Developing indicators:
issues in the use of quantitative data about poverty

Paul Spicker

Published in *Policy and Politics*, Vol 32(4), 2004, pp. 431-40.

Abstract

Despite the familiarity bred by the routine use of indicators in public services, many commentators are vague about what indicators mean and how they should be used. Several confuse indicators with measurement. This paper argues that precise and accurate measurement is often inappropriate to the use of indicators, and that we need to accept a degree of uncertainty. A good indicator should be understood as a pointer, not a measure. It should be accessible, robust and accompanied by other pointers. By treating indicators as measures, summary indices conceal key issues, hide the values and concepts which are implicit in the exercise, and are vulnerable to mathematical accident. Using multiple indicators is sounder in principle, in methodological terms, and in practical application.

Indicators of poverty

Poverty is a complex, multidimensional concept. There is no central, agreed core of meaning; it is possible to identify many discrete clusters of meaning. They include definitions concerned with material circumstances, like need, multiple deprivation and standard of living; definitions concerned with economic position, including lack of resources, inequality, and class; definitions concerned with social relationships, including social exclusion, dependency, low status, lack of entitlement and lack of basic security; and a normative view, which understands poverty as unacceptable hardship. (Spicker, 1999) There are circumstances in which people might suffer from most, or all, of these problems simultaneously, but there are no necessary links; people in need are not necessarily in a situation different from others in the same society, people who have low resources are not necessarily excluded in need, and a person is dependent is not necessarily someone who lacks entitlement to resources.

Any of these definitions might be operationalised and measured, but that is not the same as identifying poverty as a whole. There is not a whole thing to measure. It may be possible to look at the elements of poverty separately, measuring the constituent factors, giving special emphasis to circumstances where these factors overlap. Alternatively, patterns of poverty can be deciphered and interpreted from a range of information. Indicators are pieces of information which are suggestive of those patterns.

Indicators

A social indicator can be seen as

> "a statistic of direct normative interest facilitating concise, comprehensive and balanced judgment about the condition of major aspects of society." (US Dept of Health, Education and Welfare, cited Carley, 1981)

Agencies want useful information. They do not necessarily want very precise information, because precision in identifying problems cannot be matched by precise policy responses, but they want information which is valid - that is, information which can meaningfully be related to the issues. They usually want to apply information which is easily available, because new information can be expensive to generate. And they want information is sufficient to work out what they have done, and what effect their policies have.

Often the data available to governments are inadequate. Indicators serve to patch together a rough understanding of issues. This is particularly true in developing countries, where surveys are difficult to mount and data are often incomplete unreliable and difficult to replicate. Many of the figures used by the UNDP for the Human Development Index are invented to fill in the gaps: they assume that missing values are worth 25%, on the basis that a country which is unable to say whether its people have access to water or education probably has a sizeable shortfall. (UNDP, 1999) Contrast a widely used, and much criticised, indicator: the World Bank estimates "poverty" as happening when people's income is a dollar or two dollars a day. The indicator is criticised in no less than six chapters of a recent edited collection on *World Poverty*. (Townsend, Gordon, 2002) Understood as a measure, a dollar a day cannot be defended. Poverty cannot be summed up in a single measure; the standard is way too low; income is not enough; it is not really possible to say what a dollar a day means in many societies; and the standard is not genuinely comparable. This is all true, but the criticisms miss the point. A dollar a day is not much like a measure of poverty, but it is useful. It is easy to understand, accessible, and cheap. It gives some idea of whether problems are getting better or worse.

The term "indicator" means what it says, and it was chosen for that reason. (Brand, 1975) An indicator is a signpost or pointer. Our understanding of social issues is flawed, definitions are questionable, recording is imperfect and data are incomplete. The best we can hope for is a compromise. Indicators are used as a practical means of managing information.

Indicators and measures

Indicators are not measures. An indicator is a way of representing something, not necessarily the thing itself. One thing can point to another. The unemployment rate is an indicator of economic activity; a child's weight is an indicator of developmental progress; premature mortality is an indicator of ill health. The assumption that any of these was a measure of the issues they indicate, however, would be misguided.

Measures can be indicators, and indicators can be measures. A good measure needs to be accurate - that is, it needs to reflect the nature of the thing it is measuring - and an accurate measure will co-vary exactly with the thing it measures. A good indicator makes people aware of the issues it is indicating, and, because it is supposed to reflect the characteristics of the problems it identified, it ought to have a consistent relationship with it. Like a good measure, then, a good indicator will follow the pattern of the thing it indicates: if income is an indicator of poverty, it should go up when poverty goes down, and it should go down when poverty goes up. Where measures are too expensive or difficult to obtain, indicators may be used as proxies: once an issue has been measured, the results of the measure can be used as a yardstick or "gold standard" against which the indicator can be checked.

It helps indicators to be accurate, but accuracy is not a primary criterion. Accuracy is not sufficient for an indicator, because the usefulness of an indicator depends on how well it points to the issue which is being indicated, not on how well it measures what it purports to measure. Assessments of unemployment or disability are useful for understanding aspects of poverty, because they are fairly loose terms and they tend to cover a wide range of related problems. Refining the definitions of the terms - for example, by removing people in training from the lists of unemployed people, or including people at lower levels of disability - may well improve their accuracy, but it does not strengthen their relationship to poverty. Equally, accuracy is not a necessary criterion, because once one has accepted that the correspondence between the indicator is imperfect, it is usually more important to look for cross-validation from other sources than it is to refine the measure.

Because both indicators and measures should co-vary with the issue under study, the uses of indicators and measures are often similar. It is often difficult to produce appropriate measures in practice, however, and sometimes accurate and precise measurement has to be sacrificed. Good measures can be bad indicators. For example, criminal convictions are more precisely measured and more accurately recorded than reported crime, and reported crime is more precisely measured and probably more accurate than survey reports on crime, but in each case the less well defined figure is more useful as a general social indicator. Conversely, good indicators can be bad measures: no-one should imagine that income, low birthweight or housing tenure are measures of poverty, but they are effective pointers to the issues.

Many commentators treat the issues around indicators as equivalent to the problems of measurement. A recent article in the *Journal of Social Policy* comments, for example, that

> "low income does not serve as a valid indicator of exclusion, because it fails to identify households experiencing distinctive levels of life style deprivation. Various studies have indeed found a substantial proportion of those on low incomes are not suffering from deprivation, which some households above income poverty lines do experience such deprivation. One response to the foregoing difficulties would be to dispense with income and go directly to the measurement of deprivation. However, to do so without understanding why current income proves to be such a poor predictor of deprivation would undermine the theoretical, policy and normative bases of any ensuing measures." (Wheelan et al, 2003, p.2)

The statements make sense if income is being considered as a measure of poverty, because it is not a good measure. But that does not mean that it is not a good indicator, or that it is a "poor predictor". On the contrary, the choice of words acknowledges that as income increases, the incidence of deprivation falls - we move from a "substantial proportion" of false positives to "some" false negatives. That is inexact, but indicators do not need to be exact. Income is quantifiable, the information is available, it has a comprehensible relationship with poverty and there is a strong, visible association with its subject.

I began this paper after attending three meetings in close succession: a European Union conference paper on indicators of social inclusion (Atkinson et al 2002), a consultation by the British Department of Work and Pensions on *Measuring Child Poverty* (DWP 2002), and a session presented by Edinburgh Council on "Holistic indicators of social inclusion" (Edinburgh City Council 2002). In each case, the problems the meetings were addressing were about the refinement of indicators. All three argued, in different ways, for more precise and reliable measurement of data. That is a defensible call, but it is not necessarily consistent with the

objectives of producing indicators.

The EU study on indicators of social inclusion requires indicators to "identify the essence of the problem and have a clear and accepted normative interpretation" (Atkinson et al 2002, p 21). "Identifying the essence of the problem" implies that the indicator relates directly to the problem under study. Many indicators, however, relate only indirectly to social issues. Infant mortality does not identify the essence of ill health, but it is strongly associated with it. Similarly, income does not "identify the essence" of poverty, and it is not equivalent to any of the clusters of meaning identified at the outset.

The Edinburgh project began with the objective "to develop and test a set of pilot indicators which are capable of measuring the true reality of social exclusion" (Edinburgh City Council, 2002). That is beyond the scope of indicator research, because indicators depend on a compromise between the desire to measure reality and the practical need to develop effective signposts towards it. The limitation is recognised in the final report.

The same gap between aspiration and practice can be found in the DWP document on "Measuring Child Poverty". It argues that "A good measurement approach should aim to ... encompass the different definitions of child poverty". It is probably true that a good measurement approach would do this, but there has to be some doubt as to whether this is what the DWP really wished to achieve. The DWP were examining options for developing a simple, summary figure which could help to guide policy. That is the role of an indicator rather than a measure. Measures can, of course, act as indicators in certain circumstances, so it is not self-evidently wrong to try to have both. But if an indicator, rather than a measure, is what is needed, it is not necessarily appropriate to try to "encompass" issues rather than pointing in the right direction. Like "identifying the essence" of a problem, this is asking to do the wrong thing. Low birthweight does not begin to "encompass" issues of poverty, and income hardly does so, but that does not mean they are not good indicators. If we were to take this literally, it would exclude some of the most useful and practical approaches available.

Indicators and quantification

Indicators are generally quantitative. In theory, it should be possible for indicators to be qualitative. The main difficulty with qualitative indicators is the problem of attaching value to the indicator. The primary purpose of indicators is operationalisation - the translation of a concept into terms which can be applied to the problem in hand. For an indicator to work, it needs to vary with the circumstances which are being indicated. It has to register change in a way which makes it possible to say if something is better or worse. A qualitative variable - one which records factors as X, Y and Z rather than 1, 2 or 3 - may convey the information, but it does not do so directly. This tends to lead us in the direction of quantification, because only numbers give a sense of change.

The emphasis on numbers leads to certain problems. Monty Python once featured a sketch of a society of gentlemen dedicated to putting one thing on top of another. At times, it seems as if members of the society have migrated to the statistical division of certain branches of government. We all know that two plus two equals four, which makes it difficult for some people to view two incommensurate figures - e.g., an area with 30% unemployment and 6% chronic sickness - without adding the numbers together. (The example is not as fanciful as it

might appear - the Scottish Office's measure of urban deprivation in the 1980s did as much.)

It is not necessarily the case that they can be treated like numbers by being added together, averaged or otherwise subjected to calculation. Numbers are ordinal (two is greater than one) and aggregative (two plus two is four). Many social problems, however, are incommensurate - neither ordinal nor aggregative. Housing is not self-evidently more or less important than education, and a person with three problems is not necessarily worse off than someone with one. In the OPCS studies of disability, the points scheme used to measure disability was disregarded after the largest three problems were entered; the experts who validated it felt, probably rightly, that after the three largest problems were taken into account, any others had only marginal weight. (OPCS, 1988)

Indicators and aggregation

Most of the indicators we are familiar with are aggregates, averages based on aggregate data, or proxies for aggregation. Figures like the unemployment rate, crime figures or homelessness applications are put together from a range of source material about individual circumstances. Typically, because enumeration is expensive and liable to systematic biases, some indicators are constructed from sample data or surveys, like labour market studies, and many official statistics are estimated, like assessments of local populations between census dates or current expenditure on social security benefits. It is possible, in principle, to use illustrative returns, like the responses of a citizens' panel as indicators of public opinion. All of these examples, however, rely on a theoretical link between survey evidence and enumeration - the belief that if the topic could be enumerated accurately, the survey evidence would be equivalent to it. In that sense, aggregation is the normal procedure.

The emphasis on aggregation prompts the observation that the way to improve the indicator is to refine the accuracy and precision of constituent measures. This does not necessarily follow. In the first place, one of the effects of improving accuracy is to focus on indicators that are more consistently quantifiable; these are not necessarily the best signposts. Second, the attempt to make things more precise can create its own problems, increasing complexity and the costs of data collection. This is one of the central vices of health measurement scales: the search for precision has led to proliferation of different assessments. Streiner and Norman comment: "you may conclude that none of the existing scales is quite right, so it is appropriate to embark on one more scale to add to the confusion in the literature." (Streiner, Norman, 1989, p.4) Third, even if the constituent parts are finely identified, the effect of aggregation is still to round the figures out.

The validity of indicators

There are two main alternative approaches to the selection of valid indicators. The first, and probably the best, method depends on establishing a theoretical link between the indicator and the thing which is indicated. If we want to use indicators to understand social exclusion, Levitas comments, we need both to have an understanding of the nature of exclusion and the causal relationships it has with other factors. (Levitas, 2000, p.365) We may well select indicators because of their (presumed) theoretical relationship to the factors under study. Townsend's 1987 Deprivation Index, which has stood up reasonably well to comparisons with other indices over time, is based on four indicators as characteristic of different elements of

deprivation. (Townsend, 1987)

This is not the only approach, however. The alternative route is to identify an observed relationship. Indicators are valid if they indicate what they are supposed to indicate. They can point in the right direction without necessarily pointing exclusively to the right issue. Once we establish that a factor is generally associated with the area of concern - for example, the link of housing tenure with low incomes, or social class with ill-health - we can use one factor as an indicator of another.

Levitas is particularly critical of this approach:

> "Rather than moving, as research ideally should, from definition to operationalisation to data collection, the process is largely reversed: we move from available data to an implicit definition in the flawed data sets which already exist, and which never needs to be closely scrutinised." (Levitas, 2000, p.366)

She is right: the character of the data tend to drive definition, rather than the other way round. But the same pitfalls can be found in other measurement and research surveys: examples in poverty research are the driving out of previous understandings of poverty by a focus on income and household budgets, and the use of individualised poverty research to invalidate understandings of area deprivation in the 1970s and 80s. There are reasons for beginning with available data, rather than a theoretical understanding. There is a tendency in poverty research for researchers to say, like Humpty-Dumpty, that terms mean what they say they mean. (Ryan, 1986) Good theorisation relies on observation; definitions and operationalisation have to be founded in usage. Using data to form ideas - the characteristic approach of much indicator research - is not necessarily wrong, and it may be more valid than the alternatives.

Indicators are valid when they form a reasonably consistent relationship to the issue they are being used to mark out. Measures generally reflect the characteristics of the issues they are measuring: measures are more valid when they reflect the issues they are measuring more accurately and precisely. The same is not necessarily true for indicators. Some indicators cover much more broader ground than is required to consider an issue: broad-brush indicators like GDP per capita, income or urbanisation are useful in understanding poverty, but they are sufficiently different from it to mean that greater refinement of the figures does not necessarily help refine understanding of the issues. The effect of refining the figures - making them more accurate and precise - may even make them less helpful. This was strikingly the case in the alteration of unemployment counts in the UK in the 1980s, where successive downward alterations in the statistical basis of the figures did very little to improve understanding of unemployment. (Atkinson, Micklewright, 1989)

Conversely, some indicators cover much more specific insights to issues. If broad-brush figures offer mirrors to nature, where everything is reproduced, others offer miniatures: a small section is examined in detail. (The terms "mirror" and "miniature" are not drawn from social science; I have filched them from Anthony Trollope, the Victorian novelist.) In mirrors, we may be interested in the pattern or distribution of issues in a population. In miniatures, we may have identified special circumstances which we think illustrate more general issues. It is questionable whether human rights issues can be judged in the round from the treatment of prisoners, travellers and the victims of sexual assault, but a study which focused on these issues would not self-evidently be making a false start. Indicators are, after all, about signposts, not the thing itself, and the validity of using selected signposts depends more on the relationships of

the elements than the ability of the these elements to encompass the issue overall.

In principle, the validity of indicators can be reinforced by combining indicators together. Issues which are difficult to identify or research are generally focused on by "triangulation" - approaching the problems in several different ways. The same principle is widely used in measurement: researchers into attitudes, for example, do not waste their time looking for a single question that encapsulates every shade of attitude, and attitudinal questions do not claim to be precise or accurate on an individual basis. They ask questions in bundles, so that they can cross-reference the answers. The effect of using multiple alternatives is to identify a common area of co-variation, reducing the impact of individual factors while reinforcing the points which are common across a range of factors. Multi-item indices represent complex concepts and issues better than single indicators can. They are more reliable than single indicators, because they reflect the movement of a constellation of topics which vary together, rather than the fluctuations in a single factor. Using multiple indicators helps to identify underlying issues and trends, and reduces the risk that any specific element will distort the overall picture. However, indicators are not necessarily intended to be precise or fully to represent concepts or issues, and gains in the elimination of errors may need to be offset against other issues. There is the risk that the effect of combining a series of indicators may be, not to confirm the importance of issues, but to mask them. Many of the indicators routinely used in appraisals of poverty refer to relatively minor issues. In the measurement of child poverty, in particular, major issues affecting large numbers of people, like income, poor health and social exclusion, may dwarf smaller numbers in important problem areas, like school exclusions or serious unintentional injury.

Alternative approaches to the construction of indicators

There are three main ways of presenting indicators: headline indicators, summary indices, and multiple indicators.

Headline indicators

The first approach is to use a "headline" indicator: a simple, selective view. Commonly used examples are the use of income inequality as an indicator for poverty, infant mortality as an indicator of general health in developing countries, or the growth rate as a proxy for economic development. Neither describes the issues fully and precisely, but they serve the turn, and even if there are exceptions they can be defended in broad terms.

> The argument for headline indicators is given by the DWP in the following terms: "a single indicator is appealing because of clarity. A headline or proxy indicator is needed in order to monitor progress ... in the absence of any other measure the *Households Below Average Income* measure will be used by politicians, the media and others as the principal measure of progress ..." (Lisa Harker, cited DWP, 2002, p 19.)

The argument seems to be, then, that if a headline indicator is not devised expressly, one will be used anyway - such as the indicator of economic distance, based on the EU's approach, currently presented in HBAI.

What makes a good headline indicator? First, the indicator has to be available and (at least apparently) comprehensible. Second, it has to have a fairly general application, pointing to more than it appears. Infant mortality is strongly associated with adult health, economic growth

with welfare, and income with command over resources. (There are obvious exceptions: for example, economic growth can imply reduced welfare in conditions of aggravated inequality.) Third, measures to improve the headline indicator should have associated benefits in other areas. A high rate of infant mortality is usually indicative of poor nutrition coupled with infective and parasitic diseases, and measures which address those issues have a much wider implication for health.

Most headline indicators of poverty are related to income. Income is a valuable indicator, for several reasons. It is relatively easily measurable. It is strongly associated with many other indicators of welfare, though there are reservations to make about this: there is a tendency in some research to jettison indicators of welfare which do not show the same associations (Wheelan et al, for example, eliminate housing and environmental indicators from their analysis on that basis: Wheelan et al, 2003, p.4), which implies a certain circularity in the claim of a strong association. Increased income is usually indicative of increased welfare, though once again this is not necessarily true: in developing countries, a child who leaves school to sell things in the street will have a higher income, and will add to GDP, but this is suggestive of lower welfare. More seriously, *The Economist* points out, "a disease that kills millions of children and old people can produce a rise in GDP per head if those aged 15-45, the most economically productive members of society, are still standing but there are fewer people in total to share the wealth." (Economist, 2003)

The main problem with the use of income as a headline indicator has been the tendency for the focus on income to drive out other forms of discussion. Many commentators make no distinction between concepts of poverty and income thresholds. This tendency has been evident from the earliest researches into poverty. When Charles Booth undertook his early researches on poverty in Victorian London, he identified poverty primarily through observation and reports on life style, but criticism led to him producing illustrative household budgets, and the production of such budgets dominated discourse on poverty research for most of the next hundred years. The risk of using a strong, valid headline indicator is, then, that the indicator will be taken to be the thing itself, with consequences for the focus of concern and the direction of policy.

Summary indices

The second option is to use a summary index. Examples might be the Index of Multiple Deprivation, or the Human Development Index used by the United Nations. A summary index consists of a set of indicators which are compiled in order to produce a single composite indicator. In order to construct a composite, the material which is being constructed has to be standardised. Common methods are the use of percentages or proportions; indices of urban deprivation have been based on Z scores, which are based on the relative position of a proportion within the overall distribution, and chi-squared distributions (see e.g. Robson et al, 1995).

There are several methods of combining scores mathematically. One is commonly done by averaging or adding the scores together, for example in the UNDP's Human Development Index. The EU study argues that "the portfolio of indicators should be balanced across different dimensions" and that "the weight of single indicators in the portfolio should be proportionate" (Atkinson et al, 2002, pp 24, 25). In the context of poverty, which is a

complex, multidimensional set of issues, the process of calculation has the disadvantage that composite calculations may lead to one factor disguising the effects of changes in another. For example, an index of area deprivation including pensioners and unemployment - such as that used by the Department of the Environment for many years (DoE, 1983) - might find changes in one outweighed or neutralised by changes in the other. The magic of averages allows variations in incommensurate data to cancel each other out.

An alternative method for combining scores is to assign weightings in a points scheme. Points schemes crop up in many, sometimes surprising, places - like the allocation of public housing, admission to university, the assessment of industrial injuries or penalties for motoring offences - where normative judgements have to be made on a consistent basis. An example in the construction of social indicators is the Jarman index, which is used in the health service to pay GP's for certain social factors affecting the workload in their practice. Jarman sought responses from GPs about workload from their practices and processed the material after normalising data and attributing weights. (Jarman, 1983) The OPCS studies of disability made use of similar principles to allow some consistent judgments to be made between assessors. (OPCS, 1988) The Index of Multiple Deprivation uses normative weighting to combine scores between different domains. (Social Disadvantage Research Centre, 2003, p.36)

The principal method for combining material is to calculate weights mathematically. Multivariate analysis is the method of choice. It identifies quantitative relationships between variables and assigns values according to mathematical formulae. The Local Deprivation Index (Robson et al, 1995) and its successor, the Index of Multiple Deprivation (Social Disadvantage Research Centre, 2003) use factor analysis. The IMD applies this in specific domains - for example, in the assessment of health and disability, crime, and education and skills for children and young people. Noble et al give an exemplary summary of the rationale:

> "It is hypothesised that an underlying factor exists at an ecological level that makes these different states likely to exist together in a local area. This underlying factor cannot be measured directly but can be identified through its effect on individuals ... These variables need to be combined at an ecological level to create an area score. Fundamentally this score should measure, as accurately as possible, the underlying factor.
>
> There are a number of problems in achieving this goal. The variables: [1] are measured on different scales, [2] have different levels of statistical accuracy, [3] have different distributions, [4] may or may not apply to the same individual and [5] measure, to different degrees, the underlying factor imperfectly. ... The premise ... is that the underlying factor is imperfectly measured by each of the variables in the dataset but that the variables that are most highly correlated with the underlying factor will also be highly correlated with the other variables." (Social Disadvantage Research Centre, 2003, p.53)

The complexity of multivariate analysis generates its own kind of problems. First, the maths depend on certain assumptions about the data - for example, that the data need to be normally distributed. Despite the common reassurance found in statistics textbooks, that numbers will tend to normality as sample sizes increase, I have never encountered any social data which do. Outliers have to be pruned away, and missing values have to be swum round. Before they can be processed, data have to be transformed to make them reasonably compatible with the

statistical assumptions. (The standard transformations are logarithmic, square root and reciprocal, and even then there may be some gap between the theory and the practice. Noble et al use a combination of exponential transformations with a"shrinkage" of area data, adjusting data for variability within and between areas to reduce the impact of different sizes of area on apparent proportions.) Second, the variables are supposed to be independent. Unfortunately, in practice they are usually interdependent and it is difficult to distinguish effects. The construction of the equation is built on a particular relationship, at a particular point in time. The computer will normally begin with the strongest relationship and weed out others which seem not to make a difference. When there is this sort of overlap - "multicollinearity" - it matters crucially which factor goes into the analysis first. Third, the significance of relationships is determined by how well the line fits, not by the impact of one variable on another. In practice, the slope of the curve in practice is usually much more important than the precision with which it is drawn, but most analyses focus on the latter, not the former. Lastly, hardly anybody understands what on earth is going on in these formulae. This is sometimes seen as an advantage - a "technological fix" to silence political opposition - but it is not necessarily helpful.

Irrespective of the specific method used to develop composite indices, there are some general points which apply to all of them.

Validity. Indices have to represent or indicate what they are supposed to represent, and cross-validation is difficult. An index can be cross-validated with issues which are not part of it, like income or benefit receipt, or by the association of items within the index itself. This leads to the same kinds of problem associated with multivariate analysis: the items should be correlated with each other, but not too highly, because a very strong association suggests that the same issue is being counted twice. This is a matter of judgement.

Reliability. As measures, indices are often held to be more reliable than individual figures, because the effects of fluctuation in individual variables are compensated for by others. The same may be true of indices used as indicators, but reliability is not necessarily the sign of a good indicator. The validity of indices will depend on their association with the subject under study, which is a theoretical issue rather than a technical one.

There may however be problems with reliability. Indices which are reliable within a particular social context, or at a certain period, are not necessarily transferable to other circumstances. The Jarman index was intended to reflect social patterns of deprivation which were likely to increase GPs' workloads. Twenty years later, other factors have emerged (like drug dependency, which for those practices which deal with it imposes an extraordinary burden of support and administration.)

Quantification. The construction of indices tends to presume linear mathematical relationships, whether this occurs through aggregation or some kind of multivariate analysis. General problems of aggregation and quantification were noted in the previous sections, and for indices the assumption of mathematical relationships between incommensurate types of data creates further problems.

Inclusion and exclusion of relevant factors. Exclusions lead to important issues being ignored: housing standards, for example, now rarely feature in indices of deprivation. Over-inclusion can lead to excessive weight being given to particular factors.

Weighting. Factors have to be given appropriate weights, which depends partly on appropriate quantification, and partly on normative judgement. Normative points

schemes have at least the advantage of making the process explicit, but all weighting, including that made through averaging or multivariate analysis, implicitly reflects some kind of judgement, and it should be recognized that weights are neither self-evident nor beyond argument.

Norms and values. In some cases, the norms and values contained in summary indices will be evident: it is difficult to present material on human rights, for example, without a fairly explicit statement of value. However, because summary indices tend to conceal their constituent elements, they also tend to conceal the norms and values implicit in the constituent judgments. Some of these effects are indirect: for example, the implication of emphasising long-term unemployment is often to emphasise the disadvantage experienced by males, who are more likely to be counted in those figures than women. Some are direct: the choices of inclusion, exclusions and weighting have an immediate effect on the relationships identified through such indices.

Multidimensional indicators

It is possible to retain several dimensions of indicators. An example is the use of indicators in the government series *Opportunity for all* (Cm 5260, 2001). The indicators for children and young people, as an example, are these:

Improving family incomes

Children in workless households
Low income (three indicators)

Early years and education

Key stage 1 attainment (7 year olds)
Key stage 2 attainment (11 year olds)
16 year olds with one GCSE
19 yerar olds with level 2 qualification
Truancies
School exclusion
Attainment of children looked after by local authorities

Quality of life

Housing conditions
Infant mortality
Smoking rates (pregnant women and children 11-15)
Serious unintentional Injury
Re-registrations on Child Protection Register

Transition to adult life

Teenage conceptions
Teenage parents not in education, employment or training
16-18 year olds in education.

There are three main arguments for using multiple indicators. The first is methodological. Multiple indicators are used for cross-confirmation, or triangulation. If indicators are concerned with complex problems, multiple indicators help to examine a problem from

different perspectives.

The second argument is practical. Multiple indicators offer more detailed, disaggregated information. Local and voluntary agencies engaged in policy making are increasingly required to provide baseline information and indicators as evidence of their effectiveness; the more detail that is made available to them, the better able they are to respond. That is the source of the demand for neighbourhood statistics, which have made current indicators available at a detailed local level, and for the use of several domains in the Index of Multiple Deprivation. The same argument applies at every level of the policy making process.

Third, there is an argument from principle. When dealing with complex, multidimensional issues, the effect of aggregating and simplifying is to reduce the complexity at the expense of minor issues, which are over-ridden by weightier ones. Whether this matters depends on the minor issues. Poverty is a complex, multidimensional set of issues, but they have something important in common: they all make a normative claim for attention. If a concept like "poverty" is aggregated as a whole, specific, lower order elements - like gender inequalities, homelessness, or educational attainment - are likely to be ignored.

The main disadvantage of using multiple indicators is that they are complex, and maybe too complex, for easy digestion. The presentation can be simplified by classification or profiling. The indicators in *Opportunity for all* are classified - figures for children, old people and so forth are presented on separate pages. Profiling works by classifying material hierarchically, but without eliminating or disguising lower orders of data. (Spicker, 2001) Neither classification nor profiling, however, achieves the level of simplification offered by summary indices. The Index of Multiple Deprivation combines the elements by linking the advantages of cross-confirmation and multiple criteria with the capacity to disaggregate information in a series of different categories or "domains" of concern.

The EU study makes an essentially political argument against using a broad range of indicators. They write:

> "No set of indicators can be exhaustive, and there are costs in terms of lost transparency from having too extensive a range of indicators. Too large a set of indicators risks losing credibility, if member states can simply pick and choose." (Atkinson et al, 2002, p.24)

Julio Boltvinik also points to another implication of using multiple indicators: the more indicators which are used, the more people are going to be defined as poor. (Boltvinik 1996, p 290) His evidence for saying this is empirical, but the point could be anticipated theoretically. If indicators are not collinear, there will be for each an area which does not overlap with others. If the indicators refer repeatedly to the same issues, the marginal difference made by each additional indicator will be limited; but if they identify different problems, the population they refer to will increase. Conversely, a smaller range of indicators defines a smaller population as poor. Whether this is an advantage or disadvantage depends on the political perspective of the observer.

There is a danger in adopting an explicitly political approach. A set of figures selected on this basis can be altered with limited, focused policy intervention. The use of school exclusions, for example, has been affected by policies making exclusions more difficult. Effectively, indicators may be treated as performance targets. In the process, they may distort the policies

they are supposed to be signposting.

Implications for policy

This paper has made a case for approaching indicators in a particular way. It argues that refined measurement is often inappropriate; that we need to accept a degree of uncertainty in indicators; and that indicators have to be understood with the caution that this uncertainty merits. A good indicator should be understood as a pointer, not a measure. It should be accessible, robust and accompanied by other pointers. The kinds of summary index favoured in these policy documents are fraught with problems. They conceal key issues, hide the values and concepts which are implicit in the exercise, and are vulnerable to mathematical accident.

The DWP consultation paper on child poverty proposes four alternative approaches to providing "headline measures" of poverty. The options are

- "a small number of multi-dimensional headline indicators"
- an index of a small number of indicators
- a measure of "consistent poverty", and
- a core set of indicators of low income and consistent poverty.

These options have the common problems of all summary indices. They will compound and so conceal a range of issues, like income, health and social exclusion. Relatively "minor" issues currently contained in *Opportunity for All*, like serious unintentional injuries, exclusion for school or teenage parents in education, disappear from the analysis, because they cannot be allowed to distort the overall picture. Some values and weights are concealed, such as the limited weight given to housing. And any of the options is vulnerable to influence of the size, number and association of problems.

The approach advocated for analysis of social exclusion in the European Union is more elegant, and more satisfactory. The team proposes a hierarchy of different types of indicators, including a limited core for all countries, and second- and third- level indicators used for different purposes. The highest level, level 1, will consist of a few selected indicators, used in each country of the EU. Level 2 indicators are intended to "support" the key indicators and will again be collected across the EU. Level 3 indicators will be developed within the member states. (Atkinson et al, 2002) This approach makes it possible to retain some of the richness of detail which is offered by multiple indicators, though the process of selection removes some of the opportunities for cross-referencing material and triangulating approaches to issues. The same process of selection, however, also has the effect of rejecting information which might have otherwise been used in policy formation, and it is noteworthy that the team's recommendations tend to be concerned with a limited understanding of social exclusion, based primarily on material deprivation rather than social relationships.

There are good arguments for the relatively sophisticated approach of the Index of Multiple Deprivation, but it is also vulnerable to many of the objections made of mathematical techniques in this field. The best approach is probably the simplest: the classified presentation of multiple indicators favoured by *Opportunity for all*. This approach is informative, offering information which can be adapted to different contexts and uses; robust, because it is not as dependent as the alternatives on the assumptions or calculations made in presentation; and flexible, because it leaves scope to incorporate new information as it becomes available. If recent documents are a guide, however, the trend seems to be away from this approach.

References

A B Atkinson, J Micklewright, 1989, Turning the screw, in A Dilnot, I Walker (eds) *The economics of social security*, Oxford: Oxford University Press.

A B Atkinson, B Cantillon, E Marlier, B Nolan, 2002, *Indicators for social inclusion in the European Union*, Oxford: Oxford University Press.

J Boltvinik, 1996, La pauvreté en Amérique latine, *Revue internationale des sciences sociales* 148 pp 279-295.

J Brand, 1975, The politics of social indicators, *British Journal of Sociology* 26 pp 78-90

M Carley, 1981, *Social measurement and social indicators*, London: Allen and Unwin.

Cm 5260, 2001, *Opportunity for all*, London: The Stationery Office.

DoE (Department of the Environment, Inner Cities Directorate), 1983, *Information note no 2: Urban Deprivation.*

DWP (Department for Work and Pensions), 2002, *Measuring child poverty: a consultation document*, London: DWP.

The Economist, Epidemics and economics, *The Economist* 12th April 2003, p.80.

Edinburgh City Council, 2002, *Holistic indicators for social inclusion*, at http://www.onecity.org.uk/HolisticIndicators/FinalReport/

B Jarman, 1983, Identification of underprivileged areas, *British Medical Journal* 286 pp 1705-1709.

OPCS (Office of Population Censuses and Surveys), 1988, *The prevalence of disability among adults in Britain*, London: HMSO

A Ryan, 1986, Poor relatives, New Society 18th April 1986, p 25.

Social Disadvantage Research Centre, *Scottish Indices of Deprivation 2003*, Oxford: Social Disadvantage Research Centre.

P Spicker, 1999, Definitions of poverty: eleven clusters of meaning, in D Gordon, P Spicker (eds), *International glossary of poverty*, London: Zed Books.

P Spicker, 2001, Cross national comparisons of poverty: reconsidering methods, *International Journal of Social Welfare* 10(3) 2001 153-63.

D Streiner, G Norman, 1989, *Health measurement scales*, Oxford: Oxford University Press.

P Townsend, 1987, Deprivation, *Journal of Social Policy* 16(2) 125-146.

P Townsend, D Gordon (eds), 2002, *World poverty: new policies to defeat an old enemy*, 2002, Bristol: Policy Press.

B Robson, M Bradford, R Tye, 1995, The development of the 1991 Local Deprivation Index, in G Room (ed) *Beyond the threshold*, Bristol: Policy Press.

C Wheelan, R Layte, B Maitre, 2003, Persistent income poverty and deprivation in the European Union, *Journal of Social Policy* 32(1) pp 1-18.

UNDP (United Nations Development Programme), 1999, *Human Development Report 1999*, New York: Oxford University Press.