

2014 VIREC Database and Methods Cyber Seminar Series

# Measuring Laboratory Use and Results Using the VA DSS National Lab Data

April 7, 2014

**Presented by:**

Ann O'Hare, MA MD

Adam Batten, BA

Jeff Todd-Stenberg, BA

Daniel Bertenthal, MPH

# Session Outline

- Overview of laboratory data in the VA Decision Support System (DSS)
- Our experience using VA DSS laboratory data
- Overview of laboratory data in the VA Corporate Data Warehouse (CDW)
- Our experience using VA CDW laboratory data

# Audience Poll

- **Have you worked with lab data in DSS?**
  - Yes
  - No
  
- **Have you worked with serum creatinine data in DSS?**
  - Yes
  - No

# DSS Overview

## ■ What is DSS?

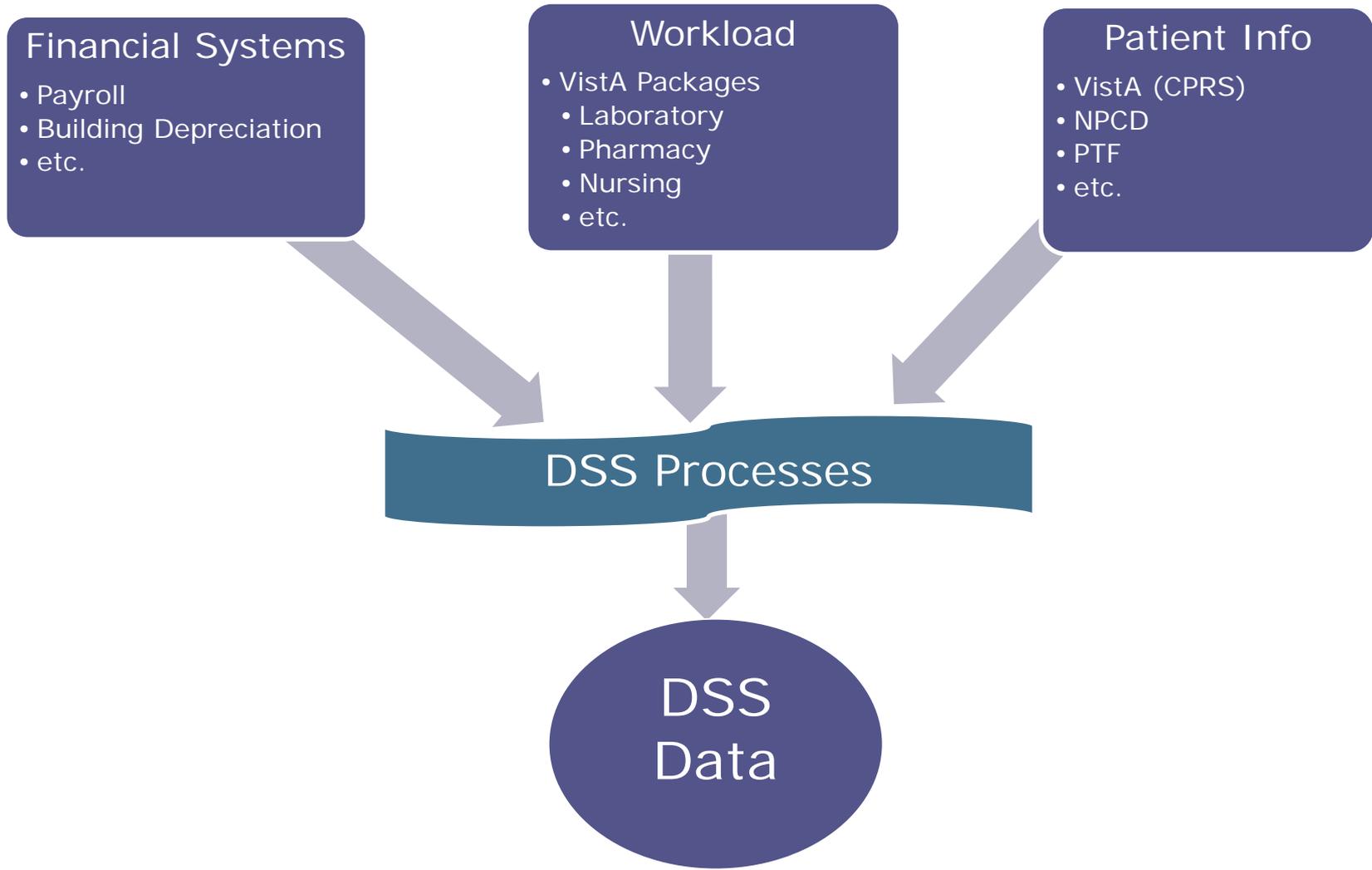
- VA's managerial cost accounting and executive information system

## ■ What is its primary purpose?

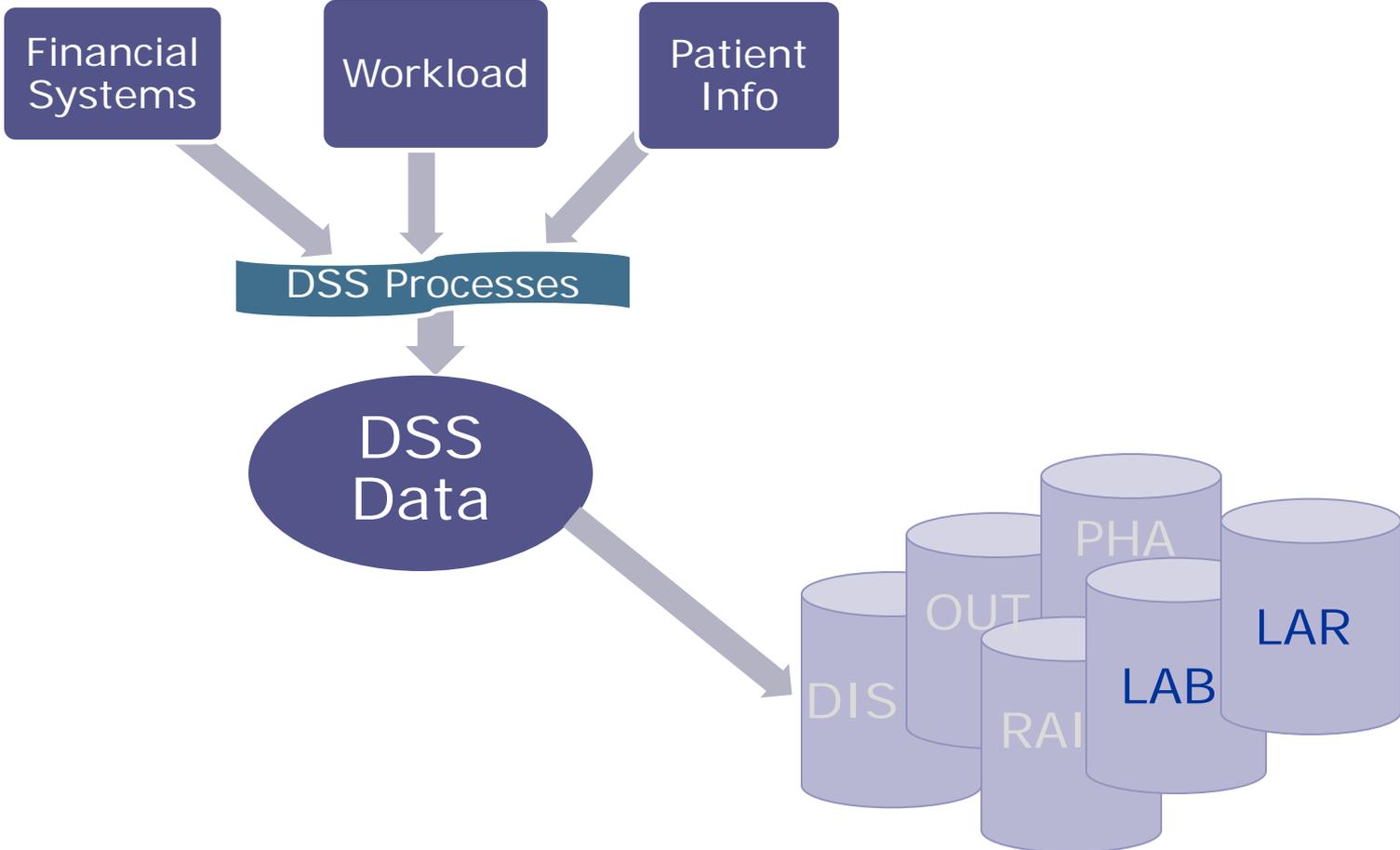
- To provide managerially-useful information (e.g., productivity measures, costs per unit of work, quality assessment) to:
  - Managers
  - Undersecretary for Health
  - Secretary
  - Congress



# DSS Source Data



# DSS National Data Extracts (NDE)



# National Data Extracts (NDEs)

## Clinical NDEs



### ■ LAB

- Workload and costs
- Test-level records

### ■ LAR

- Laboratory results for a defined list of tests (currently 91)
- Test-level records

# DSS NDEs

- **Schedule: Monthly or quarterly**
- **Cumulative year-to-date**
- **Lab data from FY2000 (LAR) or 2002 (LAB)**

# LAR TESTS used in our work

- Serum creatinine—DSSLARNO 31
- Urine albumin-to-creatinine ratio—  
DSSLARNO 56
- Hepatitis C antibody –DSSLARNO 89

# DSS NDE Data Formats

- **Reports, Data Cubes**

- VHA Support Service Center (VSSC)

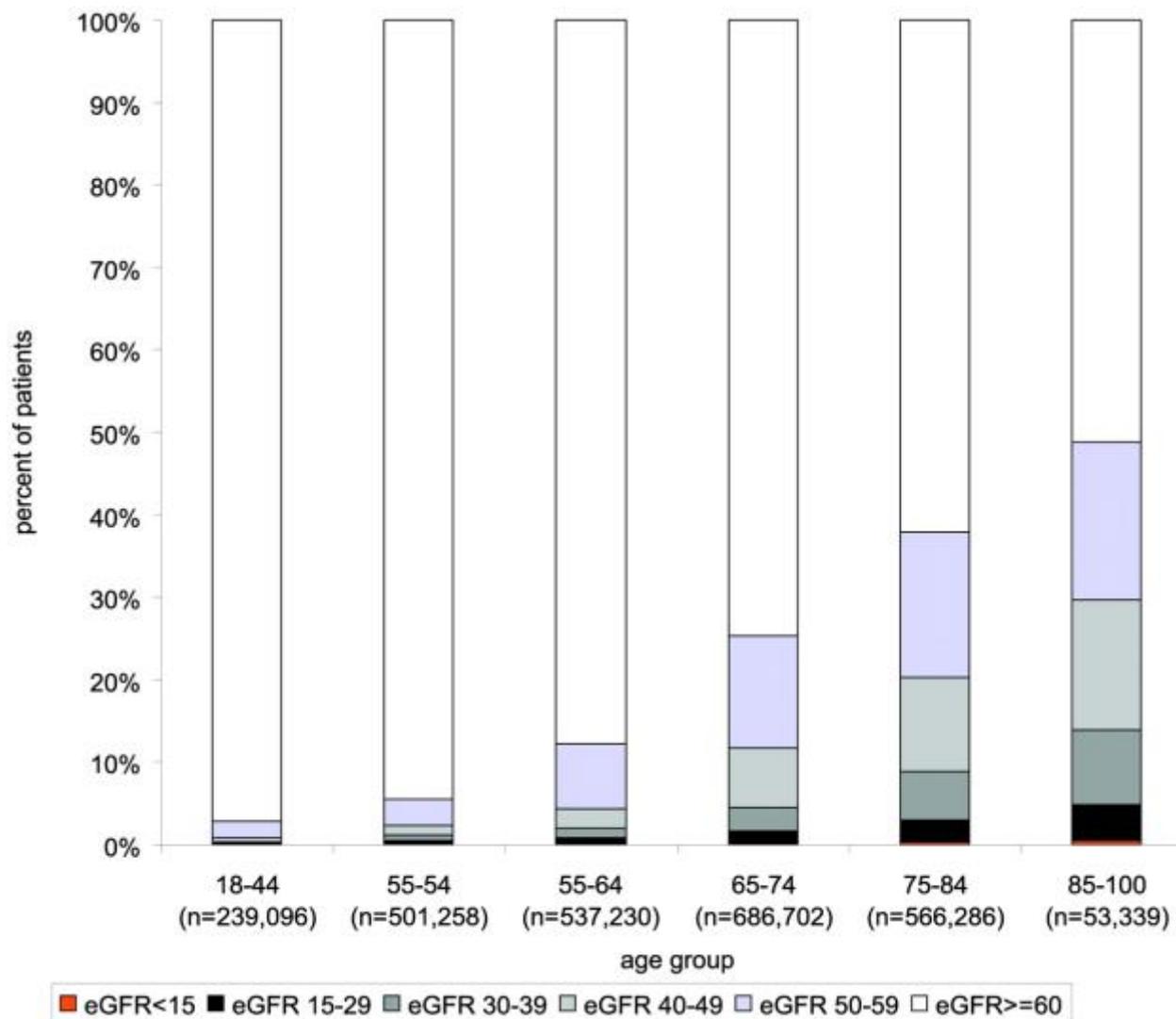
- **SAS Datasets**

- Separate LAR dataset for each VISN and each Fiscal year
- Discontinued after FY 2012
- Removed from Austin Information Technology Center (AITC) mainframe in March 2013
- LAR SAS datasets for 2000-2012 now reside in VINCI

- **SQL tables**

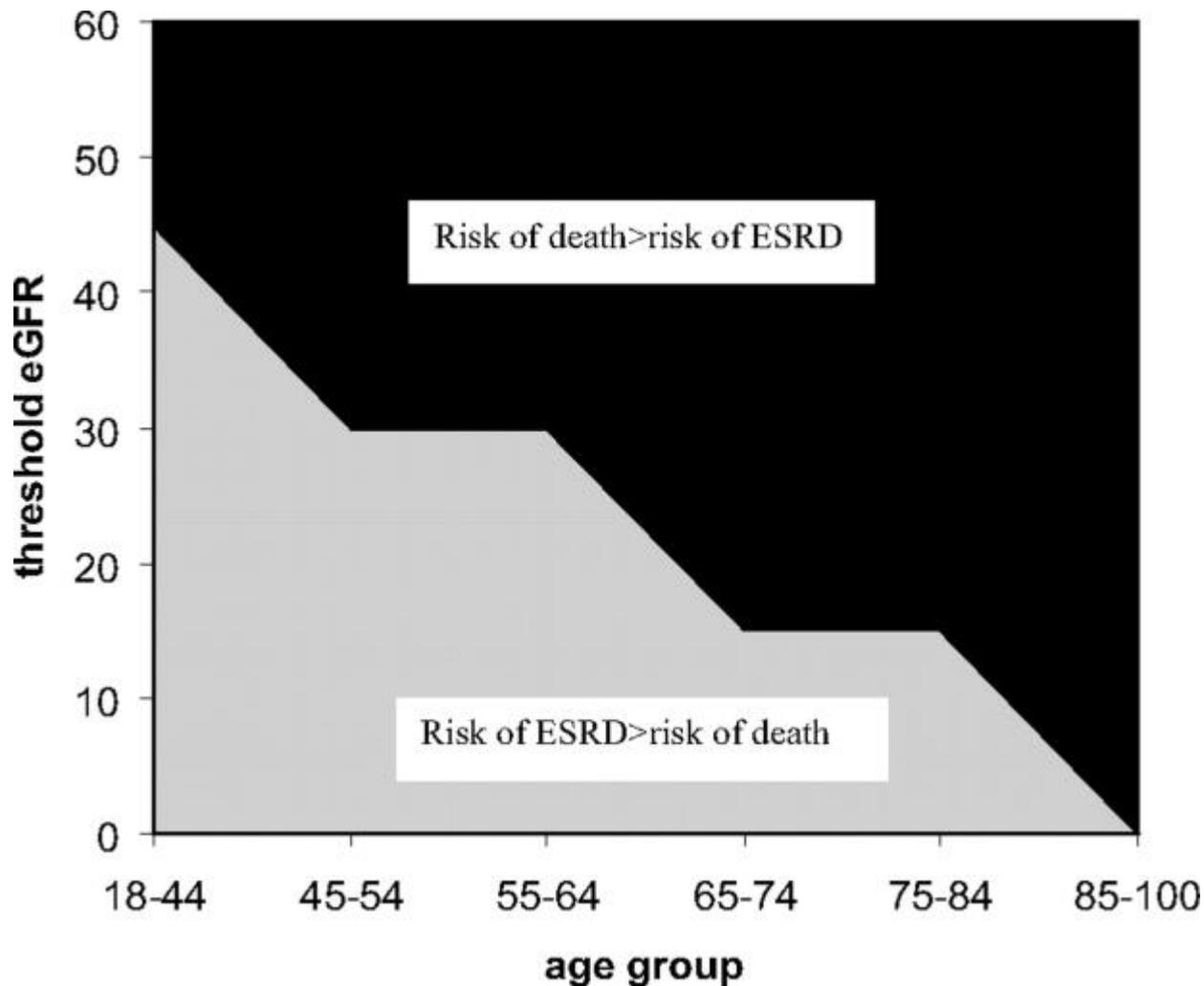
- Single SQL file for all LAR results, across fiscal years and VISNS
- Available from 2005 onward
- Reside in Corporate Data Warehouse (CDW)

## Prevalence of low estimated GFR (eGFR) by age group.

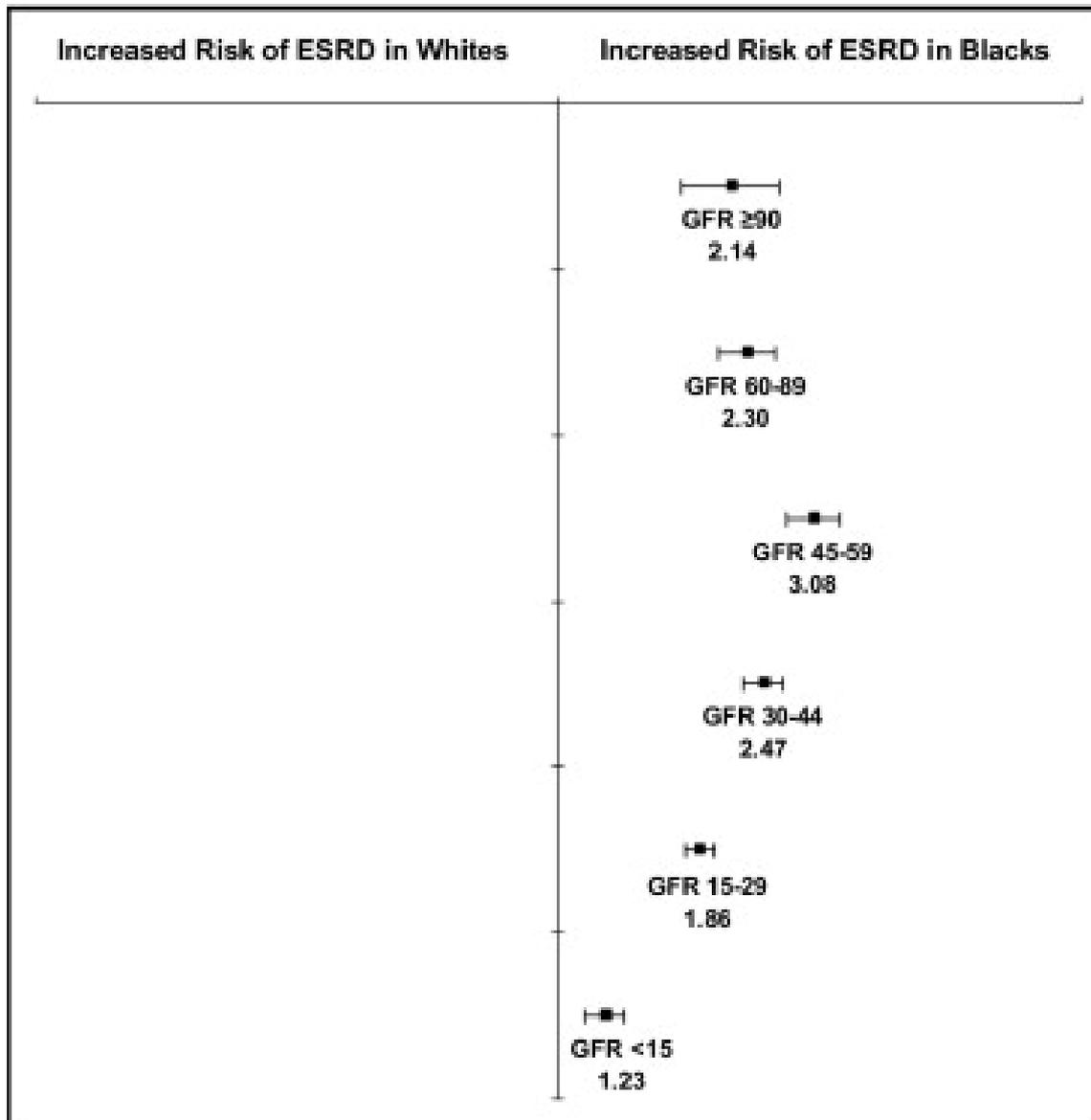


O'Hare et al. JASN 2006;17:846-853

**Baseline eGFR threshold below which risk for ESRD exceeded risk for death for each age group.**



O'Hare et al. JASN 2007;18:2758-2765



## White/Black Racial Differences in Risk of End-stage Renal Disease and Death

From: **Association of Hepatitis C Seropositivity With Increased Risk for Developing End-stage Renal Disease**  
 Tsui et al. Arch Intern Med. 2007;167(12):1271-1276.

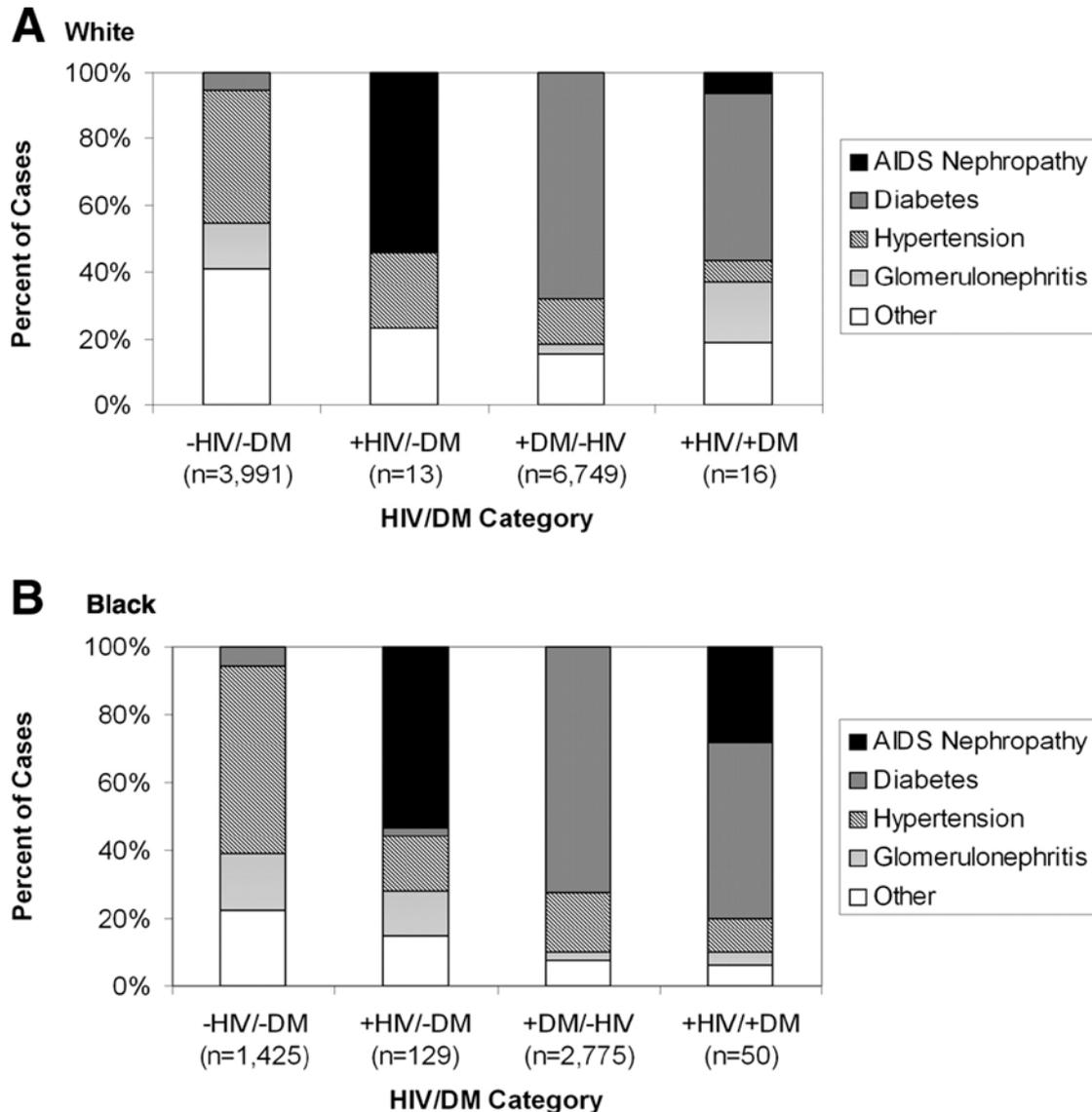
**Table 2. End-stage Renal Disease (ESRD) Incidence Rates and Hazards Ratios (HR) by Hepatitis C Virus (HCV) Seropositivity**

Variable	HCV Antibody	Observation Person-Years	ESRD Events	Unadjusted Rate per 1000 Person-Years (95% CI)	HR (95% CI)	
					Unadjusted	Adjusted*
<b>Overall</b>	Negative	<b>1 442 826</b>	<b>4393</b>	3.05 (2.96-3.14)		
	Positive	<b>178 447</b>	<b>760</b>	4.26 (3.97-4.57)	1.40 (1.29-1.51)	1.68 (1.54-1.82)
Stratified by age, y						
18-49	Negative	340 370	425	1.25 (1.14-1.37)		
	Positive	83 409	243	2.91 (2.57-3.30)	2.32 (1.98-2.72)	1.90 (1.60-2.26)
50-59	Negative	458 725	1087	2.37 (2.23-2.51)		
	Positive	69 589	333	4.79 (4.30-5.33)	2.01 (1.78-2.27)	1.91 (1.67-2.18)
60-69	Negative	321 192	1244	3.87 (3.66-4.09)		
	Positive	14 047	113	8.04 (6.69-9.67)	2.07 (1.71-2.51)	1.71 (1.40-2.08)
≥70	Negative	322 539	1637	5.08 (4.83-5.33)		
	Positive	11 403	71	6.23 (4.93-7.86)	1.22 (0.96-1.54)	1.08 (0.85-1.37)
Stratified by eGFR, mL/min per 1.73 m <sup>2</sup>						
≥60	Negative	1 219 523	308	0.25 (0.23-0.28)		
	Positive	163 846	106	0.65 (0.54-0.78)	2.56 (2.06-3.20)	2.74 (2.13-3.51)
30-59	Negative	207 484	1039	5.01 (4.71-5.32)		
	Positive	13 295	228	17.15 (15.06-19.53)	3.45 (2.99-3.98)	2.57 (2.20-3.01)
<30	Negative	15 819	3046	192.55 (185.83-199.51)		
	Positive	1307	426	326.03 (296.49-358.50)	1.62 (1.46-1.79)	1.28 (1.15-1.44)

Abbreviations: CI, confidence interval; eGFR, estimated glomerular filtration rate.

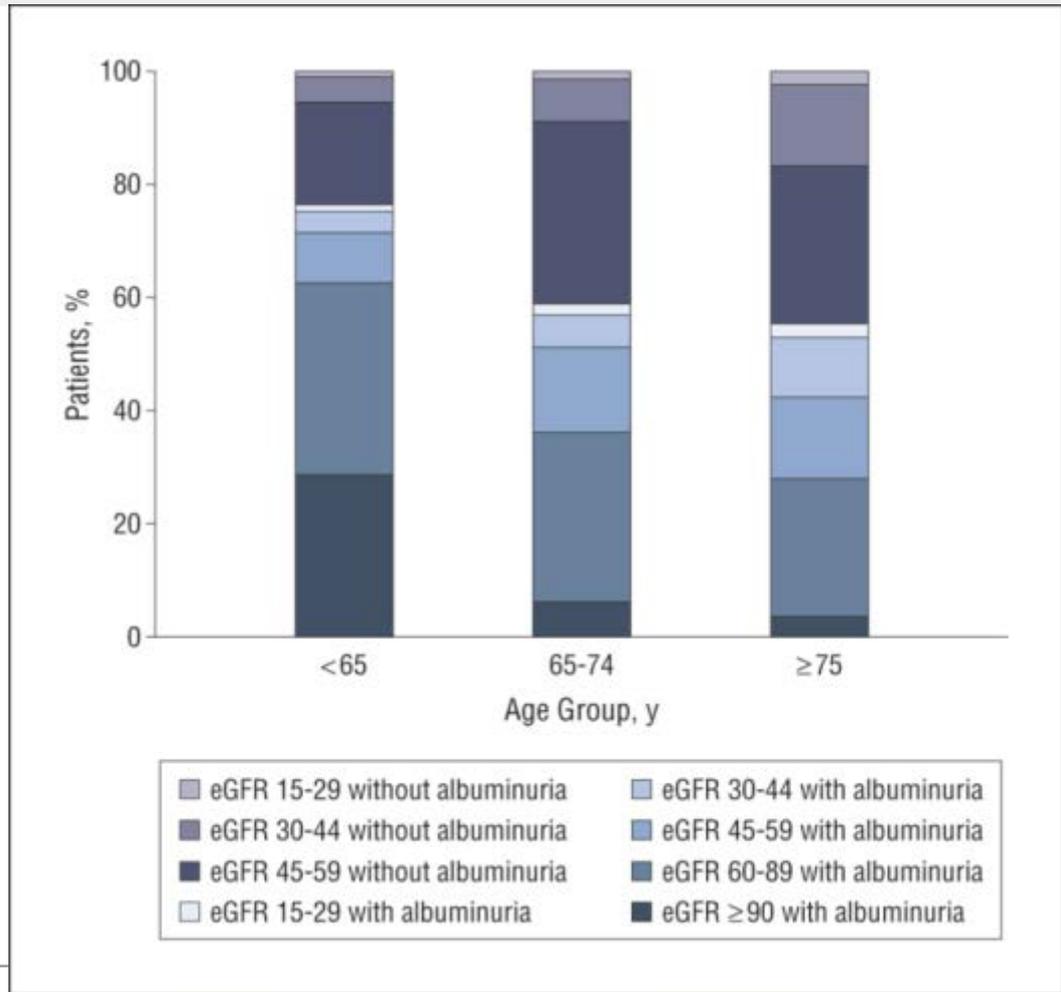
\*Adjusted for all covariates (age, sex, race/ethnicity, diabetes mellitus, hypertension, human immunodeficiency virus infection, congestive heart failure, coronary heart disease, peripheral vascular disease, chronic obstructive pulmonary disease, cerebrovascular disease, substance abuse, and baseline eGFR), minus the covariate on which stratified.

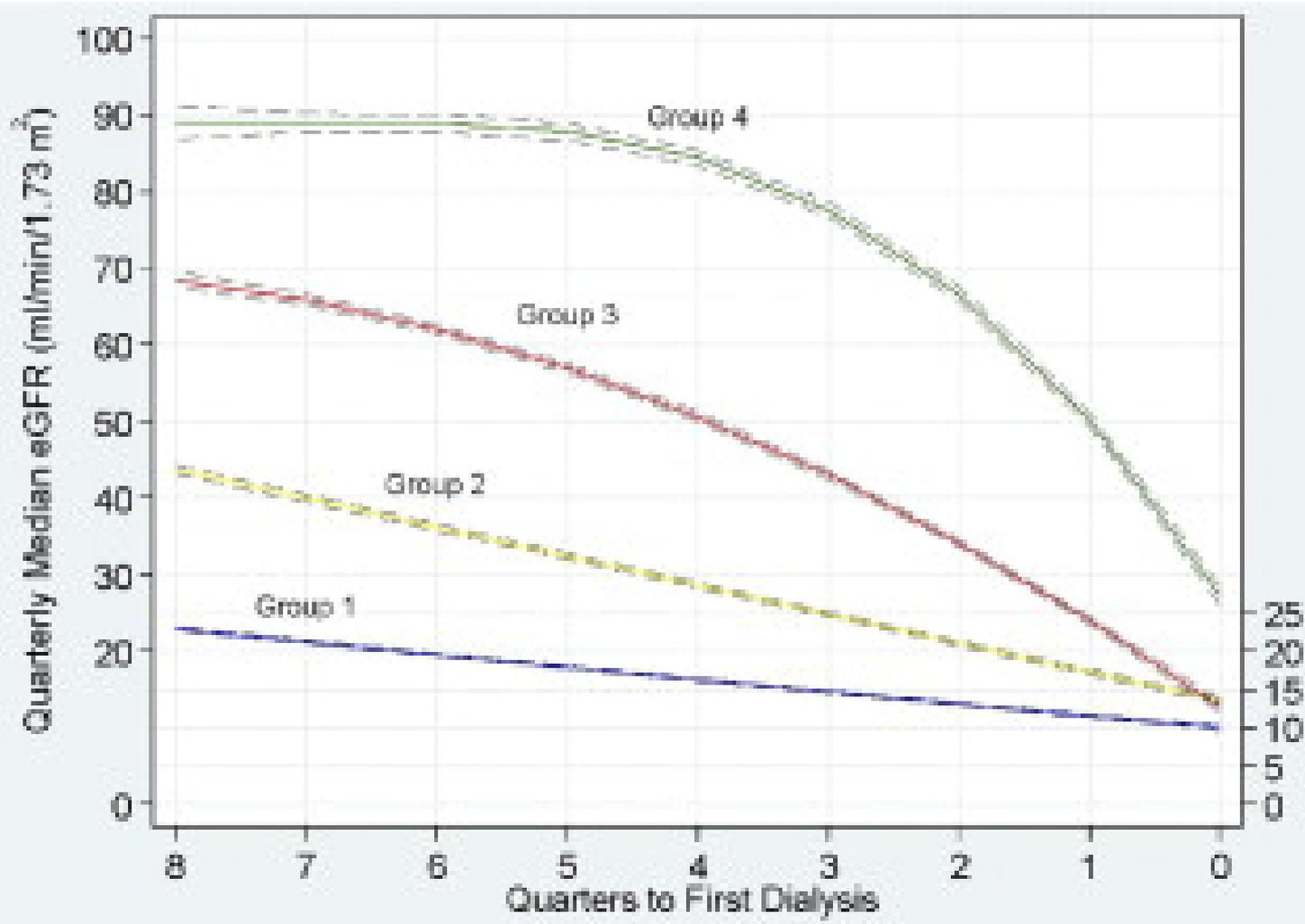
# Primary disease causing ESRD by HIV/diabetes (DM) status among white (A) and black (B) patients.



# From: Prognostic Implications of the Urinary Albumin to Creatinine Ratio in Veterans of Different Ages With Diabetes

O'Hare et al. Arch Intern Med. 2010;170(11):930-936.





## Trajectories of Kidney Function Decline in the 2 Years Before Initiation of Long-term Dialysis

O'Hare et al. American Journal of Kidney Diseases, Volume 59, Issue 4, 2012, 513 - 522

# Strengths and limitations of DSS Laboratory data

- Availability of laboratory measures allowed for major advance beyond work based on diagnostic codes
- Easy for end-user (single DSSLARNO for each test)
- Incomplete capture of all relevant lab tests at all medical centers, especially for earlier years
- Lack of transparency about mapping process ("black box"), especially for earlier years

# LOINC Codes

## <http://www.loinc.org>

- **LOINC: Logical Observation Identifier Names and Codes**
- **Highly specific - Identifies test, method of analysis, specimen source**
- **Lab results (LAR) for DSS records pulled based on LOINC, implemented nationwide back to FY 2009 onward, available in LAB from FY 2013 onward**
- **Results in better match between LAB and LAR records compared with previous method (based on test names)**
- **VistA LOINC file and DSS LOINC file have not always contained the same version of the LOINC code set.**

# Audience Poll

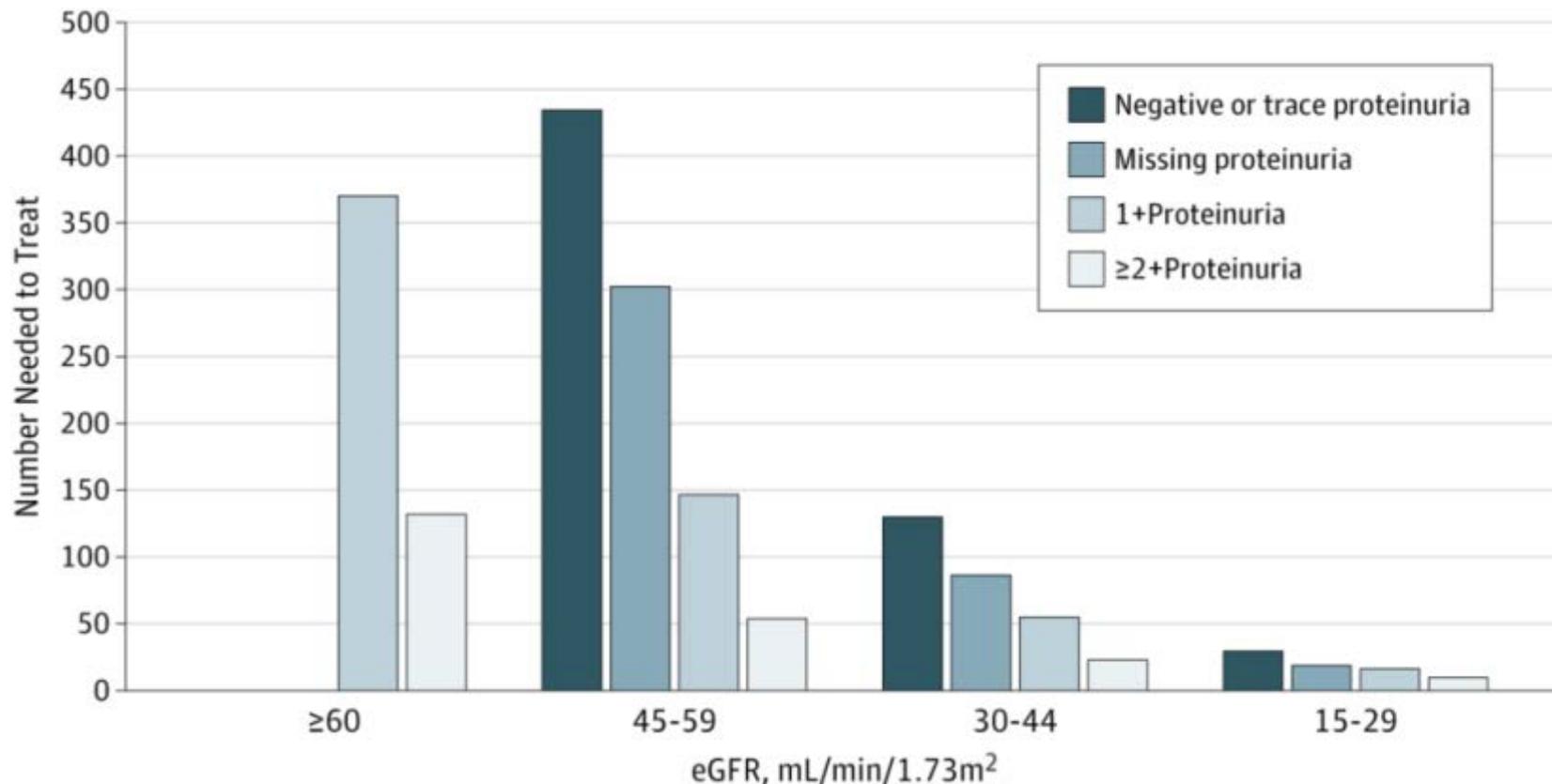
- **Have you worked with lab data in CDW?**
  - Yes
  - No
  
- **Have you worked with serum creatinine or proteinuria data in CDW?**
  - Yes
  - No

# Lab data at CDW

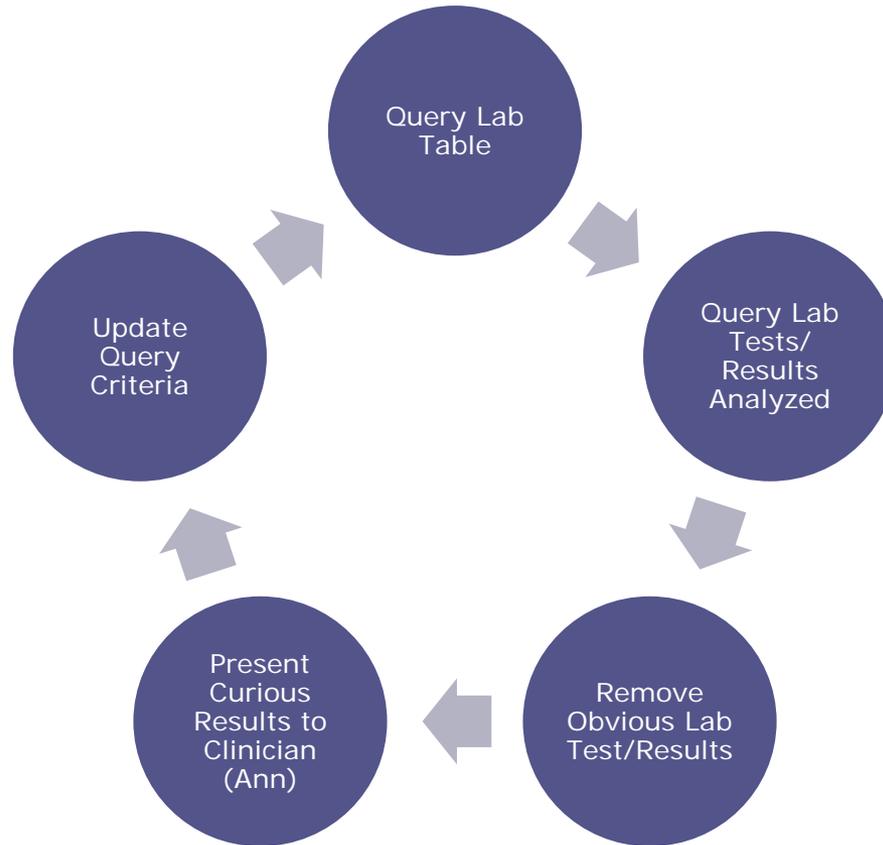
- **CDW is a parallel and increasingly relevant source of lab data in VA**
- **Both DSS and CDW lab data are derived from VistA at individual medical centers**
- **CDW includes all lab tests, not select group of tests (as in DSS)**
- **Disadvantage (or advantage) is that it follows VistA test name format at each medical center**
- **Available back to FY 2000**
- **Contains LOINC code**

**From: Interpreting Treatment Effects From Clinical Trials in the Context of Real-World Risk Information: End-Stage Renal Disease Prevention in Older Adults**

O'Hare et al. JAMA Intern Med. 2014;174(3):391-397.



# Lab Data Mapping



	B	C	D	E	F	G	H	I	J	K
1	LabChemTestName	FREQ	Missing Numeric	min	mean	median	iqr	p90	max	
2	URINE PROTEIN	12114830	10565125	0	64.6122	30	30	100	38000	
3	BUN/CREATININE RATIO	1416565	2169	-217	16.9531	16	12.4	25	310	
4	MICROALBUMIN	761995	156860	0	54.0178	7.4	2	80	64437	
5	UREA NITROGEN/CREATININE RATIO	734505	276	-37.5	15.6418	14.6	11.3	23.3	430	
6	URINE PROTEIN (UA)	503761	449626	0	77.8838	30	30	150	500	
7	PROTEIN, URINE	366745	328964	0	85.0025	30	30	133	35650	
8	MICROALBUMIN/CREATININE RATIO	363984	58443	-8400	89.8153	13.6	5.2	146.5	733000	
9	URINE PROTEIN (V2)	355594	290275	0.2	73.4863	30	30	100	300	
10	URINE CREATININE	343147	15910	-0.2	121.547	108	68.4	214.1	98317.5	
11	URINE PROTEIN (Dipstick)	244514	227082	1	1.4692	1	1	3	4	
12	URINE PROTEIN (URINALYSIS)	243078	196736	0.1	61.9557	30	30	100	2000	
13	CREATININE (ORDER URINE SEPARATELY)	235703	2549	0.1	4.3256	1.2	1	2.4	957	
14	Microalbumin	222681	38259	0	45.8758	7.9	2.5	77.1	63333.3	
15	PROTEIN, URINE	211453	169779	0	72.1782	30	15	156	6100	
16	MICROALBUMIN/CREAT RATIO	177510	30556	0	67.302	10.3	4	123	42515	
17	CREATININE, URINE	175965	8718	0	122.532	109.5	68.3	217.61	4446	
18	Urine Protein	166808	114394	0	48.1698	20	10	100	1000	
19	Protein urine	156337	123408	0	63.2836	30	30	100	1000	
20	CREATININE, URINE	153673	8712	0	135.703	109	67.9	218.2	5105	
21	MICROALBUMIN (V2)	147825	18076	0	7.6405	2.24	1	17.79	2297	
22	ALBUMIN/CREATININE RATIO	146467	40865	0	118.549	11.7	6	67.2	3166667	
23	MICROALBUMIN/CREATININE RATIO (V2)	144532	17108	0	71.0015	18	8	163.1	10948	
24	MICROALBUMIN (mg/dL)	143812	13884	0	13.4153	1.9	0.7	24.2	5551.8	
25	PROTEIN, URINE (STICK)	143635	109156	0	79.1281	30	30	100	1000	
26	URINE PROTEIN (QUAL)	131929	107564	2	84.795	30	30	300	2000	
27	PROTEIN - (dipstick urine)	121742	119768	0.2	63.188	30	30	100	300	
28	URINE MICROALBUMIN	117278	8599	0.1	15.4437	2.1	0.8	28.7	7582	
29	MICROALBUMIN (Spt Ur)	111538	27175	0.2	3.7238	1.4	0.6	10.3	133	
30	URINE PROTEIN-QUAL--O	108679	85334	30	77.6783	30	30	100	300	
31	RATIO OF MICROALBUMIN/CREATININE	104621	24589	0	36.325	11	5	93	3445	24
32	URINE DAU CREATININE	96291	1856	0	143.152	131.4	82	255.7	2860.1	

cdw\_lab\_missing.xlsx - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View Acrobat

G9 fx

	A	B	C	D
1	<b>LabChemResultValue</b>	<b>FREQ</b>	<b>Percent</b>	
2	NEGATIVE	9522344	34.66	
3	NEG	8984694	32.7	
4	TRACE	2219620	8.08	
5	Negative	1447114	5.27	
6	Neg.	1398918	5.09	
7	canc	735090	2.68	
8	comment	576438	2.1	
9	TR	521570	1.9	
10	Neg	476173	1.73	
11	Trace	398024	1.45	
12	neg	303251	1.1	
13	negative	204966	0.75	
14	NEG.	99379	0.36	
15	Tr	83356	0.3	
16	N	66106	0.24	
17	trace	55768	0.2	
18	n/a	52821	0.19	
19	POSITIVE	29101	0.11	
20	pending	27753	0.1	
21	Pos.	26281	0.1	
22	POS	23092	0.08	
23	NP	22372	0.08	
24	NORMAL	16107	0.06	
25	N/A	14065	0.05	
26	No paraprotein detected.	9977	0.04	
27	NA	8357	0.03	
28	T	7915	0.03	
29	AA	7913	0.03	
30	NO SPIKE SEEN. NORMAL PATTERN.	7732	0.03	
31	DETECTED	7522	0.03	
32	COMMENT	7462	0.03	

cdw\_lab\_missing

Ready

## Lab values

There are many spellings for the same test value.

Fortunately, the bulk of the tests have similar names

Handful of oddities .neg, +-, <1, etc...

Mysterious values taken back to Ann for review.

# Audience Poll

- **Experience level with regular expressions?**
  - Never heard of them
  - Somewhat familiar
  - Use them occasionally
  - This is your first thought in the morning  
`# ^ [><](\d+\.\?\d?)#`



```

163 DATA step1;
164   if _N_=1 then do;
165       re1=prxparse("#(\d)\+#"); /*3+*/
166       re2=prxparse("#^[><](\d+\.\d*)#"); /*>10*/
167       re3=prxparse("#.*(\d+)[\.\.]*M|O|G\D*#"); /*300MG/DL*/
168       re4=prxparse("#^\d+-\d+#"); /*100-200*/
169       re5=prxparse("#^[><](\d+\.\d*)#"); /*<25*/
170       re6=prxparse("#\+(\d)#"); /*+1*/
171       re7=prxparse("#.*T[RAT].*#"); /*TRACE*/
172       re8=prxparse("#.*POS.*#"); /*POSITIVE*/
173       re9=prxparse("#.*NEG.*#"); /*NEGATIVE*/
174       re92=prxparse("#^N[EGDRW].*#"); /*NEG*/
175       re10=prxparse("#^\+-\d{2}#"); /*+(20*/
176       re11=prxparse("#.*\d{3}\D*#"); /*300*/
177       re12=prxparse("#.*\d{4}\D*#"); /*1000*/
178       re13=prxparse("#MOD(ERATE)?#");
179       re14=prxparse("#SM(ALL)?#");
180       re15=prxparse("#L[AR](RGE)?#");
181       re16=prxparse("#\d{2}#"); /*30*/
182   end;
183   retain re1 re2 re3 re4 re5 re6 re7 re8 re9 re92 re10 re11 re12 re13 re14
184       re15 re16;
185   length one two 8 upcode $2;
186   set dp.Ppckd_cdw_labchemtestresults;
187   where trim(left(upcase(LabChemTestName))) like '%URINE%'
188       AND trim(left(upcase(LabChemTestName))) like '%PROTEIN%'
189       AND trim(left(upcase(LabChemTestName))) like '%24%'
190       AND trim(left(upcase(LabChemTestName))) not like '%12%'
191       AND trim(left(upcase(LabChemTestName))) not like '%ELECTRO%'
192       AND trim(left(upcase(LabChemTestName))) not like '%EP%'
193       AND trim(left(upcase(LabChemTestName))) not like '%LC%';
194   upcode="NA";
195   result=compress(left(upcase(LabChemResultValue)));
196   NRESULT=result*1;
197

```

Results  
REGEXP

Test Logic

# Lab Data Mapping

- **Input from clinician is invaluable**
- **Most text data is easy to map**
- **Programmers- master REGEXP, you won't be sorry.**
- **Future directions? Compare use of LOINC to text mapping? Index every possible dipstick lab test? ...**

# Summary points

- There are now several viable options for obtaining VA-wide lab data
- CDW lab data overcome many of the limitations of the DSS LAR file
- Lab data from CDW require more manipulation by end-user to transform into a usable format compared with DSS

# VIReC Internet Website

## ■ DSS Research User Guide

- <http://www.virec.research.va.gov/RUGs/RUGs-Index.htm>

## ■ VHA Corporate Data Warehouse (CDW)

- [http://www.virec.research.va.gov/CDW/Overview.htm#CDW\\_Guide](http://www.virec.research.va.gov/CDW/Overview.htm#CDW_Guide)

# VIReC Help

## ■ HSRData Listserv

- Join at the VIReC website
- Discussion among >650 data stewards, managers, and users
- Past messages in archive (on intranet)

## ■ VIReC Help Desk

- VIReC staff will answer your question and/or direct you to available resources on topics
- [VIReC@va.gov](mailto:VIReC@va.gov); 708-202-2413

# VA Intranet Websites

## ■ CDW SharePoint Site

- Send va.gov email to [VIReC@va.gov](mailto:VIReC@va.gov) for URL

## ■ VINCI SharePoint Site

- Send va.gov email to [VIReC@va.gov](mailto:VIReC@va.gov) for URL

[VIReC@va.gov](mailto:VIReC@va.gov)

# Upcoming Seminar

May 5, 2014

Assessing Race and Ethnicity

Maria Mor, PhD