2001 British Social Attitudes Survey

NOTE FOR USERS
(December 2002)

This note provides information in brief about the survey. It accompanies the final version of the main datafile (bsa01.por). For further details about the survey, see Thomson, K. *et al* (forthcoming), *British Social Attitudes Survey 2001: Technical Report*, London: National Centre for Social Research.

About the survey

The BSA survey was conducted by the *National Centre for Social Research* (NatCen). Its core-funding, provided by the Gatsby Charitable Foundation, which is one of the Sainsbury Family Charitable Trusts, and this was supplemented by grants from the Economic and Social Research Council (L219 25 2018, R000 239 188, R000 239 295, M543 285 002) and the Hera Trust. Various Government departments also supported modules in the 2000 survey: the Department for Education and Employment; the Department of the Environment, Transport and the Regions; the Department of Health; the Department of Social Security; and the Health and Safety Executive.

The survey was designed to yield a representative sample of the population in Britain aged 18+. The sample of addresses was drawn from the Postcode Address File. At each address, the interviewer established how many occupied dwelling units it contained. If there were several, one was selected at random for interview (using a Kish grid and random numbers). The interviewer then established how many adults aged 18+ lived in the (selected) dwelling unit. If there were several, one adult was selected (using a similar procedure as that used for dwelling units). The unequal selection probabilities arising from these procedures are taken into account by the weighting.

The fieldwork was conducted by NatCen. Interviews were conducted in the respondent's home, using a laptop computer. In order to increase the number of topics covered by the survey each year three versions of the questionnaire are fielded, and respondents are randomly assigned to one of the versions. All respondents answer a core set of demographic and other classificatory questions and individual modules are then carried on either one, two or all three versions. In 2001, the face-to-face interview was designed to last about one hour and was then followed by a self-completion questionnaire.

Fieldwork was carried out between June and September 2001, with a small number of interviews taking place in October and November. A summary of the response is as follows:

Issued addresses	6,200	
Of which in scope[1]	5,577	
Productive interviews	3,287	(59%)

Version C of the self-completion questionnaire included a module of questions about family and friends which were fielded as part of the *International Social Survey Programme*, of which the BSA series is a member.

---

[1] I.e. traceable, residential and occupied.

The data file should be used in conjunction with the following documentation:
- Outline of the questionnaire
- Documentation of the Blaise questionnaire program (final version dated January 2003)
- Showcards (one set per questionnaire version)
- Self-completion questionnaire (one per questionnaire version)
- Address Record Forms

Weighting

The main dataset (in common with all surveys based on samples from the Postcode Address File) must be weighted to take account of differing selection probabilities. Simplifying slightly: households are selected with equal probability, but only one person in each household is interviewed. People in small households therefore have a higher probability of selection than people in large households and the weighting corrects for this.

Please note that the data must be weighted in all analysis. The file is **not** preweighted. Before running any analysis, please use the following SPSS command:

> weight by wtfactor

(or similar, depending on the exact syntax of your version of SPSS).

Note about [Siblings] (Q1 on version C of the self-completion)

There is a problem with the data in the variable [Siblings] which can't be fully corrected. The problem arose from the layout of the questionnaire, which caused some people to enter the number of siblings incorrectly. The variable has been recoded to minimise the problem. It is, however, still likely to be the case that the number of people with 1-4 siblings is an underestimate and the number of people with 5+ siblings is an overestimate. If you are planning to use this variable, please contact NatCen to discuss the implications.

Socio-economic classifications

With the 2001 census, National Statistics have switched from SOC90 to SOC2000 for the coding of occupations. At the same time, they switched from the Social Class and Socio-Economic Group classifications to the new National Statistics Socio-Economic Classification (NS-SEC). The file contains the following variables based on the new classification:

	Respondent	Spouse/partner (if working and R not working)
SOC2000	RSOC2000	PSOC2000
NS-SEC (full)	RNSSEC	PNSSEC
NS-SEC operational categories	ROpCat	POpCat
NS-SEC analytic classes	RClass	PClass

Further information about these new classifications is available on the National Statistics web site:

> http://www.statistics.gov.uk/methods_quality/ns_sec/

It is our advice that the new classifications should be used whenever possible. However, there are some time-series analysis where the old classifications may be needed, for example, analysis of changes in the role of class over time. For this purpose we conducted a coding experiment on the BSA 2001 survey. This is described in detail in a separate note (*Examining the reliability of SOC90 and related socio-economic*

classifications after the switch to SOC2000 by Sonia Exley and Katarina Thomson). A separate date file (bsa01soc.por) is being deposited with all the variables related to this experiment, but a set of 'best estimates' have also been included on the main datafile:

	Respondent	**Spouse/partner** **(if working and R not working)**
SOC90	RNSOC90	PNSOC90
Socio-Economic Group	RNSEG	PNSEG
Socio-Economic Group compressed	RNSEGGrp	PNSEGGrp
Registrar General's Social Class	RNSocCl	PNSocCl
Goldthorpe scale	RNGH	PNGH
Goldthorpe scale compressed	RNGHGrp	PNGHGrp

The datafile does not include all the various summary versions of these classifications included on BSA in previous years. However, appendix 3 to the note by Exley and Thomson explains how these may be derived.

Publication of the survey

The results of the survey are published in: Park, A., Curtice, J., Thomson, K., Jarvis, L. and Bromley, C. (eds.) (2002) *British Social Attitudes: the 19th Report*, London: Sage.

Further information

For further information, please contact:

Katarina Thomson
National Centre for Social Research
35 Northampton Square
London EC1V 0AX
tel: 020 7549 9570
fax: 020 7250 1524
e-mail: k.thomson@natcen.ac.uk

EXAMINING THE RELIABILITY OF SOC90 AND RELATED SOCIO-ECONOMIC CLASSIFICATIONS AFTER THE SWITCH TO SOC2000

by Sonia Exley and Katarina Thomson, National Centre for Social Research

1.   Background

Social class has long been a major division within British society. Perhaps it is no longer true to say that "class is the basis of British party politics; all else is embellishment and detail" as Pulzer did in 1967[1], but it is nevertheless a major feature of many social science analyses – if only to show that its influence has declined.

The way that class is normally coded on academic and government surveys in Britain is to:
- collect job title and other details of the job,
- code the job to a 'long' list of codes,
- add certain extra information like status in employment, supervisory status and size of enterprise and derive one of the social class classifications via some kind of conversion matrix.

National Statistics (and its predecessors) is the originator of most of these classifications. From the 1991 census onwards, this 'long' list of codes was called SOC90 and from this was derived two social classifications – Socio-Economic Group (SEG) and Registrar General's Social Class. These were much older than 1991 and when SOC90 was introduced, it was mapped onto them. However, there was a consistent complaint from the academic community that these classifications were not sufficiently 'theory driven'. An alternative classification is the Goldthorpe class schema (which exists with various variations). The full versions of these classifications are shown in table 1. All of these socio-economic classifications also exist in compressed versions.

---

[1] Pulzer, P.G.J. (1967), *Political respresentation and elections*, London, p98.

Table 1   Socio-Economic Classifications (full versions)

Socio-Economic Group (SEG)	Registrar-General's Social Class	Golthorpe scale (Goldthorpe-Heath version)	National Statistics Socio-Economic Classification (NS-SEC)
Employer - large org	I (SC=1)	Service class, higher grade	Employers in large org
Manager - large org	II (SC=2)	Service class, lower grade	Higher managerial occup
Employer - small org	III (non-manual) (SC=3)	Routine non-manual employees	Higher professional occup: 'traditional' employees
Manager - small org	III (manual) (SC=4)	Personal service workers	Higher professional occup: 'new' employees
Professional worker - self-employed	IV (SC=5)	Small proprietors with employees	Higher professsional occup: 'traditional' self-employed
Professional worker - employee	V (SC=6)	Small proprietors without employees	Higher professional occup: 'new' self-employed
Intermediate non-manual - ancillary	Armed Forces	Farmers & smallholders	Lower professions & higher technical occups: 'traditional' employees
Intermediate non-manual - supervisor		Foremen & technicians	Lower professions & higher technical occups: 'new' employees
Junior non-manual		Skilled manual workers	Lower professions & higher technical occups: 'traditional' self-employed
Personal service		Semi- & unskilled manual workers	Lower professions & higher technical occups: 'new' self-employed
Foreman/supervisor - manual		Agricultural workers	Lower managerial occup
Skilled manual			Higher supervisory occup
Semi-skilled manual			Intermediate occup: clerical & administrative
Unskilled manual			Intermediate occup: sales & services
Own account worker (not professional)			Intermediate occup: technical & auxiliary
Farmer - employer/ manager			Intermediate occup: engineering
Farmer - own account			Employers in small org: non-professional
Agricultural worker			Employers in small org: agriculture
Member of the armed forces			Own account workers: non-professional
			Own account workers: agriculture
			Lower supervisory occup
			Lower technical occup: craft
			Lower technical occup: process operative
			Semi-routine occup: sales
			Semi-routine occup: service
			Semi-routine occup: technical
			Semi-routine occup: operative
			Semi-routine occup: agriculture
			Semi-routine occup: clerical
			Semi-routine occup: childcare
			Routine occup: sales & service
			Routine occup: production
			Routine occup: technical
			Routine occup: operative
			Routine occup: agriculture

Before 2001, SEG, Social Class and Goldthorpe used to be derived as follows at the National Centre for Social Research (NatCen):
- Interviewers collected job title and other details of the job.
- This was coded manually to SOC90 from paper indices.

- We then used a computer look-up program called the 'census matrix' to derive SEG and Social Class, using SOC90 and other information such as supervisory status and size of establishment.
- At some point, the ability to derive Goldthorpe had been added to this program.

See figure 1.

There are several different versions of this census matrix. There is a published paper version in the SOC90 manual, but this has rather a lot of 'holes' in it, i.e. combinations of SOC90 and supervisory status, status in employment etc that are not allowed – e.g. self-employed postman. In the nature of things, some of these 'ineligible' combinations do come up in the data, either because of some error in the data or because the real world is more complicated than National Statistics allowed for. When this happens, the case 'fails' the census matrix and the socio-economic classifications are left with an 'uncodable' value. The computer versions of the census matrix used by NatCen have been supplied by National Statistics in response to various requests and are more 'fully stuffed', i.e. have less holes, but there are still some combinations that will fail. We shall return to this point later in the paper.

Figure 1   Derivation of socio-economic classifications from SOC90
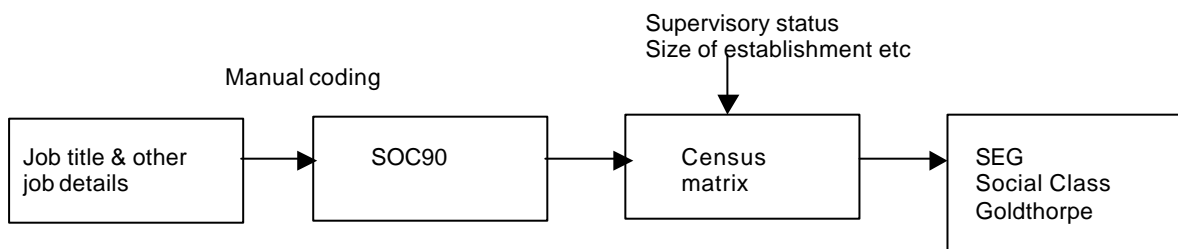
However, with the advent of 2001 census, National Statistics replaced SOC90 with SOC2000 and at the same time they discontinued Socio-Economic Group and Registrar General's Social Class – replacing them with the new National Statistics Socio-Economic Classification (NS-SEC) (see table 1). NS-SEC has some features in common with SEG, but not enough for there to be an obvious map from one to the other.

As for the Goldthorpe scale and the debate about the old classifications not being sufficiently 'theory-driven', Goldthorpe himself had some input into the development of NS-SEC. He has apparently pronounced himself satisfied with the result and is not intending to produce a map from SOC2000 to the Goldthorpe schema.

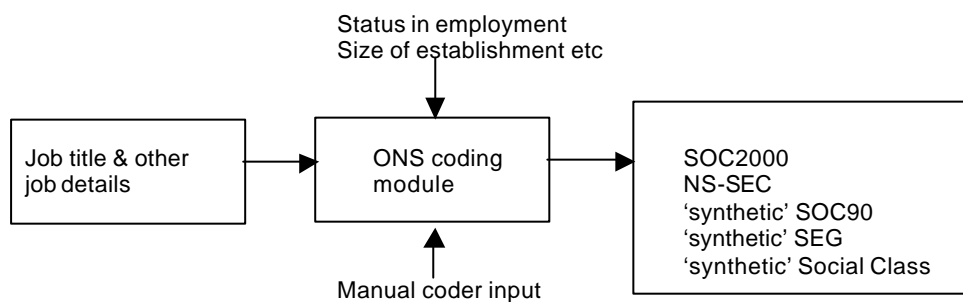So – this is all well and good on a new survey – you simply use NS-SEC from now on. But there is a problem for time-series and repeat surveys like British Social Attitudes (BSA). What if you want to do an analysis of the role of class over time? This may actually be quite important, for example, if you want to show that the role of class is declining to be replaced by education and income, as some people want to do.

One option would be to continue to code such surveys to SOC90. But:

- You would have to code to SOC90 *as well as* SOC2000 because some people who are not looking at class over time will undoubtedly want to use NS-SEC.
- This is expensive.
- The coder expertise in coding to SOC90 will gradually be lost as most surveys move over to SOC2000.
- And – SOC90 will not be updated so new job titles will become uncodable.

However, all is not lost. The way that SOC2000 is coded at NatCen is using a Blaise module called the 'ONS coding module', supplied by National Statistics. The process is set out in figure 2. National Statistics have very helpfully set this coding module up so that it also produces 'best estimates' of SOC90, SEG and Social Class – what we shall in this paper call 'synthetic' SOC90, SEG and Social Class.

Figure 2   Derivation of Socio-economic classifications from SOC2000

Status in employment
Size of establishment etc

| Job title & other job details | → | ONS coding module | → | SOC2000 NS-SEC 'synthetic' SOC90 'synthetic' SEG 'synthetic' Social Class |

Manual coder input

However, there are still two problems:
- The ONS coding module does not produce the Goldthorpe schema.
- We had some doubts about whether these synthetic SOC90, SEG and Social Class would be consistent with our earlier derivations.

2.   The BSA 2001 experiment

NatCen therefore decided to do an experiment on BSA 2001 with the help of money from the Centre for Research into Elections and Social Trends (CREST).

On BSA 2001 we went over to coding SOC2000 using the ONS coding module, but on one third of the sample (version C), we also coded SOC90 in the traditional way (using manual coding). We therefore end up with two SOC90 codes: the manually-coded one, which we shall call the traditional one in this paper, and the synthetic one.

For SEG and Social Class we have the two analogous values: manually coded SOC90 put through the census module (in the traditional way) and the synthetic versions produced

by the ONS coding module. But we also have a third, hybrid version: we took the SOC90 produced by the ONS coding module and put it through the traditional census matrix.

We did this because there are two possible links where discontinuities can enter the process:
1. the coding or mapping of SOC90
2. the derivation of SEG and Social Class from SOC90 via either the census matrix or the ONS coding module

Hypotheses:
- If there are no discontinuities to worry about, then the traditional, synthetic and hybrid versions of SEG and Social Class should be rather similar.
- If discontinuities arise primarily through the method of coding SOC90, we would expect the synthetic and hybrid SEG and Social Class to be rather similar to each other, and different from the traditional version.
- If discontinuities arise primarily through the derivation of SEG and Social Class, then we would expect the traditional and hybrid versions to be rather similar, and different from the synthetic version.
- If discontinuities arise both through the coding of SOC90 and the derivation of SEG and Social Class, we would expect all three versions to be rather different from each other.

We have spoken in terms of discontinuities. It is important to note that we are not really interested in *accuracy* here. It may be, for example, that the synthetic versions are more accurate in some deep philosophical sense than the traditional or hybrid versions, but our concern is to maintain consistency over a time-series. So accuracy doesn't really enter into it. If the traditional method was in some way flawed, we might still wish to carry on replicating that same flaw.

Another point to note is that because the ONS coding module does not supply Goldthorpe, we have only two measures for this: traditional and hybrid.

3. **Comparison of traditional, synthetic and hybrid measures**

The third of the BSA sample subjected to the double coding contained 1099 cases. Of these, 22 had never had a job and are excluded from this analysis. The rest of this paper is based on the remaining 1077 cases. All analysis in this paper is based on unweighted data as it is not the substantive results that are of interest. These variables are not included on the main BSA 2001 file deposited at the Data Archive, but there is a supplementary file (called bsa01soc) which was deposited at the same time. Appendix 2 shows all the variable names. Earlier years of BSA data have included a number of summary versions of these variables. These are not all included in the deposited datafiles, but Appendix 3 shows how they may be derived.

SOC90

SOC level analysis are rarely done on BSA and the interest in SOC is primarily in its input into the socio-economic classifications. Nevertheless, as a baseline we started by comparing our two versions of SOC90: traditional SOC90 and its synthetic counterpart. As can be seen in table 2, there were 301 cases (28 per cent) where the variables produced different values.

Table 2   Traditional SOC90 compared with synthetic SOC90

	N	%
Traditional SOC90 = synthetic SOC90	775	72
Traditional and synthetic SOC90 both uncodable	1	*
Traditional SOC90 <> synthetic SOC90	301	28
Total	1077	100

This would appear to be quite a large discrepancy, but whether it matters or not depends a lot on *how* the two SOC90 measures differ. If it is the case that discrepancies arise mainly from codings to nearby SOC90 categories, then a lot of the differences will disappear when more summary classifications are used. If, however, cases are mainly coded to vastly different codes, then the discrepancies here will remain when the socio-economic classifications are derived.

Socio-Economic Group (SEG)

The second stage of the analysis was therefore to look at the more substantively interesting socio-economic classifications derived from SOC90, starting with SEG. This analysis was done using the most compressed (7 category) version of SEG as this is the one mostly used in social science analysis. As described above, we had three versions of SEG:
- 'traditional' SEG: derived from manually coded SOC90 using the census matrix, and here treated as the 'gold standard'
- 'synthetic' SEG: produced directly by the ONS coding module
- 'hybrid' SEG: synthetic SOC 90 put through the census matrix

The first point to note is that there were many more unclassifiable cases in the manually coded version of SEG than in the synthetic version. These unclassifiable cases include cases where data was deemed to be inadequately described, not stated or missing. As mentioned earlier, in the case of traditional and hybrid SEG they also include cases where the particular combination of SOC90 with supervisory status, status in employment etc 'failed' the census matrix. The synthetic version only had 4 unclassifiable cases, whereas the hybrid version of SEG derived from synthetic SOC90 via the census matrix had the largest number of unclassifiable cases – 10% of all cases.

The traditional SEG also had a relatively large number of unclassifiable cases – 8%. This compares rather unfavourably with previous years – it was 1% in 2000. It is not entirely

clear why this should be, but could be a symptom of the loss of expertise in coding SOC90 among the coders, as most surveys had already moved onto SOC2000 already at the time when this coding was done.

Table 3   Unclassifiable cases in the different versions of SEG

	Traditional SEG from manually coded SOC90 and census matrix		Synthetic SEG direct from ONS coding module		Hybrid SEG from synthetic SOC90 and census matrix	
	N	**%**	**N**	**%**	**N**	**%**
Classifiable cases	994	92	1073	100	966	90
Unclassifiable cases	83	8	4	*	111	10
Total	1077	100	1077	100	1077	100

We then looked at whether the substantive codes arrived at differed between the versions of SEG. First, traditional SEG derived from manual SOC90 and the census matrix was compared with synthetic SEG directly from the ONS coding module. As can be seen in table 4, there were 252 cases where the variables took differing values (23 per cent of cases). This proportion is slightly lower than that for the straightforward comparison between the two versions of SOC 90, simply by virtue of being a grouped variable with a smaller number of possible values. Nevertheless, it is still a large proportion.

Table 4   Traditional SEG compared with synthetic SEG (compressed)

	N	**%**
Traditional SEG = synthetic SEG	823	76
Traditional and synthetic SEG both unclassifiable	2	*
Traditional SEG <> synthetic SEG	252	23
Total	1077	100

Second, the hybrid version of SEG derived from synthetic SOC90 via the census matrix was compared with 'gold standard' traditional SEG from manually coded SOC90 and the census matrix. We already know that this version had a lot more unclassifiable cases than either traditional SEG or the straightforward synthetic version, but how did it fare in terms of accuracy? Table 5 shows that this version proved to be more accurate than synthetic SEG taken directly from the ONS coding module. There were only 149 cases (14 per cent of cases) where the variables showed differing values.

7

Table 5   Traditional SEG compared with hybrid SEG (compressed)

	N	%
Traditional SEG = hybrid SEG	867	81
Traditional and hybrid SEG both unclassifiable	61	6
Traditional SEG <> hybrid SEG	149	14
Total	1077	100

Thus, the findings above paint a clear picture. Although the hybrid version of SEG derived via synthetic SOC90 and the census matrix has many more missing cases, it is clearly more 'accurate' in comparison with the 'gold standard' of the traditional SEG than the version that comes directly out of the ONS coding module.

Registrar General's Social Class

With Registrar General's social class, we have the same three versions as for SEG:
- 'traditional' Social Class: derived from manually coded SOC90 using the census matrix, and here treated as the 'gold standard'
- 'synthetic' Social Class: produced directly by the ONS coding module
- 'hybrid' Social Class: synthetic SOC90 put through the census matrix

In table 6 we see the same pattern in terms of unclassifiable cases as was found with the different versions of SEG. The synthetic version of Social Class direct from the coding module has substantially lower numbers of unclassifiable cases than the manually coded version, and the hybrid version has the highest numbers of unclassifiable cases. Again the number of unclassifiable cases on the traditional version of Social Class compares unfavourably with previous years (it was 1% in 2000).

Table 6   Unclassifiable cases in the different versions of Social Class

	Traditional Social Class from manually coded SOC90 and census matrix (RSOCCLA2)		Synthetic Social Class direct from ONS coding module (RSCSyn)		Hybrid Social Class from synthetic SOC90 and census matrix (RSCONS2)	
	N	%	N	%	N	%
Classifiable cases	994	92	1073	100	966	90
Unclassifiable cases	83	8	4	*	111	10
Total	1077	100	1077	100	1077	100

If the pattern found for SEG holds true, we would expect to find that the hybrid version is more accurate in terms of comparability with the traditional version than the synthetic version. Table 7 confirms that this is the case. 14 per cent of cases had different values

for the traditional Social Class and hybrid Social Class, whereas the corresponding figure where the straightforward synthetic version was compared with the traditional version was 21 per cent.

Table 7   **Traditional Social Class compared with synthetic and hybrid Social Class**

	N	%
Traditional Social Class = synthetic Social Class	844	78
Traditional and synthetic Social Class both unclassifiable	2	*
Traditional Social Class <> synthetic Social Class	231	21
Total	1077	100
	N	**%**
Traditional Social Class = hybrid Social Class	862	80
Traditional and hybrid Social Class both unclassifiable	61	6
Traditional Social Class <> hybrid Social Class	154	14
Total	1077	100

The Goldthorpe-Heath Class schema

As mentioned earlier, the Goldthorpe class schema is available only via the census matrix, i.e. it is not one of the measures derived synthetically as part of the ONS coding module. Thus, the only comparison here is between the traditional version derived from manual SOC90 coding and the census matrix, and the version derived from synthetic SOC90 plus the census matrix. For the purposes of this analysis we are using the compressed (5 category) version of the Goldthorpe schema, as amended by Anthony Heath. (This is the version normally used in analysis of BSA data).

The first point to note is that table 8 shows a substantial number of unclassifiable cases for both variables, highlighting aforementioned difficulties with the census matrix, particularly in relation to synthetic SOC90 data. Again, for reasons that are not entirely clear, the figure for the traditional method compares unfavourably with previous years. (There were 1% missing cases in 2000).

Table 8   **Unclassifiable cases in the different versions of the Goldthorpe-Heath class schema**

	Traditional Goldthorpe-Heath from manually coded SOC90 and census matrix		Hybrid Goldthorpe-Heath from synthetic SOC90 and census matrix	
	N	**%**	**N**	**%**
Classifiable cases	995	92	966	90
Unclassifiable cases	82	8	111	10
Total	1077	100	1077	100

Next, we examined the proportion of cases where the two different versions of the Goldthorpe-Heath scale differed. As can be seen from table 9, the two measures took different values in 11 per cent of cases, reinforcing the emerging pattern that variables derived via the census matrix are more accurate in terms of comparability with manually coded 'gold standard' measures, but suffer the problem of having larger numbers of unclassifiable cases.

Table 9   Traditional Goldthorpe-Heath compared with hybrid Goldthorpe-Heath (compressed)

	N	%
Traditional Goldthorpe = hybrid Goldthorpe	902	84
Traditional and hybrid Goldthorpe both unclassifiable	61	6
Traditional Goldthorpe <> hybrid Goldthorpe	114	11
Total	1077	100

4.   What do the differences between the versions actually mean

From the above analysis we can see that there are differences in certain individuals' occupational classifications according to whether coding has been manually or synthetically generated. How does this affect the way in which class composition in Britain is represented in terms of actual numbers falling into different occupational categories?

Socio-Economic Group (SEG)

Table 10 shows frequencies for different versions of compressed Socio-Economic Group alongside each other. Here it can be seen that, in most instances, proportions are roughly the same for the different versions of SEG. In particular, the traditional and hybrid versions of SEG are almost identical. However, there are some discrepancies with the synthetic SEG. In particular, synthetic SEG shows higher levels of 'intermediate non-manual workers'.

Table 10   **Frequencies for different versions of Socio-economic Group**

	Traditional SEG			Synthetic SEG			Hybrid SEG		
	N	% of all	% of valid	N	% of all	% of valid	N	% of all	% of valid
Professional, employer & manager	203	19	20	199	19	19	200	19	21
Intermediate non-manual worker	141	13	14	220	20	21	136	13	14
Junior non-manual worker	227	21	23	219	20	20	220	20	23
Supervisor, skilled manual worker, own account professional	170	16	17	179	17	17	171	16	18
Personal service, semi-skilled manual, agricul. worker	191	18	19	207	19	19	182	17	19
Unskilled manual worker	62	6	6	48	4	4	56	5	6
Member of the armed forces	-	-	-	1	0	0	1	0	9
Inadequately descr'd/Missing	83	8	na	4	0	na	111	10	na
Total	1077	100	100	1077	100	100	1077	100	100

The discrepancy between the traditional version and the synthetic version of SEG are partly to do with those cases which the failed the census matrix but which have been allocated to an SEG by the ONS coding module. Table 11 shows that of the 252 cases where traditional and synthetic SEG differed, 81 (32%) are cases where traditional SEG is unclassifiable and synthetic SEG is not. These are mainly being coded as 'professional, employers and managers' or 'intermediate non-manual' by the ONS coding module.

However, there is one curious finding in that the synthetic version of SEG (i.e. the one that comes straight out of the ONS coding module) appears to have a tendency to code cases as 'intermediate, non-manual worker' category, when the manually coded version has classified them as 'Professional, employer and manager'. This affects some 42 cases (17% of cases where the two SEGs differ).

Table 11   Traditional vs synthetic SEG (N)

Synthetic SEG → Traditional SEG ↓	Profess'l, employer & manager	Inter-mediate non-manual	Junior non-manual worker	Superv, skilled manual, own account profess'l	Personal service, semi-skilled manual, agricul. worker	Unskilled manual worker	Member of the armed forces	Inade-quately descr/ Missing	Total
Professional, employer & manager	149	42	-	5	6	-	-	1	203
Intermediate non-manual worker	11	122	5	-	3	-	-	-	141
Junior non-manual worker	1	13	207	1	4	1	-	-	227
Supervisor, skilled manual, own account profess'l	3	3	1	146	14	2	-	1	170
Personal service, semi-skilled manual, agricul. worker	1	6	4	16	159	5	-	-	191
Unskilled manual worker	-	-	-	4	18	40	-	-	62
Member of the armed forces	-	-	-	-	-	-	-	-	-
Inadequately described/Missing	34	34	2	7	3	-	1	2	83
Total	199	220	219	179	207	48	1	4	1077

Registrar General's Social Class

Similarly for Social Class, we must examine whether or not discrepancies between the different versions make a substantive difference to actual table marginals, i.e. do they affect the numbers falling into different class categories? Again, table 12 shows that we have almost identical distributions for traditional and hybrid Social Class, whereas the synthetic version differs slightly more.

Table 12   Frequencies for different versions of Registrar General's Social Class

	Traditional Social Class			Synthetic Social Class			Hybrid Social Class		
	N	% of all	% of valid	N	% of all	% of valid	N	% of all	% of valid
I	29	3	3	61	6	6	38	4	4
II	270	25	27	296	28	28	254	24	26
III (non-manual)	264	25	27	275	26	26	252	23	26
III (manual)	190	18	19	207	19	19	189	18	20
IV	178	17	18	183	17	17	174	16	18
V	63	6	6	50	5	5	58	5	6
Member of the armed forces	-	-	-	1	0	0	1	0	0
Inadequately described/Missing	83	8	na	4	0	0	111	10	0
Total	1077	100	100	1077	100	100	1077	100	100

When we look at the crosstabulation in table 13, we see that cases that were unclassifiable by the traditional method are largely being put into the higher social classes, particularly II by the ONS coding module.

Table 13   Traditional vs synthetic Social Class (N)

Synthetic Social Class → Traditional Social Class ↓	I	II	III (non-manual)	III (manual)	IV	V	Armed forces	Inadeq descr/ Missing	Total
I	25	4	-	-	-	-	-	-	29
II	11	231	11	13	2	-	-	2	270
III (non-manual)	3	9	242	2	7	1	-	-	264
III (manual)	3	5	3	163	13	3	-	-	190
IV	-	8	6	17	142	5	-	-	178
V	-	-	-	5	17	41	-	-	63
Member of the armed forces	-	-	-	-	-	-	-	-	-
Inadequately described/Missing	19	39	13	7	2	-	1	-	83
Total	61	296	275	207	183	50	1	4	1077

Goldthorpe-Heath class schema

In the case of the Goldthorpe-Heath class schema, we do not, of course, have an equivalent of the synthetic version of SEG. And given the earlier discussion, it is not surprising that the traditional and hybrid versions are very similar.

Table 14   **Frequencies for different versions of Goldthorpe-Heath**

	Traditional Goldthorpe-Heath			Hybrid Goldthorpe		
	N	**% of all**	**% of valid**	**N**	**% of all**	**% of valid**
Salariat	298	28	30	284	26	29
Routine non-manual workers	252	23	25	251	23	26
Petty bourgeoisie	78	7	8	78	7	8
Manual foremen and supervisors	63	6	6	62	6	6
Working class	304	28	31	291	27	30
Inadequately described/Missing	82	8	na	111	10	na
Total	1077	100	100	1077	100	100

The impact on substantive analysis

In order to asses the potential impact of these discontinuities on analysis, we ran a series of logistic regressions, where socio-economic group might be expected to have an impact. The full regression tables are shown in Appendix 1.[2] The independent variables were: socio-economic group, party identification, sex, age, income and highest educational qualification. Table 15 summarises the results, with a + indicating that a category is significantly more likely to score on the dependent variable (than average) and a – that it is significantly less likely.

---

[2] None of these models have left-right values as an independent variable. Normally if we were running these analyses to get substantive results, we might include this to control for the respondent's basic belief structure. However, in practically every case, the left-right scale knocks out socio-economic group altogether – and frequently most of the other variables. Since our interest is in the behaviour of socio-economic group as an independent variable, we therefore present the analyses excluding the left-right scale.

Table 15   Substantive analysis using different versions of SEG

	Conservative support			Labour support			Support for joining the Euro			Agree that benefits for unemployed are too low		
	Traditional SEG	Synthetic SEG	Hybrid SEG	Traditional SEG	Synthetic SEG	Hybrid SEG	Traditional SEG	Synthetic SEG	Hybrid SEG	Traditional SEG	Synthetic SEG	Hybrid SEG
SEG												
Profess/Emp/Manag	+	+	+		−		+		+			
Intermediate non-man												
Junior non-man												
Superv/skilled manual												
Semi-skilled/Personal	−		−									
Unskilled manual		−			+		−		−			
Inadeq/Missing									+			
Party identification												
Conservative	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	−	−	−	−	−	−
Labour	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	+	+	+	+	+	+
Liberal Democrat	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	+	+	+			
Other party	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.						
None	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	−	−	−			
Sex												
Male								+		+	+	+
Female								−		−	−	−
Age												
18-33	−	−	−	+	+	+						
34-49	−	−	−									
50-64	+	+	+	−	−	−				+	+	+
65+	+	+	+							−	−	−
Income												
Less than £10,000										+	+	+
£10,000-£17,999												
£18,000-£31,999												
£32,000+												
Unknown												
Highest educ qualif												
Degree or higher ed							+		+			
A level or equiv												
GCSE or equiv	+	+	+	−	−	−	−	−	−			
Lower than GCSE	−	−	−	+	+	+	−	−	−			

Overall, the models are pretty similar. But where there are differences they are in all cases bar one, that the traditional and hybrid SEG models throw up the same significant variables, while the synthetic version is the one that is different.

5. Reliability flag

The ONS coding module also outputs a reliability flag for SOC90. It takes a value of *zero* where mapping between SOC2000 and the synthetic version of SOC90 is believed to be *reliable*, and a value of *one* where it is believed to be *unreliable*. As seen in table 16, this flag being set to one does indeed indicate that there may be a problem with SOC90. However, it is not immediately obvious to us how this can help.

Table 16   Discrepancies in traditional and synthetic SOC90 by SOC90 reliability flag

| | SOC90 reliability flag | |
| | Reliable | Unreliable |
	%	%
Traditional SOC90 = synthetic SOC90	76	56
Traditional SOC90 <> synthetic SOC90	24	44

6. Conclusions and recommendations

The basic conclusions are, first, that there *are* inconsistencies between the traditional version of SEG and Social Class, derived from manual coding to SOC90 and the census matrix, and the 'synthetic' version of SEG and Social Class produced by the ONS coding module. Although these differences are not very great, they could affect substantive analysis. Second, as traditionally derived SEG and Social Class is rather similar to the measures we get if we run synthetic SOC90 through the census module, we can deduce that the discontinuities are at least in part to do with the varying ways that the census matrix and the ONS coding module go about getting from SOC90 to the socio-economic classifications. For all we know, the ONS coding module may be more 'accurate', but since our interest is in consistency over time-series, we would nevertheless prefer the version which is most similar to the traditionally derived measures. Add to this the fact that the Goldthorpe-Heath scale is not available from the ONS coding module.

In making our recommendations we are assuming that there are various options that are **not** open to us on time-series surveys like BSA:

- To continue manual coding to SOC90, either in parallel with SOC2000 or instead of it – too expensive, coder knowledge of SOC90 will fade, people not analysing class over time are bound to want SOC2000 and NS-SEC.
- To tell users that they can no longer do class comparisons over time – seems a strange thing to do at a time when there is so much interest in the supposed declining influence of social class on attitudes and behaviour.

So our recommendations are:

- Use the ONS coding module to get SOC2000, NS-SEC and SOC90.
- Ignore (for now) the synthetic versions of SEG and Social Class produced by the ONS coding module.
- Run synthetic SOC90 through the census matrix producing the hybrid versions of SEG, Social Class and Goldthorpe. This is the only way of obtaining Goldthorpe.
- However, a problem with the hybrid versions (and also the traditional versions once coders lose expertise in SOC90), is that there are rather a lot of unclassifiable cases. In the case of Goldthorpe, there is nothing we can do about this, but in the case of SEG and Social Class, we can top up the hybrid version with the synthetic version where there would otherwise be missing values. We call this 'new final SEG' and 'new final Social Class'.

We saw earlier that traditional SEG and synthetic SEG differed in 23% of cases and that this fell to 14% when hybrid SEG was compared with traditional SEG. As seen in table 17, if we use 'new final SEG', it rises slightly to 17%, but there are only 4 unclassifiable cases (2 of which are unclassifiable on both measures).

Table 17   Traditional SEG compared with 'new final SEG' (compressed)

	N	%
Traditional SEG = new final SEG	889	83
Traditional and new final SEG both unclassifiable	2	*
Traditional SEG <> new final SEG	186	17
Total	1077	100

There is also another option we have not looked at: National Statistics have supplied us with a conversion table to derive NS-SEC from SOC90, so we could go back and add NS-SEC to all our older datasets. This may be attractive for short time-series, but is unlikely to happen on BSA unless there is a specific demand for it.

Further information

For further information, please contact:

Katarina Thomson
National Centre for Social Research
35 Northampton Square
London EC1V 0AX
tel: 020 7549 9570
e-mail: k.thomson@natcen.ac.uk

Appendix 1: Logistics regressions using different versions of SEG

Logistic regression: Predictors of Conservative Party Support

Model 1 - Independent variables: sex, age, income, education level, traditional SEG.
Model 2 - Independent variables: sex, age, income, education level, synthetic SEG (armed forced and inadequately described/not stated excluded from analysis on account of too few cases).
Model 3 - Independent variables: sex, age, income, education level, hybrid SEG.

Category	Model 1	Model 2	Model 3
	Odds ratio (Exp(B))	Odds ratio (Exp(B))	Odds ratio (Exp(B))
Baseline odds	.317**	.297**	.311**
Socio-economic group (8 cat)			
Professional/employers/managers	1.665**	1.90**	1.744**
Intermediate non-manual	.918	1.092	1.002
Junior non-manual	1.187	1.251	1.196
Supervisor/skilled manual	1.132	1.046	1.088
Semi-skilled/personal services	.577**	.844	.625*
Unskilled manual	.598	.434*	.596
Inadequately described/not stated	1.412	–	1.180
Sex			
Men			
Women			
Age			
18-33	.590**	.573**	.594**
34-49	.773*	.766*	.768*
50-64	1.378*	1.431**	1.387*
65+	1.590**	1.591**	1.582**
Income			
Less than £9,999			
£10,000-£17,999			
£18,000-31,999			
£32,000 and above			
Unknown			
Highest educational qualification			
Degree or other higher education	.800	.798	.781
'A' level or equivalent	1.173	1.215	1.190
'GCSE' level or equivalent	1.649**	1.640**	1.672**
Lower than 'GCSE' level	.646**	.629**	.643**
	1046 cases	1041 cases	1045 cases

18

Logistic regression: Predictors of Labour Party Support

Model 1 - Independent variables: sex, age, income, education level, traditional SEG.
Model 2 - Independent variables: sex, age, income, education level, synthetic SEG (armed forced and inadequately described ⁄ not stated excluded from analysis on account of too few cases).
Model 3 - Independent variables: sex, age, income, education level, hybrid SEG.

Category	Model 1	Model 2	Model 3
	Odds ratio (Exp(B))	Odds ratio (Exp(B))	Odds ratio (Exp(B))
Baseline odds	.696**	.775**	.697**
Socio-economic group (8 cat)			
Professional/employers/managers		.670**	
Intermediate non-manual		.910	
Junior non-manual		.793	
Supervisor/skilled manual		1.000	
Semi-skilled/personal services		1.062	
Unskilled manual		1.948*	
Inadequately described/not stated			
Sex			
Men			
Women			
Age			
18-33	1.376**	1.370**	1.374**
34-49	1.150	1.177	1.154
50-64	.760*	.740*	.759*
65+	.832	.839	.831
Income			
Less than £9,999			
£10,000-£17,999			
£18,000-31,999			
£32,000 and above			
Unknown			
Highest educational qualification			
Degree or other higher education	.863	.929	.867
'A' level or equivalent	.979	1.005	.977
'GCSE' level or equivalent	.756*	.745*	.755*
Lower than 'GCSE' level	1.565**	1.436**	1.563**
	1046 cases	1041 cases	1045 cases

Logistic regression: Support for joining the Euro

Model 1 - Independent variables: sex, age, income, education level, party id, traditional SEG.
Model 2 - Independent variables: sex, age, income, education level, party id, synthetic SEG (armed forced and inadequately described ⁄ not stated excluded from analysis on account of too few cases).
Model 3 - Independent variables: sex, age, income, education level, party id, hybrid SEG derived via the Census matrix.

Category	Model 1	Model 2	Model 3
	Odds ratio (Exp(B))	Odds ratio (Exp(B))	Odds ratio (Exp(B))
Baseline odds	.303**	.355**	.316**
Socio-economic group (8 cat)			
Professional/employers/managers	1.837**		1.603**
Intermediate non-manual	1.203		.952
Junior non-manual	.930		.867
Supervisor/skilled manual	1.282		1.218
Semi-skilled/personal services	.876		.891
Unskilled manual	.404*		.395*
Inadequately described/not stated	1.071		1.764**
Party Identification			
Conservative	.651**	.681*	.647**
Labour	1.882**	1.824**	1.827**
Liberal Democrat	1.508*	1.485*	1.470*
Other party	.911	.905	.967
None	.594*	.600*	.595*
Sex			
Men		1.219**	
Women		.820**	
Age			
18-33			
34-49			
50-64			
65+			
Income			
Less than £9,999			
£10,000-£17,999			
£18,000-31,999			
£32,000 and above			
Unknown			
Highest educational qualification			
Degree or other higher education	1.533**	1.689	1.556**
'A' level or equivalent	1.163	1.202	1.174
'GCSE' level or equivalent	.732*	.722	.724*
Lower than 'GCSE' level	.767*	.682**	.756*
	1012 cases	1008 cases	1011 cases

Logistic regression: Agreement that large numbers of people falsely claim benefits

Model 1 - Independent variables: sex, age, income, education level, party id, traditional SEG.
Model 2 - Independent variables: sex, age, income, education level, party id, synthetic SEG (armed forced and inadequately described ⁄ not stated excluded from analysis on account of too few cases).
Model 3 - Independent variables: sex, age, income, education level, party id, hybrid SEG.

Category	Model 1	Model 2	Model 3
	Odds ratio (Exp(B))	Odds ratio (Exp(B))	Odds ratio (Exp(B))
Baseline odds	4.354**	4.369**	4.349**
Socio-economic group (8 cat)			
Professional/employers/managers			
Intermediate non-manual			
Junior non-manual			
Supervisor/skilled manual			
Semi-skilled/personal services			
Unskilled manual			
Armed forces			
Inadequately described/not stated			
Party Identification			
Conservative			
Labour			
Liberal Democrat			
Other party			
None			
Sex			
Men			
Women			
Age			
18-33			
34-49			
50-64			
65+			
Income			
Less than £9,999			
£10,000-£17,999			
£18,000-31,999			
£32,000 and above			
Unknown			
Highest educational qualification			
Degree or other higher education	.600**	.610**	.598**
'A' level or equivalent	.937	.934	.938
'GCSE' level or equivalent	1.491*	1.478*	1.493*
Lower than 'GCSE' level	1.192	1.188	1.193
	1012 cases	1008 cases	1011 cases

Logistic regression: Agreement that benefits for the unemployed are too low

Model 1 - Independent variables: sex, age, income, education level, party id, traditional SEG.
Model 2 - Independent variables: sex, age, income, education level, party id, synthetic SEG (armed forced and inadequately described ⁄ not stated excluded from analysis on account of too few cases).
Model 3 - Independent variables: sex, age, income, education level, party id, hybrid SEG.

Category	Model 1	Model 2	Model 3
	Odds ratio (Exp(B))	Odds ratio (Exp(B))	Odds ratio (Exp(B))
Baseline odds	.513**	.515**	.514**
Socio-economic group (8 cat)			
Professional/employers/managers			
Intermediate non-manual			
Junior non-manual			
Supervisor/skilled manual			
Semi-skilled/personal services			
Unskilled manual			
Armed forces			
Inadequately described/not stated			
Party Identification			
Conservative	.738*	.742*	.742*
Labour	1.456**	1.437**	1.454**
Liberal Democrat	.780	.793	.779
Other party	1.103	1.099	1.101
None	1.082	1.076	1.081
Sex			
Men	1.189*	1.199**	1.190*
Women	.841*	.834**	.840*
Age			
18-33	1.201	1.206	1.200
34-49	1.130	1.140	1.134
50-64	1.460**	1.443**	1.458**
65+	.505**	.504**	.504**
Income			
Less than £9,999	1.665**	1.675**	1.664**
£10,000-£17,999	1.025	1.026	1.023
£18,000-31,999	.781	.780	.780
£32,000 and above	.865	.857	.868
Unknown	.867	.871	.867
Highest educational qualification			
Degree or other higher education			
'A' level or equivalent			
'GCSE' level or equivalent			
Lower than 'GCSE' level			
	1012 cases	1008 cases	1011 cases

APPENDIX 2a: VARIABLE NAMES ON bsa01soc.por

Only the variables described below as 'new final' are included in the main BSA 2001 file as deposited at the Data Archive. A separate file (bsa01soc.por) contains all the social class variables and may be linked to the main file by means of the serial number. The table below shows the variable names used on this file.

	Respondent	Partner
SOC90:		
Traditional	RSOC[1]	n.a.
Syntethic	RSOC90sy	PSOC90sy
New final SOC	RNSOC90	PNSOC90
SEG:		
Traditional	RSEG2[1]	n.a.
Traditional <grouped>	RSEGGrp2[1]	n.a.
Hybrid	RSEGONS2	PSEGONS2
Hybrid <grouped>	RSEGOGr2	PSEGOGr2
Synthetic	RSEGsyn	PSEGsyn
Synthetic <grouped>	RSEGGrp3	PSEGGrp3
New final SEG	RNSEG	PNSEG
New final SEG <grouped>	RNSEGGrp	PNSEGGrp
Social Class:		
Traditional	RSOCCla2[1]	n.a.
Hybrid	RSCONS2	PSCONS2
Synthetic	RSCsyn	PSCsyn
New final SC	RNSocCl	PNSocCl
Goldthorpe:		
Traditional	RGHClass[1]	n.a.
Tradional <grouped>	RGHGrp[1]	n.a.
Hybrid	RGHONS2	PGHONS2
Hybrid <grouped>	RGHOGr2	PGHOGr2
New final Goldthorpe	RNGH	PNGH
New final Goldthorpe <grouped>	RNGHGrp	PNGHGrp

Notes:
1.   On version C of questionnaire only.

APPENDIX 2b: Note on derivation of 'new final' variables

The table below explains how the 'new final' (best estimate) variables were derived on the BSA 2001 file and how they will be derived in future years (when manually coded SOC90 will not be available):

Variable	Derivation on BSA 2001	Derivation in future years
RNSOC90	Traditional SOC90 where available, otherwise synthetic SOC90.	Synthetic SOC90.
PNSOC90	Synthetic SOC90.	Synthetic SOC90.
RNSEG RNSEGGrp	Traditional SEG where available, otherwise hybrid SEG. Where this fails the census matrix, synthetic SEG.	Hybrid SEG. Where this fails the census matrix, synthetic SEG.
PNSEG PNSEGGrp	Hybrid SEG. Where this fails the census matrix, synthetic SEG.	Hybrid SEG. Where this fails the census matrix, synthetic SEG.
RNSocCl	Traditional Social Class where available, otherwise hybrid Social Class. Where this fails the census matrix, synthetic Social Class.	Hybrid Social Class. Where this fails the census matrix, synthetic Social Class.
PNSocCl	Hybrid Social Class. Where this fails the census matrix, synthetic Social Class.	Hybrid Social Class. Where this fails the census matrix, synthetic Social Class.
RNGH RNGHGrp	Traditional Goldthorpe where available, otherwise hybrid Goldthorpe.	Hybrid Goldthorpe.
PNGH PNGHGrp	Hybrid Goldthorpe.	Hybrid Goldthorpe.

APPENDIX 3: DERIVATION OF SUMMARY VERSIONS OF PRE-2000 SOCIO-ECONOMIC CLASSIFICATIONS

[RManual]
Manual/non-manual status of current/last job (respondent)

		[RManual]
Non-manual		1

SOC90 codes 100-142, 152-440, 450-491, 610-613
640, 643, 651, 700-730, 732-792, 954

PLUS	SOC90 code 614 if REmpStat = 8
	SOC90 code 615 if REmpStat = 8
	SOC90 code 619 if REmpStat = 8
	SOC90 code 731 if REmpStat = 5,6,7,8

Manual 2

SOC90 codes 441, 500-599, 620-631, 641, 642, 644,
650, 652-699, 800-953, 955-996, 999

PLUS	SOC90 code 614 if REmpStat = 9
	SOC90 code 615 if REmpStat = 1,2,3,4,9
	SOC90 code 619 if REmpStat = 1,2,3,4,9
	SOC90 code 731 if REmpStat = 1,2,3,4,9

Armed forces

SOC90 codes 150, 151, 600, 601   8

Unable to classify 9

SOC90 code 614 if REmpStat = 10,11
SOC90 code 615 if REmpStat = 10,11
SOC90 code 619 if REmpStat = 10,11
SOC90 code 731 if REmpStat = 10,11

PLUS SOC90 codes 997,998

SOCIO-ECONOMIC GROUP (SEG)

The full SEG (called [RSEG2] on pre-2001 BSA files) is:

Employer - large organisation	01
Manager - large organisation	02
Employer - small organisation	03
Manager - small organisation	04
Professional worker - self-employed	05
Professional worker - employee	06
Intermediate non-manual worker - ancillary	07
Intermediate non-manual worker - supervisor	08
Junior non-manual worker	09
Personal service worker	10
Foreman/supervisor - manual	11
Skilled manual worker	12
Semi-skilled manual worker	13
Unskilled manual worker	14
Own account worker (not professional)	15
Farmer - employer/manager	16
Farmer - own account	17
Agricultural worker	18
Member of the armed forces	19
Inadequately described/not stated	20

Socio-economic group - grouped (i) - current or last job (of respondent)
Called [RSEG] on pre-2001 BSA files.

	[RSEG2]	[RSEG]
Employer/manager - large organisation	01,02	01
Employer/manager - small organisation	03,04	02
Professional worker - self-employed	05	03
Professional worker - employee	06	04
Intermediate non-manual worker	07,08	05
Junior non-manual worker	09	06
Personal service worker	10	07
Foreman/supervisor - manual	11	08
Skilled manual worker	12	09
Semi-skilled manual worker	13	10
Unskilled manual worker	14	11
Own account worker (not professional)	15	12
Farmer - employer/manager	16	13
Farmer - own account	17	14
Agricultural worker	18	15
Member of the armed forces	19	16
Inadequately described/not stated	20	17

Socio-economic group - grouped (ii) - current or last job (of respondent)
Called RSEGGrp on pre-2001 BSA files.

	[RSEG2]	[RSEGGrp]
Professional	05,06	1
Employers/managers	01-04,16	2
Intermediate (non-manual)	07,08	3
Junior (non-manual)	09	4
Skilled (manual)	11,12,15,17	5
Semi-skilled (manual)	10,13	6
Unskilled (manual)	14,18	7
Other occupation	19	8
Occupation not classifiable	20	9

Socio-economic group - grouped (iii) - current or last job (of respondent)
Called [RSEGGrp2] on pre-2001 BSA files.

	[RSEG2]	[RSEGGrp2]
Professional, employers and managers	01-06,16	1
Intermediate non-manual worker	07,08	2
Junior non-manual worker	09	3
Supervisor, skilled manual worker, own account professional	11,12,15,17	4
Personal service worker, semi-skilled manual worker, agricultural worker	10,13,18	5
Unskilled manual worker	14	6
Member of the armed forces	19	7
Inadequately described/not stated	20	8

REGISTRAR GENERAL'S SOCIAL CLASS

The full Registrar General's Social Class (called [RSSOCCla2] on pre-2001 BSA files) is:

	[RSOCCla2]
I (SC=1)	1
II (SC=2)	2
III (non-manual) (SC=3)	3
III (manual) (SC=4)	4
IV (SC=5)	5
V (SC=6)	6
Armed Forces	7
Insufficient information	8

Registrar General's Social Class - current/last job (of respondent) – grouped (i)

Called [RSOCClas] on pre-2001 BSA files:

	[RSOCCla2]	[RSOCClas]
I (SC=1)	1	1
II (SC=2)	2	2
III	3,4	3
IV (SC=5)	5	4
V (SC=6)	6	5
Armed Forces	7	8
Insufficient information	8	9

Registrar General's Social Class - current/last job (of respondent) – grouped (ii)

Called [RRGClass] on pre-2001 BSA files:

	[RSOCCla2]	[RRGClass]
I	1	1
II	2	2
IIINM	3	3
IIIM	4	4
IV	5	5
V	6	6
Insufficient information (inc Armed Forces)	7,8	9

GOLDTHORPE-HEATH CLASS SCHEME

The full Goldthorpe-Heath scheme (called [RGHClass] on pre-2001 BSA files) is:

		[RGHClass]
I	Service class, higher grade	01
II	Service class, lower grade	02
IIIa	Routine non-manual employees	03
IIIb	Personal service workers	04
IVa	Small proprietors with employees	05
IVb	Small proprietors without employees	06
IVc	Farmers and smallholders	07
V	Foremen and technicians	08
VI	Skilled manual workers	09
VIIa	Semi- and unskilled manual workers	10
VIIb	Agricultural workers	11
	Insufficient information	99

Goldthorpe-Heath class schema - compressed - current/last job (of respondent)

Called [RGHGrp] on pre-2001 BSA files:

	[RGHClass]	[RGHGrp]
Salariat (professional and managerial)	01,02	1
Routine non-manual workers (office and sales)	03,04	2
Petty bourgeoisie (the self-employed incl. farmers, with and without employees)	05,06,07	3
Manual foremen and supervisors	08	4
Working class (skilled, semi-skilled and unskilled manual workers, personal service and agricultural workers)	09,10,11	5
Insufficient information	99	9