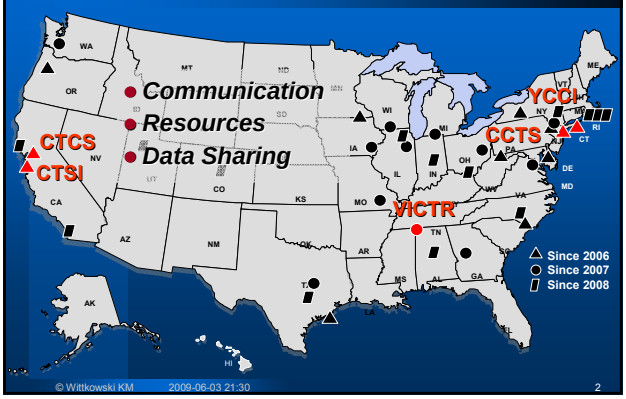
Responding to New Challenges: Statistics in Clinical and Translational Science

KNUT M. WITTKOWSKI, PhD, DSc
The Rockefeller University Hospital
Center for Clinical and Translational Science
New York, NY 10021

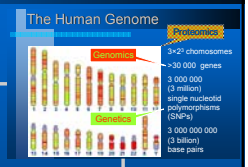
The CTSA Consortium

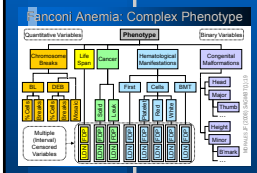


CTS: Statistical Challenges

Non-linear Systems

Basic Science

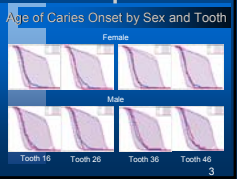




Different Scales

Clinical Research

Complex Structures



Community Impact

U-Statistics for Translational Science

Advantage over methods based on linear models:

Do not assume that all differences of the same magnitude have the same relevance.

- **Insure** that statistical results are biologically meaningful
- Are **robust** to the effect of outliers
- Do **not require** data to be transformed prior to statistical analysis

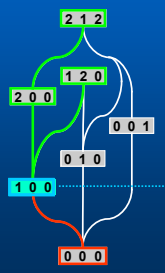
U-statistics lack of a unifying theory:

- Restricted to **uni/bi-variate data** and **simple designs**
- **Hodgepodge** of methods
- Risk of accidental **misspecification**

How to Score Multivariate Data

u scores are computed as *#(lower profiles) − #(higher profiles)*.

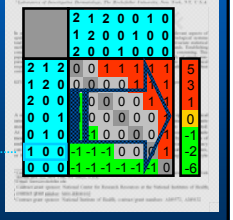
Partial Ordering Lattice



u(1 0 0) = 1−3 = −2

Pairwise Orderings

u Scores

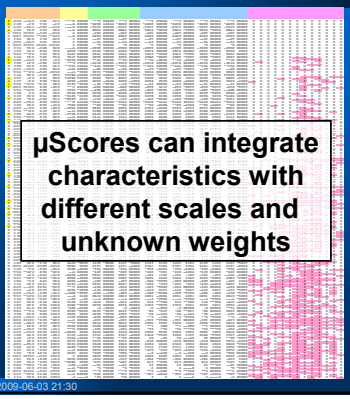


WITTKOWSKI KM (2004) *Stat Med* 23

µScores in Phenomics: FA

Fanconi Anemia Severity Indicators:

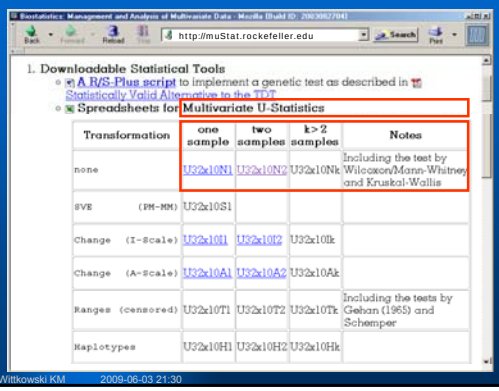
DNA Damage
Survival
Cancer / Leukemia
Bone Marrow Failure
Birth Defects



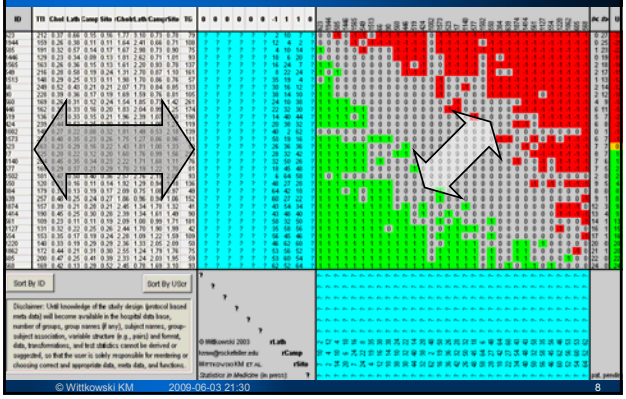
µScores can integrate characteristics with different scales and unknown weights

Patients sorted by FA severity µScores

RUH.BERD Web Tools

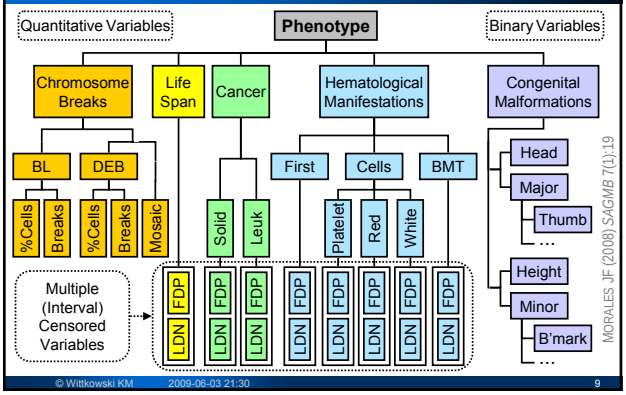


More Variables: Less Information Content



RU CCTS – UCD CTSC

Fanconi Anemia: Complex Phenotype



RU CCTS – UCD CTSC

AEs: Severity / Attribution

Severity		The adverse event ...
1	least severe	... requires none or minimal interventic
...		
4	most severe	... interfes with quality of life

Attribution	The adverse event ...
1 Unrelated	... is clearly not related ...
2 Unlikely	... is doubtfully related ...
3 Possible	... may be related ...
4 Probable	... is likely related ...
5 Definite	... is clearly related ...
	... to the investigational agent

Patient No.	Lightheadedness	Daytime sedation	Tinnitus	Dry skin/Pruritus	Constipation	Lower extremity edema	Difficulty concentrating	Paresthesias	Visual accomodation problems	Hirsutism	Skin pigmentation changes	Changes in TFTs	Bradycardia	Palpitations	Other
17	1/3			1/4	1/3				1/3		1/3				3/3 renal failure* 3?/3
12		2/3	1/4				1/4	1/4		1/3	1/4	1/4			
14	1/2	1/3		1/3	1/4			1/4							
13	1/2		2/3	1/3							1/3				
15		1/3	2/3			1/3							1/3	1/2	2/3 dyspnea on extertion
18		1/3					1/3				1/3		1/3		
19								1/2							*SAE

Tests Based on μ-Scores

The new function '**prentice.test**' integrates / extends (at twice the speed) several well-known tests:

	Conditions	Granularity	Replications	Blocks
mcnemar.test	2	2	≥2	1
SMN.pvalue	2	2	≥2	3
wilcox.test	2	≥2	≥2	1
kruskal.test	≥2	≥2	≥2	1
friedman.test	≥2	≥2	1	≥2
prentice.test	**≥2**	**≥2**	**≥0**	**≥2**

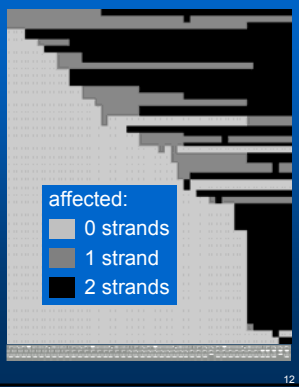
CRAN
http://cran.r-project.org/index.html

CSAN
http://csan.insightful.com/Default.aspx

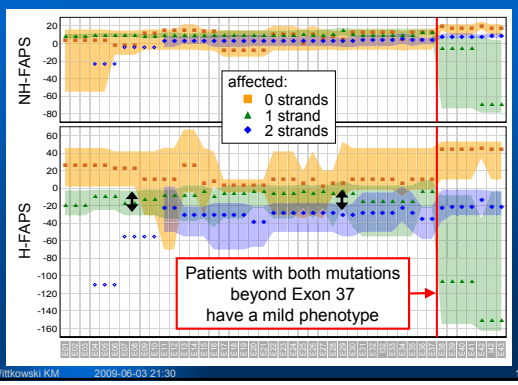
Subjects by Mutation Status

For a subject to develop FA, both strands of the same gene need to be affected.

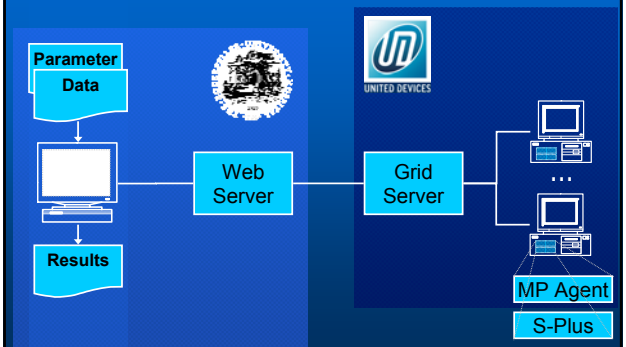
Fig.: Mutation map for 100 FANCA subjects (rows) by exons or introns with variations. The effect of frame-shift, splicing, indels, and stop codon mutations are assumed to affect the protein encoded in all downstream gene regions.



Severity by *FANC A* Mutation Status

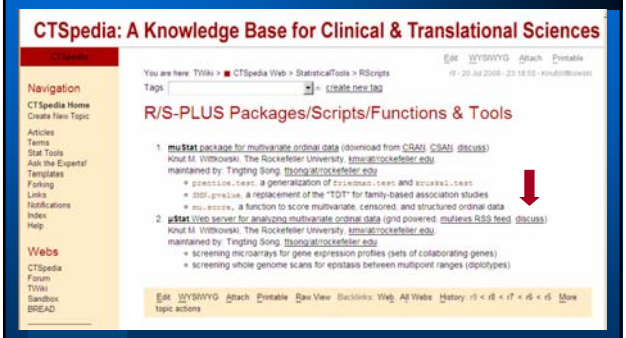


µStat: Grid Architecture



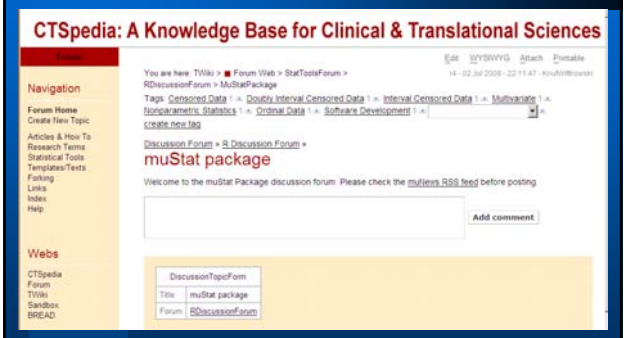
RU CCTS – UCSF CTSI

muStat available on CTSPedia



RU CCTS – UCSF CTSI

muStat Discussion Forums



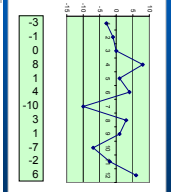
RU CCTS – UCSF CTSI

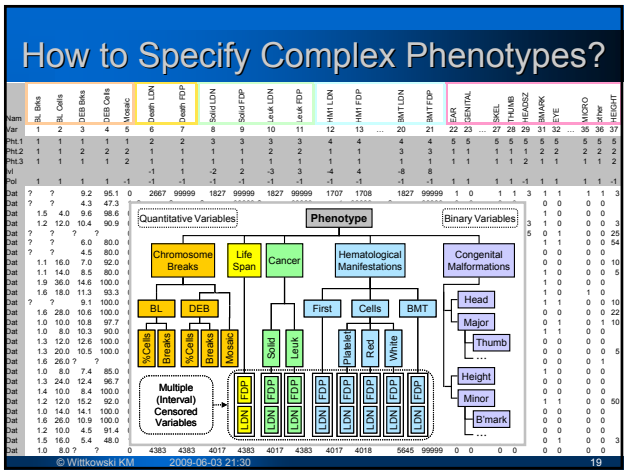
µStat Discussion Forums

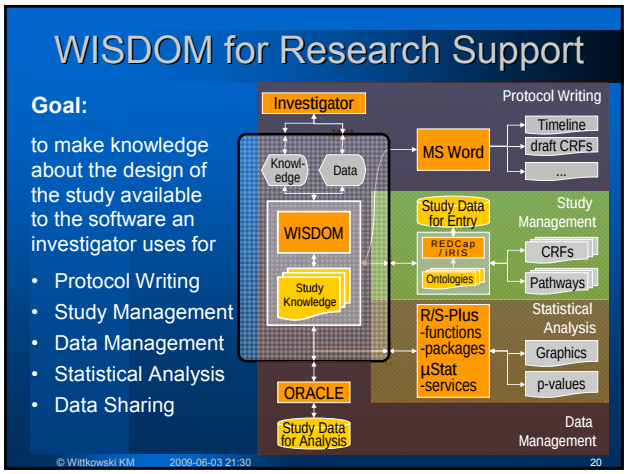
µStat Usage Metrics

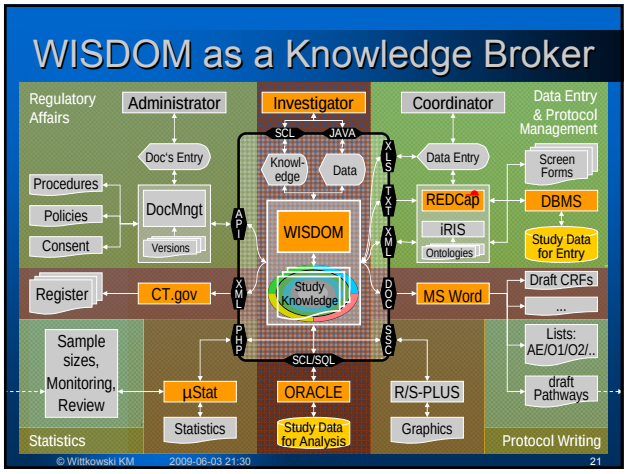
	Period	RU users	RU jobs	.edu users	.edu jobs	Other users	Other jobs	Dev users	Dev jobs
Pht1		1	1	2	2	3	3		
Pht2		1	2	3	4	5	6		
Pol		1	1	1	1	1	1		
Dat	June	2	5	0	0	1	57	1	11
Dat	July	2	20	0	0	1	9	2	2
Dat	August	2	7	0	0	4	30	1	1
Dat	September	2	25	0	0	12	82	1	6
Dat	October	2	11	1	3	1	11	1	6
Dat	November	2	12	0	0	3	50	1	7
Dat	December	1	4	0	0	0	0	1	53
Dat	January	1	21	0	0	2	51	1	4
Dat	February	2	5	1	1	2	13	1	11
Dat	March	1	2	0	0	2	24	1	5
Dat	April	1	12	0	0	2	32	1	10
Dat	May	1	18	1	1	2	35	2	2
	Total	2	140	3	5	13	381	2	119

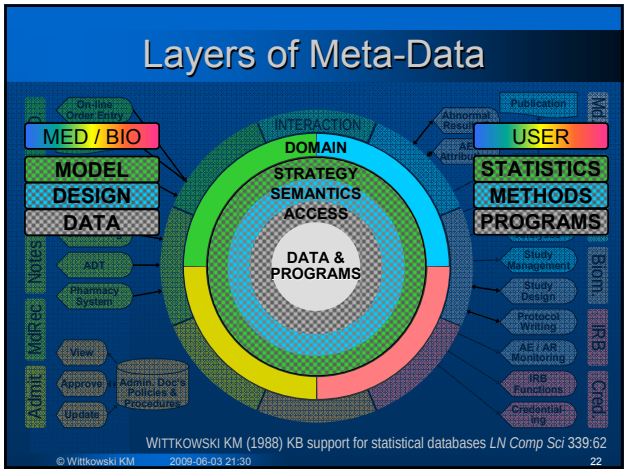
UScr

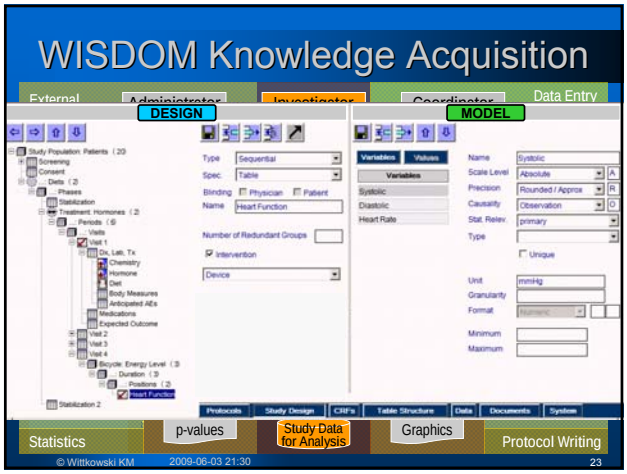


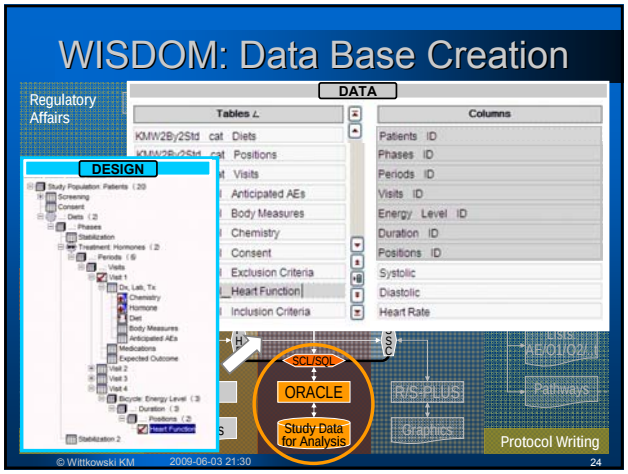

How to Specify Complex Phenotypes?
Phenotype
Quantitative Variables
Binary Variables
Chromosome Breaks
Life Span
Cancer
Hematological Manifestations
Congenital Malformations
BL
DEB
First
Cells
BMT
Head
Major
Thumb
Height
Minor
B'mark
%Cells
Breaks
Mosaic
Solid
Leuk
Platelet
Red
White
Multiple (Interval) Censored Variables
LDN
FDP
© Wittkowski KM
2009-06-03 21:30
19


WISDOM for Research Support
Goal:
to make knowledge about the design of the study available to the software an investigator uses for
Protocol Writing
Study Management
Data Management
Statistical Analysis
Investigator
Knowl-edge
Data
MS Word
Timeline
draft CRFs
...
WISDOM
Study Knowledge
Study Data for Entry
REDCap / iRIS
Ontologies
CRFs
Pathways
R/S-Plus -functions -packages
µStat -services
Graphics
p-values
ORACLE
Study Data for Analysis
© Wittkowski KM
2009-06-03 21:30
20


WISDOM as a Knowledge Broker
Regulatory Affairs
Administrator
Investigator
Coordinator
Data Entry & Protocol Management
Doc's Entry
Procedures
Policies
Consent
DocMngt
Versions
Register
CT.gov
Knowl-edge
Data
WISDOM
Study Knowledge
Data Entry
REDCap
iRIS
Ontologies
Screen Forms
DBMS
Study Data for Entry
MS Word
Draft CRFs
...
Lists: AE/O1/O2/..
draft Pathways
Sample sizes, Monitoring, Review
µStat
Statistics
ORACLE
Study Data for Analysis
R/S-PLUS
Graphics
SCL/SQL
© Wittkowski KM
2009-06-03 21:30
21

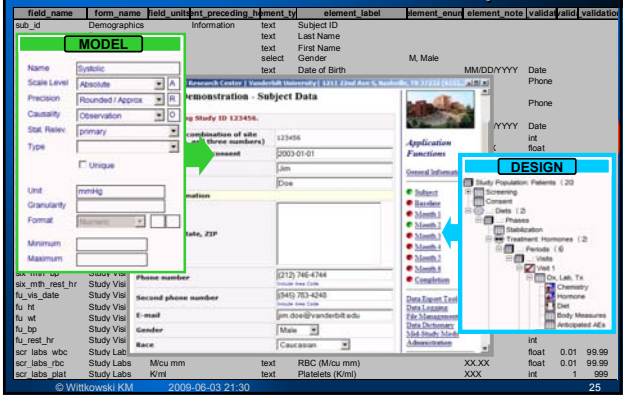

Layers of Meta-Data
MED / BIO
MODEL
DESIGN
DATA
INTERACTION
DOMAIN
STRATEGY
SEMANTICS
ACCESS
DATA & PROGRAMS
USER
STATISTICS
METHODS
PROGRAMS
WITTKOWSKI KM (1988) KB support for statistical databases LN Comp Sci 339:62
© Wittkowski KM
2009-06-03 21:30
22


WISDOM Knowledge Acquisition
Data Entry
DESIGN
MODEL
Statistics
p-values
Study Data for Analysis
Graphics
Protocol Writing
© Wittkowski KM
2009-06-03 21:30
23

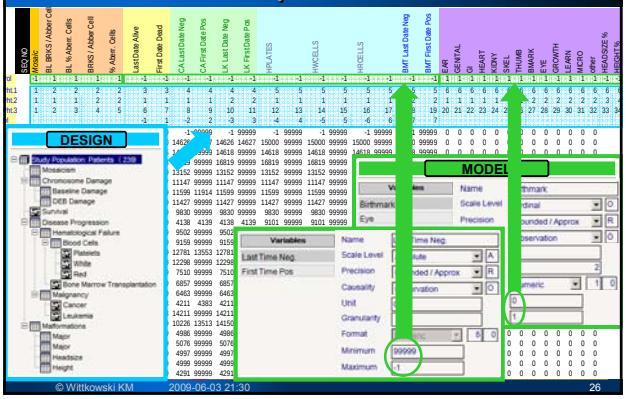

WISDOM: Data Base Creation
Regulatory Affairs
DATA
Tables
Columns
DESIGN
SCL/SQL
ORACLE
Study Data for Analysis
R/S-PLUS
Graphics
Protocol Writing
© Wittkowski KM
2009-06-03 21:30
24

RU CCTS – VICTR / YCCI

WISDOM for REDCap



WISDOM as a µStat User Interface



RU CCTS – UCSF CTSI

NIH Data Sharing Policy

"To ensure that others can use" data, meta-data is needed that documents

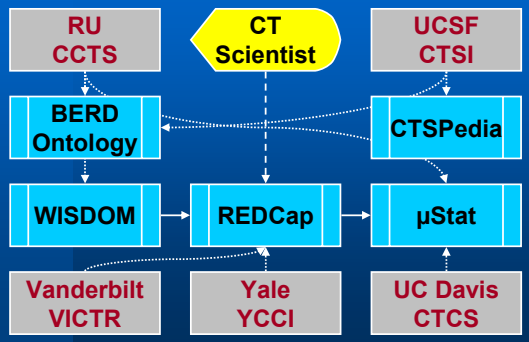
- *"definitions of variables, details about codes* — MODEL
- *the methodology and procedures used to collect the data,* — DESIGN
- *variable field locations, […].* — DATA

The precise content of only the DOMAIN documentation will vary by the

- *scientific area,* — DOMAIN
- *type of data collected,* — MODEL
- *study design, and* — DESIGN
- *characteristics of the dataset."* — DATA

http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

Collaboration on Resources



Plans to Meet Challenges

Develop

- *Best Practices and*
- *Resources*

to assist clinical and trans-lational scientists with

- *Experimental Design*
- *Protocol Writing*
- *Data Management (Oracle)*
- *Data Acquisition (REDCap)*
- *Safety Monitoring*
- *Statistical Analyses (µStat)*
- *Data Sharing*

Using the CTSA environment to

- *Develop an ontology and*
- *Build and share tools (muStat, WISDOM)*

Based on novel non-parametric approaches (u-statistics) and on biostatistical knowledge at the layers

MODEL
DESIGN
DATA