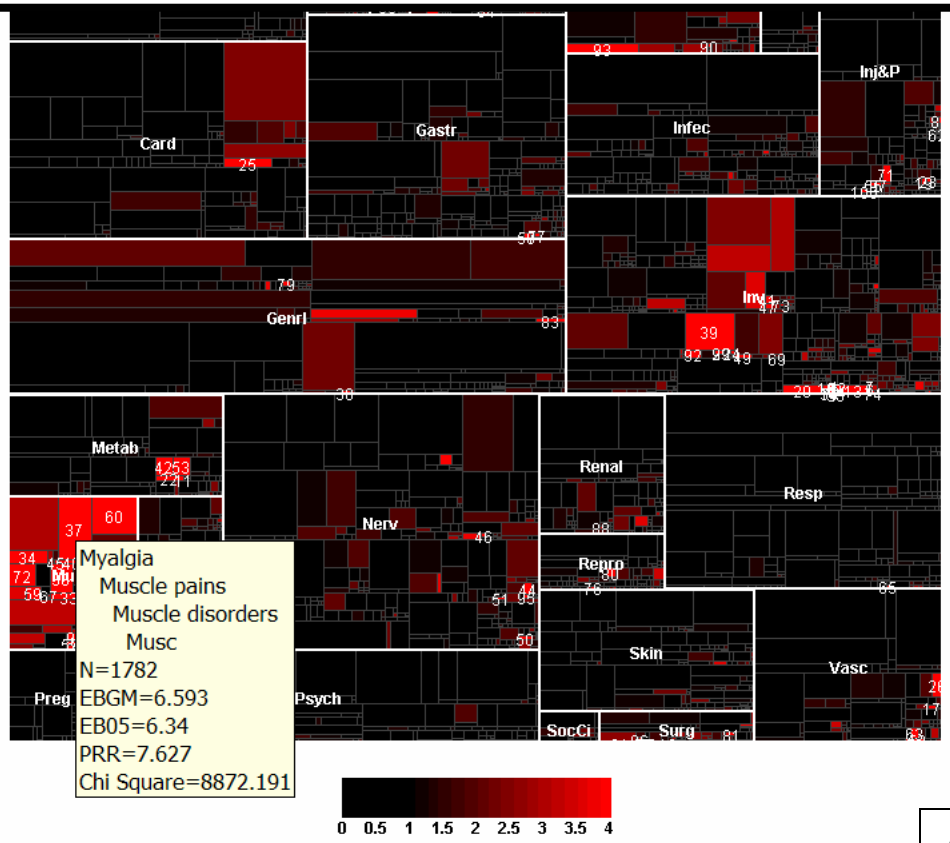
Sector Map: Cerivastatin



Dialysis
Blood and blood product treatment
Haematological and lymphoid tissue therapeutic procedures
Surg
N=184
EBGM=4.581
EB05=4.052
PRR=5.248
Chi Square=559.054

Rank	SOC	Term (PT)		
1	Musc	Rhabdomyolysis		
2	Renal	Myoglobinuria		
3	Musc	Myositis		
4	Inj&P	Muscle injury		
5	Nerv	Myasthenic syndrome		
6	Genrl	Organ failure		
7	Inv	Myoglobin blood increased		
8	Inv	Blood creatine phosphokinase increased	11.749	12173
9	Musc	Myopathy	11.270	3808
10	Renal	Chromaturia	10.882	2393
11	Inj&P	Renal injury	10.524	152

Sector Map: Atorvastatin – answers the compared to what question

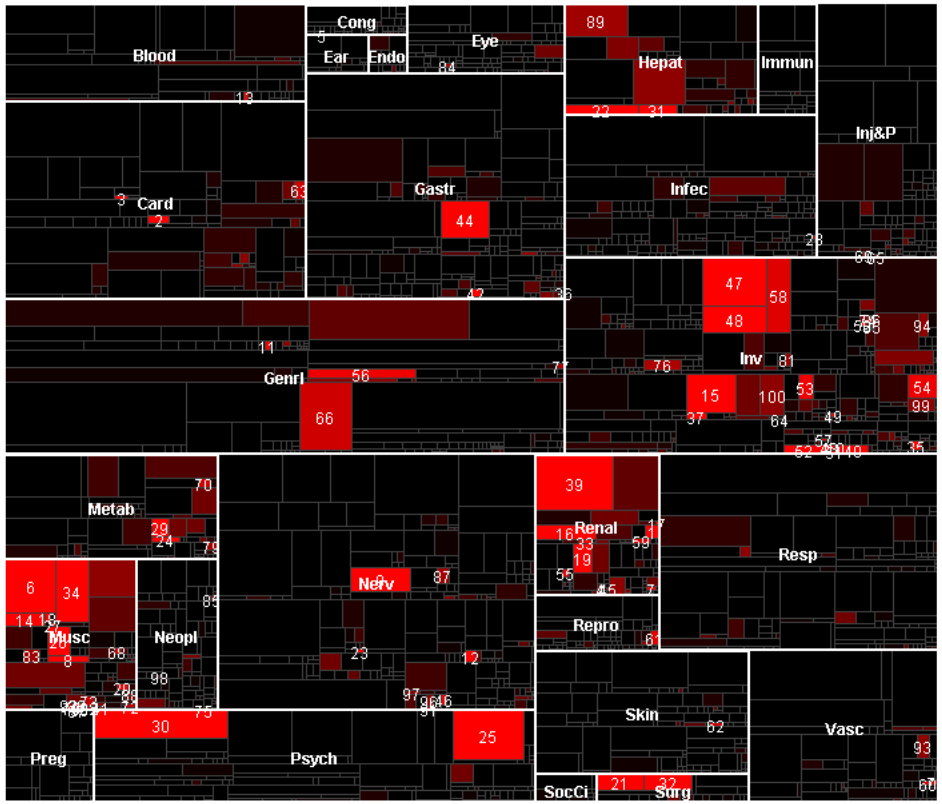


Rank	SOC	Term (PT)	EBGM	AERS cases
1	Inv	Low density lipoprotein decreased	26.914	30
2	Inv	High density lipoprotein	22.908	32
3	Inv	Low density lipoprotein	22.555	36
4	Inv	Blood cholesterol	22.166	66
5	Inv	High density lipoprotein decreased	21.263	778
6	Inv	Blood triglycerides abnormal	18.109	43
7	Inv	Blood triglycerides	17.737	53
8	Musc	Bunion	16.427	36
9	Inv	Blood creatine phosphokinase	14.996	134
10	Inv	Low density lipoprotein increased	14.765	500
11	Metab	Hypocholesterolaemia	12.385	168

<u>Answer:</u> Huge differences between cerivastatin and atorvastatin renal signals

<u>AERS cases:</u>
Number of cases for the term alone in the AERS database

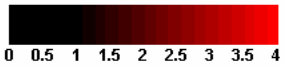
Sector Map: Gemfibrozil



Rank	SOC	Term (PT)	EBGM	AERS cases
1	Renal	Obstructive uropathy	46.801	697
2	Card	Conduction disorder	44.614	747
3	Card	Electromechanical dissociation	38.949	222
4	Renal	Myoglobinuria	30.148	52
5	Cong	Mesomelia	27.815	6
6	Musc	Rhabdomyolysis	24.710	10141
7	Renal	Urinary tract obstruction	24.325	208
8	Musc	Myositis	23.173	2364
9	Nerv	Nervous system disorder	21.087	5395
10	Inv	High density lipoprotein decreased	16.253	778
11	Genrl	Obstruction	15.782	232

<u>Answer:</u> Strong renal signals mostly due to concomitant cerivastatin use

Sector Map: Gemfibrozil w/o Cerivastatin



Rank	SOC	Term (PT)	EBGM	AERS cases
1	Cong	Mesomelia	35.548	6
2	Gastr	Rectal polyp	33.798	61
3	Musc	Rhabdomyolysis	32.977	10141
4	Inv	High density lipoprotein decreased	23.038	778
5	Musc	Myositis	21.700	2364
6	Hepat	Cholelithiasis	17.516	3842
7	Musc	Muscle necrosis	17.061	212
8	Inj&P	Hepatic rupture	14.156	31
9	Musc	Myopathy	13.912	3808
10	Blood	Splenic infarction	13.772	128
11	Musc	Polymyositis	13.100	286

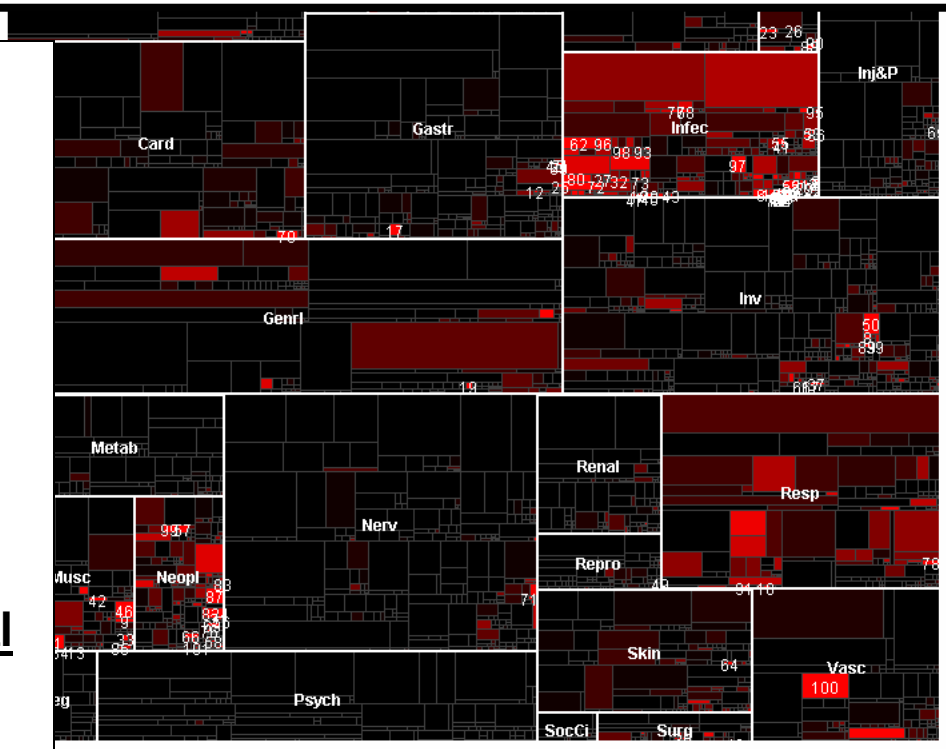
Answer: Some important renal signals remain after removing cerivastatin cases

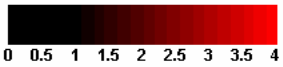
Sector Map: Infliximab, A Drug with a Different Safety Profile Pattern

<u>Answers:</u> Early detection of an extraordinary and unique array of devastating opportunistic infections, lymphoma, and SLE.

<u>This is a historical control issue:</u>

Unique and surprising, both within AERS and any real world clinical experience





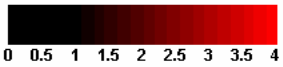
Rank	SOC	Term (PT)	EBGM	AERS cases
1	Infec	Meningitis listeria	28.383	40
2	Infec	Histoplasmosis	26.090	94
3	Infec	Histoplasmosis disseminated	24.248	33
4	Infec	Disseminated tuberculosis	23.029	188
5	Infec	Acute pulmonary histoplasmosis	22.901	7
6	Infec	Tuberculosis	22.483	753
7	Infec	Lymph node tuberculosis	21.871	39
8	Inv	DNA antibody positive	21.015	199
9	Musc	Lupus-like syndrome	20.248	383
10	Infec	Listeriosis	20.046	86
11	Infec	Tuberculosis gastrointestinal	18.939	18

<u>EBGM:</u> Signal Score; **<u>AERS cases:</u>** Number of cases for the term alone in the AERS database

Another Drug Recently in the News

Drawing Pad

Help



Rank	SOC	Term (PT)	EBGM	AERS cases
1	Renal	Urine flow decreased	24.037	134
2	Skin	Rash generalised	12.704	2503
3	Skin	Stevens-Johnson syndrome	9.422	6105
4	Skin	Dermatitis	9.345	117574
5	Skin	Erythema multiforme	8.811	5225
6	Resp	Pharyngeal oedema	6.591	1384
7	Skin	Rash pruritic	4.488	2758
8	Renal	Renal colic	4.475	1427
9	Skin	Rash papular	4.183	1257
10	Gastr	Mouth ulceration	4.127	5138
11	Resp	Throat tightness	4.127	1711

Answer: Early detection of a unique profile of serious skin events

This is quite a unique finding within this class of drugs and indication

These events do not happen very frequently

EBGM: Signal Score; **AERS cases:** Number of cases for the term alone in the AERS database

What Opportunities Does MGPS Provide?

- **Even when specific questions are not asked**
  - MGPS provides a large collection of positive and negative controls
  - Provides reminders of what other experts know that serve to assess plausibility of results
  - Provides clues to complex safety issues quickly
  - Signals important information that might be missed if the question is not asked
  - Aids in predicting future trends or behaviors (e.g., of drugs in same class)
  - Enables decision-makers to make proactive, knowledge-driven decisions

2. Hierarchical Bayesian Logistic Regression (HBLR) Data Mining

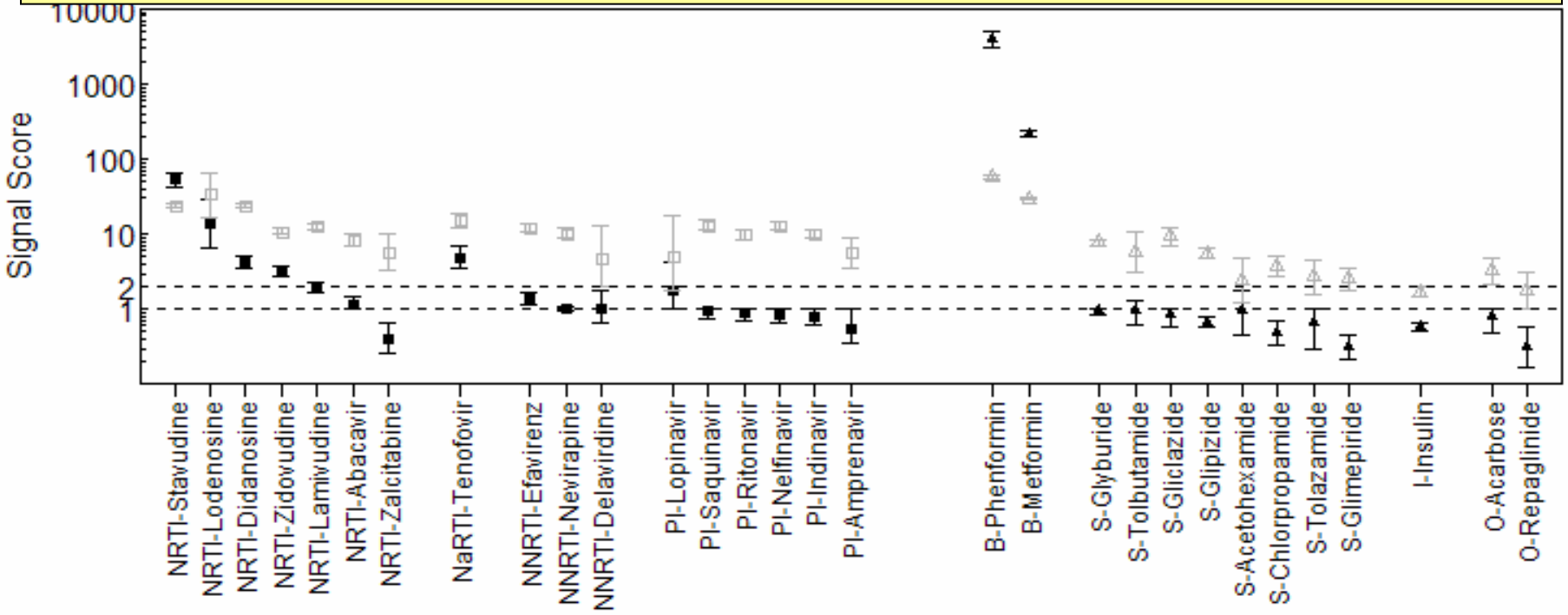
- **Unraveling the Individual Toxicities of Drugs Used in Polytherapy**

Signal Scores and Confidence Intervals for Lactic Acidosis by Drug Using MGPS (gray open symbols) and HBLR (black filled-in symbols) Grouped by Anti-HIV (Squares) and Anti-Diabetes (Triangles) Drugs

HBLR Answer: **More discriminating detection of lactic acidosis associations than MGPS for drugs that are commonly used in polytherapy for both HIV and diabetes**

Clinical knowledge help you discover some built in positive and negative controls.

These results are consistent with current knowledge about these drugs



HBLR Signal Scores are exponentials of logistic regression coefficients, interpreted as odds ratios for the association of lactic acidosis with each drug, adjusted for the presence or absence of all other drugs.

What Opportunities Does HBLR Provide?

- **More discriminating detection than MGPS for drugs that are commonly used in polytherapy**
- **Individual drug safety profiles are corrected for the confounding induced by the presence of other drugs (with their own potentially strong adverse event associations) throughout the database**
- **HBLR is a useful adjunct to MGPS in post-marketing safety assessments, especially in polytherapy regimens**

Take Home Messages

- **We need single interoperable standards in place (HHS Secretary Leavitt)**
- **Standardized data collection and better reporting**
- **Systematic/simultaneous computer approaches to greatly assist in understanding data and biases**
- **Link expertise to get the job done using the same data —** Statisticians, clinicians, epidemiologists, safety evaluators, data management professionals, and other experts

Take Home Messages (continued)

- **We need to develop together new tools to find consistent clues across databases/reference information**
- **Develop rigorous methods for reaching decisions when confronted with inconclusive but worrisome results**
- **With interoperable standards, statisticians, epidemiologists, and other analysts will be in a great position to improve drug safety and communication between all parties involved (consumers, clinicians, regulators, industry, and legislators)**

References

Levine JG, Szarfman A. Standardized Data Structures and Visualization Tools: A way to accelerate the regulatory review of the integrated summary of safety of new drug Applications. Biopharmaceutical Report, 4:12-7,1996

Szarfman A, Talarico L, Levine JG. Chapter 4.21. Analysis and Risk Assessment of Hematological Data from Clinical Trials. In Volume 4, Toxicology of the Hematopoietic System, In: Comprehensive Toxicology. 4:363-79,1997. Editors-in-chief: I.Glenn Sipes, Charlene A.McQueen, A. Jay Gandolfi. Elsevier Science Inc.

DuMouchel W. Bayesian Data Mining in Large Frequency Tables, With an Application to the FDA Spontaneous Reporting System. The American Statistician.53:177-90,1999

O'Neill, RT, Szarfman, A. Discussion: Bayesian Data Mining in Large Frequency Tables, With an Application to the FDA Spontaneous Reporting System by William DuMouchel. The American Statistician.53:190-6,1999

DuMouchel W, Pregibon D. Empirical Bayes screening formulti-item associations. In: Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining, San Diego, CA, August 26–9, 2001

O'Neill RT, Szarfman A. Some FDA Perspectives on Data mining for pediatric safety assessment. Workshop on Adverse Drug events in Pediatrics. Curr Ther Res Clin Exp. 62:650-63, 2001

Szarfman A, Machado SG, O'Neill RT. Use of Screening Algorithms and Computer Systems to Efficiently Signal Higher-Than-Expected Combinations of Drugs and Events in the US FDA's Spontaneous Reports Database. Drug Safety 25:381-92, 2002

Szarfman A, Tonning JM, Doraiswamy PM. Pharmacovigilance in the 21st Century: New Systematic Tools for an Old Problem. Pharmacotherapy. 24:1099-104,2004

Szarfman A, DuMouchel W, Fram D, Tonning JM, Almenoff J, Fleischer RD, Levine JG. Lactic Acidosis: Unraveling the Individual Toxicities of Drugs Used in HIV and Diabetes Polytherapy by Hierarchical Bayesian Logistic Regression Data Mining. 2005 FDA Science Forum. Abstract and Poster H-30 http://www.accessdata.fda.gov/scripts/oc/scienceforum/sf2005/Search/preview.cfm?abstract_id=483&backto=author

Cooper CK, Levine JG, Tonning JM, Fram D, Millstein J, Rochester G, Szarfman A. Use of Standards-Based Data and Tools to Improve the Efficiency of the NDA Safety Review. 2005 FDA Science Forum. Abstract and Poster H-03. http://www.accessdata.fda.gov/scripts/oc/scienceforum/sf2005/Search/preview.cfm?abstract_id=339&backto=author

U.S. Health Secretary Hails Promise of Information Technology. Leavitt calls for collaboration on global health information system. http://usinfo.state.gov/gi/Archive/2005/May/18-799284.html