



Negotiating the Life Course, Wave 4

Sampling Weights for Persons

Working paper only – not for citation

Anna Reimondos *

**Negotiating the Life Course Discussion Paper Series
Discussion Paper DP - 022**

April 2009

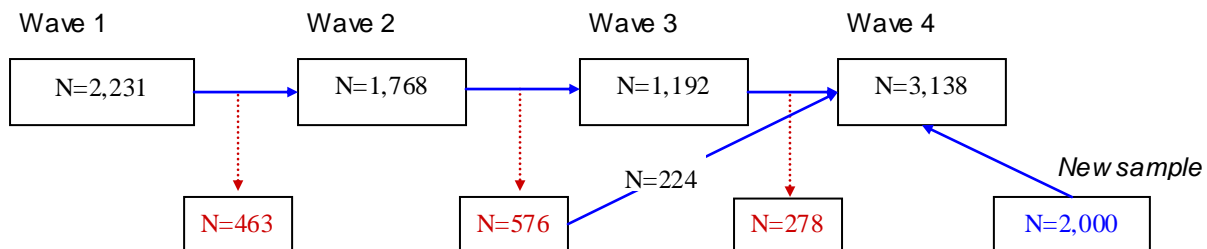
*Australian Demographic and Social Research Institute, The Australian National University

Introduction

This paper describes how the person weights for the fourth wave of the NLC data were created. The procedure for calculating weights in wave 4 was slightly different to that used in waves 2 and 3 as the wave 4 sample contained two different types of respondents; continuing respondents who had first been interviewed in wave 1 (N=1,138), plus an additional sample of new respondents who were first interviewed in wave 4 (N=2,000). The evolution of the NLC sample is displayed in Figure 1. The design and benchmark weights for these continuing and new respondents were calculated separately and then combined into one variable called '*dpswt*' (person weight for wave 4). The general steps for calculating the person weights however are the same as those used for the previous waves, as described in the discussion paper by Breusch (2003):

1. Calculate the design weight or household weighting factor (*dhhwt*), using only information on the number of eligible persons in each respondent's household. The design weight or household weighting factor takes into account the fact that when they were first selected, persons in larger households had a lower probability of being part of the survey than persons in smaller households. A modified or truncated version of the design weight is also calculated (*dhhwtm*), using an upper limit of the number of eligible people in the household of four or more.
2. Calculate the benchmarked weight or post-stratification weight by comparing the age and sex distribution of the sample (weighted by the truncated household weighting factor), with the age and sex distribution of the population using census data. For the continuing sample members the 1996 Census is used, while for the new respondents the 2006 Census is used.
3. Calculate the final person weight (*dpswt*) by multiplying the (truncated) design weight and the benchmarked weight.

Figure 1. Evolution of the NLC sample.



Continuing respondents (N=1,138)

Design weights

The design weight for the original sample members had already been calculated in Wave 1 (Breusch 2003), so the only adjustment that needed to be made for wave 4 was to rebalance the household weighting factor so that it had a mean of 1.0.

Benchmarking weights

Following the procedure used in wave 2 and wave 3, benchmarking of respondents in wave 4 was done using *information on their age/sex distribution and household size at wave 1, and the 1996 census*. At Wave 1, age was grouped into three categories (18-29, 30-44 and 45-54). There were six respondents in Wave 4 aged 55 or over when they were first interviewed in Wave 1 (id=15¹,133,297,1330,1362,2149). Following the method used in Breusch (2003:3) these cases were grouped into the 45-54 year age category.

The number of continuing respondents in each age and sex category are shown in Table 1.

Table 1. Tabulation of Wave 1 'agegrp' by adem002 (sex) and 'numhh' (truncated), Wave 4 sample

Age group	Sex	Number eligible in household				
		1	2	3	4 or more	
18-29	Male	15	48	17	4	84
	Female	31	60	11	6	108
30-44	Male	71	167	8	3	249
	Female	113	240	21	4	378
45-54	Male	40	57	16	9	122
	Female	86	68	28	15	197
Total		357	642	104	41	1,138

In some cases, the number of people in certain age/sex groups, for example the number of females aged 30-44 in households where one person was eligible, increased between waves 3 and wave 4. This is because there were 224 original sample members who were interviewed in waves 1 and 2, who did not participate in wave 3 but who were then re-interviewed again in wave 4.

¹ ID 15 had some changes made to their date of birth, after cross-checking with information in later waves. In the original file their date of birth was given as October 1945, and this was changed to November 1940 in the panel file.

The age and sex distribution of the 1996 census was used to create the benchmark weights, as shown in table 2.

Table 2. Benchmark weights for continuing respondents, Wave 4 sample.

Age group	Sex	1996 Census		NLC		1996 Census % / NLC %
		N	%	(N) weighted by hh factor	%	
18-29	Male	1,591,993	16.7	96	8.5	1.97
	Female	1,582,414	16.6	113	9.9	1.68
30-44	Male	2,033,996	21.3	239	21.0	1.02
	Female	2,083,361	21.9	364	32.0	0.68
45-54	Male	1,128,244	11.8	129	11.3	1.05
	Female	1,109,180	11.6	198	17.4	0.67
Total		9,529,188	100.00	1,138	100.0	1.00

Finally, the resulting person weights for the continuing respondents are shown in Table 3. In Wave 4, the weights for continuing respondents ranged from 0.36 to 4.27, which can be compared to the minimum and maximum values of 0.42 to 3.39 from Wave 3. As in earlier waves males are more underrepresented than females at all ages. Also underrepresented are young people, particularly those in larger households.

Table 3. Values of pswt tabulated by 'agegrp', adem002 (sex) and 'numhh' (truncated), Wave 4, original sample members.

Age group	Sex	Number eligible in household			
		1	2	3	4 or more
18-29	Male	1.07	2.14	3.20	4.27
	Female	0.91	1.82	2.73	3.63
30-44	Male	0.55	1.10	1.65	2.20
	Female	0.37	0.74	1.11	1.48
45-54	Male	0.57	1.13	1.70	2.26
	Female	0.36	0.72	1.09	1.45

New respondents (N=2,000)

Design weights

The design weight for the new respondents had to be created from the beginning, in the same way that the wave 1 weights had been created for the original sample member. The number of eligible people in the household aged 18-63 was not collected in Wave 4, however using Q421, Q422 and child table estimates were made of the number of people within household who would be eligible to participate.

Loose rules were applied to certain relationships:

- An age gap of 25 years was determined for the difference between generations
- Where the respondent was aged 18-25 and they lived in a household with other non-relative people it was assumed all other non-relative people were eligible (*e.g.* a share house)
- Where the respondent lived with parents and siblings and the respondent was aged 18-25 half of the siblings were assumed to be eligible and parents were assumed to be eligible
- Where a respondent lived with his/her children and other relative or non-relative persons it was assumed that if the respondent was over 50 the “other” would be likely to be eligible
- Where a respondent lived with his/her children and other relative or non-relative persons it was assumed that if the respondent was 40-50 or over the “other” would be unlikely to be eligible
- Where a respondent lived with his/her children and other relative or non-relative persons it was assumed that if the respondent was 30-49 the “other” would be likely to be eligible (*eg* friend staying for an extended period)

Benchmarking weights

For the new Wave 4 respondents, benchmarking was done using information on their age/sex distribution at wave 4, and the 2006 Census of the Australian population. In Wave 4, new sample members were eligible if they were aged between 18-63 so the age grouping of the new sample members involved four categories (18-29, 30-44, 45-54, 55-63), rather than three categories as was used for the continuing sample. There were however still 18 respondents who were aged 64 or over at the time of the Wave 4 interview (id=3189,3356,3414,3645,3698,3848,3993, 4199,4307,4322,4333,4459,4518,4567,4591,4729, 4768,4968) and these respondents were placed into the 55-63 age category.

The number of new sample members in the weighting categories is shown in Table 4. While the resulting benchmark weights are shown in Table 5.

Table 4. Tabulation of Wave 4 'agegrp' by ddem002 (sex) and 'numhh' (truncated), Wave 4 new sample

Age group	Sex	Number eligible in household				
		1	2	3	4 or more	
18-29	Male	22	60	83	98	263
	Female	40	183	92	120	435
30-44	Male	35	116	9	6	166
	Female	84	262	32	9	387
45-54	Male	34	72	27	11	144
	Female	44	143	50	21	258
55-63	Male	39	76	17	9	141
	Female	98	76	23	9	206
Total		396	988	333	283	2,000

Table 5. Benchmark weights for new respondents, Wave 4 sample.

Age group	Sex	2006 Census		NLC		2006 Census % / NLC %
		N	%	(N) weighted by hh factor	%	
18-29	Male	1,587,858	12.9	348	17.4	0.74
	Female	1,564,994	12.8	516	25.8	0.49
30-44	Male	2,122,507	17.3	141	7.1	2.45
	Female	2,214,794	18.0	329	16.4	1.10
45-54	Male	1,360,072	11.1	135	6.7	1.65
	Female	1,402,440	11.4	250	12.5	0.91
55-63	Male	1,009,103	8.2	123	6.2	1.33
	Female	1,010,116	8.2	158	7.9	1.04
Total		12,271,884	100.00	2,000	100.0	1.00

In Wave 4 there was deliberate oversampling of young people to deal with the ageing and attrition of young people from the original sample. The top up sample was designed to give a similar percentage of young people at Wave 4 as had originally been present in Wave 1.

The final person weight values for the new sample members are shown in Table 6. The minimum weight for new respondents was 0.22 (for females aged 18-29 in 1 person households) while the maximum weight was 4.35 (for males aged 4.35 in households with 4 or more eligible

people). Again males are more underrepresented than females, but in contrast to the continuing sample, younger respondents particularly aged 18-29 are over represented.

Table 6. Values of pswt tabulated by 'agegrp', ddem002 (sex) and 'numhh' (truncated), Wave 4, new sample members.

Age group	Sex	Number eligible in household			
		1	2	3	4 or more
18-29	Male	0.33	0.66	0.99	1.32
	Female	0.22	0.44	0.66	0.88
30-44	Male	1.09	2.18	3.26	4.35
	Female	0.49	0.98	1.46	1.95
45-54	Male	0.73	1.46	2.19	2.93
	Female	0.41	0.81	1.22	1.62
55-63	Male	0.59	1.18	1.77	2.37
	Female	0.46	0.93	1.39	1.85

After creating the person weight separately for new and old respondents, their person weight values were combined into one variable.

Appendix 1: Stata code for Wave 4.

```
*****
* Name: weights4.do
* Purpose:  Forms weights for persons in NLC, Wave 4.
* Start:   April 2009
* By:      Anna Reimondos & Sue Trevenar & Carole Heyworth
* Infile:  NLC_PanelW1-3.dta
*          NLC_PanelW4-2.dta
* Outfile: personweightsW4.dta
* Stata version 10.1
*****

clear
set memory 250m
set maxvar 7000

*In wave 4 we have continuing and new respondents.
*Calculate person weight separately for continuing and new respondents.

*****
* ORIGINAL RESPONDENTS
*****

*****
* DESIGN WEIGHT          *
*****

*-----
* Wave 1 data
*-----
use "V:\NLC_relationship_panel_vars\Final_NLC_Data\NLC_PanelW1-3.dta", clear

keep id adem002 adem017 anumhh //sex, age & number of eligible in household

*Generate the modified household weight for wave 1, maximum weight of 4 persons

gen anumhh2=anumhh if anumhh<=4
replace anumhh2=4 if anumhh>4
label variable anumhh2 "modified number of eligible in hh, max 4"
egen sumn2=sum(anumhh2)
gen ahhwtm=(anumhh2/sumn2)*_N
label variable ahhwtm "modified wave 1 hh weight"
tab ahhwtm

*-----
* Wave 4 data
*-----

*Merge in Wave 4 data.

merge id using "V:\NLC_relationship_panel_vars\original_data\NLC_PanelW4-2.dta", sort
drop _merge
drop if (ddem002==-10|id>=3000) //drop if not in wave 4, or new sample member.
count /* 1,138*/

*Rebalance wave 1 household weighting factor to average to 1.0 for continuing
respondents.
```



```

egen sumhhwtm_c=sum(ahhwtm)
gen dhhwtm_c=(ahhwtm/sumhhwtm_c)*_N
label variable dhhwtm_c "modified hh weight for continuing resps."
tab dhhwtm_c,m
sum dhhwtm_c

*-----*
* AGE GROUP *
*-----*
capture drop agegpr
gen agegpr=.
label var agegpr "age group of respondent"
label define agegpr ///
1 "18-29" ///
2 "30-44" ///
3 "45-54" ///
4 "55+"
label values agegpr agegpr

replace agegpr=1 if adem017>=18& adem017<=29
replace agegpr=2 if adem017>=30& adem017<=44
replace agegpr=3 if adem017>=45& adem017<=54
replace agegpr=4 if adem017>=55& adem017<=64

tab agegpr,m

list id if agegpr==4

*6 respondents aged over 55 from the original sample.
*Place these people in age group 3.

replace agegpr=3 if agegpr==4
tab agegpr,m

* TABLE 1
table adem002 anumhh2, by(agegpr )

*****
* BENCHMARK WEIGHT *
*****
*-----*
* Weights for Persons, weighting first by dhhwtm_c (modified form
* maximum of 4 persons/household).
*-----*

*Generate the person weight variable
capture drop pswt_c
gen pswt_c=.
label variable pswt_c "Person weight, continuing respondents"

tab agegpr ddem002 [aw=dhhwtm_c] if agegpr>=1 & agegpr<=3 , matcell(N5) row
mat list N5

* Population values from 1996 Census
* Rows `i' are age groups 1/3, columns `j' are the sex, ddem002=1/2

mat C96=(1591993,1582414\2033996,2083361\1128244,1109180)

```

```

scalar sumC96=0
scalar sumN5=0

forvalues i=1/3 {
forvalues j=1/2 {
scalar sumC96=sumC96+C96[`i',`j']
scalar sumN5=sumN5+N5[`i',`j']
}
}

matrix dir
scalar list

matrix list N5
matrix list R5
matrix list C96

matrix R5=(1,1\1,1\1,1)

forval i=1/3 {
forval j=1/2 {
mat R5[`i',`j']=((C96[`i',`j']/sumC96)/(N5[`i',`j']/sumN5))
replace pswt_c=dhhwtm_c*R5[`i',`j'] if agegpr==`i' & ddem002==`j'
}
}

sum pswt_c
tab pswt_c

*****
* TABLE 3 *
*****

table ddem002 anumhh2 if id<3000, by (agegpr) c(mean pswt_c)

keep id pswt_c
save "W:\NLC_wave4_weights\weight4_c.dta", replace

*****

*****
* NEW RESPONDENTS (N=2,000)
*****
*-----
* Wave 4 data
*-----

use "V:\NLC_relationship_panel_vars\original_data\NLC_PanelW4-2.dta", clear
drop if (ddem002==-10|id<3000) //drop if not in wave 4, or old sample member.
count /* 2,000 */

*Merge in Wave 4 information on number of eligible respondents.
merge id using "W:\NLC_wave4_weights\Wave4_weights_p3cases.dta",sort
tab _merge
keep if _merge==3 // keep only new respondents
count /* 2,000 */

*****
* DESIGN WEIGHT *
*****

```

```

gen dnumhh2=numhh if numhh<=4
replace dnumhh2=4 if numhh>4 & numhh!=.
label variable dnumhh2 "modified number of eligible in hh, max 4"
egen sumn2_n=sum(dnumhh2)
gen dhhwtm_n=(dnumhh2/sumn2_n)*_N
label variable dhhwtm_n "New resp. modified wave 4 hh weight"
tab dhhwtm_n,m
sum dhhwtm_n

*-----*
* AGE GROUP *
*-----*

capture drop agegpr
label drop agegpr
gen agegpr=.
label var agegpr "age group of respondent"
label define agegpr ///
1 "18-29" ///
2 "30-44" ///
3 "45-54" ///
4 "55-63" ///
5 "64+"
label values agegpr agegpr

replace agegpr=1 if ddem017>=18& ddem017<=29
replace agegpr=2 if ddem017>=30& ddem017<=44
replace agegpr=3 if ddem017>=45& ddem017<=54
replace agegpr=4 if ddem017>=55& ddem017<=63
replace agegpr=5 if ddem017>=64

tab agegpr,m

list id if agegpr==5

*18 respondents aged over 64 from the original sample.
*Place these people in age group 4
replace agegpr=4 if agegpr==5
tab agegpr,m

* TABLE 4
table ddem002 dnumhh2, by(agegpr )

*-----*
* Weights for Persons, weighting first by dhhwtm_n (modified form
* maximum of 4 persons/household).
*-----*

gen pswt_n=.
label variable pswt_n "Person weight, new respondents"

tab agegpr ddem002 [aw=dhhwtm_n] if agegpr>=1 & agegpr<=4 , matcell (N5) row
mat list N5

* Population values from Census 2006
* Rows `i' are agegroups 1/4, columns `j' are the sex, ddem002=1/2

```

```

mat C2006=(1587858,1564994\2122507,2214794\1360072,1402440\1009103,1010116)

scalar sumC2006=0
scalar sumN5=0

forvalues i=1/4 {
forvalues j=1/2 {
scalar sumC2006=sumC2006+C2006[`i',`j']
scalar sumN5=sumN5+N5[`i',`j']
}
}

matrix dir
scalar list

matrix R5=(1,1\1,1\1,1\1,1)

forval i=1/4 {
forval j=1/2 {
mat R5[`i',`j']=(C2006[`i',`j']/sumC2006)/(N5[`i',`j']/sumN5)
replace pswt_n=dhhwtm_n*R5[`i',`j'] if agegpr==`i' & ddem002==`j' & id>=3000
}
}
tab pswt_n
sum pswt_n

* TABLE 6

table ddem002 dnumhh2 , by (agegpr) c(mean pswt_n)

keep id pswt_n

*Save the new sample members' person weights

save "W:\NLC_wave4_weights\weight4_n.dta", replace
*****

*Combine person weight from old and new people and merge into wave 4 data.
*****

use "V:\NLC_relationship_panel_vars\Final_NLC_Data\NLC_PanelW4-2.dta", clear

merge id using "W:\NLC_wave4_weights\weight4_c.dta",sort
tab _merge
drop _merge
merge id using "W:\NLC_wave4_weights\weight4_n.dta",sort
tab _merge
drop _merge

list id pswt_c pswt_n

*Generate one combined person weight variable.
gen dpswt=.
replace dpswt=pswt_c if pswt_c!=. //information from continuing respondents
replace dpswt=pswt_n if pswt_n!=. //information from new respondents

save "W:\NLC_wave4_weights\personweightsW4.dta"

```