

Chapter five

Devising and Applying the Bristol Social Exclusion Matrix (B-SEM)

The Matrix

We defined social exclusion and deep exclusion above in the following terms:

Social exclusion is a complex process operating across several dimensions or domains. It involves both the lack or denial of resources, rights, goods and services, and the inability to participate in the normal relationships and activities, available to the majority of people in a society, whether in economic, social, cultural or political arenas. It affects both the quality of life of individuals and the equity and cohesion of society as a whole.

Deep exclusion refers to exclusion across more than one domain or dimension of disadvantage, resulting in severe negative consequences for quality of life, well-being and future life chances.

These definitions have the advantage of not narrowly limiting the causal processes at work in social exclusion, or the factors that affect life chances. These are open to empirical investigation. The operational issue is then the identification of domains and dimensions. The direct measurement of quality of life and well-being is essential.

The three main areas in which the domains fall are those of *resources*, *participation* and *quality of life*. Of course there is a sense in which all the domains reflect aspects of quality of life. However, in line with the expressed interest of the SEU in direct measures of this, the existence of other government indicator sets (specifically, the Audit Commission's local quality of life indicators and the headline sustainable development indicators), and an expanding literature on themes of well-being and happiness, we have incorporated some topics relating directly to this theme.

Resources

The first Nice criterion refers to 'access by all to resources, rights, goods and services'. This is incorporated into our composite definition of social exclusion. There is an established understanding in the policy literature that 'resources' means far more than income

(Townsend, 1979). It is reflected in traditional attempts to estimate the value of the ‘social wage’ (Sefton, 2002), but also in the bracketing together of resources goods and services. In line with the increasing recognition that family members and social networks are also a resource in both practical and emotional terms, we have divided ‘resources’ into three domains: material/economic resources; access to public and private services; and social resources.

Domain 1: Material/economic resources

The most obvious indicator of material and economic resources is of course income. The misconception that poverty is generally understood simply as a matter of low income arises from the apparent simplicity of using this as an indicator. In fact, the assessment of what constitutes ‘income’, and the collection of reliable data, are far from straightforward. There is, for example, an increasing literature on the importance of assets including, but not confined to, home ownership, which underpinned the establishing of child trust funds (for example, Kelly and Lissauer, 2000; White, 2001; White et al, 2006). Pension contributions may be treated either as assets or simply as an indicator of lowered risk of future poverty, and thus frequently appear as an indicator of social inclusion/exclusion (for example, NPI, *Opportunity for All*, Burchardt, 2000; Burchardt et al, 2002). Debt may be seen as the opposite of financial assets, and debt, problem debt and financial exclusion are also the subjects of an extensive literature (for example, Kempson and Whyley, 1999; SEU, 2004; Kempson et al, 2004; McKay and Collard, 2006). The lack of basic material necessities has also had an increasing profile in the assessment of poverty and social exclusion, most notably in the *Breadline Britain Surveys* (Mack and Lansley, 1985; Gordon and Pantazis, 1997; Gordon et al, 2000; Pantazis et al, 2006) but also in the measures used in government surveys such as FACS. Subjective poverty captures the *sense* of exclusion from ‘normal’ standards of living. The topic areas identified are listed below.

- Income. *The drivers of social exclusion* notes that “households with relatively low incomes were more likely than others to be socially excluded on all dimensions except isolation and lack of support” (Bradshaw et al, 2004, p 14). We have divided this into estimated income and components of income since it may be significant what kind of income data is provided in particular surveys.
- Possession of necessities (necessities differ for children). This is a direct measure of material deprivation, used extensively in the PSE Survey and now incorporated into the official measurement of child poverty, and into survey measures such as FACS.

- Home ownership.
- Other assets and savings. This would include child trust funds for children, private pension contributions for working-age adults, receipt of private pensions for older peoples and other assets. There is an increasing interest in asset-based welfare, and the contributions to or receipt of private pensions is a recurrent theme in social exclusion measurement.
- Debt. The UK has very high levels of personal indebtedness. It is a particular problem for those on low income (see, for example, SEU, 2004).
- Subjective poverty. People's perception of whether they live or have lived in poverty, or whether their income is below the standard needed to keep their family/home out of absolute or overall poverty is an indicator in the PSE Survey.

Domain 2: Access to public and private services

Resources, as argued above, does not refer simply to cash incomes, nor indeed to privately owned and deployable assets. The importance of access to services, whether private or public, including utilities in the home, is recognised throughout the social policy literature. It is also specifically linked to the question of social exclusion in a number of key reports and surveys (for example, Howarth et al, 2001; PSE Survey; *Monitoring Poverty and Social Exclusion*). The IMD includes a section on barriers to housing and services, including the distance to specific services. Some services (such as transport; see SEU, 2003b) are relevant to all age groups. Others, such as childcare, youth services and home care services for older people may be age specific. In applying the matrix to individual data sets, it is important to be sensitive to the range of services on which information is available for different stage of the life course:

- Public services
- Utilities
- Transport
- Private services
- Access to financial services (includes access to a bank account).

Domain 3: Social resources

The *Opportunity for All* and NPI series between them raise issues about children who are institutionalised or otherwise placed in local authority care, while earlier SEU reports have focused on young runaways, 'looked-after' children and adult offenders (SEU, 2002a, 2002b, 2003a). We have amalgamated this concern into a single category of those separated from

family and those who are institutionalised. While this does not necessarily betoken a lack of family support, it does mean that that support is not available on a day-to-day basis. There is an increasing awareness of the importance of social networks to individual well-being, including, but not confined to, the provision of social support. Questions on social support are now routinely incorporated into some surveys at national and European level. There is also a large literature on social capital, far too extensive to address here (but see Putnam, 2001). We have preferred the term ‘social resources’ to that of ‘social capital’, since the latter looks from the perspective of communities rather than individuals. Nevertheless, there are overlaps between the two. The extent and quality of social networks has also been argued to facilitate labour market participation (Demos, 1997).

- Institutionalisation/separation from family (includes looked-after children and all those in residential care, young offenders’ institutions or prison. This will not generally be captured in household surveys);
- Social support (affective and instrumental);
- Frequency and quality of contact with family members/friends/co-workers.

Participation

Participation in economic, social, cultural and political life – or its absence – is a key feature of many definitions of social exclusion, as demonstrated earlier in Chapter Two.

Domain 4: Economic participation

As we have seen, access to employment is part of the first Nice criterion, ‘Facilitating participation in employment and access by all to resources, rights, goods and services’. Employment is important as a route to access to material resources (domain 1) but there is not an absolute fit. Some in employment are poorly paid; some not in employment may have decent or even substantial incomes. It is often argued that paid work is inclusionary in its own right. But whether work is a positive, inclusionary experience depends partly on the financial rewards it brings, and partly on the nature and quality of working life. Consequently the NPI’s *Monitoring Poverty and Social Exclusion* includes data on low pay and on working conditions.

Treating those outside the labour market as thereby socially excluded should be treated with caution: in 1999 about 43% of all adults were not in a paid job, and over a third lived in a

jobless household. It is widely accepted that caring activities and unpaid work are also forms of economic participation.

This domain is relevant to all age groups. Children may be in paid work, or may be providing unpaid care, as may older people (DH, 1999). The effects of paid work and caring are not necessarily inclusionary. They may be exclusionary, as in the case of child carers who are unable to participate in the normal activities of their age group, or they may be ambiguous. These relationships can be tested empirically:

- Paid work – employed, self-employed, unemployed, non-employed
- Providing unpaid care
- Undertaking unpaid work
- Nature of working life (includes type of occupation and full-time/part-time status)
- Quality of working life (includes anti-social hours of work, nature of contract, leave entitlement, flexible working arrangements, benefits, workplace injuries).

Domain 5: Social participation

Several of the definitions in Chapter Two (definitions 6, 9, 10, 11) refer to social participation. Others refer to participation in normal activities (definitions 3, 7, 8). The Atkinson indicators include ‘social participation’ as an area where indicators need to be developed. The dividing line between social participation and economic or cultural or political participation is of course difficult: work, politics and culture may or may not involve high-quality social interaction. We have here differentiated social participation, as the PSE Survey does, in terms of participation in common social activities. The ground covered here is similar to that of ELSA and *Sure Start in later life*, but extended from older people across the whole age range: *Sure Start in later life* emphasises the importance of meaningful relationships and roles, while Townsend (1979) argued that lack of resources might prevent people from carrying out normatively defined roles as parents, grandparents, children and so on:

- Participation in common social activities
- Social roles.

Domain 6: Culture, education and skills

Two issues need to be considered here: cultural capital and cultural participation. Cultural capital is a concept that has its origins in the work of the French sociologist Pierre Bourdieu, where it is intimately linked to processes of social exclusion. These, however, are not principally about the exclusion of a disadvantaged minority from the mainstream, but about the maintenance of class domination by an elite. Cultural capital is of three kinds. *Institutionalised* cultural capital refers to such things as educational qualifications, and the topic areas identified here are principally of this kind. *Objectified* cultural capital refers to objects such as works of art, books, music and so on. *Embodied* cultural capital, however, refers to non-accredited, possibly tacit knowledge, and a range of tastes, dispositions, behaviours and demeanours acquired through participation in a privileged social environment. This latter form of cultural capital is likely to affect the extent to which institutionalised cultural capital can be ‘converted’ into economic capital by access to particular occupations or promotion within them (Bourdieu, 1984, 1997). Most empirical work on cultural capital has been in the field of education, and addresses the acquisition of formal qualifications. The topic areas that we have identified as associated with this are basic skills (which affect not just employability but have a much wider impact on social participation); educational attainment; and access to education. The conceptual relationship between cultural capital and social exclusion is addressed in Levitas (2004).

The question of cultural participation is slightly different. Of course what constitutes cultural participation is a complex area, especially in a multicultural society. But the development of an inclusionary strategy has been a major preoccupation of the Department of Culture, Media and Sport (DCMS) in recent years (for example, DCMS, 2001, 2003a, 2003b), as well as the Arts Council (Arts Council of England, 1999; Jermyn, 2001). In addition, the ESRC has recently funded a survey specifically addressing cultural capital and social exclusion that is theoretically rooted in Bourdieu’s work, and looks both at cultural leisure activities (included here as a topic area) and at cultural tastes (see *Cultural Trends*, 2004, 2006).

We have also included internet access as a separate topic in the domain of cultural participation, as it involves access to knowledge, and acknowledges the potential significance of the ‘digital divide’:

- Basic skills (literacy, numeracy, competence in English)
- Educational attainment

- Access to education (includes school exclusion, but also includes access to lifelong learning for working-age adults and older people)
- Cultural leisure activities
- Internet access.

Domain 7: Political and civic participation

Opportunity for All, Monitoring Poverty and Social Exclusion and the IMD include no indicators on political participation, and nor are such indicators proposed in the Atkinson series (see Chapter One). The Audit Commission's local quality of life indicators include the election turn-out at local level as an indicator of community involvement. Nevertheless, the definitional emphasis on exclusion from the political domain makes it important to seek indicators of political and social participation. Moreover, for older people, where the emphasis on paid work is diminished, civic activities form one of the seven dimensions derived by Barnes et al (2006) from ELSA. Secondary analysis of the BHPS also includes this dimension, as does the PSE Survey.

In general, where such information is collected, it is usually confined to asking the general population about participation in voluntary activity, group memberships and voting behaviour. We have included these topics here. However, given the focus on social exclusion, we have also included citizenship status and enfranchisement that would potentially identify more specific exclusions from the political process. Indicators of 'civic efficacy', of people's *feeling* of being able to influence decisions, would also provide some sense of subjective exclusion or its absence; a similar indicator is included in the Audit Commission's local quality of life indicators:

- Citizenship status
- Enfranchisement (voter registration and entitlement, as well as whether people voted)
- Political participation
- Civic efficacy (for example, feeling able to affect decisions)
- Civic participation, voluntary activity/membership (note that this will include active membership of faith groups).

Quality of life

All of the domains can be seen as aspects of quality of life. Consequently, many of the issues covered above are themselves incorporated into the Audit Commission's local quality of life

indicators. However, there are some additional aspects of quality of life that are not captured by the resources/participation dimensions. And, as we have seen, quality of life looms large in the indicators of social exclusion generated by ELSA, once paid work is given a lower priority.

Domain 8: Health and well-being

The IMD includes a domain for health deprivation and disability at aggregate level. The quality of life indicators similarly include a domain of health and well-being, again at aggregate level. *Monitoring Poverty and Social Exclusion* includes a range of aggregate indicators for children (low birth weight babies, infant mortality, dental health, accidental deaths and teenage conceptions); for young adults (drug users treated and suicides); for working-age adults (premature death, limiting long-standing illness, mental health and obesity); and for older people (excess winter deaths). Health is treated as a risk factor for social exclusion in Barnes et al's (2006) analysis of ELSA. 'Being healthy' and 'Enjoying and achieving' (which maps on to aspects of well-being) are two of the five dimensions in the *Every Child Matters* framework. There is thus no dispute about the significance of this domain. Many of the topics above, however, are amenable to use only at the aggregate level, not at the level of the individual or household, and are thus not appropriate for a matrix evaluating survey data. The topics selected below are all ones where data might meaningfully be sought from individuals. In addition to the more conventional indicators of physical and mental health and disability, as well as self-harm that is particularly significant for young people, we have added some potential indicators of well-being based on the literature discussed in Chapter Three. Thus the inclusion of life satisfaction, personal development, self-esteem in this domain draws on the sustainable development indicators and the work by the NEF discussed above.

- Physical health and exercise
- Mental health
- Disability
- Life satisfaction
- Personal development (including for children, but not only for them)
- Self-esteem/ personal efficacy
- Vulnerability to stigma (for example, long-term receipt of means-tested benefits)
- Self-harm and substance misuse.

Domain 9: Living environment

Homelessness and housing quality are included in most sets of indicators, and homelessness is identified as a domain in *The drivers of social exclusion*. Both the IMD and the Audit Commission's local quality of life indicators cover environmental issues and housing, as do the sustainable development indicators. Housing quality appears as an indicator to be developed in the Atkinson indicators. Housing that falls below the set standard of decency has recently been incorporated into the *Opportunity for All* series for children, young people and the over-50s, although it is mysteriously absent for adults. *Monitoring Poverty and Social Exclusion* includes several aggregate statistics on homelessness and housing quality. In addition most of these indicator sets identify neighbourhood quality, safety and/or satisfaction as an issue in disadvantage, as does Barnes et al (2006) and the PSE Survey. We have added access to open space to the more conventional indicators as it is associated with well-being.

- Housing quality
- Homelessness
- Neighbourhood safety (including traffic, atmospheric pollution, noise pollution)
- Neighbourhood satisfaction
- Access to open space (demonstrated as important to well-being).

Domain 10: Crime, harm and criminalisation

Crime occurs as a domain in the IMD, although simply as the aggregate incidence of four major groups of crime (burglary, theft, criminal damage and violence). It appears as 'community safety' in the quality of life indicators, here encompassing road traffic accidents that we have included above as an environmental consideration. The quality of life indicators also include, under 'community cohesion', an indicator of how far local residents think attacks on the basis of colour, ethnicity or religion is a problem in the area. *Monitoring Poverty and Social Exclusion* looks at victimisation rates, while *Opportunity for All* looks at crime rates in high-crime areas. *Opportunity for All* also includes figures on child protection registrations, while the *Every Child Matters* framework prioritises safety. However, it is not just children who may be subject to physical abuse within the home, but women and dependent older people. We have tried to capture the appropriate topics at an individual level in terms of exposure to harm as objective/subjective safety, exposure to bullying or harassment and discrimination.

If these topics reflect the potentially exclusionary nature of being the object of harm, there is also precedent in the literature and in the indicator sets for regarding the perpetrators of harm as potentially excluded, stigmatised or criminalised. For example, *Monitoring Poverty and Social Exclusion* includes figures for the numbers in young offenders' institutions, and the numbers of 18- to 20-year-olds with criminal records. Again, moving from the aggregate to the individual level, and extending this across all age groups, we have included as potential indicators of social exclusion having a criminal record, or an Anti-Social Behaviour Order (ASBO), or being imprisoned.

- Objective safety/victimisation (this includes actual and risk of abuse within the home, for children and adults)
- Subjective safety, for example, perceptions and fear of crime (home and neighbourhood)
- Exposure to bullying and harassment
- Discrimination
- Criminal record
- ASBO
- Imprisonment.

Risk factors

It is in the nature of the interactive process of social exclusion that many of these dimensions constitute risk factors as well as outcomes. There are additional factors that are not in themselves dimensions of social inclusion/exclusion, but which may constitute *risk factors*. In using the B-SEM to assess the scope of individual surveys, we also need to check whether they collect data on the following risk factors:

- gender
- ethnicity
- social class
- housing tenure
- household composition
- religious affiliation
- critical life events

Most, but not all, of these will be routinely collected in most household surveys. Religious affiliation is less routinely recorded, although may be very important in terms of forms of participation, especially as it intersects with gender, and in terms of discrimination. We have also included ‘critical life events’ as a risk factor. Especially in longitudinal surveys, information about these could potentially be important. Critical life events include bereavement, divorce or relationship breakdown (or divorce of parents), retirement and institutionalisation. They also include more potentially positive life events such as the formation of relationships (known to be important in lone parents leaving poverty), the birth of children or young people leaving home.

Applying the B-SEM to survey data sets

The matrix devised above is a tool for exploring the range of data on social exclusion available in existing data sets and administrative sources. Accordingly, we looked at 27 cross-sectional, and longitudinal data sets (Tables 5.1 and 5.2) to locate usable indicators of social exclusion. In each case, besides addressing whether there is data on each of the topics identified, we needed to address the coverage of the survey with particular reference to potentially excluded groups.

Table 5.1: Major cross-sectional surveys used in the application of B-SEM

Title of survey	Type
<i>Annual Population Survey</i>	Repeat
<i>British Crime Survey</i>	Repeat
<i>Expenditure and Food Survey</i>	Repeat
<i>Family Resources Survey</i>	Repeat
<i>General Household Survey</i>	Repeat
<i>General Household Survey (Elderly Individuals in Private Households Supplement)</i>	Repeat
<i>Health Survey for England</i>	Repeat
<i>Home Office Citizenship Survey</i>	Repeat
<i>Labour Force Survey</i>	Repeat
<i>ONS Omnibus Survey</i>	Repeat
<i>OPCS/ONS Survey of Psychiatric Morbidity</i>	Repeat
<i>Poverty and Social Exclusion Survey</i>	One-off
<i>Survey of English Housing</i>	Repeat
<i>Survey of Quality of Life in Older Age</i>	One-off

Table 5.2: Major panel/cohort surveys used in the application of B-SEM

Title of survey	Type
<i>Avon Longitudinal Study of Parents and Children</i>	Cohort
<i>British Cohort Study</i>	Cohort
<i>British Household Panel Survey</i>	Panel
<i>English Longitudinal Study of Ageing</i>	Panel
<i>European Communities Household Panel Survey</i>	Panel
<i>European Statistics on Income and Living Conditions (EU-SILC)</i>	Refreshed panel
<i>Families and Children Study</i>	Refreshed panel
<i>Longitudinal Study of Young People in England</i>	Cohort
<i>Millennium Cohort Study</i>	Cohort
<i>National Child Development Study (NCDS)</i>	Cohort
<i>Offending, Crime and Justice Survey</i>	Refreshed panel
<i>ONS Longitudinal Study – England and Wales</i>	Panel
<i>Youth Cohort Study</i>	Cohort

The data matrix can be found in *Appendix 7* and contains a key indicating whether the topic is covered by the survey, for each of our five population groups:

- Children
- Young People (16 to 24)
- Working Age Adults
- Older people 1 (50 to 60/65)
- Older people 2 (60/65 plus)

For each topic and population group we examined the extent of coverage (whether there was full or partial coverage of the topic), the source of the information (whether information relating to children was provided by children themselves or by adults on behalf children), and also whether the adult interviewee's information related only to themselves or their household.

A survey of surveys

Our brief also involved identifying the strengths and weaknesses of each survey data set in terms of the character of the achieved sample, the definitions used and the mode of data collection. Detailed descriptions of each of the 27 surveys reviewed can be found in *Appendix 6*. Our analysis and application of B-SEM revealed a number of possibilities for researchers interested in undertaking secondary analysis of these data sets, as well as various concerns. The main themes arising from our analyses are discussed below.

Coverage of social exclusion topics

The social exclusion domains and themes identified in the B-SEM are rarely all included in the major surveys reviewed by this study. All the surveys cover some aspects of social

exclusion but, in general, often by the very nature of their remit, they tend to focus on specific areas or topics and consequently exclude others. For example, the *Family Resources Survey* (FRS) provides very detailed and reliable information on income and has since 2005 introduced a small sub-set of questions on adult and children's material and social deprivation, but it does not include questions on social relations. The *Labour Force Survey* (LFS) and the larger *Annual Population Survey* (APS) provide extensive coverage of issues related to the labour market, and also education and training but do not include questions on material and social deprivation or other aspects of social relations. The *British Crime Survey* has very reliable data on victimisation experiences but has relatively few questions on other aspects of social exclusion. The *Home Office Citizenship Survey* (HOCS) is an excellent resource for examining issues relating to social, civic and community participation, as well as for neighbourhood satisfaction and access to services. However, the limited range of classificatory variables within this survey – and in the case of income and benefits data shortcomings in the quality of the data itself – somewhat undermine its usefulness as a resource in investigating multi-dimensional exclusion within the B-SEM framework.

There are, however, some exceptions. The PSE Survey, undertaken in 1999 with the specific aim of measuring poverty and social exclusion, represents the survey closest to the B-SEM. All themes in the B-SEM are incorporated in the survey including crime, harm and criminalisation, which many other surveys fail to include. Crucially, the PSE Survey contains detailed coverage of material and social deprivation for *both* adults and children and has uniquely identified social relations as one of the key dimensions for understanding social exclusion and consequently includes extensive questions on social support, social contact and social participation.

With the exception of the 1999 PSE Survey, the BHPS provides probably the best overall coverage of the those themes relating to the B-SEM model including material and economic resources, access to services, social resources, economic, social, cultural and civic/political participation, health and well-being and the living environment. Only with respect to crime, harm and criminalisation are there significant gaps, although in addition the deprivation items are rather restrictive, and as discussed elsewhere the adequacy of the BHPS in relation to the B-SEM framework is somewhat undermined by relatively small sample sizes.

The next best general survey in terms of topic coverage is the GHS. The survey has reliable information on income, ownership of consumer durables, economic and social participation

and support, civil and political participation, health and well-being, and also some limited data on the quality of the living environment. It does, however, contain some large omissions including an absence of questions on public and private services and also crime, harm and criminalisation. In terms of deprivation, only material goods and not social activities are covered in the GHS. Furthermore, GHS questions on consumer durables fail to differentiate between those answers where respondents lack items through choice and those going without because of affordability.

Beyond general surveys and, in relation to surveys specific to the different stages of the life course there are a few surveys containing relatively good data coverage. The *Avon Longitudinal Survey of Parents and Children* (ALSPAC), the *Families and Children Survey* (FACS) and the *Longitudinal Survey of Young People in England* (LSYPE) are all characterised by fairly comprehensive topic coverage. In particular, the FACS data provide relatively good coverage of material and social deprivation, housing and the built environment and access to services. Similarly, the LSYPE provides excellent coverage of young people's educational profiles and, as the cohort matures, of the economic and social position and social attitudes of young respondents. ALSPAC also provides good coverage of the sub-domains, although is relatively weak on income and material and social deprivation. *The English Longitudinal Survey of Ageing* (ELSA) also overlaps well with the B-SEM. The survey contains questions on all the themes included in the B-SEM, although it has stronger coverage in some areas than in others. For example, it provides relatively good coverage of material and economic resources, social, economic and political participation, and also health and well-being, but has poorer coverage of issues relating to the quality of the living environment and crime, harm and criminalisation.

In reviewing data coverage of surveys, we additionally considered the inclusion of risk factors associated with social exclusion and also critical life events that may act as important triggers for social exclusion. While all surveys universally included coverage of risk factors such as gender, age, housing tenure, social class and so on, very few surveys included religious denomination. Similarly, with the exception of some surveys such as the PSE Survey and the GHS, few routinely collected information on critical life events such as bereavement, separation and divorce or pregnancy. Furthermore, very few surveys included additional local area data such as the IMD (Index of Multiple Deprivation) or ACORN (A Classification of Residential Neighbourhood).

Coverage of the institutional and homeless populations

The most vulnerable of the socially excluded population in the UK resides in institutions or is homeless and, therefore, is not captured by any of the household surveys under review. The majority of the surveys carried out in the UK include only individuals in private households which means that a number of socially diverse, but exceptionally vulnerable, groups are missed by these surveys. These groups include all those in:

- hospitals (for example, the infirm, those with mental difficulties)
- secure accommodation such as young offenders' institutions, local authority homes for young offenders, secure hospitals, detention centres (for the young, asylum seekers), and adult prisons
- residential/institutional accommodation (for example, disabled children, older people)
- children's care homes
- people living rough on the streets.

These groups experience multiple forms of exclusion or what could be termed 'deep exclusion'. For instance, the SEU (1998) report on rough sleepers showed that many rough sleepers were on the streets because they were forced to leave home, many came from local authority backgrounds or had been in prison, and that they faced serious physical and mental health problems including alcohol and drug addictions. Similarly, the SEU (2002b) report, *Reducing re-offending by ex-prisoners*, acknowledged that many prisoners have experienced a lifetime of social exclusion. It reported that compared with the general population, "prisoners are thirteen times as likely to have been in care as a child, thirteen times as likely to be unemployed, ten times as likely to have been a regular truant, two and a half times as likely to have had a family member convicted of a criminal offence, six times as likely to have been a young father, and fifteen times as likely to be HIV positive ... 80 per cent have the writing skills, 65 per cent the numeracy skills and 50 per cent the reading skills at or below the level of an 11-year old child. 60 to 70 per cent of prisoners were using drugs before imprisonment. Over 70 per cent suffer from at least two mental disorders" (SEU, 2002b, p 6).

There have been some specialist surveys on a few of these population groups, but these have tended to be one-off surveys and are now dated. For example, the last government-funded surveys of disability, covering adults in private households, children in private households, adults in communal establishments and children in communal establishments, were carried out between 1984 and 1988 (ONS, 1989). Similarly, the last national survey of prisoners was

in 1991 (ONS), while the most recent survey of psychiatric morbidity among homeless people was in 1994 (OPCS, 1994). In 2000, the Health and Safety Executive (HSE) focused on the health and social exclusion of older people and included a sample of care home residents, but the sample did not appear to be representative of all care homes in England.

On the other hand, there are no national or large-scale surveys on some of these other vulnerable groups, for example, in relation to asylum seekers or refugee communities. Small-scale surveys have been carried out but these are unrepresentative of the total population groups (see, for example, Migration Resource Centre, 2006). Furthermore, trafficked people (including children traded as domestic servants, and women sold for prostitution), whose very existence is clandestine, are by definition impossible to survey. A separate, but related problem, is that even taking into account these omissions, household surveys may still not be representative of all groups and this will include other hidden population groups such as those without a legal right to reside in the UK (see also the section below on non-response).

Information derived from proxy informants

Many surveys we reviewed make use of proxy informants but there are often data quality issues from answers given by proxy informants. Interviews by proxy tend to fall into two categories: (a) where a carer or an adult relative living in the same household answers questions because the respondent is unavailable for interview; and (b) where an adult (usually the parent) answers questions on the circumstances relating to their children. In some surveys (for example, the LFS) up to one third of respondents will be proxy informants. This poses questions of data reliability. One study (Dawe and Knight, 1997), examining the quality of the information provided by proxy informants in the LFS, found that although there was close agreement on many questions between the answers given by proxy informants and those answers given by the subjects themselves, some questions – particularly those on income and number of hours worked – which require very detailed numerical information showed a poor overlap.

The majority of surveys containing proxy informants will involve proxies because the respondent is unavailable to take part; fewer surveys involve parents acting as proxy informants. Among those surveys requiring direct information on children and using parents as proxy informants, are the PSE Survey (parents completed questions on material and social deprivation experienced by their children); the GHS (parents answered questions related to the health and well-being of their children); and the FACS (main respondent answered

questions related to the health, well-being and education of their children; main respondent provided partner data where the partner was not available for interview).

However, a growing number of data sets now include children. Sometimes this is in the form of booster samples but more often it is because only children or young people are included in the survey's population coverage. Examples include the *Expenditure and Food Survey* (EFS) where children aged 7-15 keep a simplified expenditure diary; the HOCS, which has included children's and young people's booster samples; the LSYPE, which involved young people aged 13-14 in its first wave; and the *Offending, Crime and Justice Survey* (OCJS), which involves children as young as 10. While questions still abound in relation to the reliability of information provided by children (for example, self-reported offending studies have traditionally been seen as unreliable because children are thought to exaggerate their offences; see Coleman and Moynihan, 1996), computer-assisted interview methods (such as CAPI [computer-assisted personal interviewing] and CASI [computer-assisted self-interviewing]) may help improve the reliability of survey data involving children.

Design effects and non-response

If social survey data are to generate a balanced sample on which reliable inferences can be drawn about the wider population of interest, it is vital that respondents have an equal probability of selection and that all selected respondents agree to be interviewed. Only rarely, however, are surveys based on equal selection probabilities, either because researchers wish to deliberately over-sample specific population sub-groups to boost sample sizes (see above) or due to inadequacies in the sampling frame. For example, the postal address file, which is routinely used as a sampling frame in many large-scale UK surveys, does not indicate the number of people living at each address so it is impossible to ensure equal selection probabilities. As a result weighting is usually applied to correct for non-equal selection probabilities and differential response rates across the sample.

Nevertheless, although survey weighting is widely applied the topic remains relatively poorly understood in the social science community for a variety of reasons. First, commonly used software continue to provide biased parameter estimates and spurious levels of significance – potentially with serious policy consequences (Crockett, 2006; Dale, 2006)¹. Second, and perhaps of more fundamental significance, weighting strategies differ significantly across surveys depending on the sampling strategy adopted and the precision of parameter estimates will also vary within surveys depending on the focus of investigation. For example, where

clustering effects are present stratified samples better reflect the underlying population than a simple random sample.

However, it is only possible to obtain unbiased estimates if the data includes primary sampling units and strata that often is not the case in the datasets reviewed here. Moreover, the effect of stratification in increasing the precision of estimates depends on the relationship between the characteristic of interest and stratification criteria. Survey weighting for non-response is usually undertaken on the basis of characteristics (such as household type, tenure, social class), which are strongly associated with various dimensions of exclusion. Plewis' (2004a) analyses suggest that this may increase rather than reduce the non-response bias of parameter estimates. Clearly the problem of sample non-response is especially important in relation to the B-SEM approach since it is well known that non-response bias is associated with these characteristics.

The nature of non-response bias in relation to panel and cohort studies is further complicated by sample attrition over time. It is also well established that in the UK differential sample attrition is associated with many of those factors either associated with multi-dimensional exclusion or themselves constituting dimensions of the B-SEM approach. For example, differential attrition rates in the *1970 British Cohort Study* are associated with parental social class, employment status and educational achievement, household benefit receipt, parents' country of origin and housing tenure (Bynner, 1996). Similarly, attrition rates in the ECHP Survey vary by tenure, economic status, educational achievement, socioeconomic group (SEG), and income status, and multivariate analysis suggests that the odds of sample attrition among income-poor households in Northern Europe are around 30% higher than for non-poor households (Behr et al, 2002; Watson, 2003). However, although it is possible to weight data to account for differential attrition in panel and cohort studies, the estimation of longitudinal errors remains a contentious one.

Item non-response

Item non-response refers to missing data for respondents on individual survey variables and is a major problem because non-response is rarely missing at random (MAR). A range of advanced solutions have been proposed including 'hot deck' imputation and regression imputation (see, for example, Carpenter and Kenward, 2006). However, their application in substantive social science disciplines remains limited, and this is a particular problem in

relation to those categorical variables covered by the B-SEM approach, where straightforward imputation methods are not readily available (Dale, 2006).

Nonetheless, there is considerable evidence that simply ignoring missing data will substantially bias both parameter estimates and their statistical significance. This is a particular problem in cohort and panel studies where complete data may only be present for a relatively small subset of respondents or cases. As a result, parameter estimates may be substantially biased and sample sizes will frequently be too small to support statistically generalisable analyses. Fortunately panel and cohort data allow for robust imputation on the basis of comparison with other waves or sweeps of the dataset. Clearly, this is not so for the cross-sectional studies reviewed here.

Sample sizes

However, arguably the most important practical issue with regard to data quality in the analysis of multi-dimensional exclusion is that of sample size. As has been noted elsewhere in this report sample sizes in many of the 'large-scale' social surveys are rarely sufficient to facilitate detailed analysis of those population sub-groups most vulnerable to multi-dimensional exclusion as operationalised within the B-SEM context.

There is at present a clear trade-off between topic coverage and sample size in this respect. Many of the largest surveys covered by this review such as the APS, the LFS, the *ONS Longitudinal Study* (as well as the census Samples of Anonymised Records, or SARs) often contain sufficient samples of population sub-groups to allow for robust generalisation beyond the study sample to the wider population of interest. However, as the B-SEM data matrix illustrates, their topic coverage is often very limited since, partly as a consequence of their size, they are key multi-purpose surveys. Conversely, the most comprehensive topic coverage is associated with the national poverty surveys (for example, the 1999 PSE Survey and its predecessors) where sample sizes are usually too small to allow for robust estimation in relation to those population sub-groups most vulnerable to exclusion. Similar observations are pertinent in relation to opportunities for the spatial analysis of data relating to multi-dimensional exclusion within the B-SEM framework (which is beyond the scope of this review).

A number of strategies are available to at least partially compensate for these shortcomings in sample sizes. First, for those surveys based on a repeated cross-sectional design it is usually

possible to pool samples across waves to increase sample sizes. This is of course dependent on the extent to which variable definitions are consistent across waves, which is in itself a major issue. The need for greater harmonisation in concepts, definitions, design and processing practices is now widely recognised (for example, ONS, 2004a). However, in addition to harmonisation across surveys, harmonisation across time is clearly critical. For example, until recently household-level data in most surveys were based on the concept of ‘head of household’, a concept that has now been superseded by that of ‘household reference person’, which no longer gives priority to male partners. In practice, such definitional changes can have important effects on the distribution of key classificatory variables, such as social class, which are used in the analysis of multi-dimensional or ‘deep exclusion’.

Second, booster samples are increasingly used as means of facilitating analysis of relatively small population sub-groups. Several of the surveys covered by this review have benefited from inclusion of booster samples of this type – or from the over-sampling of groups vulnerable to poverty and exclusion as part of the overall survey design as in the 1999 PSE Survey. Surveys incorporating additional boosters for relatively small or ‘hard-to-reach’ population groups include: *British Crime Survey*, 2004-05 (minority ethnic group, young people); BHPS (Scotland, Wales); HOCS, 2003 (minority ethnic group, young people); OCJS, 2003 (minority ethnic group); *2000-01 SQOL OA* (minority ethnic group). Nevertheless, there remains considerable scope for extending this approach to investigate the social circumstances of other population groups especially at risk of exclusion including, for example, lone parents, large families, unemployed people, people with disabilities, and so on.

Administrative sources

In addition to the survey data discussed in detail above, in this report there is a vast range of administrative information held by different government departments. There has been a great deal of interest in linking different data sets, both as a research resource and for policy purposes such as fraud detection. Several publications by the GSS address the potential use of administrative data (e.g. GSS 1999a, 2000). White et al suggested that longitudinal data would be improved by linking information from administrative sources to survey data, inasmuch as this is compatible with legal and ethical requirements. They argued that much greater use is made of administrative data in other countries, notably the US, Denmark and Sweden, and suggest “a review of methodological issues connected with this type of administrative data” (GSS, 2000, p 7). Such a review was undertaken by NDS in its first year of operation,

reporting as Jones and Elias (2006). The justification and scope of the review were set out as follows:

The development of administrative data resources

Much important information about individuals and organisations is retained nowadays through the use of information technology. The capture and storage of the details of transactions between organisations and individuals has created resources with the potential to inform social scientific research directly or, via data linkage techniques, to enhance existing resources. Such data cover a wide variety of fields in both the public sector and private sectors, including demographics, consumer behaviour, education, social care and community support, crime, transport, health, taxation, social security, housing and migration. Via personal or organisational identifiers, data from different sources can be linked to generate rich resources for research purposes. However, the scope for data linkage may be restricted by ethical considerations surrounding the sensitive nature of the linked data, by legal requirements and by the resource costs of undertaking linkage.

There is a need to establish the scope, quality and potential that public and private administrative data can offer as research resources. Given the wide variety of administrative data, work in this area needs to be focussed on some important examples. Particular areas where attention could be directed include the registration and record system within the National Health Service, the Department for Work and Pensions Longitudinal Study of Benefit and Pension records and the Department for Education and Skills Pupil Level Annual Schools Census.

To progress work in this area, the following action is proposed:

An audit of the major administrative databases held by government departments and agencies will be prepared. This audit will report on the potential such data have to inform research on a range of issues (with particular emphasis on data quality and coverage), the accessibility of such data for different research purposes (data anonymisation, the need for 'safe-settings', etc), the feasibility of linking between different administrative data sources and with census and survey records. The audit will exemplify best practice linking already achieved or recently undertaken. This audit will be augmented by a study of public attitudes to data linking and the use of administrative data for research purposes. Work in this area will be undertaken as a close collaboration between the Economic and Social

Research Council, relevant government departments/agencies and the Office for National Statistics (NDS, 2006, p 10).

This section draws substantially on the results of that review. A similar review specifically focusing on social exclusion would be a major undertaking. However, many of the same points would hold, and some of the difficulties of administrative data are particularly acute in relation to social exclusion. White et al (GSS, 2000) point out that administrative data sources generally cover a narrow range of information compared to survey data. This would be a particular drawback in the analysis of social exclusion, which, as indicated above, requires a wide range of types of information.

In addition to the UK NDS report, the disadvantaged groups team at the DWP has examined the potential of linking data sets more directly focused on disadvantage (Sheppard, 2006; Pleace and Bretherton 2006).

The methodological issues about administrative data overlap with those of applying to survey data, notably questions of coverage and data quality. In addition, there are legal and ethical issues concerning the use of data for purposes other than those for which it was compiled. These are particularly acute in relation to the potential merging of data sets. Such merging also poses technical challenges. Moreover, these areas of concern are not independent: the ethical and legal difficulties may also bear on the quality of data. These issues are discussed in general terms below. We then comment on some specific data sets that may be of particular interest in relation to social exclusion.

Coverage

Coverage is an issue for administrative data sets as well as for social surveys. Jones and Elias (2006, pp 66-7) argue that administrative data is perceived as providing 100% coverage of the target population. The data sets are consequently much larger: the DWP Longitudinal Study will contain 100 times the number of cases as the LFS. Problems of attrition are sometimes argued to be minimal, although Jones and Elias (2006, p 73) point out that matching errors can lead to potentially large losses of data.

Although one might assume that because they are not based on samples, such records are in some sense more 'complete'; this is not necessarily the case. The target population of administrative data sets does not always coincide with the population in which researchers are

interested, and relevant groups may be omitted. For example, the Jobseeker's Allowance count omits those seeking work but not claiming benefits, and is thus a poor measure of unemployment compared to the LFS (which has difficulties of its own). The National Pupil Dataset (NPD) discussed below does not include pupils in the independent sector.

Data quality

The ESRC-commissioned review of longitudinal data sources endorses the principle of using administrative data sets, but again advises caution in relation to data quality:

The addition of administrative data to the datasets is seen by many of our respondents as hugely beneficial – a view which we endorse. Such identifiers as National Insurance numbers (NINOs) if added to longitudinal datasets would hugely enhance their research potential because of the wide range of DWP employment, income and benefits data that could be linked via them. However we are also conscious of the need for caution in seeing such data as some kind of panacea for augmenting longitudinal datasets and filling gaps in the longitudinal record. Apart from issues of data access, concerns about quality arise with quite a lot of the administrative information that is collected. Dialogue with data producers is needed as to how far the data can best serve the interests of a particular piece of research. We are aware that the issue of access to administrative data is under discussion by the UK Data Forum; we would encourage its members to take account of data quality issues as well as access and confidentiality (Martin et al, 2006, p 60).

Jones and Elias (2006) suggest that administrative data is perceived as more accurate than survey data, and especially less susceptible to respondents' recall errors or misreporting. However, they urge caution on this point:

The problem, potentially with administrative data is that they may be collected to varying standards and that coding standards and/or methods of transcribing data may vary across departments.... [D]ata entries may be subject to spelling mistakes, anomalous ordering or mis-recording. More generally administrative data might be of variable quality across departments and subject to missing observation, discontinuities, changing in systems of coding, and so on. (Jones and Elias, 2006, p 73)

White et al also draw attention to the particular problems of data quality that may affect administrative sources, “such as incomplete recording or documentation which is not designed for external use” (GSS, 2000, p 39).

A more comprehensive account of problems that may affect administrative data sets to an even greater extent than surveys is provided by Ruddock’s *Measuring and improving data quality* (GSS, 1999a), the final report of a GSS task force on non-sampling error. These include not only coverage errors, discussed above, but non-sampling errors can take a variety of forms, identified by Ruddock as coverage errors, non-response errors resulting from refusals to participate or answer particular questions, the accuracy of data, the impact of the context and mode of data collection, and the nature and role of the interviewer. Forms of error described by Ruddock as respondent and questionnaire errors, mode errors, and interview error (GSS, 1999a, p 33), raise basic methodological questions about the quality of data for research purposes: what was asked, of whom, in what setting and for what purpose? How was the data recorded? To what extent does it depend on observation by a third party, or proxy informants? While these issues affect all survey research, they are an even greater issue in administrative encounters, where the initial purpose of the encounter and the power relationship may be quite different, such as in applications for benefit. Scott (1990) points out that administrative records can never be treated by social scientists or researchers as neutral reporting of events. They are always shaped by political contexts and more general cultural and ideological assumptions, as well as the particular institutional contexts in which data is produced. There are also forms of error that occur in the editing and analysis of data, described as systems errors or errors in the systems for processing different kinds of data; and data handling errors described by Ruddock as errors in “data capture, editing and coding of open-ended textual responses” (GSS, 1999a, p 7) that may lead to bias and increase in variance.

Technical, ethical and legal issues

Gill (GSS, 2001) discusses the difficulties of record linking in relation to the generation of statistical information. Much of the report concerns technical procedures and difficulties. It alerts potential users to the problems of matching cases and the possibility of errors (such as where more than one individual has the same name). The merging of data sets in ways that would track individuals also poses technical difficulties and dangers of errors. However, even if these can be overcome, the legal and ethical obstacles are formidable.

Gill (GSS, 2001) also draws attention to the ethical and legal constraints concerning the protection of individual identity and the use of data for purposes other than those for which it is collected. These matters are covered both by primary legislation such as the Data Protection Act (1998) and the National Statistics Code of Practice. Gill makes the further point that where individual identity cannot be deduced from any single source, data combination may reduce this protection or increase 'disclosivity'. Similar issues are also noted by the UK NDS:

An important part of the research process is the nature of the arrangements which exist to regulate access to confidential or potentially disclosive data. High quality research requires facilities for good data access. This may involve linking between various data sources (eg linking administrative data to survey information held on individuals). Such procedures raise ethical issues relating to the confidentiality of data and the nature of consent required for linking data. Restrictions on access may also exist for legal reasons or because of the nature of undertakings to data providers given or implied during data collection. The National Data Strategy must provide guidance and common solutions to the procedures required to facilitate high quality research whilst ensuring data are used in accordance with any necessary restrictions on use. (NDS, 2005 b, p 5)

The launch report of the new Social Exclusion Task Force acknowledges the existence of statutory obstacles to data sharing, and includes an action point to overcome this: "The Government will explore how to extend data sharing in relation to the most excluded or at-risk groups, including any additional powers that may be necessary" (Social Exclusion Task Force, 2006, p 35).

Such plans are controversial. The Ministerial Committee on Data Sharing, MISC 31, set up to develop the government's strategy on data-sharing across the public sector, has already attracted criticism for announcing that information will normally be shared in the public sector, provided it is in the public interest. This is a change in the default position that a justification should be provided in terms of purpose for each instance of data sharing. Privacy campaigners and other critics have seen this as a potential threat to civil liberties and a potential breach of the European Convention on Human Rights (Ballard, 2006; Cross, 2006; Cross and Travis, 2006; Kablenet, 2006). Concern is not confined to pressure groups, but is widespread across the population:

Research by MORI has found that 60 per cent of the population are fairly or very concerned about public services sharing their personal information. Awareness of the level of information held on them by public services was also low among the general population, with 64 per cent saying they did not feel 'well informed' about the data being kept about them (MORI 2003). A more general distrust of central government, regardless of the party in power, has also been reported by successive surveys of the UK population (MORI 2002). (Pleace and Bretherton, 2006, p. 77)

The *Information sharing vision statement* (DCA, 2006) sets out to allay fears on this issue, while ethical considerations have been at the centre of discussions about developing the Work and Pensions Longitudinal Study (WLPS) (Walker, 2003; DWP, 2005a, 2005c). The political controversy is outside the remit of this report, but is relevant in terms of public trust and the potential impact on data quality. The Data Protection Register is broadly in support of data sharing but warns that the government risks losing public trust if 'reasonable expectations of privacy are not met' (Cross, 2006). Such trust rests on the perception, rather than the reality, of confidentiality. Its reduction may affect the quality of administrative data, but also runs the risk of a significant and damaging effect on the response rates and data quality in social surveys where participation is voluntary.

Jones and Elias (2006, p 78) recommend the setting up of an Administrative Data Resources Service that would address problems of access, issues of data linkage and "undertaking quality reviews of administration data resources to establish their strengths and weaknesses as research resources". We would endorse this recommendation, as there is no point in undertaking elaborate statistical analyses of temptingly large data sets without a quality assessment of all the contributing data sources.

Work and Pensions Longitudinal Study

The major data set held by the DWP that bears on social exclusion is the WPLS. The WPLS links benefit and programme records with employment records from the former Inland Revenue, now Her Majesty's Revenue and Customs (HMRC). Thus Tax, National Insurance, benefit and employment programme records are being linked, together with Pension Credit data, Housing Benefit and Council Tax benefit data. It also includes data on Welfare to Work programme participation, including age of youngest child for those on the New Deal for Lone Parents, disability, ethnicity and destination on leaving the programme. Linkage is through NINOs. Information has also been added about ISA (Individual Savings Account), PEP

(Personal Equity Plan), TESSA (Tax-Exempt Special Savings Account), private pension pots or savings accounts information from HMRC. Links to DfES data sets require the resolution of technical and legal issues in the sharing of data between government departments.

The purpose of the WPLS is to evaluate the success of existing government initiatives and to provide the evidence base for the better targeting of resources. It has existed in its present form since October 2005. A comprehensive list of possible uses is provided by the DWP (DWP, 2005b). According to Jones and Elias (2006, p 17):

The 100 per cent coverage means that WPLS is unmatched in the UK as a longitudinal data resource. As well as providing a very much enhanced facility for welfare-to-work and pension planning related policy analysis within the department, it also provides an important resource for future research work on individual labour market experiences and transitions.

In spite of the very wide range of data collated in the WPLS, there are large areas of information missing in terms of social exclusion. Notable absences (because they are not included in administrative returns) are information about health, education, material deprivation, social networks and social participation. The WPLS can provide a unique resource for looking at patterns of labour market participation, employment, benefit use and so on; but is much more limited in relation to social participation and well-being in a broader sense.

The ADMID report

The disadvantaged groups team at the DWP has considered the possibilities of linking national records with data from service providers working at local authority level to provide better information about adults facing multiple barriers to employment. The ADMID (Administrative Datasets for Measuring Impacts of Disadvantage) study was designed “to explore the possible advantages of linking project level, local authority level and city-level datasets with DWP datasets as a means to improve the range and extent of data available on ‘hard to help’ groups who are able to work” (Pleace and Bretherton, 2006). It addressed the views of “local service providers, service commissioners and service users” on the desirability, practicality and legality of data sharing.

This work is driven by two concerns. First, the well-being of those who face multiple barriers to employment (including people released from prison, early leavers from the Armed Forces, homeless people, ex-offenders and refugees, and those dependent on drugs or alcohol) and are at risk of sustained social exclusion. Second, the “high ‘lifetime’ costs to the Exchequer because benefits are claimed for sustained periods and people tend to make high use of publicly-funded services” (Pleace and Bretherton, 2006, p.1). This double imperative may be part of the reason for the report’s finding that the issue of trust is again an obstacle, both for service users and service providers.

The potential advantages of data sharing are clear:

A detailed longitudinal dataset would allow specific individuals and localities to be more accurately targeted with services. Services and programmes could also be evaluated longitudinally. The life courses of people in the target group could be better understood and services tailored accordingly. (Pleace and Bretherton, 2006, p.1)

The disadvantages concern data quality, potential stigma and, above all, issues of trust. Service providers raised questions about inaccuray and oversimplification, for example the use of categories such as drug user. Not only does this cover a range of people with different needs, but there were concerns about the label continuing to be attached to an individual in an enduring way, even when it had ceased to be applicable. Among service providers,

While many respondents saw advantages in the capacity to track individuals longitudinally, they found it difficult to reconcile this with some concerns about the implications for service users centred around the use of data to monitor individuals. (Pleace and Bretherton, 2006, p. 4)

Service users often “regarded themselves as being in a situation of surveillance by an overarching ‘state’ which processed their personal data in an arbitrary way” (p. 6) . Even free and informed consent appeared an insufficient answer because of levels of distrust. Service providers were concerned that people would refuse to engage with services if asked to sign consent forms for data sharing, and there were also concerns about resource implications for service providers.

Pleace and Bretherton concluded that ‘a combination of better education about the DPA and clear rules that provide reassurance about the uses to which shared data can be put could facilitate data sharing between people facing multiple barriers to work, local service providers and DWP’ (p. 7), and that further research is necessary. Although this final recommendation might go some way to addressing the issue of trust, it does not address the issues raised in the report about data quality.

Department for Education and Skills

The DfES has a central ‘data warehouse’ that contains a range of data on individual school pupils, schools and records from the Learning and Skills Council. There are plans to integrate information on higher education from the Higher Education Statistics Agency (HESA) and the Universities and Colleges Admissions Service (UCAS), which can be linked to individual pupils, and also to incorporate social care statistics from the Department of Health (DH). There is a new NPD on all children in the state school system: some eight million individual records are added each year. The individual pupil records are based on PLASC, the Pupil Level Annual School Census. Returns are compulsory for schools in the state sector, and optional for maintained nursery schools. It should be underlined, however, that this is not a complete national data base, as it excludes all those educated in the independent sector. The records are linked by unique pupil identifiers, so tracking of individuals and longitudinal analysis are possible. The data on individual pupils includes gender, ethnicity, age, receipt of free school meals, educational attainment, information on special educational needs, postcode and school attended. The data can be accessed through the PLASC/NPD User Group (PLUG), funded by the DfES and ESRC and managed at the Centre for Market and Public Organisation (CMPO) at the University of Bristol (Jones and Elias, 2006).

Client Records

Some of the data included in the WPLS is based on Client Records. Client Records are kept by a variety of different government and other agencies. For example, there are two separate datasets relating to clients receiving services from local authorities. One, the Supporting People Client Record, is concerned with the characteristics of the clients themselves, and is compiled from individual Client Record sheets. These sheets are generated by completion of an interview form with preset categories. The form is completed by the service provider, not the client, and in some cases comes at yet another remove through a proxy informant for the client. Such a process is inevitably subject to interviewer error, in that the interviewer is

required to categorise the client on the basis of information given, into the precoded groups (which are not mutually exclusive). It is also prone to respondent error, in that the client or their representative may consciously or unconsciously give partial or inaccurate information. A related set of administrative data is the Supporting People Local System (SPLS), which records data on service provision to this client group, although it has, as described below, suffered both technical and data quality problems. The data collected, as indicated below, is (as is the case with many administrative data sets) rather narrow for any useful multivariate analysis of social exclusion.

The Supporting People programme was launched in April 2003. It aims to provide high-quality services to vulnerable people living in the community. It does this by providing housing-related support, in order to improve the quality of life of service users and adopt interventions that prevent people requiring institutional care. The programme is administered by the top-tier authorities (150 administering authorities) who run the programme at a local authority level. The data collection, processing and preliminary statistical analysis is carried out by the Client Record Office at the Joint Centre for Scottish Housing Research (JCSHR). The 150 administering authorities and the Department for Communities and Local Government (DCLG) are also provided with reports of summary information from the Client Record Forms every quarter (produced by the JCSHR). These reports are also published on their website (www.spclientrecord.org.uk). DCLG use this information to monitor fair access to Supporting People services for those who are eligible, and to examine whether the range of support needs in their local area is being met. It can also be used by administering authorities working together to coordinate services regionally. DCLG receive the data in an SPSS format, to enable in-house analysis and answer ad hoc queries. DCLG also provide data analysis on drug-related client groups to the Home Office for the Drugs Intervention Programme on a quarterly basis. SPLS is a quarterly dataset from 150 administering authorities in England. It collects data on services to vulnerable people living in the community. SPLS is a data return from the service perspective, consisting of 29 CSV files containing data on supply, costs of services, accreditation data, quality assessment framework data, performance data and service reviews.

A copy of the Client Record Form can be found at http://ggsrv-cold.st-andrews.ac.uk/spclientrecord/ClientRecord/2006_7/Client_Record_Form%202006_7.pdf.

The Client Record Form was developed to record standard information about clients starting to receive services through the Supporting People programme in England. The Client Record

collects information on a client's characteristics (for example, their age, sex, their economic status and their ethnic origin the informant's relationship to the client), the provider type, service type, the client's primary and secondary client groups, the accommodation previously occupied and the referral type. Information is collected on all new clients accessing a Supporting People service. Information, however, is not measured if the client stops receiving the support or service. There are 21 client groups, including Young People at Risk, Young People Leaving Care and Teenage Parents.

Client Records are also held by voluntary sector agencies such as the CAB. These too are of limited value. While reasonably complete data is collected by CAB on the nature of clients' problems, only limited information is collected on clients' characteristics, and most of this is on a voluntary basis and thus not representative of the caseload as a whole. For example, employment status, the provision of unpaid care and participation in unpaid work are all optional fields. Self-reported income is similarly optional with a low level of completion, while computed income is calculated only for a small number of money advice cases. Housing tenure is also an optional field. Information about disability has recently become a compulsory item of information. Information about homelessness or discrimination is collected only where this bears on the problem for which advice is sought. No information is collected on educational levels, social and political participation, health or well-being. In terms of the quantitative analysis of social exclusion, then, the CAB database is of limited usefulness. However, there may be greater scope for qualitative analysis of these records in terms of presenting problems, surrounding difficulties and what is known of client characteristics.

All the Client Records suggest that the coverage and quality of data is a fundamental consideration in the prospects for secondary analysis.