



health care systems  
research network

The Health Care Systems Research Network  
Virtual Data Warehouse



# DEMOGRAPHICS

Version = 3.2, Implemented 5/31/2014, StdVar = &\_vdw\_demographic

## Subject Area Description

---

The DEMOGRAPHICS table contains patient/enrollee level descriptives for the people found in VDW tables. It serves as a lookup dataset for MRNs. Every MRN appearing in any other VDW file should appear in the Demographics table, even if demographics information on the person is unknown.

Variable Name	Definition	Type(Len)	Values
MRN	Medical record number is the unique patient identifier within a site and should never leave the site	char(*)	Unique to each patient at each site
BIRTH_DATE	The person's date of birth	num(4)	SAS Date
GENDER	The person's gender and/or sex; if both gender and sex are known, this variable should hold gender	char(1)	M = Male F = Female O = Other including transgendered U = Unknown
RACE1 - RACE5	The person's race. Preference is for self-reported; please see comment 1 for recording multiple race values	char(2)	HP = Native Hawaiian / Pacific Islander IN = American Indian / Alaskan Native AS = Asian BA = Black or African American WH = White MU = Multiple races with particular unknown OT = Other, values that do not fit well in any other value UN = Unknown or Not Reported
HISPANIC	Whether the person is of Hispanic origin / ethnicity	char(1)	Y = Yes N = No U = Unknown
NEEDS_INTERPRETER	Whether the person needs an interpreter to communicate with an English-only speaker	char(1)	Y = Yes N = No U = Unknown
PRIMARY_LANGUAGE	TO BE DEPRECATED 9/1/2014; primary language spoken at last contact	char(3)	<a href="#">As defined by ISO-639-2 or 'unk' for unknown</a> <a href="#">Note that value set is lowercase</a>

**Primary Key:**

MRN

**Foreign Key Relationship:**

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
[None Defined]			

**Comments**

1 - If source data indicates the person's race as "Hispanic" and there is no other known race recorded for the person, code the person's race as unknown. Please see Appendix E for guidelines on mapping local race values to the permissible value set in the VDW. When more than one race is known for a given person, assign values to the race variables in the order listed above RACE values and fill in any unused race variables with the value UN for unknown / not reported. Examples of coding RACE1 through RACE5 are illustrated here:

Example Race Coding	Race1	Race2	Race3	Race4	Race5
White only	WH	UN	UN	UN	UN
White and Pacific-Islander	HP	WH	UN	UN	UN
No Race Known	UN	UN	UN	UN	UN
African-American and Native-American	IN	BA	UN	UN	UN
Multi-racial, particular races not reported.	MU	UN	UN	UN	UN

# LANGUAGE

Version = 3.1, Implemented 3/31/2013, StdVar = &\_vdw\_language

## Subject Area Description

The LANGUAGE table contains information on the languages that patients speak and write. There is one record per person per known language. People on whom you have no language information should not be included in the table.

Variable Name	Definition	Type(Len)	Values
MRN	Medical record number is the unique patient identifier within a site and should never leave the site	char(*)	Unique to each patient at each site
LANG_ISO	A code signifying the language.	char(3)	<a href="#">As defined by ISO-639-2 or 'unk' for unknown</a> <a href="#">Note that value set is lowercase</a>
LANG_USAGE	How the person uses this language.	char(1)	S = Spoken/signed W = Written B = Both spoken and written U = Unknown
LANG_PRIMARY	For spoken languages, whether this is the person's primary spoken language.	char(1)	Y = Yes N = No U = Unknown

### Primary Key:

MRN + LANG\_ISO

### Foreign Key Relationship:

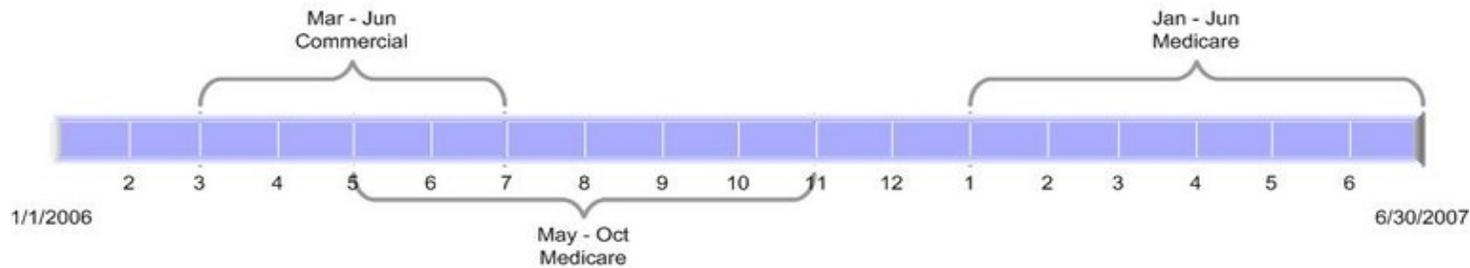
Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
MRN	DEMOGRAPHICS	MRN	No

# ENROLLMENT

Version = 3.2, Implemented 5/31/2014, StdVar = &\_vdw\_enroll

## Subject Area Description

The ENROLLMENT table contains periods of time during which we believe to have at least the partial capture of medical care information either because of enrollment in a health plan or as a patient of a care delivery system that we own or interact with. Insurance coverage is not the only basis of data capture (see ENROLLMENT\_BASIS) nor do all records necessarily indicate completeness of data capture (see OUTSIDE\_UTILIZATION). Each record represents a period of time during which the information on the included variables was true. As many records as are necessary should be added to represent changes over time. There may be many contiguous records for a single period of enrollment, in order to account for these changes. However, periods may NOT overlap one another; there should be only one record covering any given day + medical record number. The following figure illustrates this concept:



Should result in 4 records:

MRN	enr_start	enr_end	ins_commercial	ins_medicare
111	03/01/2006	04/28/2006	Y	
111	05/01/2006	06/30/2006	Y	Y
111	07/01/2006	10/31/2006		Y
111	01/01/2007	06/30/2007		Y

Variable Name	Definition	Type(Len)	Values
MRN	Medical record number is the unique patient identifier within a site and should never leave the site	char(*)	Unique to each patient at each site
ENR_START	Beginning of the period at risk for medical care capture	num(4)	SAS Date

Variable Name	Definition	Type(Len)	Values
ENR_END	End of the period at risk for medical care capture	num(4)	SAS Date
INS_MEDICAID	Whether the person had any Medicaid insurance coverage during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_COMMERCIAL	Whether the person had any commercial insurance coverage during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_PRIVATEPAY	Whether the person had any insurance coverage in a private pay plan during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_STATESUBSIDIZED	Whether the person had any state subsidized insurance coverage during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_SELFFUNDED	Whether the person had any insurance coverage through an employer group that insures itself during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_HIGHDEDUCTIBLE	<a href="#">Whether the person had any insurance coverage in a high deductible plan during the period as defined by the U.S. IRS (Pub 969) qualifying for a Health Savings Account</a>	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_MEDICARE	Whether the person had any Medicare insurance coverage, including Medicare working aged, during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_MEDICARE_A	Whether the person had Medicare Part A insurance coverage during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown

Variable Name	Definition	Type(Len)	Values
INS_MEDICARE_B	Whether the person had Medicare Part B insurance coverage during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_MEDICARE_C	Whether the person had Medicare Part C insurance coverage during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_MEDICARE_D	Whether the person had Medicare Part D insurance coverage during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_OTHER	Whether the person had insurance coverage during the period that is not otherwise included in the INS_* variables during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
PLAN_HMO	Whether the person had insurance coverage under an HMO plan during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
PLAN_POS	Whether the person had insurance coverage in a point-of-service plan during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
PLAN_PPO	Whether the person had insurance coverage in a preferred provider organization plan during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
PLAN_INDEMNITY	Whether the person had insurance coverage in a traditional indemnity plan during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
DRUGCOV	Whether the person had <i>any</i> insurance coverage that included at least some coverage for prescription drugs	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown

Variable Name	Definition	Type(Len)	Values
OUTSIDE_UTILIZATION	Is this person/period likely to have substantial health care utilization outside the HMO system under circumstances that would not result in a claim being submitted to the HMO?	char(1)	Y = Yes, we expect capture is incomplete N = No, there is no reason to suspect incomplete capture U = Unknown completeness of capture
ENROLLMENT_BASIS	The basis for the claim that we have at least partial capture of the medical care information during the period. Historically 'Insurance' was the only permissible basis	char(1)	G = Geography I = Insurance B = Both geography & insurance P = Non-enrolled patient
PCC	The primary care clinic to which the patient is paneled in administrative record.	char(*)	Same value set as UTILIZATION.FACILITY_CODE
PCP	The clinician to which the patient is paneled in administrative record	char(*)	Values link as a foreign key to UTILIZATION.PROVIDER

#### Primary Key:

MRN + ENR\_START (no overlapping time periods are permitted)

#### Foreign Key Relationship:

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
MRN	DEMOGRAPHICS	MRN	No
PCP	PROVIDER	PROVIDER	No

## Usage Notes

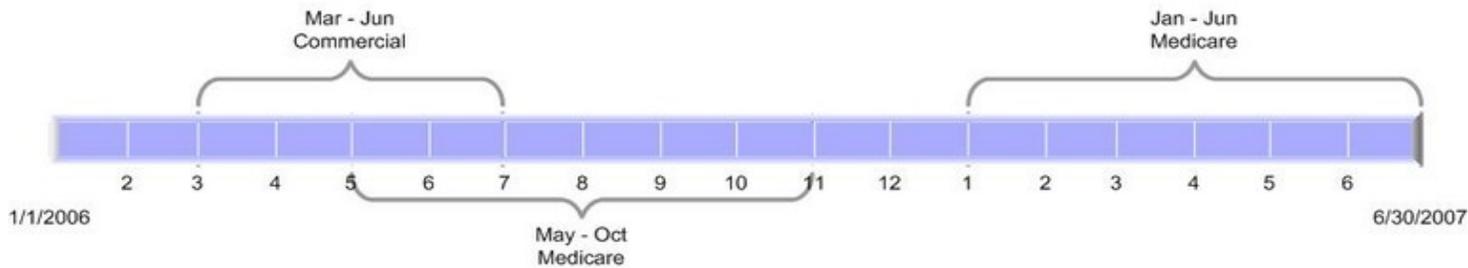
Many sites have added site-specific variables that are not a part of the specifications. At these sites, users may well find multiple records for contiguous time periods, over which none of the official specification variables change. Users will likely be interested in the %CollapsePeriods or enrollment-related macros for collapsing records in to continuous periods.

# ENROLLMENT

Version = 3.3, GOAL IMPLEMENTATION 2/1/2015, StdVar = &\_vdw\_enroll

## Subject Area Description

The ENROLLMENT table contains periods of time during which we believe to have at least the partial capture of medical care information either because of enrollment in a health plan or as a patient of a care delivery system that we own or interact with. Insurance coverage is not the only basis of data capture (see ENROLLMENT\_BASIS) nor do all records necessarily indicate completeness of data capture (see OUTSIDE\_UTILIZATION). Each record represents a period of time during which the information on the included variables was true. As many records as are necessary should be added to represent changes over time. There may be many contiguous records for a single period of enrollment, in order to account for these changes. However, periods may NOT overlap one another; there should be only one record covering any given day + medical record number. The following figure illustrates this concept:



Should result in 4 records:

MRN	enr_start	enr_end	ins_commercial	ins_medicare
111	03/01/2006	04/28/2006	Y	
111	05/01/2006	06/30/2006	Y	Y
111	07/01/2006	10/31/2006		Y
111	01/01/2007	06/30/2007		Y

Variable Name	Definition	Type(Len)	Values
MRN	Medical record number is the unique patient identifier within a site and should never leave the site	char(*)	Unique to each patient at each site
ENR_START	Beginning of the period at risk for medical care capture	num(4)	SAS Date

Variable Name	Definition	Type(Len)	Values
ENR_END	End of the period at risk for medical care capture	num(4)	SAS Date
INS_MEDICAID	Whether the person had any Medicaid insurance coverage during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_COMMERCIAL	Whether the person had any commercial insurance coverage during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_PRIVATEPAY	Whether the person had any insurance coverage in a private pay plan during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_STATESUBSIDIZED	Whether the person had any state subsidized insurance coverage during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_SELFFUNDED	Whether the person had any insurance coverage through an employer group that insures itself during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_HIGHDEDUCTIBLE	<a href="#">Whether the person had any insurance coverage in a high deductible plan during the period as defined by the U.S. IRS (Pub 969) qualifying for a Health Savings Account</a>	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_MEDICARE	Whether the person had any Medicare insurance coverage, including Medicare working aged, during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_MEDICARE_A	Whether the person had Medicare Part A insurance coverage during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown

Variable Name	Definition	Type(Len)	Values
INS_MEDICARE_B	Whether the person had Medicare Part B insurance coverage during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_MEDICARE_C	Whether the person had Medicare Part C insurance coverage during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_MEDICARE_D	Whether the person had Medicare Part D insurance coverage during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
INS_OTHER	Whether the person had insurance coverage during the period that is not otherwise included in the INS_* variables during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
PLAN_HMO	Whether the person had insurance coverage under an HMO plan during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
PLAN_POS	Whether the person had insurance coverage in a point-of-service plan during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
PLAN_PPO	Whether the person had insurance coverage in a preferred provider organization plan during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
PLAN_INDEMNITY	Whether the person had insurance coverage in a traditional indemnity plan during the period	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown
DRUGCOV	Whether the person had <i>any</i> insurance coverage that included at least some coverage for prescription drugs	char(1)	Y = Yes E = Yes, but from an External organization N = No U = Unknown

Variable Name	Definition	Type(Len)	Values
REMOVED FOR V3.3: OUTSIDE_UTILIZATION	Is this person/period likely to have substantial health care utilization outside the HMO system under circumstances that would not result in a claim being submitted to the HMO?	char(1)	Y = Yes, we expect capture is incomplete N = No, there is no reason to suspect incomplete capture U = Unknown completeness of capture
ADDED FOR V3.3: INCOMPLETE_OUTPT_RX	Is outpatient pharmacy fill data suspected to be incomplete for this person/period?	char(1)	K = There are known reasons to suspect incomplete capture N = No, there is no known reason to suspect incomplete capture X = This variable not implemented at this site
ADDED FOR V3.3: INCOMPLETE_OUTPT_ENC	Is outpatient encounter data suspected to be incomplete for this person/period?	char(1)	K = There are known reasons to suspect incomplete capture N = No, there is no known reason to suspect incomplete capture X = This variable not implemented at this site
ADDED FOR V3.3: INCOMPLETE_INPT_ENC	Is inpatient encounter data suspected to be incomplete for this person/period?	char(1)	K = There are known reasons to suspect incomplete capture N = No, there is no known reason to suspect incomplete capture X = This variable not implemented at this site
ADDED FOR V3.3: INCOMPLETE_EMR	Is electronic medical record data (e.g. social history, vital signs) suspected to be incomplete for this person/period?	char(1)	K = There are known reasons to suspect incomplete capture N = No, there is no known reason to suspect incomplete capture X = This variable not implemented at this site
ADDED FOR V3.3: INCOMPLETE_TUMOR	Is tumor data suspected to be incomplete for this person/period?	char(1)	K = There are known reasons to suspect incomplete capture N = No, there is no known reason to suspect incomplete capture X = This variable not implemented at this site
ADDED FOR V3.3: INCOMPLETE_LAB	Is lab results data suspected to be incomplete for this person/period?	char(1)	K = There are known reasons to suspect incomplete capture N = No, there is no known reason to suspect incomplete capture X = This variable not implemented at this site
ENROLLMENT_BASIS	The basis for the claim that we have at least partial capture of the medical care information during the period. Historically 'Insurance was the only permissible basis	char(1)	G = Geography I = Insurance B = Both geography & insurance P = Non-enrolled patient
PCC	The primary care clinic to which the patient is paneled in administrative record.	char(*)	Same value set as UTILIZATION.FACILITY_CODE

PCP	The clinician to which the patient is paneled in administrative record char(*)	Values link as a foreign key to UTILIZATION.PROVIDER
-----	--	--

**Primary Key:**

MRN + ENR\_START (no overlapping time periods are permitted)

**Foreign Key Relationship:**

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
MRN	DEMOGRAPHICS	MRN	No
PCP	PROVIDER	PROVIDER	No

**Usage Notes**

Many sites have added site-specific variables that are not a part of the specifications. At these sites, users may well find multiple records for contiguous time periods, over which none of the official specification variables change. Users will likely be interested in the %CollapsePeriods or enrollment-related macros for collapsing records in to continuous periods.

For the INCOMPLETE\_XXX variables, users are advised to treat N and X values as essentially identical in information content. For applications that require, for example, complete Tumor data over time, we recommend using a condition like:  
 WHERE incomplete\_tumor ne 'Y'

Users may also want to avail themselves of the enrollment\_basis variable for applications that require complete pharmacy or encounter data. When in doubt please

# ENCOUNTER

Version = 3.0, Implemented 6/15/2011, StdVar = &\_vdw\_utilization

## Subject Area Description

The ENCOUNTER table contains one record for each significant medical-related interaction between a patient and a medical provider regardless of care setting or type of encounter. All known encounters should be included whether known through claims, through clinical systems, or a third source type. Inclusion, exclusions, classification, and uniqueness of an encounter is dependent on the type of service and care setting. For a precise description, please refer to the Utilization Implementation Guidelines on the private HCSRN website.

Variable Name	Definition	Type(Len)	Values
MRN	Medical record number is the unique patient identifier within a site and should never leave the site	char(*)	Unique to each patient at each site
ADATE	Encounter or admit date for inpatient or institutional stay. If encounter/admit date is unknown, then use the first date of a claim.	num(4)	SAS date
DDATE	Discharge date for inpatient and overnight encounters. If discharge date is unknown, then use the last date of a claim.	num(4)	SAS date
PROVIDER	Identifies the provider most responsible for the encounter.	char(*)	Unique to each provider at each site. If unknown, then use value "UNK"
ENC_ID	Uniquely identifies the encounter. The value should not change at update. An encounter should be unique across MRN, ADATE, ENCTYPE, PROVIDER, ENCOUNTER_SUBTYPE, FACILITY_CODE and appointment time.	char(*)	Unique to each encounter at each site.
DRG_VERSION	Identifies the version of the Diagnostic Related Group value (DRG_VALUE). Expected for hospital and some institutional stays but populate for all encounters where known.	char(1)	A = CMS-DRG (used prior to 10/1/07) B = MS-DRG (used post 10/1/07)
DRG_VALUE	The Diagnostic Related Group value. Used for hospital encounters. Using leading zeros for codes less than 100.	char(3)	<a href="#">Values maintained by Centers for Medicare &amp; Medicaid Services</a>
ENC_COUNT	The number of visits associated with this encounter. Value may be greater than one when a claim indicates a number of visits, but dates of visits are not specified as may occur in a bundled claim e.g. of dialysis services.	num(4)	Positive integer value

Variable Name	Definition	Type(Len)	Values
ENCTYPE	The type of encounter. Valid associated ENCOUNTER_SUBTYPE values are in brackets	char(2)	<p data-bbox="1417 191 1948 293">IP = Acute inpatient hospital stay including inpatient stays, same-day hospital discharges, hospital transfers when patient was admitted into hospital, acute inpatient psych, and detox stays. [SUB = AI]</p> <p data-bbox="1417 329 1919 378">ED = Emergency department excluding urgent care [SUB = HA, OC]</p> <p data-bbox="1417 414 1944 521">AV = Ambulatory visit including outpatient clinics, same-day surgeries, observation beds, urgent care visits, and same-day ambulatory hospital encounters. Excludes emergency department. [SUB = OC, OB, SD, HA, UC, RH, DI, OT]</p> <p data-bbox="1417 557 1780 581">TE = Telephone encounter [SUB = OT, HH]</p> <p data-bbox="1417 610 1896 634">EM = E-mail including secure messaging [SUB = OT, HH]</p> <p data-bbox="1417 670 1934 777">IS = Non-acute institutional stays including hospice, SNF, rehab, nursing home, residential, overnight non-hospital dialysis and other non-hospital stays [SUB = HS, SN, NH, RH, DI, OT]</p> <p data-bbox="1417 813 1948 886">OE = Other non-overnight encounter including hospice, home health, SNF, and other non-hospital visits [SUB = HS, HH, SN, RH, DI, OT]</p> <p data-bbox="1417 922 1948 961">LO = A lab only encounter that cannot be matched to another encounter [SUB = OC, OT]</p>

Variable Name	Definition	Type(Len)	Values
ENCOUNTER_SUBTYPE	Further specification of the type of encounter. Valid associated ENCTYPE values are in brackets	char(2)	AI = Acute inpatient stay excluding observation bed [TYPE=IP] OB = Observation bed [TYPE = AV] OC = Outpatient clinic visit [TYPE = AV, LO, RO, ED] SD = Same-day surgery [TYPE = AV] UC = Urgent care [TYPE = AV] HA = Hospital ambulatory; outpatient care at hospital excluding same-day surgery and observation beds [TYPE = AV, ED] HS = Hospice [TYPE = IS, OE] HH = Home health [TYPE = OE, TE, EM] SN = Skilled nursing facility [TYPE = IS, OE] NH = Nursing home including intermediate care facilities [TYPE = IS] RH = Rehab [TYPE = IS, AV, OE] DI = Dialysis [TYPE = IS, AV, OE] OT = Other non-hospital [TYPE = IS, OE, TE, EM AV, LO, RO]
FACILITY_CODE	A code indicating the facility, hospital, or clinic in which the encounter occurred.	char(*)	Unique to each facility at each site. If unknown, then use value "UNK" A = Alive E = Expired U = Unknown
DISCHARGE_DISPOSITION	The disposition of the patient at discharge for events with discharge.	char(1)	

Variable Name	Definition	Type(Len)	Values
DISCHARGE_STATUS	The status of the patient at discharge for events with discharge.	char(2)	AF=Adult Foster Home AL =Assisted Living Facility AM=Against Medical Advice AW=Absent without leave EX=Expired HH=Home Health HS=Hospice HO=Home / Self Care IP=Other Acute Inpatient Hospital NH=Nursing Home (Includes ICF) OT=Other RS=Residential Facility RH=Rehabilitation Facility SH=Still In Hospital SN=Skilled Nursing Facility UN=Unknown
ADMITTING_SOURCE	The location from which the patient was admitted for events with admit.	char(2)	AV=Ambulatory Visit ED=Emergency Department AF=Adult Foster Home AL =Assisted Living Facility HH=Home Health HS=Hospice HO=Home / Self Care IP=Other Acute Inpatient Hospital NH=Nursing Home (Includes ICF) OT=Other RS=Residential Facility RH=Rehabilitation Facility SN=Skilled Nursing Facility UN=Unknown
DEPARTMENT	The department where the encounter took place. This is not necessarily the specialty of the clinician providing services.	char(4)	If department information is unavailable in the source data (as may happen in claims data), specify as "UNK". See Appendix A - DEPARTMENT for values

**Primary Key:**

ENC\_ID

---

**Foreign Key Relationship:**

---

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
MRN	DEMOGRAPHICS	MRN	No
PROVIDER	PROVIDER	PROVIDER	No

# DIAGNOSIS

Version = 3.1, Implemented 10/31/2012, StdVar = &\_vdw\_dx

## Subject Area Description

The DIAGNOSIS table contains all recorded diagnoses associated with the encounters indicated in the ENCOUNTERS table with the exception of admitting diagnoses for inpatient stays.

Variable Name	Definition	Type(Len)	Values
MRN	Medical record number is the unique patient identifier within a site and should never leave the site	char(*)	Unique to each patient at each site
ENCTYPE	Refer to the ENCTYPE variable on the ENCOUNTER table for definition, type, length, and value set. This variable's redundancy is to improve querying performance.		
ENC_ID	Foreign key to the ENCOUNTER table uniquely identifying the encounter.	char(*)	Unique to each encounter at each site.
PROVIDER	Refer to the PROVIDER variable on the ENCOUNTER table for definition, type, length, and value set. This variable's redundancy is to improve querying performance.		
DIAGPROVIDER	Identifies the provider that made the diagnosis. If unknown, set the value equal to the PROVIDER variable.	char(*)	
ADATE	Refer to the ADATE variable on the ENCOUNTER table for definition, type, length, and value set. This variable's redundancy is to improve querying performance.		
DX	The diagnosis made. For ICD diagnosis coding, include decimal points in the value.	char(*)	<a href="#">ICD diagnosis values are set by CMS, others may be defined locally</a>
DX_CODETYPE	The coding set used in the DX variable for this record.	char(2)	07='ICD-7-CM' (including 'ICD-7') 08='ICD-8-CM' (including 'ICD-8') 09='ICD-9-CM' (including 'ICD-9') 10='ICD-10-CM' (including 'ICD-10') 11='ICD-11-CM' (including 'ICD-11') OT='Other'
ORIGDX	The diagnosis code as reported in source data without standardization or cleaning.	char(*)	

Variable Name	Definition	Type(Len)	Values
PRINCIPAL_DX	For hospital admissions, whether this diagnosis is the principal discharge diagnosis of the encounter. The principal diagnosis indicates the main reason why the patient was admitted to the hospital for care and the value on which a DRG is assigned.	char(1)	P = Principal diagnosis N = Not principal diagnosis X = Unknown or not classifiable
PRIMARY_DX	Whether this diagnosis is the primary diagnosis of the encounter. The primary diagnosis is the most serious or resource intensive diagnosis and is the primary reason for the procedures being rendered.	char(1)	P = Primary diagnosis S = Secondary diagnosis X = Unknown or not classifiable

**Primary Key:**

ENC\_ID + DX + ORIGDX + DIAGPROVIDER + DX\_CODETYPE

**Foreign Key Relationship:**

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
MRN	DEMOGRAPHICS	MRN	No
ENC_ID	ENCOUNTER	ENC_ID	No
PROVIDER	PROVIDER	PROVIDER	No
DIAG_PROVIDER	PROVIDER	PROVIDER	No

# PROCEDURE

Version = 3.0, Implemented 6/15/2011, StdVar = &\_vdw\_px

## Subject Area Description

---

The PROCEDURE table contains all recorded procedures that were actually performed associated with the encounters indicated in the ENCOUNTERS table.

Variable Name	Definition	Type(Len)	Values
MRN	Medical record number is the unique patient identifier within a site and should never leave the site	char(*)	Unique to each patient at each site
ENCTYPE	Refer to the ENCTYPE variable on the ENCOUNTER table for definition, type, length, and value set. This variable's redundancy is to improve querying performance.		
ENC_ID	Foreign key to the ENCOUNTER table uniquely identifying the encounter.	char(*)	Unique to each encounter at each site.
PROVIDER	Refer to the PROVIDER variable on the ENCOUNTER table for definition, type, length, and value set. This variable's redundancy is to improve querying performance.		
PERFORMINGPROVIDER	Identifies the provider that performed the procedure. If unknown, set the value equal to the PROVIDER variable.	char(*)	
ADATE	Refer to the ADATE variable on the ENCOUNTER table for definition, type, length, and value set. This variable's redundancy is to improve querying performance.		
PROCDATE	The actual date when the procedure was performed. Unknown values should be coded as missing/null.	num(4)	SAS Date
PX	Code identifying the procedure that was performed	char(*)	Value set is dependent on the PX_CODETYPE value and is set by external organizations.
ORIGPX	The procedure code as reported in source data without standardization or cleaning.	char(*)	

Variable Name	Definition	Type(Len)	Values
PX_CODETYPE	Identifies the coding scheme used in the PX variable to identify the procedure performed.	char(2)	09 = ICD-9-CM 10 = ICD-10-CM 11 = ICD-11-CM C4 = CPT-4 H4 = HCPCS-4 RV = Revenue code LO = Local homegrown OT = Other
PXCNT	Number of times that the procedure was performed during this encounter.	num(4)	Positive integer value
CPTMOD1	First modifier to a CPT procedure code use to communicate special circumstances related to the performance of a procedure.	char(2)	<a href="#">Value set is set by the American Medical Association and proprietary</a>
CPTMOD2	Second modifier to a CPT procedure code use to communicate special circumstances related to the performance of a procedure.	char(2)	<a href="#">Value set is set by the American Medical Association and proprietary</a>
CPTMOD3	Third modifier to a CPT procedure code use to communicate special circumstances related to the performance of a procedure.	char(2)	<a href="#">Value set is set by the American Medical Association and proprietary</a>

**Primary Key:**

ENC\_ID + PROCDATE + PERFORMINGPROVIDER + ORIGPX + PX + PX\_CODETYPE + CPTMOD1-3

**Foreign Key Relationship:**

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
MRN	DEMOGRAPHICS	MRN	No
ENC_ID	ENCOUNTER	ENC_ID	No
PROVIDER	PROVIDER	PROVIDER	No
PERFORMINGPROVIDER	PROVIDER	PROVIDER	No

# PROVIDER

Version = 3.0, Implemented 6/15/2011, StdVar = &\_vdw\_provider\_specialty

## Subject Area Description

The PROVIDER table contains provider level descriptives for the providers found in VDW tables. Every provider appearing in any other VDW file should appear in the PROVIDER table, even if demographics information is unknown.

Variable Name	Definition	Type(Len)	Values
PROVIDER	The unique provider identifier.	char(*)	Unique to each provider at each site
SPECIALTY	The provider's specialty.	char(3)	See Appendix C - SPECIALTY for value list
SPECIALTY2 - SPECIALTYN	Optional variables (as many as desired) to indicate when a provider has multiple specialties.	char(3)	See Appendix C - SPECIALTY for value list
PROVIDER_TYPE	The position, job, title, or role of the provider.	char(3)	See Appendix C - PROVIDER_TYPE for value list
PROVIDER_BIRTH_YEAR	The year that the provider was born.	num(4)	Leave missing if unknown
PROVIDER_GENDER	The provider's gender and/or sex; if both gender and sex are known, this variable holds gender.	char(1)	M = Male F = Female O = Other including transgendered U = Unknown
PROVIDER_RACE	Refer to the RACE1 variable on the DEMOGRAPHICS table for definition, type, length, and value set.		
PROVIDER_HISPANIC	Refer to the HISPANIC variable on the DEMOGRAPHICS table for definition, type, length, and value set.		
YEAR_GRADUATED	The year that the provider graduated from medical/nursing/technical school.	num(4)	Leave missing if unknown

### Primary Key:

PROVIDER

### Foreign Key Relationship:

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
[None Defined]			

# PHARMACY

Version = 3.0, Implemented 6/15/2010, StdVar = &\_vdw\_rx

## Subject Area Description

---

The PHARMACY file contains data on medications dispensed in the outpatient setting. Dispensings in the inpatient setting are not included. Incomplete or unfilled medication orders are also excluded. Rows are unique on the combination of patient, NDC, dispense date, and prescribing provider. When multiple dispensings occur for the same patient for the same drug on the same day from the same provider, the amount dispensed and days supplied of the drug should be summed. In the rare event that these multiple dispensings were prescribed by multiple providers, select only one to attribute as prescribing provider. Compound drugs are may be represented differently across sites. If a drug of interest is a compound drug, users may want to check with individual sites on their handling of these dispensings.

Variable Name	Definition	Type(Len)	Values
MRN	Medical record number is the unique patient identifier within a site and should never leave the site	char(*)	Unique to each patient at each site
RXDATE	Date that medication was dispensed	num(4)	SAS Date
NDC	The National Drug Code in the HIPAA/CMS/NCPDP standard 5-4-2 configuration without dashes.	char(11)	
RXSUP	Number of days of medication supplied	num(4)	Positive values
RXAMT	Number of units (e.g. pills, tablets) dispensed	num(4)	Positive values
RXMD	The provider that prescribed this medication	char(*)	

### Primary Key:

MRN + RXDATE + NDC + RXMD

### Foreign Key Relationship:

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
MRN	DEMOGRAPHICS	MRN	No
NDC	EVERNDC	NDC	No
RXMD	PROVIDER	PROVIDER	No

# EVERNDC

Version = 3.1, Implemented 2/15/2013, StdVar = &\_vdw\_everndc

## Subject Area Description

---

The EVERNDC table is a lookup table containing all National Drug Codes (NDC) found in the PHARMACY table at each site. That is, pharmacy dispensings at occurring at site A will have at least one corresponding record in site A's EVERNDC table but not necessarily any records in site B's EVERNDC table.

Variable Name	Definition	Type(Len)	Values
NDC	The National Drug Code in the HIPAA/CMS/NCPDP standard 5-4-2 configuration without dashes	char(11)	
NDC_FDA	The original FDA National Drug Code in 4-4-2, 5-3-2, or 5-4-1 format including dashes	char(12)	
GENERIC	The generic name of the drug in all upper case with form, strength, and other labels expunged	char(105)	
BRAND	The brand name of the drug in all upper case with form, strength, and other labels expunged	char(105)	
AHFS1 - AHFS7	The American Society of Health-System Pharmacists drug classification code. If code is in the 6 digit format, pad with two zeros at the end of the code. Store up to seven different AHFS codes across these variables	char(8)	<a href="#">Value set is set by AHFS and considered proprietary.</a>
GPI	The Generic Product Identifier drug classification code	char(14)	<a href="#">Value set is set by Medi-Span and is considered proprietary.</a>
UNIT_OF_MEASURE	The unit of measure as reported in source data without standardization or cleaning except for storing values in all upper case	char(*)	Example values are 2000, U/4ML, GM/15ML, %/5GM, -400 UNIT
STRENGTH	The drug strength as reported in source data without standardization or cleaning except for storing values in all upper case	char(*)	Example values are 99.99%, 9G, 9MCG/0.3ML, 9000 UNIT

Variable Name	Definition	Type(Len)	Values
DOSAGE_FORM	The dosage form as reported in source data without standardization or cleaning except for storing values in all upper case	char(*)	Example values are VIAL, WAFER, SOLN SEQ, SPIRIT
OBSOLETE_DT	The date the NDC was made obsolete, only if the drug is reinstated with a different NDC/Generic combination at a later time	num(4)	SAS Date
REINSTATED_DT	The date the NDC was reinstated, only if the drug was reinstated with a different NDC/Generic combination	num(4)	SAS Date

**Primary Key:**

NDC + GENERIC

**Foreign Key Relationship:**

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
[None Defined]			

**Usage Notes**

For multi-site projects, a locally created list of NDCs rarely captures all NDCs used at all participating HCSRN sites. A two-step approach that uses the EverNDC table will help compile a complete list of all NDCs for the studied drug. The lead analyst of a project can first write a distributed program to query the EverNDC tables across sites. They may use generic names, brand names, AHFS codes (if available), or GPI codes (if available) to identify the NDCs in use at other sites. They can then collate the NDCs identified from all participating sites to compile a final list of NDCs for the study. Compound drugs are represented differently in the data across sites. If the drug you are looking for may be a compound drug, please check with individual sites to understand how they are handling these data.

# LAB

Version = 3.1, Implemented 9/1/2014, StdVar = &\_vdw\_lab

## Subject Area Description

The LAB table contains the results of laboratory tests performed on patients. If a test is not resulted or if the result is unknown for whatever reason (e.g. specimen not sufficient), then test should not appear in the table.

variable name	definition	type(len)	values
MRN	Medical record number is the unique patient identifier within a site and should never leave the site	char(*)	Unique to each patient at each site
TEST_TYPE	VDW-specific classification of lab test types.	char(20)	See Appendix D - Lab Test Types
LOINC	Logical Observation Identifiers Names and Codes (LOINC) for this test.	char(10)	<a href="#">Value list managed by Regenstrief. Do not pad with leading zeros.</a>
STAT	Indicates the immediacy of the test.	char(1)	E=Expedite S=Stat R=Routine U=Unknown or missing
PT_LOC	Location of the patient when the lab specimen was obtained.	char(1)	I=Inpatient O=Outpatient E=Emergency Department H=Home U=Unknown or missing
RESULT_LOC	General location where the result was completed. Particularly whether the lab was completed in a certified laboratory.	char(1)	L = Lab P = Point of care (e.g. home, provider office) Code nulls and unknowns as "L"
SPECIMEN_ID	Used to uniquely identify a collected specimen which may ultimately be used to obtain multiple lab results.	char(*)	

Variable Name	Definition	Type(Len)	Values
			BLOOD SERUM PLASMA SR_PLS = serum and/or plasma PPP = Platelet Poor Plasma CSF = cerebral spinal fluid URINE STOOL NSWAB = nasal swab (including nose) NWASH = nasal wash NPH = nasopharyngeal swab NPWASH = nasopharyngeal wash THRT = throat, oropharyngeal swab SPUTUM BAL = bronchoalveolar lavage (BAL) BALBX = BAL biopsy OTHER NS = not specified [Null] = missing value
SPECIMEN_SOURCE	The source and/or method for collecting the specimen.	char(6)	
LOCAL_CD	The lab test type code as reported in source data without standardization or cleaning. This code indicates in the data source system which test was performed.	char(*)	
BATTERY_CD	The battery code as reported in source data without standardization or cleaning. A battery consists of a grouping or series of tests performed.	char(*)	
PX	Refer to the PX variable on the Procedure file for type and length.	Refer to the PX variable on the Procedure file for value set EXCEPT missing is also an acceptable value on the lab result table.	
PX_CODETYPE	Refer to the PX_CodeType variable on the Procedure file for type and length.	Refer to the PX_CodeType variable on the Procedure file for value set EXCEPT missing is also an acceptable value on the lab result table.	
ORDER_ID	Uniquely identifies the order for this lab test. In the future, may link to a table of orders.	char(*)	

Variable Name	Definition	Type(Len)	Values
ORDER_DT	Date that the lab test was ordered.	num(4)	SAS date
LAB_DT	Date that the specimen was collected.	num(4)	SAS date
LAB_TM	Time that the specimen was collected.	num(4)	SAS time
RESULT_DT	Date that the test was resulted.	num(4)	SAS date
RESULT_TM	Time that the specimen was resulted.	num(4)	SAS time
RESULT_C	The result of the test stored in a character field standardized by the HCSRN.	char(8)	See Appendix D - Lab Test Types
MODIFIER	Modifies the value stored in the RESULT_C field.	char(2)	TX = Text EQ = Equal LT = Less than LE = Less than or equal to GT = Greater than GE = Greater than or equal to
RESULT_UNIT	The units of the resulted test in a character field. For some TEST_TYPE values, units have been standardized.	char(11)	See Appendix D - Lab Test Types for standardized units
NORMAL_LOW_C	The lowest value still considered normal for this test.	char(8)	

Variable Name	Definition	Type(Len)	Values
MODIFIER_LOW	Modifies the value stored in the NORMAL_LOW_C field.	char(2)	TX = Text EQ = Equal LT = Less than LE = Less than or equal to GT = Greater than GE = Greater than or equal to Null if NORMAL_LOW_C is null
NORMAL_HIGH_C	The highest value still considered normal for this test.	char(8)	
MODIFIER_HIGH	Modifies the value stored in the NORMAL_HIGH_C field.	char(2)	TX = Text EQ = Equal LT = Less than LE = Less than or equal to GT = Greater than GE = Greater than or equal to Null if NORMAL_HIGH_C is null
ABN_IND	Indicates whether the test result is abnormal.	char(2)	AB = Abnormal C = Critical AH = Abnormal high CH = Critical high AL = Abnormal low CL = Critical low IN = Inconclusive NL = Normal UK = Unknown or missing

Variable Name	Definition	Type(Len)	Values
ORDER_PROV	Identifies the provider that ordered the lab test.	char(*)	
ORDER_DEPT	The department or specialty in which the order took place.	char(4)	See Appendix A - DEPARTMENT for values
FACILITY_CODE	A code indicating the facility, hospital, or clinic in which the lab order originated.	char(*)	Value set should match ENCOUNTER.FACILITY_CODE
ROW_ID	A unique row identifier of the LAB table that allows for the joining to the LAB_NOTES table.	char(*)	

**Primary Key:**

No primary key defined

**Foreign Key Relationship:**

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
MRN	DEMOGRAPHICS	MRN	No
ORDER_PROV	PROVIDER	PROVIDER	No

**Usage Notes**

The missingness of ORDER\_DT, LAB\_DT, and RESULT\_DT varies greatly. Users are advised to coalesce values across these date variables in the order first of LAB\_DT, second RESULT\_DT, and last ORDER\_DT.

# LAB NOTES

Version = 3.0, Implemented 5/15/2011, StdVar = &\_vdw\_lab\_notes

## Subject Area Description

---

The LAB NOTES table includes text notes that may accompany lab test results. If a note is long, it may span over several records in the table.

Variable Name	Definition	Type(Len)	Values
ROW_ID	A site specific identifier that enables linkage between the LAB NOTES and LAB tables.	char(*)	
RESULT_NOTE	Contains in text the result note.	char(80)	
NOTE_TYPE	Identifies the type of note in the RESULT_NOTE field.	char(1)	R = Result L = Low normal value H = High normal value N = Note
LINE	A line counter starting at value one and incrementing as needed to accommodate the entire RESULT_NOTE + NOTE_TYPE	num(4)	Positive integer value

### Primary Key:

ROW\_ID + NOTE\_TYPE + LINE

### Foreign Key Relationship:

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
ROW_ID	LAB	ROW_ID	No

# VITAL SIGNS

Version = 3.1, Implemented 9/12/2013, StdVar = &\_vdw\_vitalsigns

## Subject Area Description

The VITAL SIGNS table includes physiological measures taken by health professionals during encounters. Additional measures on anthropometry and tobacco use are also included.

Variable Name	Definition	Type(Len)	Values
MRN	Medical record number is the unique patient identifier within a site and should never leave the site	char(*)	Unique to each patient at each site
ENC_ID	Foreign key to the ENCOUNTER table uniquely identifying the encounter; unlinkable records should contain null/missing values	char(*)	Unique to each encounter at each site
MEASURE_DATE	The date on which these vital signs were measured	num(4)	SAS Date
MEASURE_TIME	The time at which these vital signs were measured	num(4)	SAS Time
ENCTYPE	Refer to the ENCTYPE variable on the ENCOUNTER table for definition, type, length, and value set		
HT	The height measured in inches of the patient at this time	num(8)	Positive real number
WT	The weight measured in pounds of the patient at this time	num(8)	Positive real number
SYSTOLIC	The systolic blood pressure measured in mmHg of the patient at this time	num(4)	Integer between 0 and 999
DIASTOLIC	The diastolic blood pressure measured in mmHg of the patient at this time	num(4)	Integer between 0 and 999

Variable Name	Definition	Type(Len)	Values
BP_TYPE	The type of blood pressure taken	char(1)	R = Rooming O = Orthostatic M = Multiple E = Extended
POSITION	The patient's position for orthostatic blood pressure measurements	char(1)	1 = Sitting 2 = Standing 3 = Supine Null = Unknown
HT_RAW	The height measurement as reported in source data without standardization or cleaning	char(*)	Examples may specific values, ranges, or categories
WT_RAW	The weight measurement as reported in source data without standardization or cleaning	char(*)	Examples may specific values, ranges, or categories
BMI_RAW	The patient's body mass index measurement as reported in source data without standardization, calculation, or cleaning	char(*)	
HEAD_CIR_RAW	The patient's head circumference measurement as reported in source data without standardization or cleaning	char(*)	
SYSTOLIC_RAW	The systolic blood pressure measurement as reported in source data without standardization or cleaning	char(*)	Examples may specific values, ranges, or categories
DIASTOLIC_RAW	The diastolic blood pressure measurement as reported in source data without standardization or cleaning	char(*)	Examples may specific values, ranges, or categories
RESPIR_RAW	The respirations in breaths per minute measurement as reported in source data without standardization or cleaning	char(*)	
TEMP_RAW	The patient's body temperature measurement as reported in source data without standardization or cleaning	char(*)	
PULSE_RAW	The heartbeats per minute measurement as reported in source data without standardization or cleaning	char(*)	

**Primary Key:**

No primary key defined

**Foreign Key Relationship:**

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
MRN	DEMOGRAPHICS	MRN	No
ENC_ID	ENCOUNTER	ENC_ID	No

## Usage Notes

---

Body Mass Index (BMI) is not a cleaned variable on the VITAL SIGNS table. Users are recommended to calculate BMI at the time of actual need making use of the VDW Standard Macros. The RAW variables' primary purpose for capturing categories or ranges that may be recorded in the system (e.g. for BMI\_RAW values like "overweight", or "25.0-25.9" or for DIASTOLIC\_RAW values like "HYPERTENSION" or "100-109").

## SOCIAL HISTORY

version = 3.2, implemented 9/17/2014, stavar = &\_vw\_social\_hx

### Subject Area Description

The SOCIAL HISTORY table contains behavioral measures taken by health professionals during clinic visits, over the telephone, or via questionnaires. These measures include the use of tobacco, alcohol, and illegal drugs as well as sexual behavior and contraceptive use. Because of the range of tobacco products available and the variability in the level of detail recorded, precision and capture of tobacco related variables may vary both across and within sites. Consistency of response is also not guaranteed over time.

Social history measures may carry special privacy concerns. Beyond the nature of these variables, the use of free text fields in any content area warrants additional care as they may contain personal health information. Users are encouraged to consult with privacy experts and/or experienced users before extracting data if they are unfamiliar with these additional data sensitivities.

MRN	Medical record number is the unique patient identifier within a site and should never leave the site	char(*)	Unique to each patient at each site
CONTACT_DATE	Date of encounter or date of social history data collection	num(4)	SAS date, missing is not allowed
ENC_ID	Foreign key to the ENCOUNTER table uniquely identifying the encounter; unlinkable records should contain null/missing values	char(*)	SAS date, missing values allowed
EDUCATION_YEARS	Number of years of education completed recorded as free text	char(*)	Free text, missing values allowed
SEX_ACTIVE	Whether the person is sexually active	char(1)	Y = Yes N = No W = Not currently (Was) X = Not asked U = Unknown or missing
SEX_FEMALE_PARTNER	Whether the person has female sex partners	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing

Variable Name	Definition	Type(Len)	Values
SEX_MALE_PARTNER	Whether the person has male sex partners	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing
SEX_COMMENT	Comment about the person's sexual activity	char(*)	Free text, missing values allowed
BC_ABSTINENCE	Whether the person uses abstinence as birth control	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing
BC_CONDOM	Whether the person uses condoms as birth control	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing
BC_DIAPHRAGM	Whether the person uses a diaphragm as birth control	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing
BC_IMPLANT	Whether the person uses an implant as birth control	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing
BC_INJECTION	Whether the person uses injections as birth control	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing
BC_INSERTS	Whether the person uses inserts as birth control	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing
BC_IUD	Whether the person uses an intrauterine device (IUD) as birth control	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing

Variable Name	Definition	Type(Len)	Values
BC_PILL	Whether the person uses birth control pills	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing
BC_RHYTHM	Whether the person uses the rhythm method as birth control	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing
BC_SPERMICIDE	Whether the person uses spermicide as birth control	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing
BC_SPONGE	Whether the person uses a sponge as birth control	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing
BC_SURGICAL	Whether a surgery (e.g. vasectomy, hysterectomy) significantly reduces the chance of contraception	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing
ALCOHOL_USE	Whether the person is an alcohol drinker	char(1)	Y = Yes N = No Q = Quit/Former X = Not asked U = Unknown or missing
ALCOHOL_OUNCES_WEEK	The number of ounces of alcohol consumed per week	char(*)	Free text, missing values allowed
ALCOHOL_COMMENT	Comment about the person's alcohol use	char(*)	Free text, missing values allowed

Variable Name	Definition	Type(Len)	Values
ILLCIT_DRUG_USE	Whether the person uses illicit drugs	char(1)	Y = Yes N = No Q = Quit/Former X = Not asked U = Unknown or missing
ILLCIT_DRUG_USE_FREQ	Frequency of illicit drug use as free text	char(*)	Free text, missing values allowed
ILLCIT_DRUG_USE_COMMENT	Comment about the person's illicit drug use	char(*)	Free text, missing values allowed
IV_DRUG_USE	Whether the person uses IV drugs	char(1)	Y = Yes N = No Q = Quit/Former X = Not asked U = Unknown or missing
TOBACCO_USE	Whether the person is a user of tobacco	char(1)	Y = Yes / Current I = Infrequent N = Never Q = Quit / Former P = Passive X = Not Asked U = Unknown or missing
TOBACCO_USE_YEARS	The number of years the person used tobacco	char(*)	Free text, missing values allowed
TOBACCO_PACKS_DAY	Number of packs smoked per day as free text	char(*)	Free text, missing values allowed
TOBACCO_SMOKELESS_USE	Whether the person uses smokeless tobacco	char(1)	Y = Yes / Current I = Infrequent N = No / Never used Q = Quit / Former X = Not asked U = Unknown or missing

Variable Name	Definition	Type(Len)	Values
TOBACCO_SMOKELESS_START_DATE	Date the person started using smokeless tobacco	num(4)	SAS Date, missing allowed
TOBACCO_SMOKELESS_QUIT_DATE	Date the person quit using smokeless tobacco	num(4)	SAS Date, missing allowed
TOBACCO_SMOKING_USE	Whether the person uses smoking tobacco	char(1)	E = Current everyday S = Current some days H = Heavy smoker L = Light smoker Q = Former Y = Smoker, current status unknown N = Never smoked X = Never assessed P = Passive U = Unknown
TOBACCO_SMOKING_START_DATE	Date the person started using smoked tobacco	num(4)	SAS Date, missing allowed
TOBACCO_SMOKING_QUIT_DATE	Date the person quit using smoked tobacco	num(4)	SAS Date, missing allowed
TOBACCO_CHEW	Whether the person uses chew tobacco	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing
TOBACCO_CIGARETTES	Whether the person uses tobacco cigarettes	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing
TOBACCO_CIGARS	Whether the person uses tobacco cigars	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing
TOBACCO_PIPES	Whether the person uses tobacco pipes	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing

Variable Name	Definition	Type(Len)	Values
TOBACCO_SNUFF	Whether the person uses snuff tobacco	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing
TOBACCO_UNSPECIFIED	Whether the person uses an unspecified form of tobacco	char(1)	Y = Yes N = No X = Not asked U = Unknown or missing
TOBACCO_COMMENT	Comment about the person's tobacco use	char(*)	Free text, missing values allowed

**Primary Key:**

No primary key defined

**Foreign Key Relationship:**

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
MRN	DEMOGRAPHICS	MRN	No
ENC_ID	ENCOUNTER	ENC_ID	No

## Usage Notes

Data on marijuana use is most often contained as "free text" in the illicit\_drug\_comment variable, although the tobacco\_comment field may also contain pertinent information. However, relevant information may also be contained in a medical center's chart notes or 'problem lists,' which would not then be transferred into the VDW Social History table.

With the increasing decriminalization of marijuana, many may no longer consider it to be "illicit," and thus be uncertain where to record this information, particular in states where it has been legalized. Investigation of and familiarity with individual site's recording/data-population processes may be important to your needs.

Default settings for some variables may vary across sites, and may result in some values being more heavily weighted than others. Check with the site for default values if there are questions on value distribution. For example, one source system has a default value of No for Tobacco\_Chew, rather than Unknown or Not Asked.

Comment data at some sites may have restricted access. Site data managers should be consulted for rules governing access.

Date values may be outside of expected parameters. Query results involving dates should be carefully reviewed for extreme outliers.

# DEATH

Version = 3.2, Implemented 5/31/2014, StdVar = &\_vdw\_death

## Subject Area Description

---

The DEATH table contains one record per person in the DEMOGRAPHICS table for whom there is some belief that the person may be dead. When sources of death provide conflicting information, site data managers should make local determinations as to which source to use, collating information when possible, and reflecting their confidence in the observation overall using the CONFIDENCE variable.

Variable Name	Definition	Type(Len)	Values
MRN	Medical record number is the unique patient identifier within a site and should never leave the site	char(*)	Unique to each patient at each site
DEATHDT	The date that the person most likely died.	num(4)	SAS date, missing values allowed
DTIMPUTE	Indicates whether the death date is imputed and if so, how it was imputed.	char(1)	M = Month of date imputed D = Day of date imputed B = Both month & day imputed N = Date not imputed
SOURCE_LIST	A list of all sources of data that report this death regardless of any discrepancies in other variables (e.g. DEATHDT).	char(8)	A concatenated string of letters in descending order of reliability indicating all sources that report this person's death B = Social Security Admin N = National Death Index S = State Death records T = Tumor registry E = Encounter data P = Patient data M = Membership data O = Other
CONFIDENCE	Based on all information available, the level of confidence that this person is in fact dead.	char(1)	E = Excellent F = Fair P = Poor

### Primary Key:

MRN

### Foreign Key Relationship:

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
MRN	DEMOGRAPHICS	MRN	No

# CAUSE OF DEATH

Version = 3.2, Implemented 5/31/2014, StdVar = &\_vdw\_cause\_of\_death

## Subject Area Description

---

The CAUSE OF DEATH table contains one record per person per known cause of death.

Variable Name	Definition	Type(Len)	Values
MRN	Medical record number is the unique patient identifier within a site and should never leave the site	char(*)	Unique to each patient at each site
COD	The cause of death diagnosis mortality code. For ICD diagnosis coding, include decimal points in the value.	char(6)	
DX_CODETYPE	Refer to the DX_CODETYPE variable on the ENCOUNTER table for definition, type, length, and value set.		
CAUSETYPE	The type of cause of death.	char(1)	I = Immediate/Primary U = Underlying C = Contributory O = Other
SOURCE_LIST	Refer to the SOURCE_LIST variable on the DEATH table for definition, type, length, and value set.		
CONFIDENCE	Refer to the CONFIDENCE variable on the DEATH table for definition, type, length, and value set.		

### Primary Key:

MRN + COD + DX\_CODETYPE + CAUSETYPE

### Foreign Key Relationship:

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
MRN	DEMOGRAPHICS	MRN	No

# TUMOR

Version = 3.2, Implemented 9/30/2014, StdVar = &\_vdw\_tumor

## Subject Area Description

The TUMOR table contains records of documented neoplasms (typically malignant) diagnosed in patients as indicated in a tumor registry. There is one record per separately diagnosed neoplasm. Diagnoses of neoplasms in sources other than registries should not be included.

Variable Name	Definition	Type(Len)	Values
MRN	Medical record number is the unique patient identifier within a site and should never leave the site	char(*)	Unique to each patient at each site
SEQUENCE	As defined by NAACCR items #560 and #380. If both item numbers are known and discrepant, use item #560.	char(2)	As defined by NAACCR items #560.
DXDATE	<a href="#">As defined by NAACCR item #390</a>	num(4)	SAS date
DXDATE_IMPUTE	Indicates whether the DXDATE has been imputed, and if so, how it was imputed	char(1)	M = Month of date imputed D = Day of date imputed B = Both month & day imputed N = Date not imputed
DXYEAR	The year of DXDATE	num(4)	
ICDOSITE	<a href="#">As defined by NAACCR item #400</a>	char(4)	<a href="#">As defined by NAACCR item #400</a>
SS1977	As defined by NAACCR item #760. The derived version of summary stage (item #3010) is also allowed.	char(1)	<a href="#">As defined by NAACCR item #760</a>
SS2000	As defined by NAACCR item #759. The derived version of summary stage (item #3020) is also allowed.	char(1)	<a href="#">As defined by NAACCR item #759</a>
STAGEGEN	The general / summary stage of the tumor independent of year of staging guidelines. This variable is the coalesce of SS1977 and SS2000.	char(1)	<a href="#">As defined by NAACCR item #760</a>
DER_SS2000F	<a href="#">As defined by NAACCR item #3050</a>	char(1)	<a href="#">As defined by NAACCR item #3050</a>
AJCC_GRP	Best available AJCC stage	char(2)	0, 0A, 0S, 1, 1A, 1B, 1C, 2, 2A, 2B, 2C, 3, 3A, 3B, 3C, 4, 4A, 4B, 4C, 88, 99, OC, XX
STAGEAJ	AJCC summary stage or best AJCC stage	char(4)	As defined by AJCC
AJCC_FULL_DER	Expanded staging elements including neoadjuvant flags	char(30)	Examples include ypt3a ypN1a cM0 Stage: IIIB and pT3 pN0 cM0 Stage: II

Variable Name	Definition	Type(Len)	Values
AJCC_ED	AJCC Staging Scheme Edition	char(1)	0 = Not staged 1 = First Edition 2 = Second Edition 3 = Third Edition 4 = Fourth Edition 5 = Fifth Edition 6 = Sixth Edition 7 = Seventh Edition 8 = Not applicable (no AJCC scheme) 9 = Unknown edition
MORPH	<a href="#">As defined by NAACCR item #522</a>	char(4)	<a href="#">As defined by NAACCR item #522</a>
BEHAV	<a href="#">As defined by NAACCR item #523</a>	char(1)	<a href="#">As defined by NAACCR item #523</a>
GRADE	<a href="#">As defined by NAACCR item #440</a>	char(1)	<a href="#">As defined by NAACCR item #440</a>
DXAGE	<a href="#">As defined by NAACCR item #230</a>	num(3)	<a href="#">As defined by NAACCR item #230</a>
BDATE	<a href="#">As defined by NAACCR item #240</a>	num(4)	<a href="#">As defined by NAACCR item #240</a>
GENDER	<a href="#">As defined by NAACCR item #220</a>	char(1)	<a href="#">As defined by NAACCR item #220</a>
RACE1 - RACE5	<a href="#">As defined by NAACCR item #160-164</a>	char(2)	<a href="#">As defined by NAACCR item #160-164</a>
HISPANIC	<a href="#">As defined by NAACCR item #190</a>	char(1)	<a href="#">As defined by NAACCR item #190</a>
CLASS	<a href="#">As defined by NAACCR item #610</a>	char(2)	<a href="#">As defined by NAACCR item #610</a>
VITAL	<a href="#">As defined by NAACCR item #1760</a>	char(1)	<a href="#">As defined by NAACCR item #1760</a>
DCAUSE	<a href="#">As defined by NAACCR item #1910</a>	char(6)	<a href="#">As defined by NAACCR item #1910</a>
DOD	The person's date of death as reported or derived from the tumor registry.	num(4)	SAS date
DT_FU	<a href="#">As defined by NAACCR item #739</a>	num(4)	<a href="#">As defined by NAACCR item #739</a>
LATERALITY	<a href="#">As defined by NAACCR item #410</a>	char(1)	<a href="#">As defined by NAACCR item #410</a>
IDPLAN	A unique VDW implementing site identifier.	char(2)	As defined by the CODE documented on the HCSR Portal website

Variable Name	Definition	Type(Len)	Values
DATA_SOURCE	Indicates the source of data for this record.	char(4)	For the first two positions use: LO = Local registry SE = Regional SEER registry ST = State operated registry The final two positions are locally defined to further differentiate source. Use values 'XX' for these positions if they are not needed at your site.
DCNFRM	<a href="#">As defined by NAACCR item #490</a>	char(1)	<a href="#">As defined by NAACCR item #490</a>
DSTZ	<a href="#">As defined by NAACCR item #780</a>	char(3)	<a href="#">As defined by NAACCR item #780</a>
DAJC1T_P	<a href="#">As defined by NAACCR item #880</a>	char(5)	<a href="#">As defined by NAACCR item #880</a>
DAJC1N_P	<a href="#">As defined by NAACCR item #890</a>	char(5)	<a href="#">As defined by NAACCR item #890</a>
DAJC1M_P	<a href="#">As defined by NAACCR item #900</a>	char(5)	<a href="#">As defined by NAACCR item #900</a>
DAJC1T_C	<a href="#">As defined by NAACCR item #910</a>	char(5)	<a href="#">As defined by NAACCR item #910</a>
DAJC1N_C	<a href="#">As defined by NAACCR item #950</a>	char(5)	<a href="#">As defined by NAACCR item #950</a>
DAJC1M_C	<a href="#">As defined by NAACCR item #960</a>	char(5)	<a href="#">As defined by NAACCR item #960</a>
DSRG_FAC	<a href="#">As defined by NAACCR item #670</a>	char(2)	<a href="#">As defined by NAACCR item #670</a>
DRAD_FAC	<a href="#">As defined by NAACCR item #690</a>	char(2)	<a href="#">As defined by NAACCR item #690</a>
DCHM_FAC	<a href="#">As defined by NAACCR item #700</a>	char(2)	<a href="#">As defined by NAACCR item #700</a>
DHRM_FAC	<a href="#">As defined by NAACCR item #710</a>	char(2)	<a href="#">As defined by NAACCR item #710</a>
DBRM_FAC	<a href="#">As defined by NAACCR item #720</a>	char(2)	<a href="#">As defined by NAACCR item #720</a>

Variable Name	Definition	Type(Len)	Values
DOTH_FAC	<a href="#">As defined by NAACCR item #730</a>	char(2)	<a href="#">As defined by NAACCR item #730</a>
HTE_FAC	As defined per NAACCR item #3250 except that the source of information is from the reporting facility only	char(2)	As defined per NAACCR item #3250
R_MOD_FAC	As defined per NAACCR item #1570 except that the source of information is from the reporting facility only	char(2)	As defined per NAACCR item #1570
R_VOL_FAC	As defined per NAACCR item #1540 except that the source of information is from the reporting facility only	char(2)	<a href="#">As defined per NAACCR item #1540</a>
B_MOD_FAC	As defined per NAACCR item #3200 except that the source of information is from the reporting facility only	char(2)	As defined per NAACCR item #3200
B_VOL_FAC	As defined per NAACCR item #1540 except that source of information is from the reporting facility only and for boost radiation rather than regional radiation	char(2)	<a href="#">As defined per NAACCR item #1540</a>
DNDI	<a href="#">As defined by NAACCR item #820</a>	char(2)	<a href="#">As defined by NAACCR item #820</a>
DNDX	<a href="#">As defined by NAACCR item #830</a>	char(2)	<a href="#">As defined by NAACCR item #830</a>
DTMRK1	<a href="#">As defined by NAACCR item #1150</a>	char(1)	<a href="#">As defined by NAACCR item #1150</a>
DTMRK2	<a href="#">As defined by NAACCR item #1160</a>	char(1)	<a href="#">As defined by NAACCR item #1160</a>
DTMRK3	<a href="#">As defined by NAACCR item #1170</a>	char(1)	<a href="#">As defined by NAACCR item #1170</a>
CLN_STG	<a href="#">As defined by NAACCR item #970</a>	char(4)	<a href="#">As defined by NAACCR item #970</a>
EOD	Extent of disease	char(12)	
DT_SURG	<a href="#">As defined by NAACCR item #1200</a>	num(4)	<a href="#">As defined by NAACCR item #1200</a>
DT_CHEMO	<a href="#">As defined by NAACCR item #1220</a>	num(4)	<a href="#">As defined by NAACCR item #1220</a>
DT_HORM	<a href="#">As defined by NAACCR item #1230</a>	num(4)	<a href="#">As defined by NAACCR item #1230</a>

Variable Name	Definition	Type(Len)	Values
DT_RAD	<a href="#">As defined by NAACCR item #1210</a>	num(4)	<a href="#">As defined by NAACCR item #1210</a>
DT_RAD_END	<a href="#">As defined by NAACCR item #3220</a>	num(4)	<a href="#">As defined by NAACCR item #3220</a>
DT_BRM	<a href="#">As defined by NAACCR item #1240</a>	num(4)	<a href="#">As defined by NAACCR item #1240</a>
DT_SYS	Date any systemic treatment began.	num(4)	SAS date
DT_HTE	Date of hematologic transplant and endocrine procedures	num(4)	SAS date
DT_OTH	<a href="#">As defined by NAACCR item #1250</a>	num(4)	<a href="#">As defined by NAACCR item #1250</a>
R_N_SURG	<a href="#">As defined by NAACCR item #1340</a>	char(1)	<a href="#">As defined by NAACCR item #1340</a>
R_N_CHEMO	<a href="#">As defined by NAACCR item #1440</a>	char(2)	<a href="#">As defined by NAACCR item #1440</a>
R_N_HORM	<a href="#">As defined by NAACCR item #1450</a>	char(1)	<a href="#">As defined by NAACCR item #1450</a>
R_N_RAD	<a href="#">As defined by NAACCR item #1430</a>	char(1)	<a href="#">As defined by NAACCR item #1430</a>
R_N_BRM	A placeholder for indicating why the patient did not receive immunotherapy as a part of first course of therapy.	char(1)	No value set defined
R_N_HTE	Reason no hematologic transplant or endocrine ablation was given.	char(1)	0 = HTE administered 1 = HTE not recommended 2 = HTE contraindicated; autopsy only cases 5 = HTE not administered, pt died 6 = Reason unknown for no HTE 7 = Patient/pt`s guardian refused HTE 8 = HTE recommended, unk if administered 9 = Unk if HTE recommended/admin; death cert only

Variable Name	Definition	Type(Len)	Values
R_N_OTH	A placeholder for indicating why the patient did not receive some other form of treatment as a part of the first course of therapy.	chart(1)	No value set defined
DSRG_SUM	As defined by NAACCR item #1290, #1646, and #1640. If multiple items are known, preference is given first to #1290, then #1646, and lastly #1640.	char(2)	As defined by NAACCR item #1290
DRAD_SUM	<a href="#">As defined by NAACCR item #1360</a>	char(2)	<a href="#">As defined by NAACCR item #1360</a>
DCHM_SUM	<a href="#">As defined by NAACCR item #1390</a>	char(2)	<a href="#">As defined by NAACCR item #1390</a>
DHRM_SUM	<a href="#">As defined by NAACCR item #1400</a>	char(2)	<a href="#">As defined by NAACCR item #1400</a>
DBRM_SUM	<a href="#">As defined by NAACCR item #1440</a>	char(2)	<a href="#">As defined by NAACCR item #1440</a>
DOTH_SUM	<a href="#">As defined by NAACCR item #1420</a>	char(2)	<a href="#">As defined by NAACCR item #1420</a>
HTE_SUM	<a href="#">As defined by NAACCR item #3250</a>	char(2)	<a href="#">As defined by NAACCR item #3250</a>
R_MOD_SUM	<a href="#">As defined by NAACCR item #1570</a>	char(2)	<a href="#">As defined by NAACCR item #1570</a>
R_VOL_SUM	<a href="#">As defined by NAACCR item #1540</a>	char(2)	<a href="#">As defined by NAACCR item #1540</a>
R_DOSE	<a href="#">As defined by NAACCR item #1510</a>	char(5)	<a href="#">As defined by NAACCR item #1510</a>
B_MOD_SUM	<a href="#">As defined by NAACCR item #3200</a>	char(2)	<a href="#">As defined by NAACCR item #3200</a>
B_VOL_SUM	As defined per NAACCR item #1540 except for boost radiation rather than regional radiation.	char(2)	<a href="#">As defined per NAACCR item #1540</a>
B_DOSE	<a href="#">As defined by NAACCR item #3210</a>	char(5)	<a href="#">As defined by NAACCR item #3210</a>
CS_SZ	<a href="#">As defined by NAACCR item #2800</a>	char(3)	<a href="#">As defined by NAACCR item #2800</a>

Variable Name	Definition	Type(Len)	Values
CS_EXT	<a href="#">As defined by NAACCR item #2810</a>	char(3)	<a href="#">As defined by NAACCR item #2810</a>
CS_NODES	<a href="#">As defined by NAACCR item #2830</a>	char(3)	<a href="#">As defined by NAACCR item #2830</a>
CS_NODES_EVAL	<a href="#">As defined by NAACCR item #2840</a>	char(1)	<a href="#">As defined by NAACCR item #2840</a>
CS_METS	<a href="#">As defined by NAACCR item #2850</a>	char(2)	<a href="#">As defined by NAACCR item #2850</a>
CS_METS_EVAL	<a href="#">As defined by NAACCR item #2860</a>	char(1)	<a href="#">As defined by NAACCR item #2860</a>
SSF1 - SSF25	Up to 25 different site specific factors. Definition and values are dependent on ICDOSITE and are documented in the NAACCR standards starting with item #2880 for SSF1.	char(3)	As defined by NAACC standards. Refer to the NAACCR data dictionary for value set.
DER_T6	<a href="#">As defined by NAACCR item #2940</a>	char(3)	<a href="#">As defined by NAACCR item #2940</a>
DER_T6_D	<a href="#">As defined by NAACCR item #2950</a>	char(1)	<a href="#">As defined by NAACCR item #2950</a>
DER_N6	<a href="#">As defined by NAACCR item #2960</a>	char(3)	<a href="#">As defined by NAACCR item #2960</a>
DER_N6_D	<a href="#">As defined by NAACCR item #2970</a>	char(1)	<a href="#">As defined by NAACCR item #2970</a>
DER_M6	<a href="#">As defined by NAACCR item #2980</a>	char(3)	<a href="#">As defined by NAACCR item #2980</a>
DER_M6_D	<a href="#">As defined by NAACCR item #2990</a>	char(1)	<a href="#">As defined by NAACCR item #2990</a>
DER_T7	<a href="#">As defined by NAACCR item #3400</a>	char(3)	<a href="#">As defined by NAACCR item #3400</a>
DER_T7_D	<a href="#">As defined by NAACCR item #3402</a>	char(1)	<a href="#">As defined by NAACCR item #3402</a>
DER_N7	<a href="#">As defined by NAACCR item #3410</a>	char(3)	<a href="#">As defined by NAACCR item #3410</a>
DER_N7_D	<a href="#">As defined by NAACCR item #3411</a>	char(1)	<a href="#">As defined by NAACCR item #3411</a>

Variable Name	Definition	Type(Len)	Values
DER_M7	<a href="#">As defined by NAACCR item #3420</a>	char(3)	<a href="#">As defined by NAACCR item #3420</a>
DER_M7_D	<a href="#">As defined by NAACCR item #3422</a>	char(1)	<a href="#">As defined by NAACCR item #3422</a>
PAL_FAC	<a href="#">As defined by NAACCR item #3280</a>	char(1)	<a href="#">As defined by NAACCR item #3280</a>
PAL_SUM	<a href="#">As defined by NAACCR item #370</a>	char(1)	<a href="#">As defined by NAACCR item #370</a>
RECUR_DT	<a href="#">As defined by NAACCR item #1860</a>	num(4)	<a href="#">As defined by NAACCR item #1860</a>
RECUR_TYPE	<a href="#">As defined by NAACCR item #1880</a>	char(2)	<a href="#">As defined by NAACCR item #1880</a>
RECUR_FL	<a href="#">As defined by NAACCR item #1861</a>	char(2)	<a href="#">As defined by NAACCR item #1861</a>

**Primary Key:**

MRN + SEQUENCE

**Foreign Key Relationship:**

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
MRN	DEMOGRAPHICS	MRN	No

**Usage Notes**

NAACCR definitions and values sets change over time. Each record reflects the definitions and values as they were set at the time of diagnosis (DXDATE).

# CENSUS LOCATION

Version = 3.1, Implemented 2/28/2014, StdVar = &\_vdw\_census\_loc

## Subject Area Description

---

The CENSUS LOCATION table contains geographic location markers for patient residencies over time.

Variable Name	Definition	Type(Len)	Values
MRN	Medical record number is the unique patient identifier within a site and should never leave the site	char(*)	Unique to each patient at each site
LOC_START	The date on which the person's tenure at this location began.	num(4)	SAS date
LOC_END	The date on which the person's tenure at this location ended.	num(4)	SAS date
GEOCODE	The concatenation of FIPS codes for State, County, Tract, Block Group, and Block	char(15)	<a href="#">FIPS values are set by the National Institute of Standards &amp; Technology and are public information</a>
GEOCODE_BOUNDARY_YEAR	The census year for which the GEOCODE applies.	num(8)	Year values (e.g. 2000, 2004)
GEOLEVEL	The most granular geographic level of the GEOCODE indicating the specificity of the match made. Also called 'map flag'.	char(1)	B = Block G = Block group T = Tract C = County Z = Zip code U = Unable to be appended P = Address is post office

Variable Name	Definition	Type(Len)	Values
MATCH_STRENGTH	A code indicating the type of match achieved by the geocoding software.	char(1)	0 = No coordinates 1 = Zip +0 centroid 2 = Zip +2 centroid 3 = Zip +4 centroid 4 = Shape path centroid 5 = Street address position 6 = Point zip centroid X = Street intersection
LATITUDE	The latitude of the location	num(8)	Value between -90 and +90 measured in degrees
LONGITUDE	The longitude of the location	num(8)	Value between -180 and +180 measured in degrees
GEOCODE_APP	The name of the application used to geocode this location	char(*)	Free text field

**Primary Key:**

MRN + LOC\_START (no overlapping time periods are permitted)

**Foreign Key Relationship:**

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
MRN	DEMOGRAPHICS	MRN	No
GEOCODE + CENSUS_YEAR	CENSUS DEMOG	GEOCODE + CENSUS_YEAR	No

# CENSUS DEMOGRAPHICS

Version = 3.1, Implemented 2/28/2014, StdVar = &\_vdw\_census\_demog

## Subject Area Description

The CENSUS DEMOGRAPHICS table contains US Census information in a given year, Census data source, and geocode.

Variable Name	Definition	Type(Len)	Values
CENSUS_YEAR	Year the census data was collected	Num(8)	Year values (e.g. 2000, 2010)
CENSUS_DATA_SRC	Source of the data in the record	Char(16)	Census, ACS5, ESRI, etc.
GEOCODE	Concatenation of the FIPS codes for State, County, Tract, Block Group and Block	Char(15)	The first two characters signify the state, the next three signify the county, the next six signify the tract, the next 1 signifies the block group, and the final 3 signify the block.
STATE	State Code--a two-character numeric code assigned to US states, districts, territories and protectorates.	char(2)	Any valid state code, as used by the USCB.
COUNTY	County Code--a 3-character numeric code assigned to census counties.	char(3)	Any valid county code, as used by the USCB.
TRACT	Tract Code--a six-character numeric code assigned to census tracts.	char(6)	Any valid tract code, as used by the USCP.
BLOCKGP	Block Group Code--a 1-character numeric code assigned to census block groups.	char(1)	Any valid block group code, as used by the USCB.
BLOCK	Block code--a 3-character numeric code assigned to census blocks.	char(3)	Any valid block code, as used by the USCB.
ZIP	Zip Code--a five-character numeric code assigned by the US Postal Service to various regions where it delivers mail.	char(5)	Any valid zip code, as assigned by the US Postal Service
EDUCATION1	Less than 9th grade	num(8)	Any proportion between 0 and 1.
EDUCATION2	9th - 12th grade	num(8)	Any proportion between 0 and 1.

Variable Name	Definition	Type(Len)	Values
EDUCATION3	high school graduate	num(8)	Any proportion between 0 and 1.
EDUCATION4	some college, no degree	num(8)	Any proportion between 0 and 1.
EDUCATION5	associate degree	num(8)	Any proportion between 0 and 1.
EDUCATION6	bachelor degree	num(8)	Any proportion between 0 and 1.
EDUCATION7	graduate or professional degree	num(8)	Any proportion between 0 and 1.
EDUCATION8	Doctorate degree	num(8)	Any proportion between 0 and 1.
MEDFAMINCOME	Median Family Income	num(8)	Any integer.
FAMINCOME1	less than \$10,000	num(8)	Any proportion between 0 and 1.
FAMINCOME2	\$10,000 - \$14,999	num(8)	Any proportion between 0 and 1.
FAMINCOME3	\$15,000-\$19,999	num(8)	Any proportion between 0 and 1.
FAMINCOME4	\$20,000-\$24,999	num(8)	Any proportion between 0 and 1.
FAMINCOME5	\$25,000-\$29,999	num(8)	Any proportion between 0 and 1.
FAMINCOME6	\$30,000-\$34,999	num(8)	Any proportion between 0 and 1.
FAMINCOME7	\$35,000-\$39,999	num(8)	Any proportion between 0 and 1.
FAMINCOME8	\$40,000-\$44,999	num(8)	Any proportion between 0 and 1.
FAMINCOME9	\$45,000-\$49,999	num(8)	Any proportion between 0 and 1.
FAMINCOME10	\$50,000-\$59,999	num(8)	Any proportion between 0 and 1.
FAMINCOME11	\$60,000-\$74,999	num(8)	Any proportion between 0 and 1.
FAMINCOME12	\$75,000-\$99,999	num(8)	Any proportion between 0 and 1.
FAMINCOME13	\$100,000-\$124,999	num(8)	Any proportion between 0 and 1.
FAMINCOME14	\$125,000-\$149,999	num(8)	Any proportion between 0 and 1.

Variable Name	Definition	Type(Len)	Values
FAMINCOME15	\$150,000-\$199,999	num(8)	Any proportion between 0 and 1.
FAMINCOME16	\$200,000+	num(8)	Any proportion between 0 and 1.
FAMPOVERTY	Proportion of family households in the geography with below-poverty level income.	num(8)	Any proportion between 0 and 1.
MEDHOUSINCOME	median household income	num(8)	Any integer.
HOUSINCOME1	< \$10,000	num(8)	Any proportion between 0 and 1.
HOUSINCOME2	\$10,000-\$14,999	num(8)	Any proportion between 0 and 1.
HOUSINCOME3	\$15,000-\$19,999	num(8)	Any proportion between 0 and 1.
HOUSINCOME4	\$20,000-\$24,999	num(8)	Any proportion between 0 and 1.
HOUSINCOME5	\$25,000-\$29,999	num(8)	Any proportion between 0 and 1.
HOUSINCOME6	\$30,000-\$34,999	num(8)	Any proportion between 0 and 1.
HOUSINCOME7	\$35,000-\$39,999	num(8)	Any proportion between 0 and 1.
HOUSINCOME8	\$40,000-\$44,999	num(8)	Any proportion between 0 and 1.
HOUSINCOME9	\$45,000-\$49,999	num(8)	Any proportion between 0 and 1.
HOUSINCOME10	\$50,000-\$59,999	num(8)	Any proportion between 0 and 1.
HOUSINCOME11	\$60,000-\$74,999	num(8)	Any proportion between 0 and 1.
HOUSINCOME12	\$75,000-\$99,999	num(8)	Any proportion between 0 and 1.
HOUSINCOME13	\$100,000-\$124,999	num(8)	Any proportion between 0 and 1.
HOUSINCOME14	\$125,000-\$149,999	num(8)	Any proportion between 0 and 1.
HOUSINCOME15	\$150,000-\$199,999	num(8)	Any proportion between 0 and 1.
HOUSINCOME16	\$200,000+	num(8)	Any proportion between 0 and 1.
HOUSPOVERTY	%Households with below-poverty level income	num(8)	Any proportion between 0 and 1.
POV_LT_50	<50% of poverty level	num(8)	Any proportion between 0 and 1.

Variable Name	Definition	Type(Len)	Values
POV_50_74	Between 50 and 74.99% of poverty level	num(8)	Any proportion between 0 and 1.
POV_75_99	Between 75 and 99.99% of poverty level	num(8)	Any proportion between 0 and 1.
POV_100_124	Between 100 and 124.99% of poverty level	num(8)	Any proportion between 0 and 1.
POV_125_149	Between 125 and 149.99% of poverty level	num(8)	Any proportion between 0 and 1.
POV_150_174	Between 150 and 174.99% of poverty level	num(8)	Any proportion between 0 and 1.
POV_175_184	Between 175 and 184.99% of poverty level	num(8)	Any proportion between 0 and 1.
POV_185_199	Between 185 and 199.99% of poverty level	num(8)	Any proportion between 0 and 1.
POV_GT_200	> 200% of poverty level	num(8)	Any proportion between 0 and 1.
RA_NHS_WH	%white, non-hispanic	num(8)	Any proportion between 0 and 1.
RA_NHS_BL	%black or african american, non-hispanic	num(8)	Any proportion between 0 and 1.
RA_NHS_AM	%american indian or alaska native, non-hispanic	num(8)	Any proportion between 0 and 1.
RA_NHS_AS	%asian, non-hispanic	num(8)	Any proportion between 0 and 1.
RA_NHS_HA	%native hawaiian or other pacific islander, non-hispanic	num(8)	Any proportion between 0 and 1.
RA_NHS_OT	%other, non-hispanic	num(8)	Any proportion between 0 and 1.
RA_NHS_ML	%two or more races, non-hispanic	num(8)	Any proportion between 0 and 1.
RA_HIS_WH	%white, hispanic	num(8)	Any proportion between 0 and 1.
RA_HIS_BL	%black or african american, hispanic	num(8)	Any proportion between 0 and 1.
RA_HIS_AM	%americian indian ans alaska native, hispanic	num(8)	Any proportion between 0 and 1.
RA_HIS_AS	%asian, hispanic	num(8)	Any proportion between 0 and 1.
RA_HIS_HA	%native hawaiian or other pacific islander, hispanic	num(8)	Any proportion between 0 and 1.
RA_HIS_OT	%other, hispanic	num(8)	Any proportion between 0 and 1.
RA_HIS_ML	%two or more races, hispanic	num(8)	Any proportion between 0 and 1.

Variable Name	Definition	Type(Len)	Values
HOUSES_N	Number of housing units in geography	num(8)	Any integer.
HOUSES_OCCUPIED	Proportion of houses_n that are occupied.	num(8)	Any proportion between 0 and 1.
HOUSES_OWN	Proportion of occupied housing units that are occupied by owners.	num(8)	Any proportion between 0 and 1.
HOUSES_RENT	Proportion of occupied housing units that are occupied by renters.	num(8)	Any proportion between 0 and 1.
HOUSES_UNOCC_FORRENT	Proportion of unoccupied housing units that are for rent.	num(8)	Any proportion between 0 and 1.
HOUSES_UNOCC_FORSALE	Proportion of unoccupied housing units that are for sale	num(8)	Any proportion between 0 and 1.
HOUSES_UNOCC_RENTSOLD	Proportion of unoccupied housing units that are rented or sold, but still unoccupied.	num(8)	Any proportion between 0 and 1.
HOUSES_UNOCC_SEASONAL	Proportion of unoccupied housing units that are used seasonally or other occasional use.	num(8)	Any proportion between 0 and 1.
HOUSES_UNOCC_MIGRANT	Proportion of unoccupied housing units that are used for migrant workers.	num(8)	Any proportion between 0 and 1.
HOUSES_UNOCC_OTHER	Proportion of unoccupied housing units that are vacant for other reasons.	num(8)	Any proportion between 0 and 1.
ENGLISH_SPEAKER	Proportion of people over age 5 that speak only English or speak english "very well"	Num(8)	Any proportion between 0 and 1.
SPANISH_SPEAKER	Proportion of people over age 5 who speak only spanish or speak spanish "very well"	Num(8)	Any proportion between 0 and 1.
BORNINUS	Proportion of people over age 5 that were born in the US	Num(8)	Any proportion between 0 and 1.
MOVEDINLAST12MON	Proportion of households that moved in the last 12 months	Num(8)	Any proportion between 0 and 1.
MARRIED	Proportion of people over age 15 who are married	Num(8)	Any proportion between 0 and 1.
DIVORCED	Proportion of people over age 15 who are divorced	Num(8)	Any proportion between 0 and 1.
DISABILITY	Proportion of people over age 18 living with any disability	Num(8)	Any proportion between 0 and 1.

Variable Name	Definition	Type(Len)	Values
UNEMPLOYMENT	Proportion of civilian noninstitutionalized population between 18 and 64 who are unemployed	Num(8)	Any proportion between 0 and 1.
UNEMPLOYMENT_MALE	Proportion of civilian noninstitutionalized males between 18 and 64 who are unemployed	Num(8)	Any proportion between 0 and 1.
INS_MEDICARE	Proportion of people covered by Medicare	Num(8)	Any proportion between 0 and 1.
INS_MEDICAID	Proportion of people covered by Medicaid	Num(8)	Any proportion between 0 and 1.
HH_NOCAR	Proportion of households with no car (owner and renter occupied)	Num(8)	Any proportion between 0 and 1.
HH_PUBLIC_ASSISTANCE	Proportion of households on public assistance	Char(1)	Any proportion between 0 and 1.
HMOOWNER_COSTS_MORT	Proportion of households with monthly owner costs > 50% HH income, in homes with mortgages	Num(8)	Any proportion between 0 and 1.
HMOOWNER_COSTS_NO_MORT	Proportion of households with monthly owner costs > 50% HH income, in homes without mortgages	Num(8)	Any proportion between 0 and 1.
HOMES_MEDVALUE	Median value of homes	Num(8)	Any integer.
PCT_CROWDING	Proportion of households with >= 1 person per room	Num(8)	Any proportion between 0 and 1.
FEMALE_HEAD_OF_HH	Proportion of households headed by females (no male present)	Num(8)	Any proportion between 0 and 1.
MGR_FEMALE	Proportion of female management occupations		Any proportion between 0 and 1.
MGR_MALE	Proportion of male management occupations	Num(8)	Any proportion between 0 and 1.
RESIDENTS_65	Proportion of population over 65	Num(8)	Any proportion between 0 and 1.
SAME_RESIDENCE	Proportion of persons in same residence since year 2005	Num(8)	What years does ACS provide for this?

**Primary Key:**

GEOCODE + CENSUS\_YEAR

**Foreign Key Relationship:**

Source Variable (Foreign Key)	Target Table	Target Variable (Primary Key)	Orphans allowed?
[None Defined]			

## Appendix A - DEPARTMENT

A list of permissible values for the DEPARTMENT variable on the ENCOUNTERS table

DEPARTMENT Value	Description
ACUP	Acupuncture
ALGY	Allergy
AMBU	Ambulance Services
ANES	Anesthesiology
AUD	Audiology
CARD	Cardiology
CASR	Cast Room
CHEM	Chemical and Alcohol Dependency
CHIR	Chiropractic
CMHL	Community Health
CRIT	Critical Care Medicine
CRMG	Care Management
DENT	Dental
DERM	Dermatology
DIAL	Dialysis
DME	Durable Medical Equipment
EDUC	Education
ENDO	Endocrinology
ENT	Otolaryngology
ER	Emergency Room
FP	Family Practice
GEN	Genetics
GER	Gerontology/Geriatrics
GI	Gastro-Intestinal Medicine
HAP	Health Appraisals

DEPARTMENT Value	Description
HEP	Hepatology
HH	Home Health
HOSP	Hospital Care
HSPC	Hospice
ICF	Intermediate Care Facility
IM	Internal Medicine
IMUN	Immunology
IND	Industrial Medicine
INF	Infectious Disease
INFU	Infusion Center
IR	Injection Room
LAB	Laboratory
MH	Mental Health
NATU	Naturopathy
NEPH	Nephrology
NEUR	Neurology
NEWB	Newborn
NRSG	Neurosurgery
NUCL	Nuclear Medicine
NUT	Nutrition
OBGN	Obstetrics/Gynecology
OCTH	Occupational Therapy
ONC	Oncology
OPTH	Ophthalmology
OPTO	Optometry
ORTH	Orthopedics
OST	Osteopathy
PAL	Palliative Care

DEPARTMENT Value	Description
PATH	Pathology
PC	Primary Care
PEDS	Pediatrics
PERI	Perinatology
PHYS	Physiatry
POD	Podiatry
PSRG	Plastic Surgery
PT	Physical Therapy
PULM	Pulmonary Medicine
RAD	Radiology
RADT	Radiation Therapy
RECT	Recreational Therapy
REHB	Rehabilitation
RESP	Respiratory Therapy
RHEU	Rheumatology
RN	Registered Nurse
SNF	Skilled Nursing Facility
SPOR	Sports Medicine
SPTH	Speech Therapy
SURG	General Surgery
TRAN	Transplant
URG	Urgent Care
URO	Urology
OTH	Other
UNK	Unknown

Appendix C - PROVIDER\_TYPE and SPECIALTY Value Sets

PROVIDER_TYPE	Description
001	ACUPUNCTURIST
002	ANESTHESIOLOGY ASST
003	AUDIOLOGIST
004	AUDIOLOGY ASSISTANT
005	CARDIAC REHAB THERAPIST
006	CASE MANAGER
007	CERTIFIED NURSE SPECIALIST
008	CERTIFIED REG NURSE ANESTHETIST
009	CHAPLAIN
010	CHEM DEP COUNSELOR, ADULT
011	CHEM DEP COUNSELOR, CHILD/ADOL
012	CHIROPRACTOR
013	CIGNA
014	CLINICAL NURSE SPECIALIST
015	COUMADIN NURSE
016	CPAP SPECIALIST
017	DENTAL ASSISTANT
018	DENTIST
019	DIABETIC NURSE
020	DIETETIC TECH
021	DIETITIAN
022	EEG TECHNICIAN
023	ER TECH
024	EXERCISE PHYSIOLOGIST
025	FELLOW
026	HEALING TOUCH
027	HEALTH EDUCATOR
028	HOME HEALTH AIDE
029	HP PHYSICIAN
030	HYGIENIST
031	INDEPENDENT LAB/X-RAY

SPECIALTY	Description
ADO	Adolescent Medicine
AER	Aerospace Medicine
ALC	Chemical Dependency
ALL	Allergy
ANC	Ancillary Services
ANE	Anesthesiology
ATH	Sports Medicine
AUD	Audiology
BON	Bone And Mineral
CAR	Cardiology
CAV	Cardiovascular Surgery
CHR	Chiropractor
CLC	Clin Cardiac Electrophysiology
COL	Colon & Rectal Surgery
COM	Complimentary & Alternative Medicine
CON	Continuing Care
CRI	Critical Care
DEN	Dentistry
DER	Dermatology
DOR	Dor
EDU	Medical Education
EME	Emergency Medicine
EMI	Emi
END	Endocrinology
ENT	Otolaryngology
FAM	Family Medicine
FLX	Flexible
GAS	Gastroenterology
GEN	Genetics
GER	Gerontology
HAN	Hand Surgery

PROVIDER_TYPE	Description
032	INF CNTRL PRACT
033	INFUSION THERAPIST
034	INTERPRETER
035	LAB ASSISTANT
036	LAB TECHNICIAN
037	LACTATION CONSULTANT
038	LIC FAM THERAPIST
039	LIC SOCIAL WORKER
040	MASSAGE THERAPIST
041	MEDICAL ASSISTANT
042	MIDLEVEL
043	MIDWIFE
044	MUSIC THERAPIST
045	NOT APPLICABLE
046	NURSE
047	NURSE CLINICIAN
048	NURSE PRACTITIONER
049	OCCUPATIONAL THERAPIST
050	OPHTHALMIC ASSISTANT
051	OPHTHALMIC TECHNICIAN
052	OPTICIAN
053	OPTOMETRIST
054	ORTHO PHY ASST
055	ORTHOPAEDIC
056	OSTEOPATH
057	PERSONAL CARE ATTENDANT
058	PHARM TECH
059	PHARMACIST
060	PHYSICAL THERAPIST
061	PHYSICIAN
062	PHYSICIAN ASSISTANT
063	PODIATRIST
064	PSYCH TECHNICIAN

SPECIALTY	Description
HOM	Home Health
HOS	Hospital
HYM	Hyperbaric Medicine
HYP	Hypertension
IMG	General Internal Medicine
INF	Infectious Disease
LAB	Laboratory
MEN	Mental Health
MGM	Care Management
MID	Midlevel
MUL	Multispecialty
NEH	Nephrology
NEO	Neonatology
NES	Neurosurgery
NEU	Neurology
NOB	No Boards
NUM	Nuclear Medicine
NUR	Nurse
NUT	Nutrition
OBO	Gynecologic Oncology
OBS	Obstetrics - Gynecology
OCM	Occupational Health
ONC	Oncology
ONS	Surgical Oncology
OPH	Ophthalmology
OPL	Optical
OPT	Optometry
ORA	Oral Surgery
ORD	Orthodontia
ORT	Orthopedics
OTO	Otolaryngology
PAI	Pain Management
PAT	Pathology

PROVIDER_TYPE	Description
065	PSYCHOLOGIST
066	PSYCHOTHERAPIST
067	PSYCHOTHERAPIST, ADULT
068	PSYCHOTHERAPIST, CHILD/ADOL
069	PT ASSISTANT
070	RADIATION THERAPIST
071	RADIOLOGY TECHNICIAN
072	RECREATIONAL THERAPIST
073	REGISTERED NURSE
074	REHAB ASSISTANT
075	REHAB THERAPIST
076	REIKI
077	RESIDENT
078	RESOURCE
079	RESPIRATORY THERAPIST
080	SOCIAL WORKER
081	SPEECH PATHOLOGIST CERTIFIED
082	STUDENT
083	SURG TECH
084	SURGEON ASSISTANT
085	THERAPEUTIC AQUATIC SPECIALIST
086	ULTRASOUND TECHNICIAN
888	OTHER
999	UNKNOWN

SPECIALTY	Description
PED	Pediatrics
PES	Pediatric Subspecialty
PEY	Perinatology
PHA	Pharmacy
PHT	Physical Therapy
PHY	Physiatry
PLA	Plastic Surgery
POD	Podiatry
PRE	Preventive Medicine
PRO	Prosthodontia
PSY	Psychiatry
PUB	Public Health
PUL	Pulmonary Medicine
RAD	Radiology
REH	Rehabilitation Medicine
RES	Respiratory Therapy
RHE	Rheumatology
ROP	Radiation Oncology
SLC	Sleep Center
SOC	Social Services
SPP	Speech Patholgy
SUR	Surgery
TEE	Teen Clinic
TOX	Medical Toxicology
TRN	Transportation/Non-Emergency
TRS	Transplant Surgery
UNK	Unknown
URG	Urgent Care
URO	Urology
VAS	Vascular Surgery