Analytic Guidelines:
Creating Disability Identifiers Using the Washington Group Short Set on Functioning - Enhanced (WG-SS Enhanced) Stata Syntax

Introduction

The Stata syntax for the WG Short Set on Functioning - Enhanced (WG-SS Enhanced) is extracted from the syntax developed for the WG Extended Set on Functioning (WG-ES).

Disability identification for the WG-SS Enhanced is based on a level of inclusion that is at least one domain/question is coded A LOT OF DIFFICULTY or CANNOT DO AT ALL – or – for the domains Anxiety or Depression, the highest level of difficulty on a four-point scale.

The WG-SS Enhanced comprises the WG-SS PLUS Upper body functioning, Anxiety and Depression: a total of 12 questions over 8 domains of functioning.

NOTE: For data analysis, use your standard weighting and estimation techniques.

The Stata syntax is based on the *variable labels* indicated in the table below. The complete WG-SS Enhanced module includes more questions than appear in this table. Disability status is determined through difficulty in the basic, universal activities *without* the use of assistive technology or other assistance. Questions on the use of medication for symptoms of anxiety or depression are not included among the analytic variables considered for the syntax.

Only those questions/variables below are used in the determination of disability identifiers.
Ensure that you use the same *variable labels* OR revise the Stata syntax to reflect the *variable labels* in your database.

The WG-SS is administered as part of the U.S. National Health Interview Survey (NHIS). The data used to prepare these guidelines come from the 2013 NHIS.

> The **Washington Group Implementation Documents** cover the tools developed by the Washington Group on Disability Statistics (WG) to collect internationally comparable disability data on censuses and surveys. The documents address best practices in implementing the Short Set, Extended Set, Short Set – Enhanced, the WG / UNICEF Child Functioning Modules for children 2-4 and 5-17 years of age, and the WG / ILO LFS Disability Module, as well as other WG tools. Topics include translation, question specifications, analytic guidelines, programming code for analyses, the use of the tools for the purposes of disaggregation, and more.
>
> To locate other WG Implementation Documents and more information, visit the Washington Group website: http://www.washingtongroup-disability.com/.

Note to users of the NHIS: the variable names in the NHIS data file and documentation may differ from those used in this document; e.g., the self-care domain variable referenced as SC-SS in this document is referred to as UB_SS in the NHIS data file and documentation.

WG Extended Set Questions	Variable Label	Response Pattern
VISION		
1. **Do you have difficulty seeing even if wearing glasses?**	VIS_SS	1
COMMUNICATION		
2. **Using your usual language, do you have difficulty communicating (for example understanding or being understood by others)?**	COM_SS	1
HEARING		
3. **Do you have difficulty hearing even if using a hearing aid?**	HEAR_SS	1
COGNITION		
4. **Do you have difficulty remembering or concentrating?**	COG_SS	1
SELF-CARE		
5. **Do you have difficulty with (self-care such as) washing all over or dressing?**	SC_SS	1
UPPER BODY		
6. **Difficulty raising 2 liter bottle of water from waist to eye level?**	UB_1	1
7. **Degree of difficulty using hands and fingers**	UB_2	1
MOBILITY		
8. **Do you have difficulty walking or climbing stairs?**	MOB_SS	1
AFFECT (ANXIETY)		
9. **How often feel worried, nervous, or anxious?**	ANX_1	2
10. **Level of feelings when last felt worried, nervous, or anxious?**	ANX_3	3
AFFECT (DEPRESSION)		
11. **How often do you feel depressed?**	DEP_1	2
12. **How depressed you felt last time you were depressed?**	DEP_3	3

NOTE: **Red** refers to the Washington Group Short Set (**WG-SS**).

Red plus **Green** questions are included in **WG-SS Enhanced**.

Response patterns:

	Pattern 1	Pattern 2	Pattern 3*
1	No difficulty	Daily	A little
2	Yes, Some difficulty	Weekly	A lot
3	Yes, A lot of difficulty	Monthly	Somewhere in between a little and a lot
4	Cannot do at all	A few times a year	
5		Never	
7	Refused	Refused	
8	Not ascertained	Not ascertained	
9	Don't know	Don't know	

* IN THE SYNTAX BELOW, NOTE THAT ITEMS WITH RESPONSE PATTERN 6 (ANX_3 AND DEP_3) ARE RECODED TO PLACE "SOMEWHERE BETWEEN" NUMERICALLY IN-BETWEEN "A LITTLE" AND "A LOT".

The Stata syntax presented below includes a couple of elements that were particular to the content of the WG-SS Enhanced.

First, it was important to determine single domain-specific identifiers for those domains of functioning that included multiple questions (upper body functioning, anxiety and depression). For example, upper body functioning includes two questions, each eliciting specific and unique actions: difficulty raising a bottle of water from waist to eye level (arms/shoulders), and difficulty using hands and fingers. Those two questions were analyzed and combined to produce a single upper body indicator with four levels of difficulty ranging from 1 - low difficulty to 4 - high difficulty – not unlike the categorical responses to the single WG-SS questions: no difficulty, some difficulty, a lot of difficulty and cannot do at all. As with the upper body domain, other WG-SS Enhanced domains anxiety and depression have different response patterns that do not readily 'translate' into the usual WG response pattern. For these domains of functioning, a similar 4-scale response pattern was produced and annotated as level 1 through 4, where 1 is the lowest level of difficulty and 4 is the highest.

Second, individual domain indicators were assessed to determine the appropriate cut-off for inclusion into an overall disability identifier – for the purposes of estimating prevalence and disaggregating outcome indicators by disability status.

NOTE:
For all variables, codes (7) *Refused*, (8) *Not Ascertained*, and (9) *Don't know*, are recoded to ***Missing***.

Stata WG-SS Enhanced Syntax Annotated with Output Tables

Actual Stata syntax is indented and are in Bold text.
NOTE: For data analysis, use your standard weighting and estimation techniques.

The syntax below produces **frequency distributions** on individual domain questions – **cross-tabulations** on multiple domain questions, and calculates INDICATOR variables for domains with multiple questions – for use in the determination of disability identifiers.

Step 1: Generate frequency distributions on each of the six domain variables.

VIS_SS is the WG-SS **Vision** question.

> **gen** Vision=VIS_SS **if inlist**(VIS_SS, 1, 2, 3, 4)
> **replace** Vision=. **if inlist**(VIS_SS, 7, 8, 9)
> **tabulate** Vision

Vision: Degree of difficulty seeing

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No difficulty	13690	79.0	81.6	81.6
	Some difficulty	2708	15.6	16.2	97.8
	A lot of difficulty	**333**	**1.9**	**2.0**	**99.8**
	Cannot do at all	**36**	**.2**	**.2**	**100.0**
	Total	16767	96.8	100.0	
Missing		559	3.2		
Total		17326	100.0		

HEAR_SS is the WG-SS **Hearing** question.

> **gen** Hearing=HEAR_SS **if inlist**(HEAR_SS, 1, 2, 3, 4)
> **replace** Hearing=. **if inlist**(HEAR_SS, 7, 8, 9)
> **tabulate** Hearing

Hearing: Degree of difficulty hearing

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No difficulty	13680	79.0	81.6	81.6
	Some difficulty	2753	15.9	16.4	98.0
	A lot of difficulty	**310**	**1.8**	**1.8**	**99.9**
	Cannot do at all	**23**	**.1**	**.1**	**100.0**
	Total	16766	96.8	100.0	
Missing		560	3.2		
Total		17326	100.0		

http://www.washingtongroup-disability.com/

MOB_SS is the WG-SS **Mobility** question.

> **gen** Mobility=MOB_SS **if inlist**(MOB_SS, 1, 2, 3, 4)
> **replace** Mobility=. **if inlist**(MOB_SS, 7, 8, 9)
> **tabulate** Mobility

Mobility: Degree of difficulty walking or climbing steps

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No difficulty	13424	77.5	80.1	80.1
	Some difficulty	2165	12.5	12.9	93.0
	A lot of difficulty	**792**	**4.6**	**4.7**	**97.7**
	Cannot do at all	**380**	**2.2**	**2.3**	**100.0**
	Total	16761	96.7	100.0	
Missing		565	3.3		
Total		17326	100.0		

COM_SS is the WG-SS **Communication** question.

> **gen** Communication=COM_SS **if inlist**(COM_SS, 1, 2, 3, 4)
> **replace** Communication=. **if inlist**(COM_SS, 7, 8, 9)
> **tabulate** Communication

Communication: Degree of difficulty communicating using usual language

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No difficulty	15874	91.6	94.7	94.7
	Some difficulty	745	4.3	4.4	99.2
	A lot of difficulty	**94**	**.5**	**.6**	**99.7**
	Cannot do at all	**43**	**.2**	**.3**	**100.0**
	Total	16756	96.7	100.0	
Missing		570	3.3		
Total		17326	100.0		

SC_SS is the WG-SS **Self-care** question.

> **gen** Self_Care=SC_SS **if inlist**(SC_SS, 1, 2, 3, 4)
> **replace** Self_Care=. **if inlist**(SC_SS, 7, 8, 9)
> **tabulate** Self_Care

Self_Care: Degree of difficulty with self-care

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No difficulty	16029	92.5	95.7	95.7
	Some difficulty	544	3.1	3.2	98.9
	A lot of difficulty	**114**	**.7**	**.7**	**99.6**
	Cannot do at all	**68**	**.4**	**.4**	**100.0**
	Total	16755	96.7	100.0	
Missing		571	3.3		
Total		17326	100.0		

COG_SS is the WG-SS **Cognition** question.

> **gen** Cognition=COG_SS **if inlist**(COG_SS, 1, 2, 3, 4)
> **replace** Cognition=. **if inlist**(COG_SS, 7, 8, 9)
> **tabulate** Cognition

Cognition: Degree of difficulty remembering or concentrating

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No difficulty	13719	79.2	81.9	81.9
	Some difficulty	2632	15.2	15.7	97.6
	A lot of difficulty	**382**	**2.2**	**2.3**	**99.9**
	Cannot do at all	**20**	**.1**	**.1**	**100.0**
	Total	16753	96.7	100.0	
Missing		573	3.3		
Total		17326	100.0		

UPPER BODY

Step 2. Generate frequency distributions and cross-tabulations for Upper body domain questions and determine the Upper Body Indicator.

UB_1 is *Difficulty raising 2 liter bottle of water from waist to eye level.*
UB_2 is *Difficulty using hands and fingers*

> **gen** UB_1_R=UB_1 **if inlist**(UB_1, 1,2,3,4)
> **replace** UB_1_R=. **if inlist**(UB_1, 7,8,9)
>
> **gen** UB_2_R=UB_2 **if inlist**(UB_2, 1,2,3,4)
> **replace** UB_2_R=. **if inlist**(UB_2, 7,8,9)

First, calculate frequency distributions on the two extended set questions.

tabulate UB_1_R

UB_1_R: Diff raising 2 liter bottle of water from waist to eye level

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No difficulty	15677	90.5	93.6	93.6
	Some difficulty	743	4.3	4.4	98.0
	A lot of difficulty	167	1.0	1.0	99.0
	Cannot do at all	166	1.0	1.0	100.0
	Total	16753	96.7	100.0	
Missing		573	3.3		
Total		17326	100.0		

tabulate UB_2_R

UB_2_R: Degree of difficulty using hands and fingers

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No difficulty	15199	87.7	90.7	90.7
	Some difficulty	1229	7.1	7.3	98.1
	A lot of difficulty	255	1.5	1.5	99.6
	Cannot do at all	70	.4	.4	100.0
	Total	16753	96.7	100.0	
Missing		573	3.3		
Total		17326	100.0		

Step 3. Generate a cross-tabulation of the two Upper body Extended Set questions: UB_2_R and UB_1_R.

The syntax below produces a cross-tabulation of the two Extended Set questions: UB_1_R: *Difficulty raising a 2 liter bottle of water from waste to eye level* UB_2_R: *Difficulty using hands and fingers* to determine a single UPPER BODY INDICATOR (UB_INDICATOR).

tabulate UB_2_R UB_1_R

		UB_1_R: Diff raising 2 liter bottle of water from waist to eye level				
UB_2_R: Difficulty using hands and fingers		No difficulty	Some difficulty	A lot of difficulty	Cannot do at all	Total
Degree of difficulty using hands and fingers	No difficulty	14786	309	58	44	15197
	Some difficulty	782	355	51	40	1228
	A lot of difficulty	98	73	51	33	255
	Cannot do at all	9	5	7	49	70
Total		15675	742	167	166	16750

Step 4. Create an UPPER BODY INDICATOR (UB_INDICATOR) based on the two additional upper body questions UB_2_R and UB_3_R.

Syntax below creates UB_INDICATOR based on the distribution in the cross-tabulation above.

> **gen** UB_INDICATOR=4 **if** (UB_1_R==4 | UB_2_R==4)
> **replace** UB_INDICATOR=3 **if** UB_INDICATOR ~=4 & (UB_1_R==3 | UB_2_R==3)
> **replace** UB_INDICATOR=2 **if** UB_INDICATOR ~=4 & UB_INDICATOR~=3 & ///
>   (UB_1_R==2 | UB_2_R==2)
> **replace** UB_INDICATOR=1 **if** UB_INDICATOR~=4 & UB_INDICATOR~=3 & ///
>   UB_INDICATOR~=2 & (UB_1_R==1 | UB_2_R==1)
> **tabulate** UB_INDICATOR

UB_INDICATOR

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	**1.00**	14790	85.4	88.3	88.3
	2.00	1448	8.4	8.6	96.9
	3.00	331	1.9	2.0	98.9
	4.00	187	1.1	1.1	100.0
	Total	16756	96.7	100.0	
Missing		570	3.3		
Total		17326	100.0		

ANXIETY
Step 5.Generate frequency distribution on ANX_1.

First, calculate frequency distributions on ANX_1: *How often do you feel worried, nervous or anxious*?

> **gen** ANX_1_R=ANX_1 **if inlist**(ANX_1, 1,2,3,4,5)
> **replace** ANX_1_R=. **if inlist**(ANX_1, 7,8,9)
> **tabulate** ANX_1_R

ANX_1_R: How often feel worried, nervous, or anxious?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Daily	1632	9.4	9.8	9.8
	Weekly	1872	10.8	11.2	21.0
	Monthly	1558	9.0	9.3	30.4
	A few times a year	4898	28.3	29.4	59.7
	Never	6714	38.8	40.3	100.0
	Total	16674	96.2	100.0	
Missing		652	3.8		
Total		17326	100.0		

Step 6. The syntax below recodes ANX_3R into ANX_3Y
1) to create a NOT ASKED category based on those who responded NEVER to ANX_1_R and
2) to place "SOMEWHERE BETWEEN" numerically in-between "A LITTLE" and "A LOT".

```
gen ANX_3Y=1 if ANX_3R==1
replace ANX_3Y=3 if ANX_3R==2
replace ANX_3Y=2 if ANX_3R==3
replace ANX_3Y=. if inlist(ANX_3R, 7, 8, 9)
```

Recode ANX_3Y to 0 (not asked) if ANX_1 is 5 (Never)

```
replace ANX_3Y=0 if ANX_1==5
tabulate ANX_3Y
```

ANX_3Y: Level of feelings last time felt worried/nervous/anxious

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not asked	6714	38.8	40.3	40.3
	A little	5700	32.9	34.2	74.5
	In between a little and a lot	3076	17.8	18.5	92.9
	A lot	1176	6.8	7.1	100.0
	Total	16666	96.2	100.0	
Missing		660	3.8		
Total		17326	100.0		

Step 7. Generate a cross-tabulation of the anxiety Extended Set questions: ANX_1_R and ANX_3Y.

The syntax below produces a cross-tabulation of ANX_1_R: *How often you felt worried, nervous or anxious* (a measure of frequency) and ANX_3Y: *The level of those feeling the last time you felt worried, nervous or anxious* (a measure of intensity) – used to determine a single ANXIETY INDICATOR (ANX_INDICATOR).

```
tabulate ANX_3Y ANX_1_R
```

ANX_3Y: Level of feelings last time felt worried, nervous or anxious	ANX_1_R: How often feel worried, nervous or anxious?					
	Daily	Weekly	Monthly	A Few Times A Year	Never	Total
Not asked	0	0	0	0	6714	6714
A little	489	887	897	3417	0	5690
In between a little and a lot	589	725	535	1221	0	3070
A lot	548	256	123	248	0	1175
Total	1626	1868	1555	4886	6714	16649

Step 8. Create an ANXIETY INDICATOR (ANX_INDICATOR) based on the two anxiety questions ANX_1_R and ANX_3Y.

Syntax below creates ANX_INDICATOR based on the distribution in the cross-tabulation above.

> **gen** ANX_INDICATOR=1 **if** (ANX_3Y <= 4 & (ANX_1_R==4 | ANX_1_R==5))
> **replace** ANX_INDICATOR=2 **if** ((ANX_1_R==3) | (ANX_1_R < 3 & ANX_3Y==1) | ///
>   (ANX_1_R==2 & ANX_3Y==2))
> **replace** ANX_INDICATOR=3 **if** ((ANX_1_R==1 & ANX_3Y==2) | ///
>   (ANX_1_R==2 & ANX_3Y==3))
> **replace** ANX_INDICATOR=4 **if** (ANX_1_R==1 & ANX_3Y==3)
> **replace** ANX_INDICATOR=. **if** (missing(ANX_1_R) | missing(ANX_3Y))
> **tabulate** ANX_INDICATOR

ANX_INDICATOR

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	**1.00**	11600	67.0	69.7	69.7
	2.00	3656	21.1	22.0	91.6
	3.00	845	4.9	5.1	96.7
	4.00	548	3.2	3.3	100.0
	Total	16649	96.1	100.0	
Missing		677	3.9		
Total		17326	100.0		

DEPRESSION

Step 9. Generate frequency distribution on DEP_1.
First, calculate frequency distributions on DEP_1: *How often do you feel depressed?*

> **gen** DEP_1_R=DEP_1 **if inlist**(DEP_1, 1,2,3,4,5)
> **replace** DEP_1_R=. **if inlist**(DEP_1, 7,8,9)
> **tabulate** DEP_1_R

DEP_1_R: How often do you feel depressed?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Daily	756	4.4	4.5	4.5
	Weekly	926	5.3	5.6	10.1
	Monthly	1038	6.0	6.2	16.3
	A few times a year	4012	23.2	24.1	40.4
	Never	9929	57.3	59.6	100.0
	Total	16661	96.2	100.0	
Missing		665	3.8		
Total		17326	100.0		

Step 10. The syntax below recodes DEP_3R into DEP_3Y to place "SOMEWHERE BETWEEN" numerically in-between "A LITTLE" and "A LOT". It also creates the category NOT ASKED, if DEP_1 is NEVER (1)

```
gen DEP_3Y=1 if DEP_3R==1
replace DEP_3Y=3 if DEP_3R==2
replace DEP_3Y=2 if DEP_3R==3
replace DEP_3Y=. if inlist(DEP_3R, 7,8,9)

replace DEP_3Y=0 if DEP_1==5
tabulate DEP_3Y
```

DEP_3Y: Level of feelings last time felt depressed

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not asked	9929	57.3	59.6	59.6
	A little	3775	21.8	22.7	82.3
	In between a little and a lot	2016	11.6	12.1	94.4
	A lot	935	5.4	5.6	100.0
	Total	16655	96.1	100.0	
Missing		671	3.9		
Total		17326	100.0		

Step 11. Generate a cross-tabulation of the depression Extended Set questions: DEP_1_R and DEP_3Y.

The syntax below produces a cross-tabulation of DEP_1_R: *How often do you feel depressed* (a measure of frequency) and DEP_3Y: *The level of those feeling the last time you felt depressed* (a measure of intensity) – used to determine a single DEPRESSION INDICATOR (DEP_INDICATOR).

```
tabulate DEP_3Y DEP_1_R
```

DEP_3Y: Level of feelings last time felt depressed	DEP_1_R: How often do you feel depressed?					
	Daily	Weekly	Monthly	A Few Times A Year	Never	Total
Not asked	0	0	0	0	9929	9929
A little	161	346	548	2708	0	3763
In between a little and a lot	209	384	378	1042	0	2013
A lot	381	191	112	248	0	932
Total	751	921	1038	3998	9929	16637

Step 12. Create a DEPRESSION INDICATOR (DEP_INDICATOR) based on the two depression questions DEP_1_R and DEP_3Y.

Syntax below creates DEP_INDICATOR based on the distribution in the cross-tabulation above.

```
gen DEP_INDICATOR=. if (missing(DEP_1_R) | missing(DEP_3Y))
replace DEP_INDICATOR=1 if (DEP_3Y <= 4 & (DEP_1_R==4 | DEP_1_R==5))
replace DEP_INDICATOR=2 if ((DEP_1_R==3) | (DEP_1_R < 3 & DEP_3Y==1) | ///
       (DEP_1==2 & DEP_3Y==2))
replace DEP_INDICATOR=3 if ((DEP_1_R==1 & DEP_3Y==2) | ///
       (DEP_1_R==2 & DEP_3Y==3))
replace DEP_INDICATOR=4 if (DEP_1_R==1 & DEP_3Y==3)
tabulate DEP_INDICATOR
```

DEP_INDICATOR

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	**1.00**	13927	80.4	83.7	83.7
	2.00	1929	11.1	11.6	95.3
	3.00	400	2.3	2.4	97.7
	4.00	381	2.2	2.3	100.0
	Total	16637	96.0	100.0	
Missing		689	4.0		
Total		17326	100.0		

Creating Disability Status Indicator for the WG-SS Enhanced

WG-SS Enhanced: WG-SS + Upper Body-indicator + Anxiety (level 4) + Depression (level 4)

The syntax below calculates the WG Short Set ENHANCED Disability Indicator (**SS_E**) based on the 12 questions at the recommended cut-off. The level of inclusion is: at least 1 domain/question is coded A LOT OF DIFFICULTY or CANNOT DO AT ALL for the six short set question; severity levels 3 or 4 for the Upper body-Indicators; and severity level 4 for Anxiety- and Depression-Indicators.

```
gen SS_E=2
replace SS_E=. if (missing(Vision) & ///
        missing(Vearing) & ///
        missing(Mobility) & ///
        missing(Communication) & ///
        missing(SELF_CARE) & ///
        missing(Cognition) & ///
        missing(UB_INDICATOR) & ///
        missing(ANX_INDICATOR) & ///
        missing(DEP_INDICATOR) )
 replace SS_E_3=1 if ((Vision==3 | Vision==4) | ///
        (Hearing==3 | Hearing==4) | ///
        (Mobility==3 | Mobility==4) | ///
        (Communication==3 | Communication==4) | ///
        (SELF_CARE==3 | SELF_CARE==4) | ///
        (Cognition==3 | Cognition==4) | ///
        (UB_INDICATOR==3 | UB_INDICATOR==4) | ///
         ANX_INDICATOR==4 | DEP_INDICATOR==4)
tabulate SS_E
```

SS_E: WG-SS Enhanced Disability Indicator based on 8 domains and 12 questions

		Frequency	Percent	Valid Percent	Weighted Estimate
Valid	WITHOUT DISABILITY	14393	83.1	85.8	87.7
	WITH DISABILITY	2384	13.8	14.2	12.3
	Total	16777	96.9	100.0	
Missing		549	3.2		
Total		17326	100.1		