



VSAC Users' Forum

Thursday, June 25, 2015
2:00 – 3:00 PM



Today's Agenda

Today's Topic

Overview of the New Value Set Similarity and Overlap Tools

Announcements

- VSAC Code System Updates: RxNorm 2015-06
- Coming Soon: Downloaded files and API will include value set purpose statements. Excel, Text, and XML formats will change because of this. Upcoming release notes will describe the changes to these files.



Identical Value Sets in VSAC

Jaccard Score	Value Set 1 Name	CS: #of Codes	Value Set 2 Name	Measures (VS 1)	Measures (VS 2)
1	Hospital measures-Hip arthrop	ICD10PCS:57	Hip Replacement Surgery	CMS171v4 CMS172v4	CMS108v3 CMS190v3
1	Hospital Measures - AMI	SNOMEDCT:27	Hospital Measures-Acute or B	CMS60v3 CMS100v3	CMS53v3 CMS60v3
1	Hospital Measures-Caesarean	ICD9CM:5	Cesarean Section	CMS171v4	CMS113v3
1	Hospital Measures-Caesarean	ICD10PCS:3	Cesarean Section	CMS171v4	CMS113v3
1	Pregnancy	SNOMEDCT:63	Pregnancy Dx	CMS64v4	CMS69v3
1	Dietary Recommendations	ICD10CM:1	Below Normal Follow up	CMS22v3	CMS69v3
1	Dietary Recommendations	ICD10CM:1	Above Normal Follow-up	CMS22v3	CMS69v3
1	Dietary Recommendations	ICD9CM:1	Below Normal Follow up	CMS22v3	CMS69v3
1	Below Normal Follow up	ICD10CM:1	Above Normal Follow-up	CMS69v3	CMS69v3
1	Other Proliferative Retinopathy	SNOMEDCT:6	Retrolental Fibroplasias	CMS133v3	CMS132v3
1	Cataract, Posterior Polar	ICD10CM:1	Cataract, Congenital	CMS132v3	CMS132v3
1	Radiation Therapy II	CPT:5	Radiation Treatment Manage	CMS140v3	CMS157v3
1	Ophthalmological Services	CPT:4	Diabetes Visit	CMS138v3 CMS142v3 CMS167v3 CMS143v3	CMS148v3



Highly Similar Value Sets in VSAC

Jaccard Score	Value Set 1 Name	CS: #of Codes	Value Set 2 Name	Measures (VS 1)	Measures (VS 2)
0.944	Pregnancy Dx	ICD9CM:1109	Complications of Pregnancy,	CMS69v3	CMS153v3
0.944	Asthma	ICD10CM:18	Asthma	CMS145v3 CMS144v3	CMS26v2
0.936	Pregnancy	ICD9CM:1102	Complications of Pregnancy,	CMS64v4 CMS61v4	CMS153v3
0.932	Mental Disorders	ICD10CM:725	Psychiatric/Mental Health Pa	CMS108v3	CMS111v3 CMS55v3 CMS32v4
0.913	Chemotherapy	CPT:21	Chemotherapy Administration	CMS140v3	CMS157v3
0.889	Hospital Measures-Acute or Ev	ICD10CM:8	Acute Myocardial Infarction	CMS53v3 CMS60v3	CMS164v3 CMS182v4
0.881	Functional Status Assessment f	LOINC:79	Functional Status Assessment	CMS66v3	CMS56v3
0.875	Hospital Measures - AMI	ICD10CM:14	Myocardial Infarction	CMS60v3 CMS100v3 CMS53v3 CMS30v4	CMS145v3
0.875	Chemotherapy for Colon Canc	RXNORM:8	Leucovorin	CMS141v4	CMS52v3
0.87	Peritoneal Dialysis	CPT:22	Hemodialysis	CMS147v4 CMS65v4	CMS147v4 CMS65v4
0.857	Major Depression Including Re	ICD9CM:12	BH Condition Involving Unipo	CMS159v3 CMS160v3	CMS169v3
0.857	Delivery - Procedure	CPT:12	Delivery Live Births	CMS158v3	CMS153v3
0.854	Bipolar Disorder	ICD9CM:40	BH Condition Involving Bipola	CMS160v3 CMS159v3	CMS169v3



Example of Similar Value Sets

Name	#codes	OID	Name	#codes	OID
ACE Inhibitor or ARB	151	2.16.840.1.113883.3.526.2.39	ACE inhibitor or ARB Clone 1	129	2.16.840.1.113883.3.2.2016

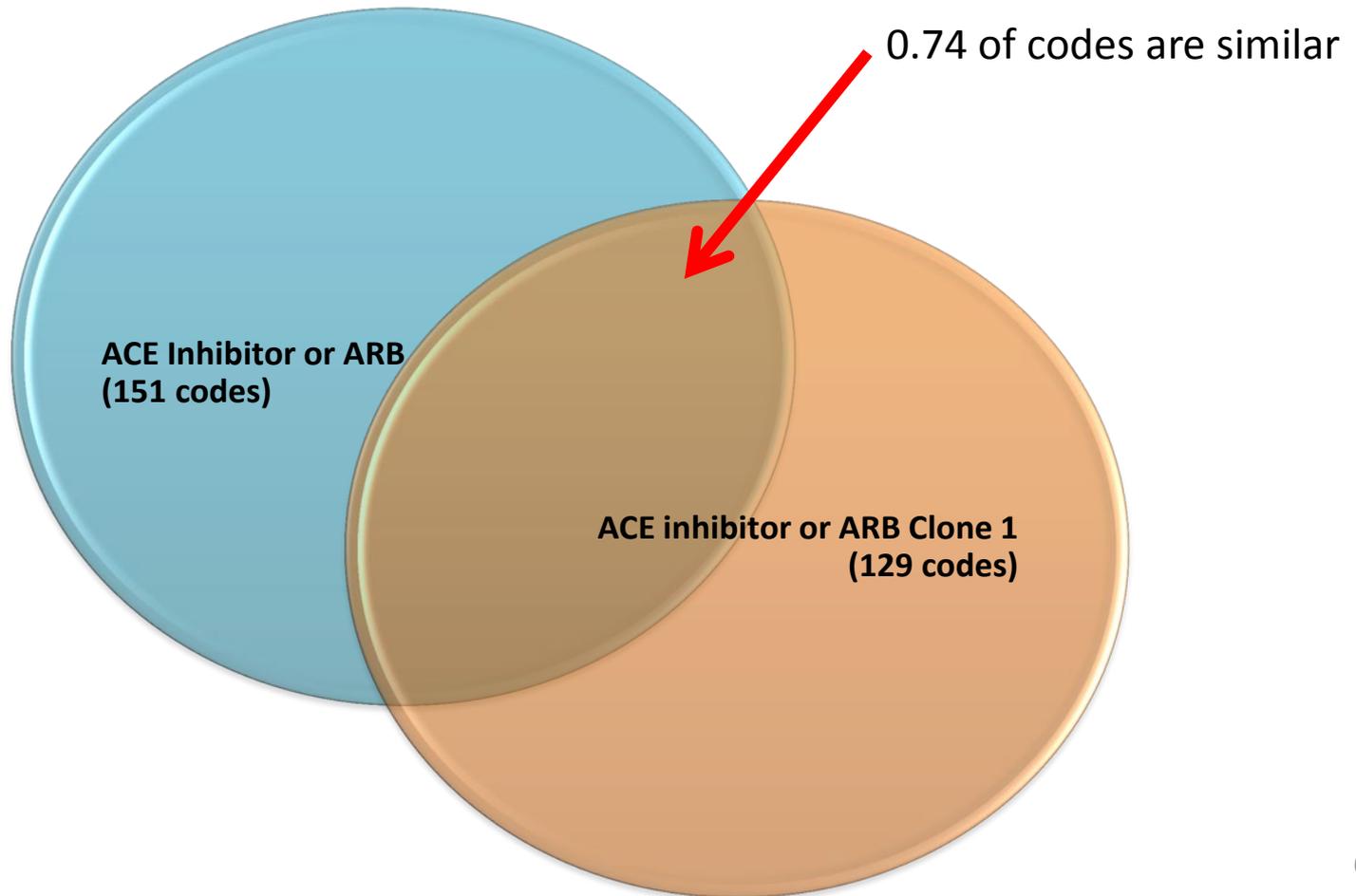
Similarity Results for Threshold: 0.85

Similar Value Sets					
OID	Name	# of Codes	Jaccard Index	Sorenson Inde	Cosine Index
2.16.840.1.113883.3.2.2016	ACE inhibitor or ARB Clone 1	129	0.74	0.85	0.85

Page 1 of 1 | 10 | View 1 - 1 of 1



Intersection of Common Codes between Two Value Sets





Similarity Algorithms Can Miss Overlaps

Similarity

2.16.840.1.113883.3.526.2.39 **“ACE Inhibitor or ARB”**

is 74% similar to

2.16.840.1.113883.3.2.2016 **“ACE inhibitor or ARB Clone 1”**

Overlap

2.16.840.1.113883.3.526.2.39 **“ACE Inhibitor or ARB”**

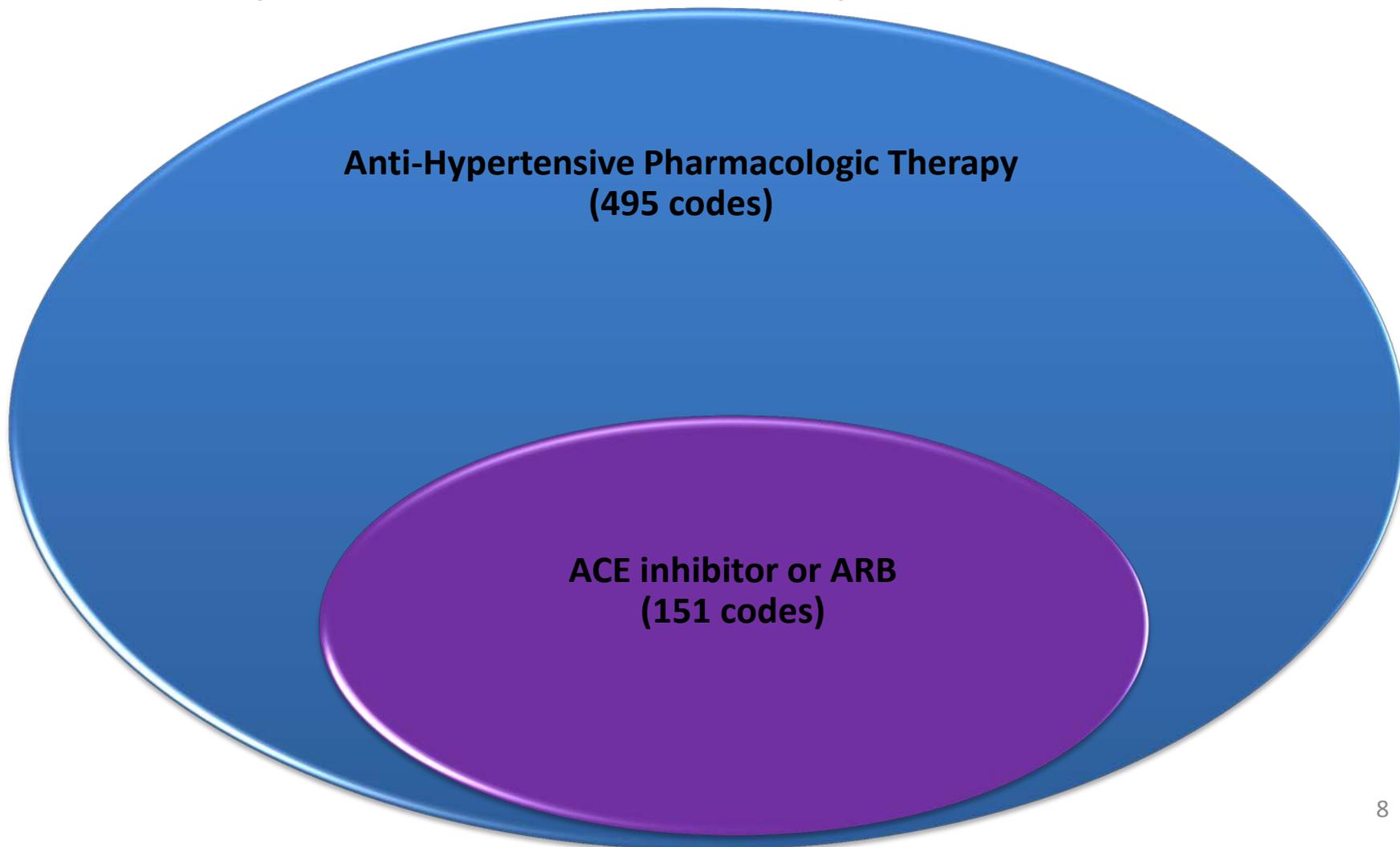
has 100% overlap with (but only 30% similarity to)

2.16.840.1.113883.3.600.1476 **“Anti-Hypertensive Pharmacologic Therapy”**

Similarity algorithms with high thresholds will miss this overlap because “ACE Inhibitor or ARB” is comparatively a subset of the larger “Anti-Hypertensive Pharmacologic Therapy,” and its codes are only 30% of the larger superset “Anti-Hypertensive Pharmacologic Therapy.”



Overlap of Subset and Superset Value Sets





Value Set Similarity

Value Set Similarity

OID: 2.16.840.1.113883.3.526.2.39

Name: ACE Inhibitor or ARB

Code System: RXNORM

of Codes: 151

Threshold:

[Generate Indices](#)

Similarity Results for Threshold: 0.50

Similar Value Sets

OID	Name	# of Codes	Jaccard Index	Sorenson Inde	Cosine Index
2.16.840.1.113883.3.2.2016	ACE inhibitor or ARB Clone 1	129	0.74	0.85	0.85
2.16.840.1.113883.3.600.1476	Anti-Hypertensive Pharmacologic Therapy	495	0.31	0.47	0.55

Page 1 of 1 | 10 | View 1 - 2 of 2

[Close](#)



Value Set Overlap

Value Set Overlap

OID: 2.16.840.1.113883.3.526.2.39

Name: ACE Inhibitor or ARB

Code System: RXNORM

of Codes: 151

Overlap to Superset

Threshold:

Generate Index

Overlap Results for Threshold: 0.85

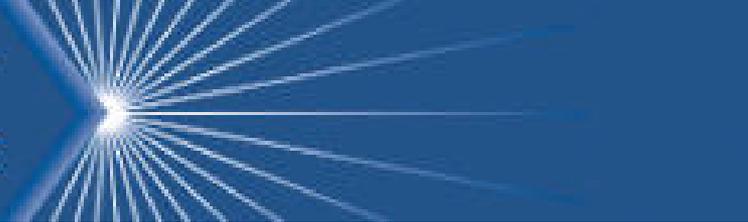
Value Sets for Superset Overlap

OID	Name	# of Codes	Overlap Index
2.16.840.1.113883.3.600.1476	Anti-Hypertensive Pharmacologic Therapy	495	1.00

Page 1 of 1 | 10 | View 1 - 1 of 1

Overlap to Subset

Close



Use Similarity Tools in VSAC Authoring

Welcome Search Value Sets **Authoring** Download Help

Value Sets ACE Inhibi... x

Create a Value Set ? Group Existing Value Sets ?

My Value Sets

Search All Value Sets

Query: 2.16.840.1.113883.3.526.2.39 Search [Export Search Results\(Excel\)](#)

Value Sets									
Download Excel Refresh Clear Filter									
Page 1 of 1 20 View 1 - 1 of 1									
<input type="checkbox"/>	Name	Type	Code System	Steward	Author	OID	Latest Version	Updated Date	Status
<input type="checkbox"/>	ACE Inhibitor or ARB	Extensional	RXNORM	AMA-PCPI	AMA-PCPI A	2.16.840.1.113883.3.526.2.39	20150430	2015-04-30	Published

Page 1 of 1 20 View 1 - 1 of 1

My Program Releases



Harmonization → Check Similarity

Welcome Search Value Sets **Authoring** Download

Value Sets ACE Inhibi... ✕

Harmonization Report

Name: ACE Inhibitor or
OID: 2.16.840.1.113883.3.526.2.39
Status: Published
Note:

Purpose: [Show](#)

Type: Extensional
Steward: AMA-PCPI
Author: AMA-PCPI Author
Code System: RXNORM
Code System Version: 2015-01

430

Check Similarity...
Check Overlap...

Value Set Code List

+ Import Validate Page 1 of 8 20 View 1

<input type="checkbox"/>	Code	Descriptor	TTY	Property
<input type="checkbox"/>	1000001	Amlodipine 5 MG / Hydrochlorothiazide 25 MG / Oln	SCD	Generic,P
<input type="checkbox"/>	1011718	aliskiren 150 MG / valsartan 160 MG Oral Tablet	SCD	Generic,P



Harmonization → Check Similarity

Value Set Similarity

OID: 2.16.840.1.113883.3.526.2.39

Name: ACE Inhibitor or ARB

Code System: RXNORM

of Codes: 151

Threshold:

[Generate Indices](#)

Similarity Results for Threshold: 0.50

Similar Value Sets

OID	Name	# of Codes	Jaccard Index	Sorenson Inde	Cosine Index
2.16.840.1.113883.3.2.2016	ACE inhibitor or ARB Clone 1	129	0.74	0.85	0.85
2.16.840.1.113883.3.600.1476	Anti-Hypertensive Pharmacologic Therapy	495	0.31	0.47	0.55

Page 1 of 1

View 1 - 2 of 2

[Close](#)



Harmonization → Check Overlap

Welcome Search Value Sets **Authoring** Download

Value Sets ACE Inhibi... ✕

Harmonization Report

Name: ACE Inhibitor or
OID: 2.16.840.1.113883.3.526.2.39

Status: Published

Note:

Purpose: [Show](#)

Type: Extensional
Steward: AMA-PCPI
Author: AMA-PCPI Author

Code System: RXNORM
Code System Version: 2015-01

Value Set Code List

+ Import Validate Page 1 of 8 20 View 1

<input type="checkbox"/>	Code	Descriptor	TTY	Property
<input type="checkbox"/>				
<input type="checkbox"/>	1000001	Amlodipine 5 MG / Hydrochlorothiazide 25 MG / Oln	SCD	Generic,P
<input type="checkbox"/>	1011718	aliskiren 150 MG / valsartan 160 MG Oral Tablet	SCD	Generic,P



Harmonization → Check Overlap

Value Set Overlap [Close]

OID: 2.16.840.1.113883.3.526.2.39 **Name:** ACE Inhibitor or ARB
Code System: RXNORM **# of Codes:** 151

▼ **Overlap to Superset**

Threshold: *0.85 **Generate Index**

Overlap Results for Threshold: 0.85

Value Sets for Superset Overlap			
OID	Name	# of Codes	Overlap Index
2.16.840.1.113883.3.600.1476	Anti-Hypertensive Pharmacologic Therapy	495	1.00

Page 1 of 1 10 View 1 - 1 of 1

► **Overlap to Subset**

Close



Jaccard Similarity

Jaccard=intersection/union

#codes common to both value sets

Total codes in value set A plus total codes in value set B

Winnenburg R, Bodenreider O.

Metrics for assessing the quality of value sets in clinical quality measures.

AMIA Annu Symp Proc 2013:1497-1505.

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900160/>

https://en.wikipedia.org/wiki/Jaccard_index

Jaccard scores can be misleading or inaccurate once the code counts differ by >10%



Sorenson Similarity

Gives greater weight to codes common between value sets than to those codes found in only one value set.

$2 \times (\text{\#codes common to both value sets})$

Total codes in value set A plus total codes in value set B

https://en.wikipedia.org/wiki/S%C3%B8rensen%E2%80%93Dice_coefficient

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2373804/>



Cosine Similarity

‘Cosine similarity is commonly used in data mining and information retrieval as a measure of the resemblance between data sets; i.e. how similar or alike these are. It is an important concept used in Vector Space Theory and affine models.’

<https://irthoughts.wordpress.com/2015/04/10/a-cosine-similarity-tool-and-companion-tutorial/>

‘Cosine similarity is a measure of similarity between two vectors of an inner product space that measures the cosine of the angle between them. The cosine of 0° is 1, and it is less than 1 for any other angle. It is thus a judgment of orientation and not magnitude: two vectors with the same orientation have a cosine similarity of 1, two vectors at 90° have a similarity of 0.’

https://en.wikipedia.org/wiki/Cosine_similarity



What should you do if you discover similar, identical or overlapping value sets?

- **Use the previously existing value sets if possible.**
- **Study the similar value sets to discover codes you may have missed or disagree with.**
- **Notify the steward of the similar value set.**



Questions?

Send comments, questions, ideas to:

custserv@nlm.nih.gov

(place VSAC in the subject line)

Upcoming VSAC Users' Forum:

July 16, 2015