

### Appendix 3: Forensic notes

(unedited, in order of surveys tackled, direct from logs kept during processing of files from Essex and attempts at restoration using SPSS for Windows)

#### **Fifth form survey 1981 (processed Oct 2002 and Oct 2004)**

2 Oct 2002 Converted **fifth.dat** from Essex WP6.1 to MSWord \*.txt format as **fifthdat.txt**

Converted **fifth.sps** from very old SPSS syntax to SPSS11 for Windows syntax (mainly input format changes to read as alpha and convert to numeric and changes in value labels to get rid of brackets and replace with single primes) Ran a few test jobs on sexism and other scales (not saved) and left as initial \*.sav file with no derived variables. Currently saved as **fifthx.sps** and **fifthx.sav** on c:\jfh\fifth and backed up on floppy.

Some multiple response specifications written. Scaled variables were initially in short jobs for teaching purposes for use one at a time. Have to watch problem of permanent recoding of items used in batteries to generate attitude measures: might be better to save derived variables separately using **save out .../keep....** and merge files at a later stage.

References:

Paul Ahmed, Harriet Cain and Alan Cook **Playground to Politics: a study of values and attitudes among fifth formers in a North London comprehensive school** Report on 2<sup>nd</sup> year project for BA Applied Social Studies (Social Research) Polytechnic of North London 1982

John Hall and Alison Walker, **User manual for Playground to Politics: a study of values and attitudes among fifth formers in a North London comprehensive school** Survey Research Unit, Polytechnic of North London 1982 (mimeo 40 pp – codebook, questionnaire, coding notes)

Note: Latest version is SPSS portable file **fifthx.por** (Feb 2004, 107kb) now saved in sub-folder **fifth** in folder **PNL\_SRU** in **desktop**

**Need to generate a flysheet for this study as per QoL, Trinians etc**

**Also need variable and value labels for spread data on card 4.**

**JFH 16 Oct 2004**

## Quality of Life: First Pilot survey March 1971 (processed Jan – Feb 2004)

Resuscitation attempt 15-16 Jan 2004

Data received from Essex as a concatenation of SPSS setup files and data files, although the survey was originally deposited as ready to use SPSS saved files.

In order to recreate these files it was necessary to be the person who created the original data (me!) and to know why and how an original set of multipunched Hollerith cards (2 per case) was exploded into 6 lines of data per case (multiple response questions and also more than one variable per column!!) and use of upper and lower zones on the cards as well as digits 0-9.

This was an absolute nightmare as SPSS syntax has completely changed and the data had to be rewritten using **data list** and different conventions for reading alpha data then converting it to numeric. It has taken the best part of two days.

Most of the data seems to have been captured with most of the labelling and missing value info, but there is a lot of checking and tidying up to do. In 1973 SPSS was very primitive and everything is in upper case (including value labels) with most variable names in the VARxxx to VARyyy convention.

Much of the data was first read in as single column alpha to circumvent the use of upper and lower zones ('+' and '-') and of multipunching in the same column, then converted to numeric and then (if necessary) reassembled as multicolumn numeric. Sounds horrifyingly cumbersome but is actually quicker if you know what you're doing. Later the alpha variables will be dumped and the remaining variables reordered to follow the original questionnaire order, apart from the multipunching. These can be left as spread out data on records 3 to 6 and a file of multiple response specifications can be created from which sections can be copied into analysis runs.

17 Jan. 04

Couldn't get SPSS to read the data from **d247** yesterday, so copied contents of data file into spss job and ran with:

```
begin data  
.....data set on 6 cards.....  
end data.
```

This worked.

Unearthed original documentation, including interviewer instructions, data layout info and Users' Manuals (SSRC reprographics request 1973), together with PNL printout of labels (18 Nov 77). Manuals include raw frequency counts for all variables in the original file.

Time taken in trying to use original SPSS jobs probably might have been better spent starting from scratch using more recent facilities such as lower case letters and sequential variable naming with other than **VARxxx TO VARyyy**.

Positional variable naming retained.

Gradual piecemeal restoration of file, but frustrating. SPSS frequency counts don't appear to allow codes and labels on same table. Must investigate this. Awkward work, but frequencies so far seem to tally with original. Main problem is checking which code values have been or need to be declared as missing

## Value labels

Lower case letters introduced in value labels as these are neater. Also some original labels were written up as two blocks of 8 characters to keep output reasonably clear and tidy with SPSS then limits of 20 characters (only 16 printed as column headers). These restrictions no longer apply, except when using **mult response**.

## Variable labels

Later file construction conventions at SSRC and PNL mean that some of these do not comply for easy relating to questionnaire (eg Q8 Anomy scale). These should be changed to include question number at beginning plus some indication of content of question.

Thus: **VAR124 ANOMY MEASURE Q8A** might usefully be changed to  
**VAR124 Q8a Most people will go out of their way** or longer as SPSS no longer limited  
to 40 characters, eg **VAR124 Q8a (Anomy scale) Most people will go out of their way**

Perhaps a better example would be **VAR144 STATE OF HEALTH** changed to  
**VAR144 'Q.12 Your general state of health'**.

## Variable names

For the moment I'm leaving them in **VARddd** format, but it would save typing for analysis if they were in **vddd** format. That's 2 key depressions saved for every variable typed, and they can be done in lower case as well.

Gone through checking and deleting alpha columns whose multipunched codes were spread out on cards 3-6.

Next check that frequencies for converted variables from alpha **Vxxx** to numeric **VARxxx** are a) all present and b) the same as per original user manual. If so the alpha **Vxxx** variables can also be deleted. Then got to find a way of saving the file with all the variables back in the original order. Some converted variables need to be kept in order to generate derived variables such as duration of interview (in minutes). Some **VARxxx** variables can be renamed in the case of common demographic variables such as SEX AGE etc. These will be kept together in a block at the end of the file to make analysis easier.

Some codes grouped on original (because of very small numbers) have been left ungrouped here eg VAR144 where only 2 respondents gave their health as "poor". Also SPSS does not print totals for empty categories (or at least I haven't found a way to force this). Thus the frequency count for complete dissatisfaction with health (Code 0) has no respondents, but should have been included in the table. Same problem with satisfaction with friendships and police and courts (VAR162 and VAR169): table is truncated as there are no R's on scale point 1.

Manual has 54 R's on 0 code for job satisfaction (VAR149) but this file only has 2. Check that this tallies with numbers with jobs (either self or partner).

Can't get SPSS to reorder variables in the file using **save out filename /keep varlist**

Just managed to do this by copying file to drive **a:** then using a **get file** command to retrieve it with **/keep** etc and save the resultant file to QL1 area on **c:**

One or two missing values added, but the whole of the spread out multipunches needs to be looked at again in the light of practice developments at PNL and later versions of SPSS. The best thing would be a complete set of mult response specifications kept on file for downloading into particular SPSS jobs, bearing in mind that there are limits on the number of implied variables than can be used at any one time. It is also better for analysis to use separate codes for each response and use these for labelling the first variable in a mult response job. (SPSS only looks at the first variable in the list for value labels) In binary mode the variable labels need to be clear as to the nature of the variable. Unless duplicate sets of mult response variables are kept on file (wasteful) either convention will require recodes to spread binary 1's out to 2,3,4 etc or to recode 2,3,4 etc to 1 for binary analysis. This will not be a problem, but will be time consuming. Anyway that's a job for later!

I've been at this all day today, but at least I've cracked it for now.

Also had to generate intermediate variables to get alpha coded to numeric as (**convert**) doesn't work if recode is into original vars.

24 Jan 2004

Ran off a set of overall satisfaction items to send to Roger Jowell at NatCen to compare with their new European stuff.

2 February 2004

Checking over versions for deposit at Essex. Most further work will involve changing case of letters in labels. Missing values added for qq9-23.

Errors found: (amend manual accordingly)

---

page	variable	amendment
5	Var144	there are 2 cases with code 1 (Poor) and 21 with code 2 (Fair) total = 23
	Var149	there are only 2 cases with code 0, but 52 blank (not asked) in data
9	var230	summary only: age is coded as actual years
16	var306-318	Labels not clear: change
20	var365	180 should be 179 (1 code blank in data)
	var374	174 should be 173
	var368	90 should be 89
21	var420 to 427	should be corrected to var421-428 in manual.
28	566,576	NA (No) declared as missing, but could be recoded and used as "No"

---

No var232 or 233. Multipunched? Where now? Spread out on 250ff I think.

Checked means and correlations for Abrams & Hall paper. They're not the same, so tried using a weighting procedure, but they're still not the same. God knows where we got these from. Could have been LSE or also RSL. Doesn't make much difference to the rank order of values, but it's a bit worrying all the same.

## Quality of Life: Second Pilot survey Oct-Nov 1971 (processed Jan – Feb 2004)

:

Latest system file: **qlukpilot2.sav**

Same procedure as for QL1, with same problems for alpha recodes. Much quicker this time, but tedious having to make manual alterations to variable lists to resolve it.

File saved in original format with all capital letters and VARddd except where edited manually, but SPSS case insensitive for varnames..

There's a weight statement at the end which gives whole sample 2 but London 3. Not sure whether to put this in or not. **Have put it in as weight**

11 Feb. 04

Added spread out multipunches from card 3. Put all variables in card order. Computed sdscore and anomy. Computed weight = 2 for every case except 3 for London. Run off unweighted frequency counts for all variables. Can't get Adobe to print manual on single pages, so son Richard has printed off pp 15-92 and will post. Need to check frequency count against manual. Then this one can be put to bed as portable file. **Done**

There do not appear to be any derived variables other than **sdscore** and **anomy** in the original setup files, so it may be worth creating some standard variables such as **sex, age, class** etc plus a set of overall satisfaction ratings to tally with the same variables eg **life, health, job** etc in the other surveys. The latter are all on 1-7 scales for comparison with ISR studies by Campbell et al. Codes on the items in the sd scale have been reversed so that 7 = high/good for scaling purposes.

Saved to dsk:e as **ql2ukpilot.por**

All frequencies to be checked against manual pp 19 ff **Done, but nearly went blind comparing spo with pdf files**

All rating scales have codes 8 and 9 as missing, but in the manual these seem to have been condensed to 9. The combined missing totals tally OK.

Some NA and DK codes don't tally, though the totals missing do tally. This may be because of later logical checks. Usually this amounts to only a single case. It may also be due to the way DK was coded. Check that this was consistently 9 or was sometimes 8. **Not worth it: leave**

Dollar signs in value labels need changing to £ signs. **Done**

Card 1 frequencies OK apart from comments above. Missing values won't affect any analysis as they're all declared anyway.

**Var272** week of interview: code 5 = ??? Too many to be missing, so could be Nov 7-13. Have entered this as **??Nov 7-13??**

**Var273** in output is grouped as **var/10** for conurbations, but manual has full list on p.23 (also in sampling appendix) Have generated **var273** as **var273\*10+var274** and put labels in to match manual.

Reversed items from semantic differential scales have no missing values (because they were outside the original command having been converted. A new missing values statement should sort it, but the data have code 8 whilst manual has 9 for DK. **Done**

No value labels on var252 ff (be careful as codes are reversed on alternate items to retain scaling properties) . **Done**

What happened to var259 (newspaper readership)? . **Done**

**Sdscore** and **anomy** are simple sums of items in their respective scales, but strictly speaking they need reducing by the number of items in the scale to yield a true zero point. However, I've left them in their crude state for now.

For Essex, strip out derived variables, expand var273 to full borough codes (? Add labels?) so that data set matches manual as published. **Done**

First edition of file for Essex is **qlukpilot2.por** with all variables from **case** to **var399** plus four additional variables, **sdscore**, **anomy**, **conurb** and **weight**. Partial setup file **ql2newvars.sps** for additional variables and labelling, including additional value labels for variables using response cards A, B and C. No further work envisaged on this file for some time, but this will involve changing all labelling to lower case, and generation of standard sets of derived variables. Also an unweighted frequency count for all variables **ql2freq.spo** **Done**

Lot of piddling fiddly work on some incomplete labels and the odd missing value, but I think it's all there now for the first release. **Erratum** on data layout sheet: **var264** and **var265** (sex and age group) transposed. Age and class ditributions very even, probably because of quota restrictions. Should the data be reweighted to take account of this, even though it's a quota sample?

## Attitudes and Opinions of Senior Girls – Feb 1973 (processed Jan – March 2004)

There was no information on the questionnaire which could be used for data layout and data-preparation (it would have made for cluttered presentation and in any case there was no room!). The questionnaires were manually coded in-house by Eleanor Clutton-Brock and the data transferred to (?pre-printed?) coding sheets (can't remember unless there's an example extant), then punched on to 80-column Hollerith cards (3 per questionnaire).

This makes it difficult to work direct from the questionnaire when performing data management and analysis, so if there is no data guide sheet, then one needs to be produced. Otherwise variable labels need to be checked to ensure the question number is included.

Most of the questions were single response pre-coded on the questionnaire and these were single-punched on cards 1 and 2. Codes for some multiple response questions were multipunched, but facilities for handling multi-punching of columns were not available in SPSS at the time, and so codes for multiple response questions on readership of newspapers and magazines were spread out and single-punched on card 3 for input to a very early version of SPSS. Some data seems not to be present ( Questions 6, 7 and 8: 'O' and 'A' level subjects taking/taken, and pupil's interest therein. This seems odd, but unless any other documentation comes to light, we must assume the data lost or the questions not coded in the first place.)

### Restoration of files

Although the final version of the SPSS saved file was submitted to Essex on a mag tape, this has not been preserved. A later version kept at the Polytechnic of North London seems to have suffered the same fate, as the tape archive available only goes back to 1986. This a great pity, as **export** and **import** would have saved a great deal of time and tears. The author is not yet completely *au fait* with the Windows version, but has managed to recreate a new saved file from the original setup files supplied by Essex.

\*\* Mark Abrams and John Hall **Attitudes and Opinions of Girls in Senior Forms**  
SSRC Survey Unit, March 1973 (mimeo 20pp)

[NB Author hasn't worked out to do footnotes yet, or superscript characters]

Since 1973 there have been many subsequent releases of SSPS, not just for mainframe, but also for PC and most recently for Windows. The Windows release 11 has now got most, but not all, of the facilities of mainframe release 4. SPSS syntax has completely changed, and so many setup jobs simply will not work. Thus (with apologies to Ronald Searle) the file supplied defined data thus:

```
RUN NAME    TRINIANS CREATION
PAGESIZE    NOEJECT
FILE NAME    TRINIANS
VARIABLE LIST  FORM NUMBER MONEY YEARBORN MONTH VAR111 TO VAR119 JOB1 TO JOB5
              JOBAT25 SUCCESS1 SUCCESS2 LIKELY FATHER MOTHER PARENTS WEEKENDS
              SISTERS BROTHERS ELDEST VAR142 TO VAR176 VAR205 TO VAR234 VAR237
              TO VAR266 VAR270 TO VAR276 VAR305 TO VAR312 VAR314 VAR317 TO
              VAR339 VAR341 TO VAR349 VAR353 TO VAR364
INPUT MEDIUM  INDATA
INPUT FORMAT  FIXED(F1.0,F2.0,1X,F3.1,F1.0,F2.0,9F1.0,6A2,2A1,43F1.0/
              4X,30F1.0,2X,30F1.0,3X,7F1.0/
              4X,8F1.0,1X,F1.0,2X,23F1.0,1X,9F1.0,3X,12F1.0)
N OF CASES    216
```

But this had to be changed to:

data LIST records 3

```
/1 FORM 1 NUMBER 2-3 MONEY 5-7 (1) YEARBORN 8 MONTH 9-10 VAR111 TO VAR119 11-19
xOB1 TO xOB5 xOBAT25 20-31 (a)
xUCCESS1 xUCCESS2 32-33 (a)
LIKELY FATHER MOTHER PARENTS WEEKENDS
  SISTERS BROTHERS ELDEST VAR142 TO VAR176 34-76
/2 VAR205 TO VAR234 5-34 VAR237 TO VAR266 37-66 VAR270 TO VAR276 70-76
/3 VAR305 TO VAR312 5-12 VAR314 14 VAR317 TO VAR339 17-39
VAR341 TO VAR349 41-49 VAR353 TO VAR364 53-64.
```

A second problem was trying to read the data from an external file. On my machine, SPSS could not find the data file specified, or did not like the way it was defined. Eventually it was quicker to copy the raw data file into the setup job and run it with **begin data** and **end data**. The eventual saved file was generated over several runs.

In the original version of SPSS it was possible to read in variables in alpha format and then **recode** them with a (**convert**) keeping the same variable names. This is no longer permitted as **string** variables (as they are now called) can only be converted into a new set of variables. Therefore the first letter of the initial variables to be read as strings was changed to x (eg **JOB1** was read in as **xOB1**) to create intermediate variables and a later **recode (convert)**ed them **into** the original names as specified in 1973; the intermediate variables were then deleted from the file.

This entailed modifications to the data transformation commands which were tedious rather than complicated. The variable labels and value labels needed modification to get rid of single primes and full stops, which took several runs as they were quite difficult to spot, but with the sheer speed of SPSS it was quicker to run jobs and look at the error reports, then delete the output file without saving it.

**SPSS still generates far too much output and could do with a facility for automatically keeping only two editions of output files, or at least having a prompt “Do you want to keep the output?” instead of clicking on the x and then answering a question.**

Also in 1971 there were no facilities for lower case letters or for automatic variable generation other than by **VARxxx TO VARyyy**. Later releases allowed names with any letter of the alphabet, but still only in capital letters (eg **Q1 to Q10**): nowadays lower case letters are allowed for names in setup jobs, but will be printed as capitals in output. **There is still no facility for generating names by e.g. Q1a to Q1g.**

The author has a distinct preference for operating via syntax files rather than ‘point and click’ on a menu, which horrifies him and is confusing and exasperating to use because not all the information needed is displayed in the view. Because at SSRC and later at PNL he and his colleagues were handling large numbers of surveys and even larger numbers of SPSS runs he developed a system for naming of files in which file names indicated what kind of run it was and file extensions what kind of file. Thus:

<b>TRINIANS.DAT</b>	contains raw data for the Trinians survey	
<b>TRINIANS.SPS</b>	would be a SPSS setup file generating output file	<b>TRINIANS.LST</b>
<b>TRINIANS.SYS</b>	would be the saved system file	
<b>TRINIANS.DOC</b>	would be a documentation file for the Trinians survey.	

and so on for **RECODE1.SPS RECODE2.SPS VARLABS.SPS VALLABS.SPS**

For a full explanation and of SSRC/SU and PNL/SRU conventions for variable naming, see file **NAMES.DOC**

**FREQ1.SPS** and **TAB1.SPS** generate **FREQ1.LST** and **TAB1.LST** (frequencies and tabulations) Even the extension names have been changed over the years, so even though **.sps** is the same, **.lst** became **.lis** and then **.spo**, whereas **.dat** now seems to indicate a WordPerfect file and **.doc** a file for MS-Word! Self-evidently jobs like **FACTOR.SPS** and **ANOVA.SPS** are easy to find in a directory and indicate the contents better than **SYNTAXddd** etc. At least two and sometimes three copies of all files were backed up on mag tape, and in cases where significant and substantial changes had been, there would be two or three previous editions of each file backed up as well.

SPSS for Windows doesn't like this convention for names and extensions, but it doesn't take long to learn to leave the extensions off and use the default SPSS (implied) extensions.

So far this restoration has taken 15 hours on 17 Jan and 5 hours on 18 Jan. and even more on subsequent days. The file has all the original variable and value labels in block capitals, except where some editing has been done. A first frequency count has thrown up some variables which have unexpected values or values with no labels, plus a few values still to be declared as missing. Also, the variable labels need to be checked to make sure the question numbers are included, as otherwise analysis would be a nightmare as the only documentation so far available is an unannotated questionnaire.

At least this now exists, but caused problems when printing from .pdf as the printer kept having a memory overflow and two of the pages wouldn't fit properly, so even this is now a scissors and paste job!

[NB Should the relevant bits of the transformations and labels be included here (if I can find them all!) or as an appendix? Originals are on **d951.sps**, amendments (perhaps not all) on **syntax2.sps**]

JFH Sunday 18 January 2004 12:50 hrs

Tidied up missing values which, though declared seemed not to work and sorted value labels for some variables where full stop abbreviations made SPSS stop working. Like I said, tedious, but at least it's done. Current labelling very ugly and might have been quicker to retype the lot with decent lower case printing for output. File needs rearranging to get variables in a logical order, or at least questionnaire needs annotating by hand to indicate variable names.

Phase 1 complete at last!

JFH Sunday 18 January 2004 1500 hrs

Printed up some preliminary documentation last night from SPSS setup files and output from **data list** and **display**. There seem to be some variables missing, so need to check original data. Variables were not declared in questionnaire order for some (probably perfectly good) reason. Marked up copy of questionnaire with varnames and data locations. Some of these will need to be changed to conform to PNL-SRU conventions, and it would be useful to have at least rudimentary user manual with full question text, coding instructions, data locations and transformations, plus a frequency count (raw n only, but how to do this with SPSS **frequencies** which gives everything but the kitchen sink!)

JFH Monday 19 Jan

Tue 20 Jan. 04

Renamed variables from VARxxx and mnemonics to vddd (except derived vars) and reordered variables into order as entered. **This is not the same as the questionnaire.** Deleted superfluous and

intermediate variables and added a couple of labels. Must find out how coding was done for Q2 Weeklies and others: also data for Q3 enjoyment of Folio.

### Wed 21 Jan

Checked original data files to see what was coded where for multipunching. There is some, but apparently nothing for qq6-8. Printout of data file does not retain fixed width columns, so very difficult to read. Easier to use SPSS to write out a new data file. Our full conventions would have left a space after the serial number and a blank column somewhere in the middle of each card so that a printout will reveal codes that have slipped forwards or backwards (easily done when punching long lists of digits). This would be done separately for each card so that the blanks show up as a blank vertical column. Can't remember who did the spreading out, or where, but probably Jim Ring, who had by then joined SSRC/SU from LSE.

Thu 22 Jan. 04

Amendments to log of work done (confidentiality). Must really edit setup files to use lower case letters for labelling. If I could work out how to do it, the info on the data editor is enough to create a codebook key, but frequencies produces too much, if all we need is the raw codes and counts.

### Fri 23 Jan. 04

Had a shot at multiple response tables, but SPSS won't do recodes into same vars, so had to create new vars for newspaper readership etc. Also Sundays and monthlies have been given labels in common, so needed to split these.

Being lazy, I've been trying to find quick ways of doing things, which is frustrating, but I'm learning my way round the editing facilities of SPSS and Word, and using whichever is quicker for me. So I find it's quicker to copy chunks of text out of SPSS setup files into Word, use that to change cases (usually whole file from upper to lower) and make mass substitutions to put some capitals back, then save as a .txt file. Latter can then be copied into a .sps file and run. Main problem is keeping track of all the changes and filenames, but am using old conventions of **varlab...** and **vallab....** for these plus **mult...** for multiple response setups. There's a lot of complex programming and trial and error in some of these, but there's no real need to include them in the main documentation except for SPSS buffs to show a few tricks of the trade.

The basic data set has multiple responses spread out as binary data in 1's and 0's, but for some applications the 1's need to be recoded to an ordinary coding sequence of 1 to n. In the former case tabulations can be done in binary format and the tables make sense, but only if the var label includes the code reference: in the latter, it is only necessary to put value labels on the first variable in the group = list even though this may seem bizarre to the novice user as all codes except the first one will not exist for the first variable. Question is whether to save the converted variables and labelling on the main file (eg by using **Mddd** instead of **Vddd** to indicate part of a set of variables for use in **mult response**

Hopefully have now managed to get file into presentable and usable format. One or two more mult response lists to sort out, but some base vars need checking first to see what's in there. Also the var sequence doesn't match the questionnaire sequence for precoded responses, but this may be due to in-house coding. Not sure who did this: could have been Sara herself or a trainee researcher, Eleanor Clutton-Brock.

To produce a multiple response frequency table in binary mode..

```
mult response /group = Dailies 'Daily newspapers read'
                (v305 to v314 (1))
                /freq dailies.
```

Group DAILIES Daily newspapers read  
(Value tabulated = 1)

Dichotomy label	Name	Count	Pct of Responses	Pct of Cases
Q2 Daily papers - Express	V305	29	10.2	13.4
Q2 Daily papers - Mail	V306	23	8.1	10.6
Q2 Daily papers - Mirror	V307	5	1.8	2.3
Q2 Daily papers - Morning Star	V308	1	.4	.5
Q2 Daily papers - Sun	V309	1	.4	.5
Q2 Daily papers - Telegraph	V310	55	19.4	25.5
Q2 Daily papers - Times	V311	86	30.3	39.8
Q2 Daily papers - Guardian	V312	46	16.2	21.3
Q2 Daily papers - None read	V314	38	13.4	17.6
	Total responses	284	100.0	131.5

0 missing cases; 216 valid cases

But an attempt to produce the alternate format with...

```
recode v305 (1=1)/v306(1=2)/v307(1=3)/v308(1=4)/v309(1=5)/v310(1=6)/v311(1=7)/v312(1=8)/v314(1=0).
value labels v305 1 'Daily Express'
              2 'Daily Mail'
              3 'Daily Mirror'
              4 'Morning Star'
              5 'Sun'
              6 'Daily Telegraph'
              7 'Times'
              8 'Guardian'
              0 'None'.
mult response /group = Dailies 'Daily newspapers read'
                (v305 to v314 (0,8))
                /freq dailies.
```

produces exactly the same table and so the following is needed....

```
do repeat
  x1=v305 to v314
  /x2=m305 to m312 m314.
compute x2 = x1.
end repeat.
recode m305 (1=1) /m306(1=2) /m307(1=3) /m308(1=4) /m309(1=5) /m310(1=6)
       /m311(1=7) /m312(1=8) /m314(1=0).
missing values m305 to m314 (0).
if v314=1 m314=9.
value labels m305 1 'Daily Express'
              2 'Daily Mail'
              3 'Daily Mirror'
              4 'Morning Star'
              5 'Sun'
```

- 6 'Daily Telegraph'
- 7 'Times'
- 8 'Guardian'
- 9 'None'.

mult response /group = Dailies 'Daily newspapers read'  
(m305 to m314 (0,9)).

Group DAILIES Daily newspapers read

Category label	Code	Count	Pct of Responses	Pct of Cases
Daily Express	1	29	10.2	13.4
Daily Mail	2	23	8.1	10.6
Daily Mirror	3	5	1.8	2.3
Morning Star	4	1	.4	.5
Sun	5	1	.4	.5
Daily Telegraph	6	55	19.4	25.5
Times	7	86	30.3	39.8
Guardian	8	46	16.2	21.3
None	9	38	13.4	17.6
		-----	-----	-----
	Total responses	284	100.0	131.5

0 missing cases; 216 valid cases

The scales at the end need to be adjusted to give a true zero point, by subtracting the number of items in the scale from the score.

## Quality of Life Survey (Urban Britain) 1973 (processed Jan – Feb 2004)

Real problems reading data. Alpha data included ‘/’ characters, but not reported in error or processing messages. After several attempts and getting blank saved file, realised what was happening and converted all ‘/’ to ‘£’ in raw data. This worked.

File restored in 3 stages so far (easier to keep control)

- 1 Read in alpha data from cards 1-5
- 2 Convert alpha to numeric
- 3 Further changes with compute and recode

Major problem with repeated shut-down of SPSS. After a couple of hours, tracked this down to a recode list with two variable names separated from their labels by a hyphen, not a space or comma. SPSS should surely have picked this up? Replacing hyphens with spaces solved the problem.

Next stage is to add data from cards 6-9.

Lot of fannying about, but got it done eventually. SPSS makes a new file when using data list, so can't use it to amend existing file. Is there an ADD DATA LIST command? All saved on QL3UK

24 Jan

Construct single setup file from several piecemeal sequential setup files.

Found quite a lot of ‘.’ characters in labels, especially ‘Q. etc....’ which have now been eliminated. Some data corrections to var456 (all coded 33 but needed changing) have been entered manually into the data editor as **seqnum** no longer available as a keyword. Fortunately the SPSS line numbers are the same as the serial numbers.

Labels needed for VAR743 to VAR753

Load of vars called RECddd etc., but they are not in the user manual. May be stuff used for **Norman Perry\***, but there are no recodes with them, so ??? Finished up with double the number of cases, so start over!! All alphas recoded to numeric, alphas deleted and vars put in questionnaire order as per manual.

Think it's all sorted now. Also put derived vars on file, but these aren't in the manual, so must decide what to do with them. This has taken all day on Sat.

15 Feb. 04 File has all the original derived variables in at the end. REC864 is not a duplicate of var864, it's a recode to take account of no local paper on var862.

File had **sexkid1** to **agekid8**, but have renamed them as per manual as **var916** etc. Can't think why these were spread out with spaces between or started in col 16. Added labels for health symptoms **var743** to **var753**

All variables in file now labelled.

Current file has JFH's working derived variables, but perhaps for general release these should be in a separate file or at least signposted for users. They're much more convenient to use, especially when using the varxxx to varyyy convention.

E3 needs to be recoded and labelled for leisure wants etc. **Done**

Get from var406 ff

P50 E1 code 3 should be 291 not 191

Latest file is **e:qluk73jfh.por** or **....\ql3\qluk1973-2.sav**

Must sort out E3 as it's too complicated for students. Var347 ff **Done**

P 20 var369 to var369 should be var347 to var369

Tried this: totals tally for codes 2–5, but not for 1. Why? Ditto for “want to do more often”. Codes 1 and 2 tally, but nothing else. Looks like complex conditional transformations needed. Something wrong here anyway as Yes totals are sometimes lower than the follow-up totals. Think the layout on p20 is misleading: the Yes goes with E3c not E3b so Yes to E3b is the **sum** of Yes, No DK, so the IF clauses need to be done **before** the recodes to condense the time spent codes. Got it down to a few cases, and the totals tally if 98\* is included. Need to split this off now. So far, so good. Got it! It was original ‘/’ in data, but needed changing to ‘£’ then pick up ‘£’ in recodes. This involved reading in raw data for cols 347 to 369 in alpha format then running three separate recode commands to generate three sets of variables for qq E3a-c. This is probably too big to put in basic public version so had better be a supplementary file (or setup file) Setup file is **E3sort.sps**, data file is **E3sort.sav** and frequency check output is **E3freq.spo**. This file has been merged with the main file, and the intermediate alpha variables **ar347** to **ar369** stripped out.

Labels missing on anomy and sdscale items; these are now added. SD scale items not reversed on raw data, but have in the **.sav** file. The manual is confusing (p39) as the frequencies are correct, but the labels need switching or vice versa. Would it be better to have 2 files, one as per manual and the other as a supplement? **Some missing values are 10 and 55; odd, but left them as they match manual. Same argument for var476 where 0,10,55,1 need recoding to 1,2,3,4 as they're not even in order!! Done this.**

Two variables **workstat** and **occstat** should be the same, but they aren't. The labels on output for **workstat** don't match the ones in the data file either!! Kept both for now.

Check coding at g6b: should 98 be 1? Coding for H5 doesn't match manual. Ditto J7. System file is all binary. Decide what to do, but it will mean changing all the labels or having a special label for binary and using recode. Ditto newspapers at Q.L

## Quality of Life in Britain 1975 (processed Feb 2004)

Got most of this done, but problems with labels (v363) so check against manual. **Sorted**

Hopefully sorted out. File stored as QL4UK6. Need to find codes for VAR363.

Whole stack of value labels misplaced: must start again. Stuff on consumer goods seems to have got on to all the 0-10 scales. **Got rid of them**, but now have to find correct value labels. This caused serious problems, but got round it by specifying labels for all of these as ( ' ') which SPSS reported as an error, but it worked!

Some missing values not declared.

Some odd values in some vars.

var244

Value labels needed for:

var150 var244

Whole string of variables disappeared VAR308ff. Recreated them with a data list and saved the whole thing as **ql4uk7.sav**.

Why won't SPSS let me start over from the original data list? It looks as if it's working, but doesn't actually read the data in when it's doing **begin data** Think this is because I should have done

File...

New...

Data..

No derived variables in this data set, but there were some in the PNL version, and I'm sure the instructions for these are in the user manual. Can't find my QL4 user manual for now (unless it's in the pdf files) , but have found questionnaire, show cards, interviewing and coding instructions.

**Found it now**

**There's some really fantastic stuff in here**, especially given the history of the last 30 years. Pity little of it ever got reported, but we were in the middle of being closed down and made redundant. It would be wonderful to repeat some of the questions today.

**Some labels in here are misleading and should be changed.**  
(eg on var722 pets in house) see **petcheck** runs:

Need to do something with var150 2-digit codes for single change most wanted to house: can be grouped by first digit into smaller generic codes. Value labels for var244 var450.

**Ditto for var634 to var640 (too long: leave alone)**

Latest file **ql4uk8.sav**

Derived variables pp 56 ff Better to use **compute** than **count** because of missing values? This has been done on this file, or missing values have been accounted for in conjunction with **count**.

**Recorded** 10=11 and 0=10 for var707 to var720 to yield more logical sequence for tabulation.

Got catastrophic error in SPSS whilst exporting file to disk. Can't reproduce it, but I think it was to do with overlapping names in either value labels or missing values lists.

```
_ASSERT(qvalid) failed in svqfil
```

```
>Error # 91
```

```
>An SPSS program error has occurred: Programmer's assertion failed. Please
```

```
>note the circumstances under which this error occurred, attempting to
```

```
>replicate it if possible, and then notify SPSS Technical Support.
```

```
>This is an error from which SPSS cannot recover.
```

```
>The SPSS run will terminate now.
```

```
export out 'e:qluk1975.por'
```

```
  /keep serial to var964 symptoms limit anxiety to trust
```

```
  affgen constr noise nuisance .
```

Error in data file on **var513**: need to swap 1 and 0 over. May mean **TRUST** not right either. **Done**, also **trust** recalculated and new **sav** file saved.

### Quality of Life: Sunderland 1973 (processed April 2004)

17 April 2004-04-17

basic data file created. Check ql3gb files and run some, but some odd recodes (eg var114 var115 1=4 makes spouse = child!)

18 April

Results all wrong when using national setup file. Checked data supplied and found only 8 cards per case, so data for sex and age of children may be lost. Preliminary checks on frequencies seem OK. Got most of this up, but still some missing values and var and value labels to add.

Latest file is **sund1check.sav**

### Quality of Life: Stoke (processed June 2004)

14 June 2004

First shot at creating stoke file using copies of Sunderland setup. Something odd about var372 as recoded once  $GT30 = var372 - 20$ , but another setup has GT9 ditto. There are cases with value 79 which must originally have been 99, therefore missing. Think I'm right, but will now have to go back to raw data to unscramble the 99's from the 0's! Created file **var372.sav** to merge.

**Done.** Current saved file is **stoke1.sav**

Still got to split leisure items as per QL3GB. Check QL3GB log for this: may need to change '/' to '£' in raw data. There's at least one full stop in there as well!