

**icd9** — ICD-9-CM diagnosis and procedure codes

<a href="#">Description</a>	<a href="#">Quick start</a>	<a href="#">Menu</a>	<a href="#">Syntax</a>
<a href="#">Options</a>	<a href="#">Remarks and examples</a>	<a href="#">Stored results</a>	<a href="#">References</a>
<a href="#">Also see</a>			

## Description

The `icd9` and `icd9p` commands are a suite of commands for working with ICD-9-CM diagnosis and ICD-9-CM procedure codes from the 16th version (effective October 1998) and later. To see the current version of the ICD-9-CM diagnosis codes and any changes that have been applied, type `icd9 query`.

`icd9 check`, `icd9 clean`, and `icd9 generate` are data management commands. `icd9 check` verifies that a variable contains defined ICD-9-CM diagnosis codes and provides a summary of any problems encountered. `icd9 clean` standardizes the format of the codes. `icd9 generate` can create a binary indicator variable for whether the code is in a specified set of codes, a variable containing a corresponding higher-level code, or a variable containing the description of the code.

`icd9 lookup` and `icd9 search` are interactive utilities. `icd9 lookup` displays descriptions of the diagnosis codes specified on the command line. `icd9 search` looks for relevant ICD-9-CM diagnosis codes from key words given on the command line.

`icd9p` may be used in place of `icd9` for any command above to obtain results for procedure codes.

## Quick start

Determine if ICD-9-CM diagnosis codes in `diag1` are invalid and store reasons in `invalid`

```
icd9 check diag1, generate(invalid)
```

Standardize display of codes in `diag2` to remove all periods and align codes by padding with spaces

```
icd9 clean diag2, pad
```

Create `descr3` as the diagnosis code prepended to short description of diagnosis code in `diag3`

```
icd9 generate descr3 = diag3, description long
```

Create `diabetes` as an indicator for a diabetes diagnosis in `diag4` using ICD-9-CM codes 250.xx

```
icd9 generate diabetes = diag4, range(25000/25093)
```

Look up descriptions for ICD-9-CM diagnosis codes E827.0 to E828.9

```
icd9 lookup E8270/E8289
```

Note: For any command above, `icd9p` may be used in place of `icd9` to obtain results for procedure codes.

## Menu

Data > ICD codes > ICD-9

## Syntax

Verify that variable contains defined codes

```
{icd9|icd9p} check varname [if] [in] [, any list generate(newvar) ]
```

Clean variable and verify format of codes

```
{icd9|icd9p} clean varname [if] [in] [, dots pad]
```

Generate new variable from existing variable

```
{icd9|icd9p} generate newvar = varname [if] [in] , category
```

```
{icd9|icd9p} generate newvar = varname [if] [in] , description [long end]
```

```
{icd9|icd9p} generate newvar = varname [if] [in] , range(codelist)
```

Display code descriptions

```
{icd9|icd9p} lookup codelist
```

Search for codes from descriptions

```
{icd9|icd9p} search ["text"] [{"text"} ...] [, or]
```

Display ICD-9 code source

```
{icd9|icd9p} query
```

*codelist* is

<i>icd9code</i>	(the particular code)
<i>icd9code*</i>	(all codes starting with)
<i>icd9code/icd9code</i>	(the code range)

or any combination of the above, such as 001\* 018/019 E\* 018.02. *icd9codes* must be typed with leading 0s. For example, type 001 (diagnosis code) or 01 (procedure code); typing 1 will result in an error.

## Options

Options are presented under the following headings:

*Options for icd9[p] check*  
*Options for icd9[p] clean*  
*Options for icd9[p] generate*  
*Option for icd9[p] search*

**Warning:** The option descriptions are brief and use jargon. Please read the *Remarks and examples* before using the `icd9` or `icd9p` command.

### Options for `icd9[p] check`

`any` tells `icd9[p] check` to verify that the codes fit the format of ICD-9-CM codes but not to check whether the codes are defined.

`list` specifies that `icd9 check` list the observation number, the invalid or undefined ICD-9-CM code, and the reason the code is invalid or if it is undefined code. When specified with `icd9p check`, the `list` option references ICD-9-CM procedure codes.

`generate(newvar)` specifies that `icd9 check` create a new variable containing, for each observation, 0 if the observation contains a defined code or is missing. Otherwise, it contains a number from 1 to 10. The positive numbers indicate the kind of problem and correspond to the listing produced by `icd9 check`. The values are labeled with the Stata-defined value label `__icd_9`.

### Options for `icd9[p] clean`

`dots` specifies that the period be included in the final format. If `dots` is not specified, then all periods are removed.

`pad` specifies that `icd9[p] clean` pad the codes with spaces, front and back, to make the (implied) dots align vertically in listings. Specifying `pad` makes the resulting codes look better when used with most other Stata commands.

### Options for `icd9[p] generate`

`category`, `description`, and `range(codelist)` specify the contents of the new variable that `icd9[p] generate` is to create. You do not need to `icd9[p] clean varname` before using `icd9[p] generate`; it will accept any ICD-9-CM format or combination of formats.

`category` generates a new variable that also contains ICD-9-CM codes. The resulting variable may be used with the other `icd9[p]` subcommands. For procedure codes, the category code is the first two characters. For diagnostic codes, the category code is the first three characters, except for E-codes when it is the first four characters.

`description` creates *newvar* containing descriptions of the ICD-9-CM codes.

`long` is for use with `description`. It specifies that the code be prepended to the text describing the code.

`end` modifies `long` (specifying `end` implies `long`) and places the code at the end of the string.

`range(codelist)` creates a new indicator variable equal to 1 when the ICD-9-CM code is in the range specified and equal to 0 otherwise.

## Option for `icd9[p]` search

or specifies that ICD-9-CM codes be searched for descriptions that contain any word specified with `icd9[p]` search. The default is to list only descriptions that contain all the words specified.

## Remarks and examples

[stata.com](http://www.stata.com)

Remarks are presented under the following headings:

[Using `icd9` and `icd9p`](#)  
[Verifying and cleaning variables](#)  
[Interactive utilities](#)  
[Creating new variables](#)

If you have not yet read [Introduction to ICD coding](#) in [\[D\] `icd`](#) before using the `icd9` and `icd9p` command.

## Using `icd9` and `icd9p`

The ICD-9-CM coding system includes diagnosis and procedure codes. Some examples of diagnosis codes are 552.3 (Diaphragmatic hernia with obstruction) and E871.0 (Foreign object left in body during surgical operation). Some example of procedure codes are 01.2 (Craniotomy and craniectomy) and 55.23 (Closed renal biopsy).

Many datasets record (and some people write) codes without the period, for example, diagnosis code 550.1 may appear as 5501. The `icd9` and `icd9p` commands understand both ways of recording codes. The commands are also insensitive to codes recorded with or without leading and trailing blanks. For E-codes and V-codes, the `icd9` commands are case-insensitive. All the following codes are acceptable formats.

diagnosis	procedure
001	27.62
001.	72
00581	32.6
552.3	97.11
E800.2	872
e8002	5523
v82.2	08.51

**Important note:** What constitutes a valid code changes between versions. For the rest of this entry, a defined code is any code that is currently valid, was valid at some point since version 16 (V16, effective 1 October 1998), or has meaning as a grouping of codes. The list of valid codes and their associated descriptions is from the U.S. Centers for Medicare and Medicaid Services (CMS). These codes are jointly maintained and distributed by the U.S. Centers for Disease Control and Prevention's National Center for Health Statistics and by CMS ([Centers for Disease Control and Prevention 2013](#)).

In `icd9` and `icd9p`, descriptions that end with an asterisk (\*) are used to denote codes that are invalid for medical coding purposes but are defined as a category code or a subcategory code that has been further subdivided. For example, diagnosis code 001 (Cholera) is invalid without a fourth digit but is defined as a category code, so its description appears as `cholera*`. CMS does not distribute short descriptions of category and subcategory codes that are defined but not valid for coding. To ensure that Stata reports that these codes are defined, we added them to the datasets that `icd9` and `icd9p` use with a description of `*`.

Codes that were valid in the past, but no longer are, have descriptions that end with a hash mark (#). For example, the diagnosis code 645.01 was deleted between V16 and V18. It remains a defined code, and its description appears as `prolonged preg-delivered#`.

To view the current version of ICD-9-CM diagnosis codes in Stata, its source, and a log of changes that have been made to the list of ICD-9-CM codes since the `icd9` commands were implemented, type

```
. icd9 query
ICD9 Diagnostic Code Mapping Data for use with Stata, History
(output omitted)
V32
Dataset obtained 26aug2014 from
<http://www.cms.gov/Medicare/Coding/ICD9ProviderDiagnosticCodes/
> codes.html>, by selecting the 'Version 32...' file. Can be gotten
directly via
<http://www.cms.gov/Medicare/Coding/ICD9ProviderDiagnosticCodes/
> Downloads/ICD-9-CM-v32-master-descriptions.zip>. After unzipping, the
useful file name is "CMS32_DESC_SHORT_DX.txt (there are other files we
did not use)."
09oct2014: V32 put into Stata distribution
BETWEEN V31 and V32: There were no additional codes.
BETWEEN V31 and V32: 0 codes were deleted.
BETWEEN V31 and V32: There were no description changes.
(output omitted)
```

`icd9p` query provides information about changes to procedure codes.

Throughout the remainder of this entry, we use `nhds2010.dta`, an extract of adult same-day discharges from the 2010 National Hospital Discharge Survey (NHDS). Below, we [describe](#) the data and [list](#) the first five observations for the diagnosis and procedure code variables.

```
. use http://www.stata-press.com/data/r14/nhds2010
(Adult same-day discharges, 2010 NHDS)
. describe
Contains data from http://www.stata-press.com/data/r14/nhds2010.dta
  obs:      2,210      Adult same-day discharges, 2010
  vars:      15        30 Jan 2015 15:03
  size:     83,980     (_dta has notes)
```

variable name	storage type	display format	value label	variable label
ageu	byte	%8.0g	ageu	Units for age
age	byte	%8.0g		Age
sex	byte	%8.0g	sex	Sex
race	byte	%8.0g	race	Race
month	byte	%8.0g		Discharge month
status	byte	%8.0g	status	Discharge status
region	byte	%8.0g	region	Region
atype	byte	%8.0g	atype	Type of admission
dx1	str5	%9s		Diagnosis 1
dx2	str5	%9s		Diagnosis 2
dx3	str5	%9s		Diagnosis 3 (imported incorrectly)
dx3corr	str5	%9s		Diagnosis 3 (corrected)
pr1	str4	%9s		Procedure 1
wgt	int	%12.0g		Frequency weight
recid	float	%9.0g		Order of record (raw data)

Sorted by: recid

```
. list recid dx1 dx2 dx3 pr1 in 1/5
```

	recid	dx1	dx2	dx3	pr1
1.	84	4414	99811	14275	3834
2.	105	25013	3572	-2506	
3.	255	51909	1489	-V146	
4.	651	9678	E8528	8	
5.	696	V271	64421	16564	7359

## Verifying and cleaning variables

`icd9[p]` `check` verifies that *varname* contains defined ICD-9-CM codes and, if not, provides a full report on the problems. It is a good idea to begin with this command and fix any potential problems before proceeding to other `icd9[p]` commands. However, the `check` subcommand is also useful for tracking down problems when any of the other `icd9[p]` commands tell you that the “variable does not contain ICD-9 codes”.

`icd9[p]` `clean` modifies the variable to ensure consistency and to make subsequent output look better. This is not strictly necessary because all `icd9[p]` commands work equally well with cleaned or uncleaned codes. `icd9[p]` `clean` also can be used to verify that the codes in a variable conform with the ICD-9-CM format, without checking to see if the codes are defined.

### ► Example 1: Checking the validity of a variable

We noticed when we listed our data that `dx3` appears to be padded with dashes instead of spaces. As a preemptive step, we replace the dashes with spaces by using the `substr()` function because the `icd9[p]` commands ignore spaces.

```
. replace dx3=substr(dx3,"-", " ",.)
(1,009 real changes made)
. list recid dx1 dx2 dx3 pr1 in 1/5
```

	recid	dx1	dx2	dx3	pr1
1.	84	4414	99811	14275	3834
2.	105	25013	3572	2506	
3.	255	51909	1489	V146	
4.	651	9678	E8528	8	
5.	696	V271	64421	16564	7359

Now that we have replaced the characters we know will be a problem, we can `icd9 check` the diagnosis variables. We add the `generate()` option so that we can identify any observations with invalid codes.

```

. icd9 check dx1, generate(prob1)
(dx1 contains defined ICD-9-CM codes; no missing values)
. icd9 check dx2, generate(prob2)
(dx2 contains defined ICD-9-CM codes; 179 missing values)
. icd9 check dx3, generate(prob3)
(dx3 contains 277 missing values)
dx3 contains invalid codes:
  1. Invalid placement of period           0
  2. Too many periods                     0
  3. Code too short                       79
  4. Code too long                        0
  5. Invalid 1st char (not 0-9, E, or V)  0
  6. Invalid 2nd char (not 0-9)          128
  7. Invalid 3rd char (not 0-9)         0
  8. Invalid 4th char (not 0-9)         0
  9. Invalid 5th char (not 0-9)         0
 10. Code not defined                    793
-----
Total                                   1,000

```

We see that all codes in dx1 are valid and all discharges have a primary diagnosis recorded. Likewise, all codes in dx2 are defined, and we see that 179 observations did not have a second diagnosis.

However, `icd9 check` reports that 1,000 of the 2,210 observations on dx3 have some sort of problem: 79 codes are too short, 128 have an invalid second character, and 793 are undefined. After some investigation, we discover that when we imported the data, we started reading from the wrong position in the file. Hereafter, we use the correctly imported variable, `dx3corr`.

```

. icd9 check dx3corr
(dx3corr contains defined ICD-9-CM codes; 356 missing values)

```

◀

Rather than typing the `icd9 check` command once for each variable, we could have checked all three simultaneously. See [Working with multiple codes](#) in [D] `icd`.

## ▶ Example 2: Standardizing the format of codes

If we plan to do any reporting with these codes later, we may want to make them more readable. Suppose we want to report the primary diagnosis and procedure for each discharge. We can use `icd9 clean` with the `dots` and `pad` options to add the period between the category code and any subsequent digits and to align the periods.

```

. icd9 clean dx1, dots pad
(2210 changes made)

```

To standardize the procedure codes, we simply use the `icd9p` analogue to `icd9`.

```
. icd9p clean pr1, dots pad
(821 changes made)
. list recid dx1 pr1 in 1/5
```

	recid	dx1	pr1
1.	84	441.4	38.34
2.	105	250.13	
3.	255	519.09	
4.	651	967.8	
5.	696	V27.1	73.59

Using `icd9[p] clean` with undefined codes will not result in an error message. So if you are using codes from a country other than the United States, the `clean` subcommand can still be used to standardize the format of your codes and check for correct placement of the period.

◀

## Interactive utilities

`icd9[p] search` looks for relevant ICD-9-CM codes from the description given on the command line, and `icd9[p] lookup` lists the descriptions of codes given on the command line. The two commands complement each other.

### ▶ Example 3: Finding diagnosis codes

Suppose that we want to identify the observations for which the primary diagnosis is congestive heart failure (CHF). As part of a quick exploratory analysis, we can use `icd9 search` to find ICD-9-CM codes that we may want to use to define our study population. We use the terms “heart failure” and “chf”. We enclose “heart failure” in quotation marks and use option `or` so that `icd9 search` looks for either term.

```
. icd9 search "heart failure" chf, or
5 matches found:
 398.91  rheumatic heart failure
 428     heart failure*
 428.0   chf nos
 428.1   left heart failure
 428.9   heart failure nos
```

Because the descriptions are abbreviated, we are concerned that some of the 428 codes may be left out. So, we use `icd9 lookup` to look up a range of codes.



```
. icd9 lookup 428*
19 matches found:
  428      heart failure*
  428.0    chf nos
  428.1    left heart failure
  428.2    *
  428.20   systolic hrt failure nos
  428.21   ac systolic hrt failure
  428.22   chr systolic hrt failure
  428.23   ac on chr syst hrt fail
  428.3    *
  428.30   diastolic hrt failure nos
  428.31   ac diastolic hrt failure
  428.32   chr diastolic hrt fail
  428.33   ac on chr diast hrt fail
  428.4    *
  428.40   syst/diast hrt fail nos
  428.41   ac syst/diastol hrt fail
  428.42   chr syst/diastl hrt fail
  428.43   ac/chr syst/dia hrt fail
  428.9    heart failure nos
```

The same result could be found by typing

```
. icd9 lookup 428/4289
```

if we knew that 428.9 was the last code in the 428 category.

◀

## Creating new variables

`icd9[p] generate` produces new variables based on existing variables containing (cleaned or uncleaned) ICD-9-CM codes. `icd9[p] generate, category` creates *newvar* containing the category code that corresponds to the code in the existing variable. `icd9[p] generate, description` creates *newvar* containing the abbreviated textual description of the ICD-9-CM code. `icd9[p] generate, range()` produces numeric *newvar* containing 1 if *varname* records an ICD-9-CM code in the range listed and containing 0 otherwise.

### ► Example 4: Creating an indicator variable

We review the list of codes we found in [example 3](#) and decide that we will use 398.91 and all of the 428 codes in our definition of a CHF diagnosis. Now we can use `icd9 generate` with the `range()` option to create an indicator variable.

```
. icd9 generate chf = dx1, range(398.91 428*)
. tabulate chf [fweight=wgt]
```

chf	Freq.	Percent	Cum.
0	563,048	97.88	97.88
1	12,192	2.12	100.00
Total	575,240	100.00	

After [tabulating](#) the results, we see that about 2.1% of all same-day discharges were for CHF in 2010.

◀

### ► Example 5: Adding descriptions to codes

Continuing [example 4](#), we may wish to know what procedures were performed for patients with CHF. We check the procedure codes in `pr1` and then generate a new variable with their descriptions. We include option `long` so that we can see the ICD-9-CM procedure code as well.

```
. icd9p check pr1
(pr1 contains defined ICD-9-CM procedure codes; 1389 missing values)
. icd9p generate pr1descr = pr1, description long
. tabulate pr1descr [fweight=wgt] if chf==1, missing sort
```

label for pr1	Freq.	Percent	Cum.
	7,185	58.93	58.93
37.22 left heart cardiac cath	1,906	15.63	74.57
92.05 c-vasc scan/isotop funct	1,027	8.42	82.99
88.72 dx ultrasound-heart	776	6.36	89.35
03.31 spinal tap	498	4.08	93.44
39.95 hemodialysis	388	3.18	96.62
34.91 thoracentesis	138	1.13	97.75
99.60 cardiopulm resuscita nos	112	0.92	98.67
37.94 implt/repl carddefib tot	110	0.90	99.57
89.44 cardiac stress test nec	52	0.43	100.00
Total	12,192	100.00	

We see that the majority of same-day discharges (58.9%) did not involve any procedure. When a procedure was performed, the most common was left heart cardiac catheterization (15.6%).

◀

### □ Technical note

The dataset that supports `icd9` and `icd9p` includes all codes that were added or deleted between V16 and the current version (V32). However, the descriptions are updated with each new version. If you are using `icd9[p]` `generate` with option `description` for codes other than the current version, please review the `icd9[p]` query log for any changes to descriptions between the version you are using and the current version.

□

### ► Example 6: Combining commands for reporting

The `icd9[p]` `generate` commands are useful in isolation, but their real power comes when they are combined. For example, suppose that we want to make a graph showing the number of discharges in each diagnosis category for ICD-9-CM chapter 4, “Diseases of Blood and Blood Forming Organs”. We could use several `generate` commands and string functions, but `icd9 generate` greatly reduces our work.

First, we extract the category code from the detailed diagnosis code. Then, because the `icd9[p]` commands work equally well with complete codes or category codes, we can use `icd9 generate` with the `range(280/289)` option to create an indicator variable for whether the discharge had a primary diagnosis in chapter 4.

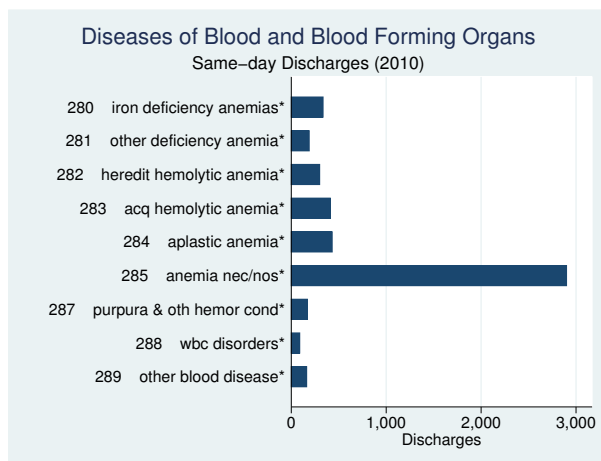
```
. icd9 generate dx1cat = dx1, main
. icd9 generate ch4 = dx1cat, range(280/289)
```

Next, we create a variable with the descriptions of the category codes in chapter 4.

```
. icd9 generate ch4des = dx1cat if ch4==1, description long
```

Finally, we use `graph hbar` to make a horizontal bar graph showing the frequencies of same-day discharges by diagnosis category.

```
. graph hbar (count) [fweight=wgt], over(ch4des) ytitle(Discharges)
> title(Diseases of Blood and Blood Forming Organs, span)
> subtitle(Same-day Discharges (2010), span)
```



See [G-2] **graph bar** for information about customizing the graph above. For more information about graphing results, see [G-2] **graph**.

◀

## Stored results

`icd9[p]` `check` stores the following in `r()`:

Scalars

<code>r(e#)</code>	number of errors of type #
<code>r(esum)</code>	total number of errors

`icd9[p]` `clean` stores the following in `r()`:

Scalars

<code>r(N)</code>	number of changes
-------------------	-------------------

`icd9[p]` `lookup` stores the following in `r()`:

Scalars

<code>r(N)</code>	number of codes found
-------------------	-----------------------

## References

Centers for Disease Control and Prevention. 2013. International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). <http://www.cdc.gov/nchs/icd/icd9cm.htm>.

Gould, W. W. 2000. `dm76: ICD-9 diagnostic and procedure codes`. *Stata Technical Bulletin* 54: 8–16. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 77–87. College Station, TX: Stata Press.

National Center for Health Statistics. 2011. International Classification of Diseases, Ninth Revision, Clinical Modification. [ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Publications/ICD9-CM/2011/](ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Publications/ICD9-CM/2011/).

—. 2012. National Hospital Discharge Survey: 2010 Public Use Data File Documentation.  
[ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Dataset\\_Documentation/NHDS/NHDS\\_2010\\_Documentation.pdf](ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHDS/NHDS_2010_Documentation.pdf).

## Also see

[D] **icd** — Introduction to ICD commands