

## Symptoms, Missing Values and Recodes

**David Fitzgerald , 2015**

The ALSWH surveys have asked about symptoms since the first survey. Usually these questions are answered well, some surveys have had a very large number of missing values – much larger than what would have been expected. Also, the values for these high-missing symptoms have looked longitudinally odd, that is when compared with values from other surveys.

### Typical Symptoms Question

*In the Last 12 Months have you had any of the following?*

*Allergies, hay fever, sinusitis,*

*Never, Rarely, Sometimes, Often*

### Standard Recoding for Scaled Questions with Never/No

ALSWH recodes missing values for questions that have a scale including a 'Never' or 'No' option. The recode will check if there are any missing values, the respondent has not checked the 'Never' option, and at least one positive option has been checked. If all these conditions are met then all the missing values are set to 'Never'. The assumption is that the respondent chose to only fill in the non-never responses and left the 'Never' blank. The symptoms questions all have a scaled response with a 'Never' option and therefore received the standard recode for missing values.

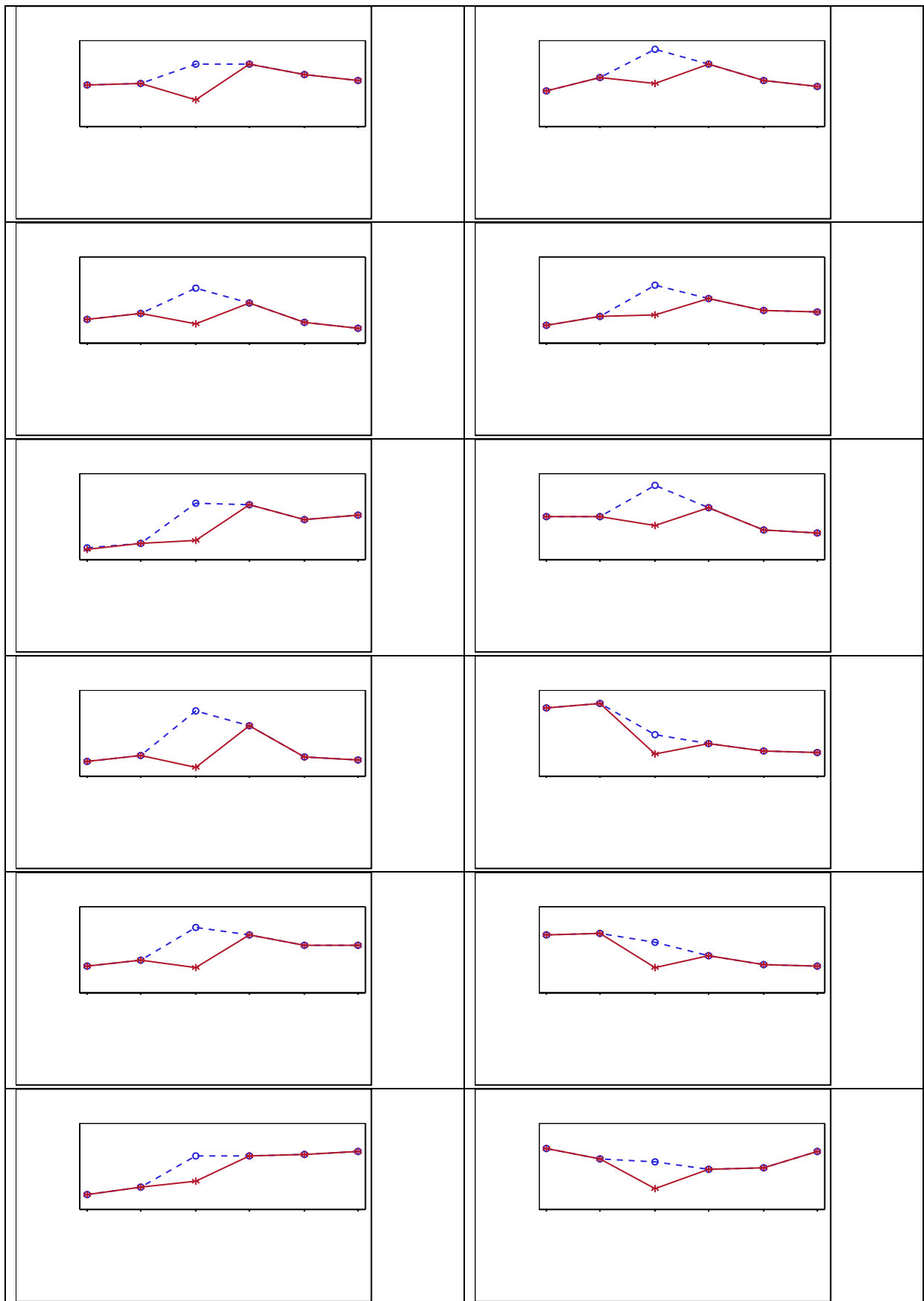
### Recoding and the Missing Rate

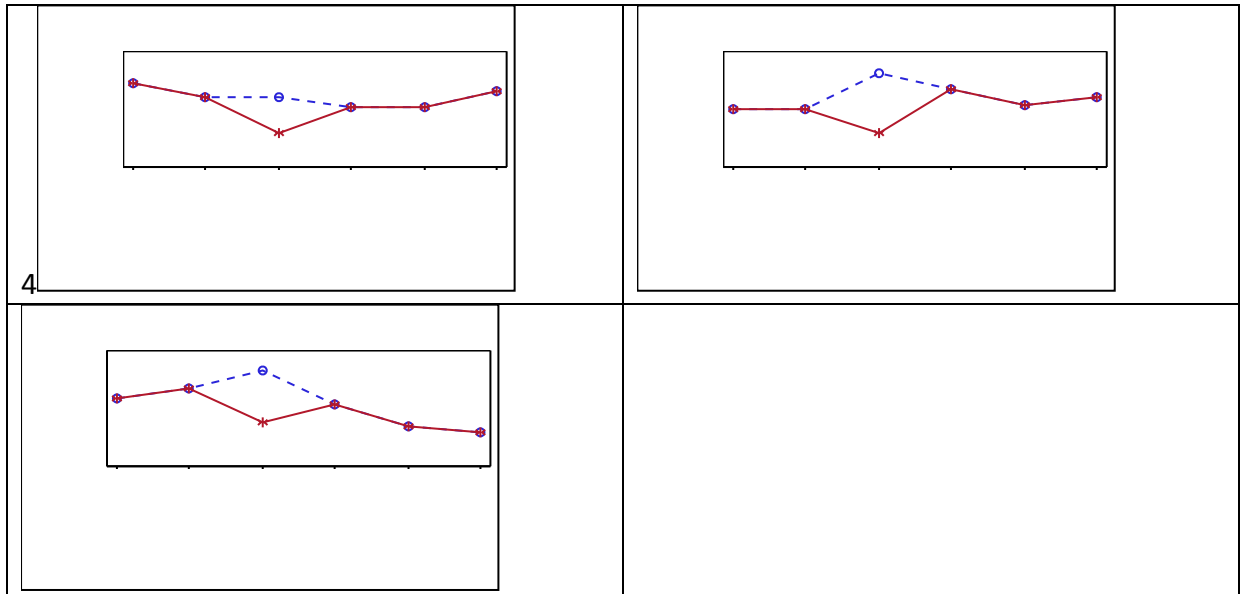
This recode had been used for all the symptom questions in the ALSWH surveys where the missing rate was around 4 or 5 %. It was not envisioned that the missing rate could get to around 40% but this is what happened with the symptom questions in a few surveys, notably the 3<sup>rd</sup> and 4<sup>th</sup> surveys for the 1946-51 cohort. When the missing rate is very high, very many records would be set to 'Never' when this was unlikely to be the true response. This is what appeared to happen in the 3<sup>rd</sup> survey of the 1946-51 cohort, as the graphs below show.

### Symptoms Graphs in 1946-51 Cohort Before and After removing the Recode

The graphs below show the proportion of 'Never' for the symptoms common to Surveys 1 to 6 in the 1946-51 cohort. The blue or dotted line is the value with the recode as initially released which included the standard recode. In 2011 it was decided to remove this recode from the 3<sup>rd</sup> survey only and re-release the survey data effectively as raw data. The red or unbroken line shows the re-released data with no recoding that replaced the initial release. Only Survey 3's data have changed.

Percentages of Never had Symptom, 1946-51 Cohort: Release 1 (Blue), Release 2 (Red)





4

### Missing Data Numbers

The numbers of missing in the raw data were very high for a few surveys, not only Survey 3. After removing the recode in Survey 3 the published data had very high missing. In Survey 4, by contrast, the raw data did also have high missing but after the recoding the published data had low missing. The table below shows the missing percentages for each symptom in Surveys 1 to 6, for both raw data and published data.

**Table of Percentage Missing for each Symptom by Survey, 1946-51 Cohort**

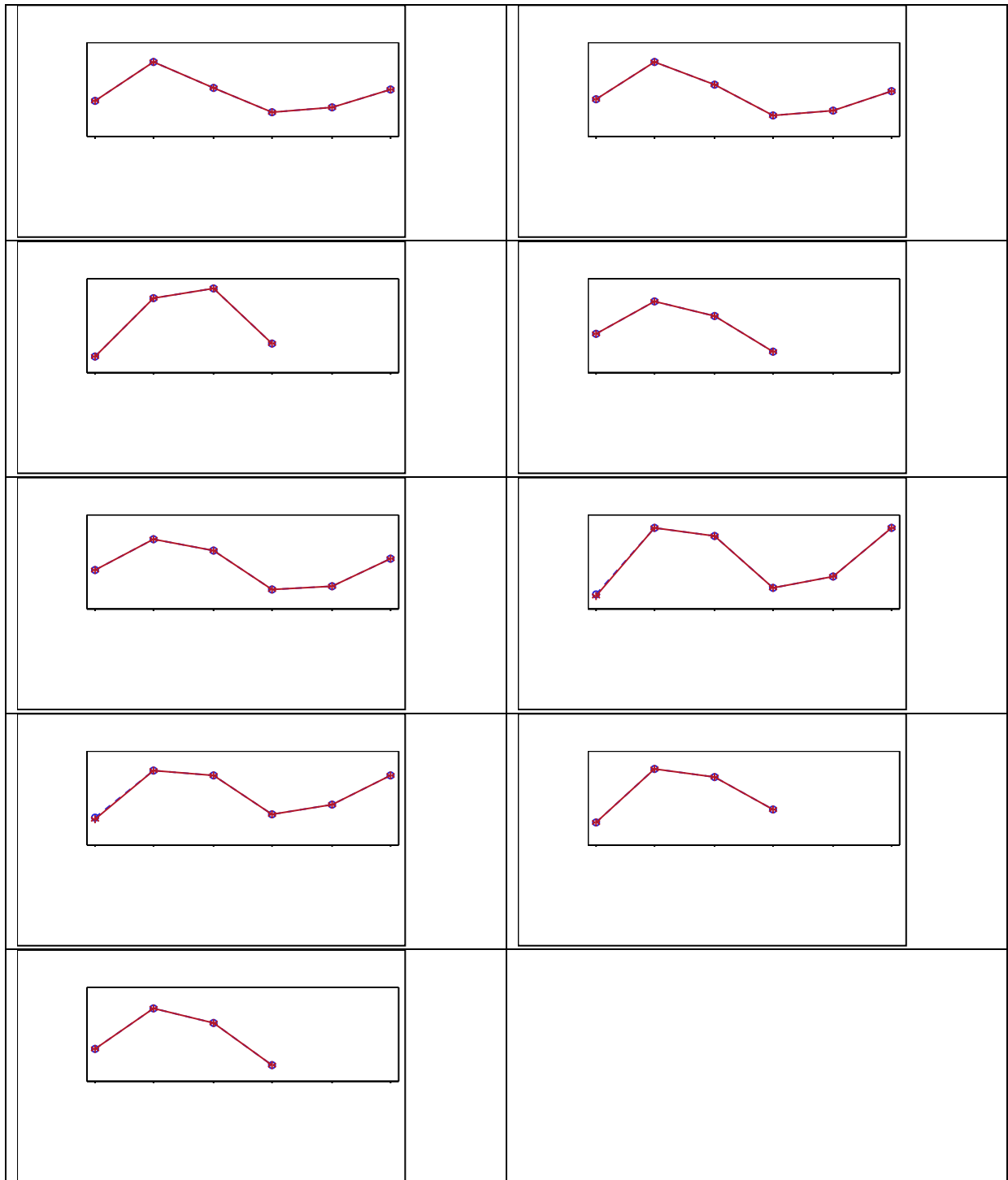
<b>Symptom</b>	Version	<b>Survey</b>					
		1	2	3*	4	5	6
		m1miss	m2miss	m3miss	m4miss	m5miss	m6miss
allergy	Raw	1	10	17	17	6	5
allergy	Output	1	8	17	3	2	2
brthdiff	Raw	1	13	34	31	11	9
brthdiff	Output	1	9	34	5	3	3
indign	Raw	1	12	30	27	9	7
indign	Output	1	8	30	5	2	2
chestp	Raw	1	14	38	35	13	11
chestp	Output	1	9	38	7	4	3
headache	Raw	1	9	21	20	7	6
headache	Output	1	7	21	3	2	1
joints	Raw	1	10	24	19	6	5
joints	Output	1	7	24	3	2	1
backpain	Raw	1	10	22	20	6	5
backpain	Output	1	7	22	3	1	1
urine	Raw	1	13	37	32	11	9
urine	Output	1	8	37	5	2	2
piles	Raw	1	12	35	31	11	9
piles	Output	1	7	35	5	2	2
bowel	Raw	1	13	36	32	11	9
bowel	Output	1	8	36	5	2	2
vagina	Raw	1	12	37	33	12	9
vagina	Output	1	7	37	5	2	2
hotflsh	Raw	1	10	23	19	8	7
hotflsh	Output	1	7	23	3	1	1
nsweats	Raw	1	11	28	25	10	8
nsweats	Output	1	7	28	4	2	2
eyesight	Raw	1	11	29	25	9	8
eyesight	Output	1	8	29	4	2	2
hearing	Raw	1	12	36	31	10	8
hearing	Output	1	7	36	4	1	1

\*\* Standard recodes removed

## 1921-26 Cohort

The 1921-26 cohort had a similar problem but the missing rare was much lower. In Survey 2 the missing rate got to 9% but otherwise it was about 5%. It is possible that Survey 2 had an inflated rate of 'Never' for each symptom.

### Percentage of Never had Symptom, 1921-26 Cohort



**Table of Percentage Missing for each Symptom by Survey, 1921-26 Cohort**

<b>Symptoms</b>	time	<b>Survey</b>					
		o1miss	o2miss	o3miss	o4miss	o5miss	o6miss
stifpain	Early	4	9	4	5	4	4
stifpain	Later	4	9	4	5	4	4
backpain	Early	4	9	4	4	4	4
backpain	Later	4	9	4	4	4	4
allergy	Early	5	9	5	6		
allergy	Later	5	9	5	6		
skinP	Early	4	9	6	5		
skinP	Later	4	9	6	5		
breath	Early	6	9	5	4	5	5
breath	Later	6	9	5	4	5	5
indiges	Early	5	9	5	5	5	5
indiges	Later	5	9	5	5	5	5
chestp	Early	5	9	6	5	6	6
chestp	Later	5	9	6	5	6	6
headache	Early	4	9	5	5		
headache	Later	4	9	5	5		
severet	Early	5	9	4	4		
severet	Later	5	9	4	4		