

CREST

**CENTRE FOR RESEARCH INTO
ELECTIONS AND SOCIAL TRENDS**

Working Paper
Number 66
September 1998

How Comparative is Comparative Research

By Roger Jowell

The *Centre for Research into Elections and Social Trends* is an ESRC Research Centre based jointly at the *National Centre for Social Research* (formerly SCPR) and the *Department of Sociology, University of Oxford*

<http://www.crest.ox.ac.uk>

CREST Paper No 66 : How Comparative is Comparative Research ?

by Roger Jowell

also published in American Behavioural Scientist, 42, 2, 168-177.

Abstract

We tend to accept lower standards of rigour in crossnational surveys than in national surveys, leading to heroic conclusions about differences between nations on the basis of obviously faulty data. Arising perhaps from a commendable wish to respect different cultural norms, even some of the most conscientious crossnational studies make the mistake of permitting considerable variations by country in the type and quality of the methods they deploy. Meanwhile, analysts of crossnational data frequently abandon offering explanations and interpretations in favour of league tables of distributions showing merely 'gee whiz' national differences. This article acknowledges the formidable obstacles in the way of achieving rigour in large-scale comparative studies and offers ten possible 'rules' to mitigate the difficulties. It suggests that bigger is usually worse and recommends routine crossnational collaboration in analysis and interpretation - not just in design, development and execution - among scholars in each of the nations under the microscope.

Introduction

The importance and utility to social science of rigorous crossnational measures is incontestable. They help to reveal not only intriguing differences between countries and cultures, but also aspects of one's own country and culture that would be difficult or impossible to detect from domestic data alone. Or, as Durkheim famously put it, "Comparative sociology is not a particular branch of sociology: it is sociology itself." The purpose of this article, however, like so much of the literature on comparative research, is to underline the pitfalls of crossnational measures, to urge greater caution in their interpretation and to suggest a number of partial remedies.

The strict standards we apply to the evaluation of national surveys are too often suspended when it comes to crossnational studies (Scheuch, 1966; Teune, 1992). We make allowances for the fact that such studies tend to be more difficult than national studies, that they involve clashes between cultural norms and compromises of methodological purity in the noble causes of mutual respect and downright practicality. These act as powerful arguments for a tolerance of design imperfections in crossnational studies, an acceptance that they employ questions or techniques that are far from functionally equivalent (Verba, 1971). On the other hand, too much tolerance of faulty methods - especially of ones that could be a serious source of error or bias - is hardly to be recommended. The demands of good science require us not to turn a blind eye to such imperfections when we come to interpreting the data arising from these studies, nor to accept at face value differences between nations that we know may well be illusory and artifactual.

These problems arise not only in respect of comparative research based on national data sources which were never designed to be compared across countries in the first place, but also in respect of studies especially designed to elicit equivalent crossnational measures. Part of the source material for this article is the International Social Survey Programme (ISSP), widely - and fairly - regarded as a conscientious, rigorous and successful model of a crossnational social survey. A recent methodological report on one of

its annual rounds (Park and Jowell, 1997) reveals the intractability of the problem it faces in trying to ensure consistency of method and, therefore, of its measures across its member nations. Despite its high ambitions and the relative diligence with which it applies its standards and rules, inconsistencies between countries still abound. Some of these differences may, perhaps, be dismissed as immaterial national idiosyncrasies, but others - such as subtly but critically different sample designs - cannot.

THE PRINCIPLE OF EQUIVALENCE

To some extent, of course, the problems that beset crossnational surveys are ever-present in national surveys too, since no nation is homogeneous with respect to vocabulary, modes of expression, levels of education and so on (Laboy, 1964). But to the extent that heterogeneity always presents a problem for surveys at a national level, it is a major obstacle to success in international surveys.

After all, quantitative surveys depend for their reliability on a sort of principle of equality or equivalence. So, probability of selection in a national sample should be equal (or at least known and non-zero) to satisfy the demands of representativeness. Co-operation rates must not vary greatly between subgroups to maintain that principle of equal representation. Questions should have a broadly equivalent meaning to all respondents to ensure that variations in the data derive from differences in their answers rather than in their interpretation of the questions. Coding schemas are devised to ensure that it is the codes rather than the coders that account for differences in the distribution of answers. And so on. A great deal of work in national surveys goes into the sheer process of attaining standardisation or equal treatment of respondents and potential respondents, and of different classes of respondent. Only to the extent that a national survey succeeds in these goals are its findings likely to approximate to some sort of social reality.

In trying to achieve such equality or equivalence in multinational surveys, however, researchers are always likely to be defeated to some extent by a host of major cultural differences - among which language and idiom are only one very large component. Much has been written over more than three decades about the particular problems in comparative research of cultural variations that can so easily disguise and obfuscate the desired object of measurement - from the 'courtesy bias' in South East Asian studies (Jones, 1963), through the 'spurious lexical equivalence' of denotations which disguise major differences in connotations (Deutscher, 1968; Rokkan, 1968; Cseh-Szombathy, 1985), to the call for more effective cooperation between linguists and social scientists (Hantrais and Ager, 1985). The fact is that different languages are not just equivalent means of defining and communicating the same ideas and concepts. In many respects they reflect different thought processes, institutional frameworks and underlying values (Lisle, 1985; Harding, 1996).

Researchers are also likely to be frustrated by country-specific differences in methodological and procedural habits - such as preferred modes of interviewing, deeply-ingrained but often idiosyncratic sampling methods, major differences in what are considered to be acceptable response rates, esoteric uses of visual cues, different methods of training interviewers, specialised socio-demographic classifications and so on (Mitchell, 1965). They will soon discover - if they are lucky - that certain 'standard' conceptualisations of, say, cleavages within one country (such as the left-right continuum, or the liberal-conservative one) have no direct counterpart in another, and that identical questions about concepts such as, say, strong leadership or strong government, or nationalism or religiosity, tend to be interpreted quite differently in different countries because of their different cultural, social structural and political conditions (Miller et al, 1981; Scherpenzeel and Saris, 1997; Saris and Kaase, 1997).

A great deal of comparative work is, of course, based not on tailor-made crossnational studies but on

broadly parallel data collected initially for strictly national purposes. It goes without saying that international comparisons based on these sorts of data are highly problematical. The more pressing question, however, is the extent to which taylor-made crossnational surveys actually improve the rigour of crossnational comparisons by successfully overcoming deep-rooted cultural and methodological divides. The ISSP is a useful case study.

Principles and practice

Academically-driven and well-regulated, the ISSP is a 28-nation annual social survey, (and still growing), which aspires to admirable standards of rigour and standardisation. It has high entry qualifications, each potential member nation having to guarantee to deliver certain collectively-determined technical standards. The ISSP does not therefore conform to the classic 'safari-model' of international surveys where social scientists in one or more developed countries breeze into one or more developing countries and impose their methods and questionnaires. Instead, it operates a sort of democratic model in which the member nations, having agreed a set of standards, are then supposed to apply them rigorously and consistently.

In practice, however, this does not happen as well as it should. For instance, as far as sampling is concerned, the ISSP rules are uncompromising. Each member nation is required to employ a national probability sample for its annual survey and to calculate its response rates in a standard form. Quota sampling, still so beloved in some of the ISSP member nations, is outlawed by the ISSP rules. Yet in the recent methodological survey of what member nations actually do on the ground (Park and Jowell, 1997), it turned out that, while all members did respect the basic rule of employing a form of probability sampling as a starting point, by no means all nations ended up with a recognisable probability sample at the end. Non-contacts and refusals in several countries, instead of being treated as such and being counted as categories of non-response, are in fact 'cloned' or substituted, usually by some quota method. In some cases, therefore, there is virtually no recorded non-response, since - just as in a traditional quota sample - it is normal practice to ignore (or suppress) it than to collect details of non-contacts and refusals at each stage of the process. The result is not only that the recorded response rates in these countries are improbably high, but that the 'availability biases' in their sample composition are also likely to be larger than those in other countries.

In addition, the definition of adults turns out to vary somewhat between nations, partly because - as a self-financed survey in each country - the particular module of ISSP questions is very often incorporated into another existing survey vehicle as a cost-saving measure. Some nations exclude people over the age of 74 or 75; others define adults as anyone aged 14+; most include in their universe everyone over 18 living at private addresses. These differences are hardly surprising given that ISSP rules embarrassingly fail to specify a standard definition of 'adults' and, in particular, that people over 74 still qualify as such. This is merely a manifestation, all too common in crossnational surveys, of a failure to recognise in advance just how fundamentally survey norms, like other aspects of culture, differ between countries. In this case the differences in national practice are partly remediable, since data analysts can in principle construct consistent samples across countries of people aged between 18 and 74. But it is highly doubtful that many would actually do so, and in any case the penalty for such adjustments would be to 'disenfranchise' older people altogether.

Not uncommonly, English is laid down as the ISSP's official language of discussion and questionnaire drafting. (In fact - with an admirable sense both of detail and cultural appreciation - the official language is even more closely specified as 'British English'.) Every other country then has to translate the

questions into their own functionally equivalent words and phrases. We discover, however, from responses to the methodological questionnaire, that this process too leaves something to be desired. In the first place, some of the other native English-speaking countries in the ISSP (among which are Australia, Canada, Ireland, New Zealand and the United States) make the potentially serious mistake of not bothering to translate the questionnaire at all. Secondly, only one other country bothers to back-translate into English after it has done the initial translation and then to continue the process iteratively until achieving a near-perfect match. Moreover, it is rare among ISSP member nations even to employ specialist translators, presumably because they would be too expensive. Instead, nearly all members prefer to rely on their own (usually very considerable) expertise in English. Difficult as it is for a native British English-speaker to judge, it is hard to resist wondering whether some of the national differences that emerge from ISSP data are in reality merely artifacts of doubtful translations, where the translated word or phrase has acted as a slightly different stimulus from the one intended. To the extent that this happens, some fascinating crossnational differences that emerge may be partially (or wholly) illusory.

With the best will (and the best methods) in the world, some of these difficulties may well be near insuperable. Consider even the answer categories employed in 'British English' questionnaires. Are there comfortable and familiar functional equivalents in, say, Polish, Hungarian, Bengali or Japanese of the phrases "slightly agree" and "slightly disagree", or "just a bit"? (Davis and Jowell, 1989). When the ISSP was in the throes of designing its 1991 module on religion, the Japanese delegation eventually came to the reluctant conclusion that there was no appropriate word or phrase in Japanese which approximated to the concept of "God". In the end, of course, they managed to come up with a doubtless somewhat tortuous circumlocution designed to get across the basic meaning of the Judao-Christian-Islamic concept of 'God'. But beware of data that depends on such contrivances based on collegiality! Similarly, when we in Britain wish to find out about people's political orientation on a 'left-right' or 'liberal-conservative' continuum, we cannot successfully apply the measures so familiar in many other countries. Instead we require around ten agree/disagree items - five to locate them on a left-right continuum and another five to locate them on a libertarian continuum - followed by a good deal of data manipulation in a probably vain attempt to find crossnational approximations of the same concept (Heath et al, 1990).

ISSP practice now tries to mitigate some of the problems of inappropriate translations by requiring its questionnaire drafting groups to provide accompanying notes on the conceptual framework and detailed meaning of potentially ambiguous questions in the hope that translators will resist literal transpositions in favour of functional equivalence.

Another possibly serious source of discontinuity in ISSP practice is that fieldwork periods differ widely country by country, varying by up to eighteen months. This is a direct and inevitable function of the absence of central funding for the ISSP, leading many countries to piggy-back their questionnaires onto another convenient survey which is already locally funded. On certain subjects in certain periods this could, of course, be a problem. In any event, serious data analysts need to take these period differences into account and, at the very least, reassure themselves that they are not responsible for any differences in distribution they discover.

The point of choosing the ISSP as an example of what can go wrong in crossnational studies is certainly not that the ISSP is unusually negligent in its methods and practices. Far from it. It seems to be among the best-regulated and most conscientious of crossnational surveys, collectively aware of good practice

and determined where possible to attain it. Yet it still turns out to sustain major differences in inputs between its member nations which could well account for some of the national differences in outputs it has discovered over the years. And if that is true of this academically-driven, custom-built crossnational time series, it is likely to be more true of many other comparative datasets on which we depend for our knowledge of national differences, and still more so of comparisons drawn from parallel surveys that were never even designed to fulfil a comparative purpose.

Comparative studies will, of course, always be defeated to some extent by differences between nations in matters of taxonomy and technique. Their very starting point is that important differences exist between nations in their behaviour, circumstances and attitudes. It is hardly surprising then that some of these differences should confound the achievement of standardised measures, nor that the difficulties tend to increase with the number and heterogeneity of the countries involved.

Mitigating the problem

What can we do about these problems? The unique and largely intractable measurement difficulties faced by crossnational studies are hardly a reason for abandoning them. They are far too important for that and are becoming more so with globalisation of the world economy and moves in Europe towards crossnational governance (Hantrais and Mangen, 1996). Given the well-known problems of comparative measurements, however, scholarly analysis of them ought perhaps to be rather more circumspect than it sometimes is. It would be unwise, to say the least, to draw heroic conclusions on the basis of unexpected national variations in the answers to a single question. But, recalling the difficulty of asking questions about 'God' in Japan, it is frankly unwise to draw firm conclusions about any large national variations in data in the absence of sound local knowledge of the countries concerned - relating not only to the methods employed and possible ambiguities in concepts or questions, but also to the society's social structure and history.

In the hope of attracting additional suggestions from other social scientists, there follow ten very practical rules of thumb for mitigating some of the problems referred to above, and a few others en route.

Rule 1: Social scientists should undertake not to interpret survey data (and perhaps any other data either) relating to a country about which they know little or nothing. This would tend to ensure crossnational collaboration in the interpretation as well as the design of comparative research. This self-denying pledge would help to mitigate the 'gee whiz' phenomenon of much crossnational research, when - using country as variable (often the only variable) - the analyst lights on a national variation in the pattern of answers as a revelation in its own right, simultaneously abandoning appropriate scepticism and any obligation to explain differences. The fact is that, devoid of detailed local knowledge of the social structure, history and culture of the country or countries they are investigating, an analyst is powerless to do much more than describe and gawp at the variations. Several other rules stem from this one.

Rule 2: Analysts of crossnational data should resist the temptation to compare too many countries at once. When they do, the 'marginals' inevitably become their primary focus. A practical way of mitigating this danger is to require one author or analyst for each country being compared. There may be many good reasons for expanding multinational surveys to, say, 28 countries - as in the case of ISSP - but it does not follow that all the countries should be the subject of all analyses. When that happens, explanation and interpretation soon give way to 'league tables'.

Rule 3: Crossnational surveys should pay as much attention to the choice and compilation of aggregate-level contextual variables as they do to individual-level dependent and independent variables. A comparative survey of religiosity, for instance, that did not include in its dataset aggregate-level data about the structure and prevalence of the major religions and denominations within each country would, of course, be almost impossible to interpret sensibly. Similarly, a comparative survey of social or economic inequality would be almost worthless in the absence of aggregate-level national data about, say, average incomes within different subgroups of the population and some reasonably robust comparative measure of each country's overall wealth. All too often analysts seem to compare national datasets in vacuo.

Rule 4: Social scientists contemplating or engaged in crossnational studies should be as open about their limitations as they are enthusiastic about their explanatory powers. The fact is that only certain subjects, and only certain aspects of those subjects, can successfully be measured crossnationally. Even in favourable conditions, only the lowest common factors between countries can ever be compared successfully. So, for instance, when there turns out to be no Japanese word or phrase for 'God', we should not expect Japanese respondents to negotiate questions based on a Judao-Christian-Islamic concept of a deity (assuming all the while that, say, Israel, Italy and Saudi Arabia could indeed find sufficient questions, concepts and methods in common to participate in a successful comparative study). In defining and locating common concepts to measure consistently across disparate countries and cultures, major compromises are always necessary.

Rule 5: Stringent and well-policed groundrules for comparable survey methods should become much more common in comparative studies than they are now. To avoid infringing well-established cultural norms (or, more often perhaps, organically-grown methodological 'habits') in one country or another, substantial national variations in methods are sometimes tolerated which should render comparisons invalid. Rules about sample coverage, sampling method, calculation of response rates, fieldwork periods, and so on should all be clear, unambiguous and capable of being implemented with more or less equal rigour in all the countries involved. In other words, the same technical standards as we would impose for national surveys should apply to crossnational surveys too. Arguably they should be even higher in view of the fact that country will inevitably be one of the major independent variables in the analysis of crossnational studies.

Rule 6: Analysts of crossnational data should undertake to suspend belief initially in any major inter-country differences they discover. All too often such unexpected differences turn out to be imposters - the result of a poor translation, a subtly different show card, a variation in sampling coverage, or a particular cultural cue which subtly alters the meaning of the variable in that country. Large national differences, or even small ones that have little face validity, especially when based on single items, need much stronger supporting evidence than they often get. The dictum that questions on all topics travel better in convoy applies a fortiori to those undertaking international journeys.

Rule 7: Emerging naturally from the six previous rules, crossnational surveys should ideally be confined to the smallest number of countries consistent with their aims, rather than celebrating as many nations as possible in their purview. True, analysis and interpretation could always be confined to a small selection of countries at a time, but - as noted - matters of design and content are also heavily influenced by the varying habits and circumstances of the participating countries. At its simplest, of course, the more dispersed the focus of a study, the more blurred will be its image.

Rule 8: In order to transform crossnational surveys from parallel exercises into joint ones, collective development work, experimentation, scale construction and piloting should be undertaken in all participating nations. In the absence of multinational 'proving' of the instruments, claims of functional equivalence are always suspect. Underlining Rule 7 again, the fact is that the greater the number of countries involved in a study, so the less likely it is that each one will participate in full-scale development work, pre-testing and piloting.

Rule 9: Routinely provide for secondary analysts of the data (and primary analysts for that matter) detailed methodological reports about each participating nation's procedures, methods and success rates, highlighting rather than suppressing variations. Analysts must be able either to omit certain countries from their comparisons or qualify their findings on grounds such as poor response rates, a non-conforming fieldwork period, use of a different mode of interviewing or method of sampling, and so on.

Rule 10: Routinely include methodological experiments in crossnational research. As exemplified by this article, too many assertions and too few pieces of hard evidence exist about the properties and pitfalls of multinational studies. Perhaps those engaged in them regard their very achievement as sufficient contribution, resisting further complications? But more hard evidence on different approaches to translation, employing, say, cognitive interviewing in different countries, would pay rich dividends.

If these ten rules (which are far from exhaustive) were even roughly adhered to, the situation would improve considerably. Indeed, any comparative dataset which complied with these rules would immediately transform itself from being deeply suspect to plain problematical.

NOTES

Return to [CREST Page](#)