VANDERBILT UNIVERSITY · School of Medicine

Sparsity, Irregularity, and Asynchrony in Structured Medical Data

Slides from Jacob P. VanHouten, PhD MS

EHR <u>data</u> are not (inherently) suited for research

TEST_SNAME	ENTRY_DATE	TEST_VALUE	TEST_UNIT
EOSIAB	2010-07-09T13:22:00	0.41	thou/uL
CCPG	2009-08-03T13:21:00	6	Units
PTT-pt	2010-10-27T03:20:00	66.9	sec
BUN	2009-06-04T05:37:00	15	mg/dL
NRBC	2010-07-09T13:22:00	0	/100_WBC
pH	2009-06-01T17:44:00	7.42	.
SSAIgG	2010-12-12T11:52:00	0.39	Index
WBC	2001-11-25T16:47:00	9	thou/uL
LymAbs	2005-04-24T09:16:00	3	thou/uL
Cl	2009-06-03T04:00:00	107	mEq/L
AlkP	2009-06-01T13:00:00	180	U/L
HgbA1C	2010-08-16T15:09:00	9.6	%

???

08/04/15 13:42 CBC   PCV: 32* Plt-Ct: 433*
07/31/15 09:50 CBC   WBC: 16.4* Hgb: 7.5* PCV: 24* MPV: 9.3 Plt-Ct: 211 RBC: 3.14* MCV: 78*
                      MCH: 23.9* MCHC: 30.7* RDWSD: 57.6* RDW: 20.2*
07/31/15 09:50 Differentl NRBC: 0 NRBC#: 0.02
07/30/15 13:39 CBC   WBC: 11.4* Hgb: 15.3 PCV: 47 MPV: 13.8* Plt-Ct: 205 IPF: 9.8* RBC: 4.71
                      MCV: 99* MCH: 32.5* MCHC: 32.7 RDWSD: 48.3 RDW: 13.3 RetiCt: 1.3
                      RetAbs: 0.060 IRF: 4.9 RETHE: Not Measured
07/30/15 13:39 Differentl NTAuto: 7.38 NEUTRE: 64.9 LYMPRE: 26.4 MONORE: 6.2 EOSIM: 1.9
                      BASORE: 0.3 NEUTAB: 7.38 LYMPAB: 3.00 MONOAB: 0.70 EOSIA: 0.22
                      BASOAB: 0.03 IGRE: 0.3 IGAB: 0.03 NRBC: 0 NRBC#: 0.00
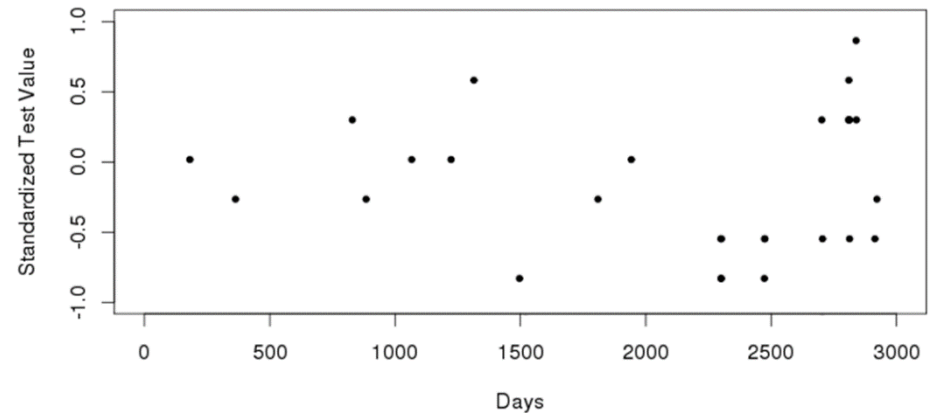07/28/15 14:09 CBC   WBC: 9.2 Hgb: 16.4 PCV: 48 MPV: Not Measured Plt-Ct: 11* ->Critical Lab
                      RBC: 5.43 MCV: 88 MCH: 30.2 MCHC: 34.5 RDWSD: 41.7 RDW: 13.2
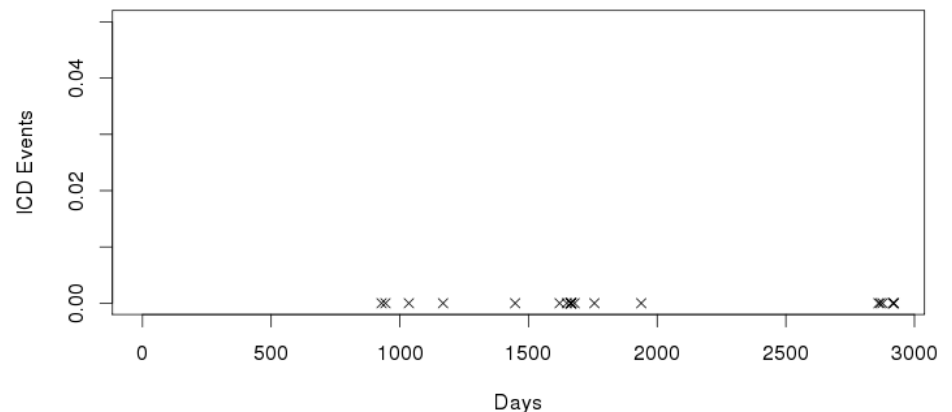07/28/15 14:09 Differentl NTAuto: 5.91

	Var 1	Var 2	Var 3	...	Var *p*	Outcome *y*
Patient 1	*val*	*val*	*val*	...	*val*	0
Patient 2	*val*	*val*	*val*	...	*val*	1
Patient 3	*val*	*val*	*val*	...	*val*	0
...	...	...	...	...	...	...
Patient *n*	*val*	*val*	*val*	*val*	*val*	1

Orientation to Examples

TEST_SNAME	ENTRY_DATE	TEST_VALUE	TEST_UNIT
EOSIAB	2010-07-09T13:22:00	0.41	thou/uL
CCPG	2009-08-03T13:21:00	6	Units
PTT-pt	2010-10-27T03:20:00	66.9	sec
BUN	2009-06-04T05:37:00	15	mg/dL
NRBC	2010-07-09T13:22:00	0	/100_WBC
pH	2009-06-01T17:44:00	7.42	.
SSAIgG	2010-12-12T11:52:00	0.39	Index
WBC	2001-11-25T16:47:00	9	thou/uL
LymAbs	2005-04-24T09:16:00	3	thou/uL
Cl	2009-06-03T04:00:00	107	mEq/L
AlkP	2009-06-01T13:00:00	180	U/L
HgbA1C	2010-08-16T15:09:00	9.6	%

ENTRY_DATE	CODE
2005-10-10T00:00:00.0000000	786.5
2000-02-15T00:00:00.0000000	786.5
2000-06-12T00:00:00.0000000	786.5
2009-11-24T00:00:00.0000000	786.5
2009-11-22T00:00:00.0000000	786.5
2007-04-15T00:00:00.0000000	786.5
2009-11-24T00:00:00.0000000	786.5
2006-09-17T00:00:00.0000000	786.5
2009-11-22T00:00:00.0000000	786.5
2002-07-29T00:00:00.0000000	786.5

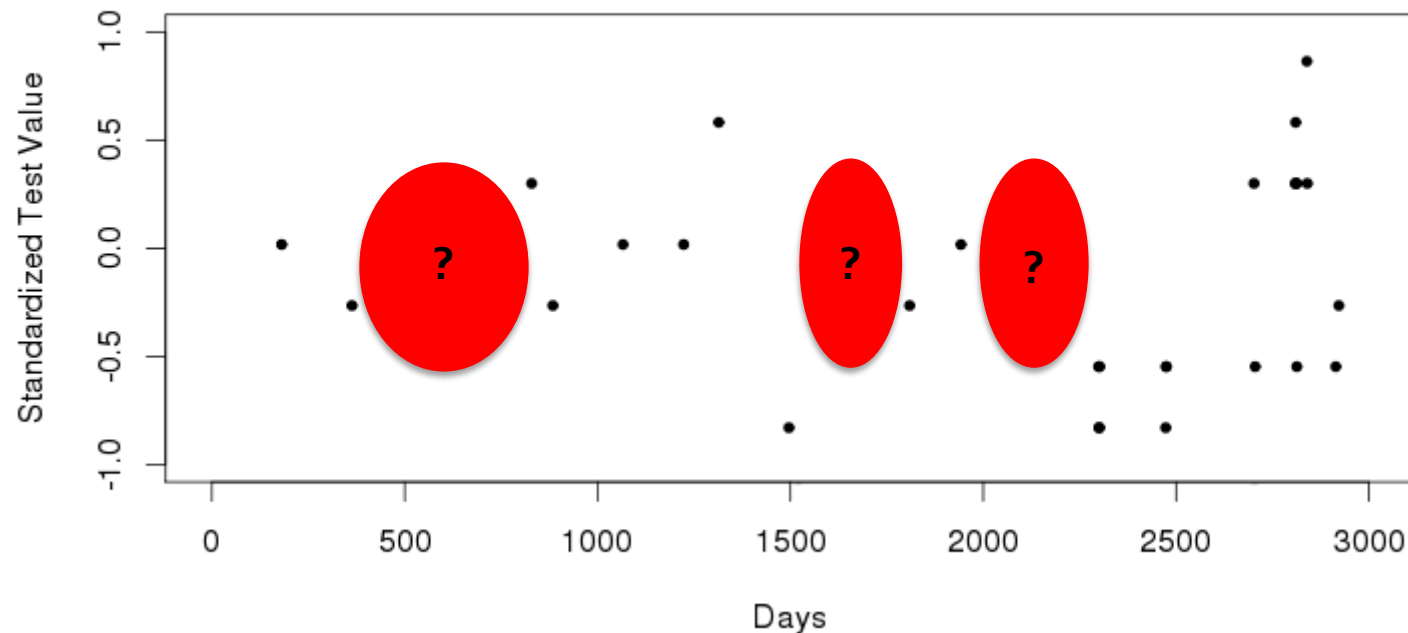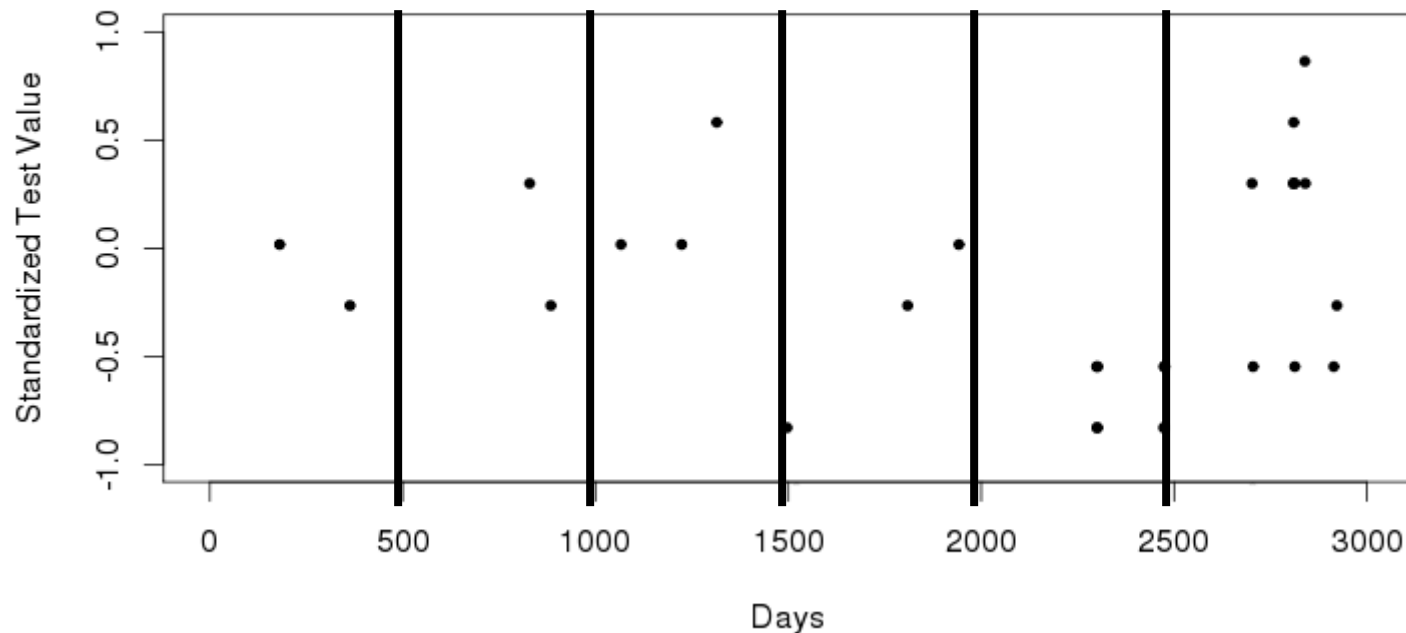
3

Problems with structured data

Sparsity   Irregularity   Asynchrony

Problems with structured data

Sparsity   Irregularity   Asynchrony

Problems with structured data