Assessing quality in cross-country comparisons of health care systems and policies: towards a set of generic quality criteria

Cacace, Mirella; Ettelt, Stefanie; Mays, Nicholas; Nolte, Ellen

European Health Policy Group, 20/21 September 2012
King’s Fund, London
(DRAFT; Cite or quote with authors’ permission)

Abstract

There is a growing body of cross-country comparisons in health care systems and policy research. However, there is little consensus as to how to assess the quality of these cross-country comparisons. This is partly due to the fact that this is a field of study and is thus approached differently by researchers from different disciplines and traditions, with different quality concerns.

Using the distinction between variable-based and case-based research, we start by identifying the pertinent quality issues related to these comparative strategies. We then present findings from an exploratory survey of health system and policy experts in the European Health Policy Group to elicit their views on the characteristics of high quality cross-country comparisons.

Bringing the insights gained by these two exercises together, we have identified the following generic quality criteria for cross-country comparisons: (1) adequate use of theory, (2) explicit selection of comparator countries, (3) rigour of the comparative research design, (4) an appreciation of the complexity of comparison at the country level, and (5) contribution to knowledge. This list may not be exclusive; however, we believe that it will help advance the discussion about quality of cross-country comparisons in health care systems and policy research.
Introduction

There has been a long-standing interest in cross-national comparisons of health care systems and policies among policy analysts and policy makers alike. However, while the body of literature in the field has expanded over time, less attention has been given to the systematic assessment of the quality of studies in the field. In particular, we argue, it remains unclear whether there is any consensus as to the meaning of ‘high quality’ in this field. Arguably, the concept of ‘quality’ itself is problematic, given that it is multi-dimensional and means different things to different disciplines and audiences. While there are assessment tools for some types of research and/or methods (e.g. Yardley, 2000), these are not designed to capture the particular challenges of cross-country comparisons of health care systems and policies. Furthermore, research on health care systems and policies constitutes a diverse interdisciplinary field of study, with much variation in relation to disciplinary perspectives, foci and levels of analyses, and methodological approaches.

In this paper, we attempt to identify potential quality criteria for cross-country comparative health policy research. We approach this in two steps: first, by distilling the quality issues related to two different comparative strategies, variable-oriented and case-oriented research, and second, by conducting an exploratory survey of health care systems and policy experts involved in the European Health Policy Group (EHPG).

We define ‘health care systems and policies’ as the organisation and governance of health care at the macro-level (countries; regions), which include “the issues related to the organizational structure, the model of financing, the regulation and planning of the system, the ways to create physical and human resources and to provide services, as well as their changes over time” (Velasco Garrido et al., 2011: 28). By ‘cross-country comparison’ we mean “an approach to knowing social reality through the examination for similarities and differences between data gathered from more than one [country]” (Elder, 2006: 340). While we acknowledge that the geographical boundaries of health care systems are not always identical with territorial or political boundaries (Pennings et al., 1999) and it is debatable whether some countries have one or several health care systems, we limit our discussion to countries as the unit of comparison.
Cross-country comparisons of health care systems and policies

There is debate as to whether cross-country comparison, and comparison in general, constitutes a distinct method to be set alongside the experimental, statistical and case study methods (Lijphart, 1971), or whether it is simply an area of interest within established disciplines (such as comparative politics) (Hantrais, 2009). Comparisons are a “broad-gauge general method, not a narrow specialised technique” (Lijphart 1971: 682). Accordingly, cross-country comparison of health care systems and policies is not a distinct method in the social sciences, but a diverse field that faces a distinctive set of challenges because it focuses on comparing ‘large macro-social units’ (Ragin 1987; 5).

A consequence of this diversity is that cross-country comparisons of health care systems and policies can be undertaken for a wide variety of purposes: to provide contextual description; to develop a classification or typology; to test a hypothesis and/or develop theory and explanations for health system behaviour; to trace processes (e.g. of policy implementation) through systems; to make predictions about future trends; and to inform policy learning in one country with reference to experience and performance in others (Lijphart, 1971; Rose, 1991; 2001; Hantrais, 2009; Marmor and Wendt, 2012). Much of the research supported by organisations such as the Commonwealth Fund (e.g. Schoen et al., 2007; 2011), the World Health Organization or the Organisation for Economic Co-operation and Development is motivated by the aspiration of policy learning. For example, identifying common policy problems and goals, and then attempting to evaluate how different countries deal with these in order to isolate notions of ‘best practice’ and to assess the transferability of policies between countries.

Distinguishing comparative strategies: variable-oriented and case-oriented cross-country comparisons

There are a number of ways of categorising and distinguishing cross-country studies. Lor (2012) helpfully distinguishes between comparative research design, namely whether a study compares a larger or smaller number of cases; comparative strategy, which draws on Ragin’s distinction between variable-oriented and case-oriented research (Ragin, 1987) (see below); general methodology, i.e. the type of methods of data collection; and meta-theory, the underlying philosophical orientation of the study (Figure 1).
A *variable-oriented* strategy as identified in Figure 1 seeks to assess the relationship between aspects of cases (variables) across a sample of observations, usually in order to be able to specify general patterns that hold for the sample as a whole, thereby enabling predictions or inferences to be drawn (Ragin, 1987). The focus of the analysis is on the variables not the cases from which they are derived. Causation is inferred from co-variation rather than from the combination of factors present (conjunctures) when a particular outcome is also present. In contrast, a *case-oriented* strategy aims to understand the dynamics of a small number of cases, selected for their analytical or theoretical significance. Causality is seen as conjunctural; that is, effects are created by the precise combinations of factors present in specific examples of a phenomenon, each taken as a whole.

The two comparative strategies tend towards different methodological preferences, although this is not deterministic (see below). Thus, variable-oriented research is often associated with quantitative methods and case-oriented research with qualitative methods, but not invariably. This distinction has its roots in the set of meta-theoretical (ontological and epistemological) assumptions made by the researcher, which broadly fall into the categories of positivism, post-positivism and interpretivism.

The distinction between variable oriented and case oriented comparisons affects the selection of case studies for comparison, with the former more likely to draw on a larger number of countries. However, the number of countries selected for comparison is likely to depend on a
number of other factors, such as the purpose of the study or the extent to which a country is seen as an entire ‘macro-social unit’, which will impact on the number of countries that can be compared confidently.

To some extent, differences in meta-theoretical assumptions correlate with the methodological preferences and practices of different disciplines, which affect the choice of research design and data collection strategies they adopt in comparative research and their propensity to structure their research using explicit ex ante theoretical frameworks. Large scale quantitative (variable-oriented) studies aim to identify universal trends and relationships by striving to establish the strength of associations, e.g. between national and demographic variables (such as age, gender, or socio-economic status) (Barbour, 2010) or other input/output relations in health care systems. For example Gerdtham and Jönsson (2000) use a dataset covering a large number of countries to explore the influence of countries’ institutional arrangements on their levels and trends in health care expenditure.

These contrast starkly with the fine-grained, comparative case studies undertaken in, say, sociology (Hantrais, 2009). Case-oriented qualitative comparative research – provided that it goes beyond the purely descriptive – aims at providing an explication of the mechanisms involved in linking variables and outcomes (Barbour, 2010). Medical sociologists, for example, are perhaps more likely to use qualitative research designs than the majority of economists or epidemiologists, who are likely to give preference to quantitative study designs. Multi-method case study (case-oriented) research is well established in management and organisational studies, policy studies, history and political science, while other disciplines may find the notion of contextualisation problematic and be more attuned to establishing causality by assessing the average effect of variables across contexts. Ellen Immergut’s (1992) book tackles a similar question to Gerdtham and Jönsson (2000) but uses a very different approach. She compares a small, carefully chosen set of Western European countries in order to explain how particular features of their political institutions (namely, the number of ‘veto points’ in each) account for health system policy divergence. Similarly, Tuohy (1999) seeks to explain the dynamics of change and lack of change in the United States, Britain and Canada by teasing out the distinctive ‘logics’ that underlie each system. Both Immergut and Tuohy analyse their selected countries not as a set of variables but as complete cases.

In the multi-disciplinary field of health system and policy research, these differences in assumptions, strategies and methods complicate any assessment of quality. In what follows
we discuss the quality issues associated with variable-based and case-based cross-country comparisons in health system and policy research. In part, these are likely to reflect quality concerns associated with different research methods. However, we here are specifically interested in examining those issues that are specific, or exacerbated by, cross-country comparisons.

**Quality issues relating to variable-oriented cross-country comparisons on health care systems and policies**

Quantitative approaches to cross-national comparisons of health care systems have most often been initiated by economists (Abel-Smith, 1963; International Labour Office, 1959). Such approaches crucially depend on the availability of a sufficiently wide range of variables across a large number of countries over time. Much of the early work undertaken from the 1960s examined healthcare expenditure and its determinants. Studies sought to understand how different methods of financing and delivering health care contributed to (the control of) overall spending levels (Kanavos and Mossialos, 1990; OECD, 1985; Schieber, 1987). This emphasis on inputs changed in the 2000s, in light of increasing concerns about costs and broader concerns about accountability of, and within, health care systems, which gained particular momentum with the publication of the World Health Report 2000 and its ranking of the world’s health systems (WHO, 2000). The report not only played an important role in stimulating a wide ranging debate on health system performance; the various criticisms that have been debated widely helped bring to light the methodological challenges inherent in conducting and interpreting international comparisons using available data (Almeida et al., 2001).

There is now a wide range of international data available that allow for, and are being employed in, cross-national comparisons (e.g. OECD, WHO, Eurostat). While such data provide a useful source to help understand variation between countries, and, potentially, the determinants of observed variation in, for example, population health, such comparisons remain problematic. This is, in part, because of limitations in the availability, quality and completeness of data, but, perhaps more importantly, because of a frequent lack of any underlying theory guiding the selection of data and approaches to analysis (Nolte et al., 2005).

One example of the difficulty of a theoretical investigation is provided by studies that have employed a simple production function approach to the relationship between inputs, a range of standard explanatory variables and health care output at country level using regression
analysis. These have generated very mixed results that are frequently difficult to explain, and often running counter to what would have been expected from other research (Nolte and McKee, 2004; Nolte et al., 2012). Other studies have examined the association between specific characteristics of different types of health care systems and selected health outcomes. For example, Elola et al. (1995) studied 17 health systems in Western Europe, distinguishing national health service (NHS) systems (e.g. Denmark, Ireland, Italy, Spain, the United Kingdom) from social security systems (e.g. Germany, Austria, The Netherlands). Controlling for socioeconomic indicators, they found, in a cross-sectional analysis, that countries with NHS systems achieved lower infant mortality rates at similar levels of GDP and health care expenditure than did social security systems.

These types of study do, however, suffer from limitations beyond the obvious ones arising from data availability and reliability, and under-developed use of theory. One major weakness relates to the cross-sectional nature of many, so failing to take account of lagged relationships (Gravelle and Blackhouse, 1987). An obvious example is cancer mortality, where death rates will often reflect treatments undertaken up to five years previously. Furthermore, a cross-sectional design will not adequately address causality and frequently studies fail to set out the theoretical plausibility of the relationships they explore, giving the impression that the model was driven by data availability rather than plausible mechanisms (Buck et al., 1999).

Importantly, although notable exceptions exist (Macinko et al., 2003), the majority of studies of this type employ indicators of population health such as life expectancy and total mortality that are influenced by many factors that lie outside the health care sector so it is frequently difficult to attribute an observed variation in population outcomes to specific activities in the health system. For example, a re-analysis of the country ranking produced by the 2000 World Health Report, using a measure of population health that more closely reflects the actions of the health system than the measure of disability-adjusted life expectancy used by the WHO, shows how the country rankings change depending which outcome measure of health care performance is used (Nolte and McKee, 2003). This analysis highlights the importance of clarifying the underlying assumptions of studies.
Quality issues relating to case-oriented cross-country comparisons on health care systems and policies

The predominant approach to ‘case-oriented’ comparative research is the comparative case study in which each ‘case’ is considered analytically as a whole. At a minimum, quality in case-oriented comparison depends on the accuracy of description across countries linked to an acute awareness of the importance of understanding the relevant context in which documents, statements, interviews and observations take place. Indeed, some writers argue that contextual description and understanding is a basic condition for all successful cross-country comparison (Landman, 2000).

Case-oriented comparative studies are typically, though not exclusively, ‘mixed method’, drawing on qualitative and quantitative methods and data from a range of sources (see Figure 1). These may be used sequentially or simultaneously. For example, quantitative analysis may be used to formulate the cross-country comparison or to corroborate the findings from qualitative research or vice versa (Hantrais, 2009: 110). The potential of using mixed methods is often overlooked in cross-national research (Brannen, 2005). However, comparative research can benefit greatly from methodological triangulation, in particular since cultural sensitivity and the need for contextualisation pose additional challenges to the quality of studies.

The nested design approach is a mixed method approach that links intensive case-study analysis with statistical analysis (Lieberman, 2005). Rothgang et al. (2005; 2010) use a nested design to examine the changing role of the state in OECD health care systems. Their multidimensional framework measures two dimensions (financing and service provision) quantitatively while the third dimension (system regulation) is analysed qualitatively. In this example, qualitative and quantitative methods were equally important and used in parallel.

The key consideration when assessing the quality of such studies is whether the sequencing and/or inter-relationship between the methods and data sources is clearly related to answering the study question and is properly followed through in the analysis rather than different methods producing disconnected pieces of evidence (O’Cathain and Thomas, 2006).

The combination of an emphasis on detailed description, contextual richness and triangulation between different methods and sources of data across a number of ‘macro-social’ units can produce major analytical problems in reducing what can become a vast array of data to manageable proportions, as well as in isolating key explanatory factors influencing
the phenomenon of interest across countries. This tends to limit comparative case-oriented studies to a relatively small number of countries which may be insufficient to develop robust analytical generalisations. Qualitative Comparative Analysis (QCA) is a response to these difficulties and to the difficulty facing conventional variable-oriented approaches to cross-country comparisons such as regression when there are often too few cases available for robust analysis either for empirical reasons or because of the theoretically driven need to study only carefully selected cases (Ragin, 1987). The approach is a method for analysing the complex causal pathways described in different case studies of the same phenomenon in as economical a way as possible. Blake and Adolino (2001) use QCA to explain the enactment or not of National Health Insurance in 20 advanced industrial countries between 1945 and 1994, by systematically examining the propositions that emerge from a synthesis of the case study literature.

Since most case-oriented comparative studies rely on some form of ‘mixed’ method approach to data collection and analysis, there is unlikely to be a single set of quality criteria applicable to all parts of each study. Indeed, it may be that, as well as assessing the quality of the ‘mixing’ of methods, each method also needs to be assessed in its own terms. On the other hand, although there are different quality assessment frameworks for qualitative and quantitative methods of research, there are also common underlying criteria (e.g. whether the aim, objectives and questions of the study are clear). For qualitative research, for example, the Critical Appraisal Skills Programme (CASP) developed a set of ten general questions that can be used as criteria for appraising qualitative studies (PHRU, 2006). These are as follows:

1. Was there a clear statement of the aims of the research?
2. Is a qualitative methodology appropriate?
3. Was the research design appropriate to address the aims of the research?
4. Was the recruitment strategy appropriate to the aims of the research?
5. Were the data collected in a way that addressed the research issue?
6. Has the relationship between researcher and participants been adequately considered?
7. Have ethical issues been taken into consideration?
8. Was the data analysis sufficiently rigorous?
9. Is there a clear statement of findings?
10. How valuable is the research?

It is easy to see that these questions and the criteria of quality that underlie them are highly relevant to assessing the quality of cross-country comparative work: Is there a clear statement of the aims of the comparison? Is it appropriate to compare in the first place? Can the research question be answered by comparison (as this is not automatically so)? Was the justification for selecting country case studies coherent? Was the data analysis rigorous and did the research add sufficient insight to be considered valuable to the intended or likely audience?

This then raises the question as to what is different and/or more challenging about high quality case-oriented international comparisons. Hantrais (2009: 50) argues that while all the same general criteria of quality in social science research apply in cross-country comparisons, they are all the more important and more difficult to achieve because of the intrinsic complexity of comparing ‘macro-social’ units. Certain criteria are likely to be particularly salient in defining quality in cross-country work. The prime candidates include:

- Sensitive and explicit attention to the significance of contextual differences and related complexity (it is easier to get this wrong and for research to be misinterpreted and misused when the contexts involve different countries)

- Sensitive and explicit attention to the potential problems of linguistic and cultural differences (including semiotics) (these make it more difficult to compare like with like; and also may make it easier to compare formal structures (e.g. ways of financing) than culture and day-to-day behaviours (e.g. how patients regard doctors and how doctors communicate with patients)

- Sensitive and explicit attention to data interpretation issues, given that similar seeming data may be produced through very different processes in different countries (again, these make it more difficult to have confidence that the same phenomena are being compared across countries, even if linguistic and definitional differences have been surmounted).

- Explicit rationale for, and relevance of, case (country) selection in relation to the goals and questions of the research.

Having reviewed the main potential criteria of quality of both variable-oriented and case-oriented cross-country comparisons in health care systems and policies raised in the literature,
most of which overlap with general research quality criteria, we turned to a group of experts to help develop a set of quality criteria that were specific to this field.

Exploratory survey of experts’ views on quality criteria for cross-country comparisons of health care systems and policies

We undertook an exploratory survey of members of the European Health Policy Group (EHPG), in order to understand how those who are routinely involved in cross-country comparisons, either as producers or as users assess studies. The EHPG is an interdisciplinary group that has met twice a year since the early 2000s. Since its inception, it has been an incubator for research on cross-country comparative health policy and systems analysis, and a locus for discussion. In 2011, we invited EHPG members by email to identify what they considered as ‘high quality’ work in international comparisons, with the opportunity to nominate up to five publications, and to explain why they regarded these as high quality comparisons (see Appendix 1 for the questionnaire). The survey was completed by eleven members, with two members cooperating, yielding a total of ten independent responses (R1-R10, below). The majority of respondents were working in academia; one respondent had a non-academic affiliation. Respondents represented a range of disciplines such as political science, health or institutional economics, sociology, law and epidemiology. The countries represented were the US, the UK, Germany, the Netherlands, and Italy. In total, respondents nominated 37 publications including books, journal articles, and reports (see Appendix 2).

Criteria of high quality cross-country comparisons suggested by EHPG members

As a general observation, respondents clearly perceived quality differences in comparative works:
“Thinking of cross-national comparisons that are worthy of celebration, I can think of a number. But I can also think of many that clog the pages of journals and books that are not worth reading under the label.” (R1).

The selected ‘high quality’ comparisons varied widely in terms of their subject matter, their disciplinary orientation and cross-country comparative approaches. In what follows, we discuss the criteria identified by EHPG members in their free-text written responses, which we have grouped into seven categories. The comparative studies that EHPG members suggested as being exemplary are listed in a separate reference list in Appendix 3 and their authors are given in brackets in the text, below.

**Ability to explain**
The ability of cross-country comparisons to explain and help understand social phenomena was valued highly by the majority of respondents. To some extent, this is likely to reflect the fact that more than half of the respondents to the survey indicated having a political science background. In this tradition, ‘high quality’ of cross-country comparisons is seen in those studies that successfully conceptualise, measure, and explain the link between independent and dependent variables, and identify relevant explanatory factors (R4, R5), for example, by examining the relationship between health policy ‘inputs’ and reform ‘outcomes’. Approaches that examine the configuration of factors were seen by some as particularly useful (R1, R3) (e.g. Tuohy 1999). Further value was placed on comparisons that empirically demonstrate how the health system interacts with social values and wider sets of policies (R8) (e.g. Marmor et al, 2009; Sen, 2011]. The ability to explain was also valued by some for its potential to ‘scatter myths’ (R1), as such studies look behind the rhetoric and challenge the claims made by policy advocates by contrasting these with de facto policy outcomes (R2) (e.g. Evans et al, 1991; Okma and Crivelli, 2010).

**Practical relevance and health policy learning**
Relevance for policy and practice, and the potential for policy learning were identified by some as criteria for the quality of cross-country comparisons (R10). Specifically mentioned were studies that develop instruments that help improve health care systems and policies and explore their consequences, such as the effects of policies implemented at the micro level on outcomes at the macro level (R7) (e.g. some cross-country comparisons of the Commonwealth Fund (R10); Nolte and McKee, 2008; Moran, 1999; 2000). Others stated that
they particularly valued the potential for policy learning by comparing approaches across systems more widely.

**Use of theory**
Competent use of theory and a strong theoretical foundation were also named by a number of respondents as criteria of quality (R6, R3, R4, R9).

However, some comparative work was seen by others as outstanding for exactly the opposite reason, namely that it does not start with a preconceived theory. In these cases, comparisons were seen as especially successful if authors inductively develop new questions and categories that advance understanding of the subject matter (R5). One respondent also stated that good comparisons generate or advance theory rather than just apply existing theoretical perspectives (R6).

**Comparative research design**
The comparative research design was seen as important by a number of respondents, in particular, the use of a systematic approach to analysis, guided by a well structured analytical framework (R4). One respondent highlighted the importance of work that is analytically coherent, which in his/her view is often associated with single authorship (R4) (e.g. Tuohy, 1999; Moran, 1999, 2000; Maioni, 1998; Immergut, 1992; 2000). Concern was expressed about studies that apply a form of “telephone directory” listing of socio-economic variables to the analysis without reference to any theories or hypotheses that might have informed the choice of the variables in the first place (R4). In contrast, a good comparative design requires a clear specification of the research question or hypotheses, the methods used and conclusions drawn (R9). It was suggested that, given the complexity of the subject, good frameworks are likely to encompass multiple dimensions of the phenomenon, to allow for a systematic exploration of differences and similarities (R7, R1) (e.g. Wendt, 2009; Wendt and Kohl, 2010). It was also noted that combining qualitative and quantitative analyses could improve the depth of analysis (R4) (e.g. Rothgang et al., 2010).

Respondents also stated different preferences for the level or unit of analysis which they related to quality. While some respondents found comparisons of entire health care systems less insightful than comparisons of specific policies or programmes (R2, R8), others had a less clear preference for micro or macro level comparative research.
Case selection
Several respondents stated that the selection of countries should reflect the questions the comparison aims to address and that these choices should be justified explicitly (R2, R3, R9) (e.g. Tuohy, 1999). If the study aims to explain a phenomenon (such as policy change or system reform), a good comparison would be expected to have identified, and be organised according to, the variables that help explain the phenomenon (R5). Bevan’s comparison of health care performance in the devolved countries in the UK was mentioned as a study that made good use of a natural experiment thereby justifying its case selection (Bevan, 2010) (R1).

Some respondents were critical of the tendency to opt for pragmatic comparisons, with case selection often reflecting researchers’ country expertise, personal contacts and command of foreign languages, rather than factors relating to answering the research question. However, this sentiment was not shared by all; one respondent stated an explicit preference for unusual choices of countries (‘off the beaten track’) (R2).

Use of data
The majority of respondents emphasised that the quality of cross-country comparisons depends on good data and the good use of data (e.g. R3, R4, R9). This criterion is likely to have different implications for different types of studies. Case studies, it was noted, require accurate description as a basic ingredient (R2). Single country reports published by the European Observatory on Health Systems and Policies were listed as a positive example, although these are not explicitly comparative (but use a common framework to enable subsequent comparison) (R4).

The issues are likely to be different for quantitative, variable oriented research, as data quality largely depends on the data available. In any case, respondents noted that the strength and weaknesses of comparative data should be adequately discussed (R4) (e.g. Connolly et al., 2010).

Cultural and contextual sensitivity
Several respondents indicated that a well-informed awareness of differences in context marks out cross-country comparisons of high quality (R6, R10). Context has many dimensions and judgement is required to decide what matters in each study. Respondents pointed to cultural, political, social, historical, institutional and functional differences in context. At one level, it is even difficult to be sure whether the use of the same term denotes the same phenomenon in
different country settings such is the pervasiveness of contextual effects. One respondent noted that good comparisons display intricate knowledge of different countries and use the comparison as a mirror to ‘estrang[e]’ from any culturally dominant perspective (R6). An example provided was Payer’s work on cultural differences in relation to health and medicine (Payer, 1988).

Quality criteria for cross-country comparative studies

If doing rigorous research is difficult then cross-country comparisons face additional challenges. They not only require the ‘normal’ internal coherence of aims and objectives, research design and conclusion of any study, but they also have to deal with the added challenges of the comparative design and the complexity of comparison in multiple country settings.

At the beginning of this paper, we have asked the question of whether it is possible to define quality criteria for cross-country comparative research on health care systems and policies. Having reviewed both quality issues in variable-oriented and case-oriented comparative strategies, and the responses of experts we conclude that it is possible to develop a set of criteria which can be used to describe the quality of comparative studies.

Earlier in this paper, we introduced Ragin’s distinction between variable oriented and case oriented studies, which we (inspired by Lor, 2012) have applied to cross-country comparisons in health care systems and policy research. We find this distinction particularly useful to avoid a potentially over-simplistic focus on aspects of quality that may result from the (sometime parochial) methodological preferences of individual research disciplines and to help identify generic criteria of study quality.

We suggest the following six criteria (Table 1) to assess the quality of cross-country comparisons, highlighting distinctions between variable oriented and case oriented comparisons, as follows:

1. Adequate use of theory: Cross-country comparisons should make adequate use of theory to inform both the research design and the comparison (i.e. selection of
countries). Theory should underpin the selection of variables (variable-based comparisons) or case study design and analysis (case-oriented comparisons). If no theory is used, the purpose of this should be explicitly stated (e.g. to allow for analytic induction).

2. **Explicit country selection**: The selection of countries for comparison should be justified and reflect the aims of the study (i.e. the purpose of the comparison).

3. **Rigour of the comparative research design**: The comparative design should be rigorous, systematic, purpose driven and in line with the aims of the study. It is evident that this criterion applies to all types of research, irrespective of whether this is a cross-country comparison or not. The issues are however slightly different for variable and case oriented comparative research, although they are not mutually exclusive. In relation to variable oriented research, the availability, completeness and quality of data are crucial; case oriented research will perhaps more strongly depend on the accuracy and richness of descriptive detail, as well as the integration of data, if different types of data are used in combination. For both types of comparisons, the strength and weakness of the data and analysis should be critically discussed.

4. **Dealing with complexity**: Both case oriented and variable oriented designs have to deal with the complexity associated with comparing large ‘macro-social units’ in cross-country comparisons. Variable oriented comparisons have to deal with selecting variables that appropriately reflect all the variables relevant to the phenomenon to be compared. Case oriented comparisons should provide the richness of contextual description adequate to fulfil the aims of the comparison.

5. **Contribution to knowledge**: Cross-country comparisons should make a relevant contribution to knowledge, while it may depend on the audience, which knowledge is considered useful. While some cross-country comparisons intend to establish general propositions and explanation, others exploit the quasi-experimental character of comparisons, suited in particular for policy learning. Variable-oriented comparisons are often seen as particularly useful in establishing the strength of associations. Case-oriented comparisons, instead, are likely to be assessed on the basis of their potential for providing an explication of the mechanisms involved in linking variables and
outcomes that are more contextually embedded. Studies can also contribute to knowledge by helping develop theory and methods.

This list may overstate the difference between different types of comparisons. It may also not be exclusive; however, we are confident that these criteria cover substantial ground and are sufficiently generic to apply to the broad field of health care systems and policy research.

Publication and discussion of the list of criteria should advance the development of a greater degree of consensus in this field on what constitutes high quality research. We also hope that others will build on and amend this list through using it to shape their own research and to judge the quality of others’ research.

**Conclusion**

Our contribution started from the observation that there is little consensus on how to approach quality in cross-country comparisons on health care systems and policies. Comparisons are a diverse field that faces a distinctive set of challenges because it focuses on comparing ‘large macro-social units’. Furthermore, as research on health care systems and policies also is a multi-disciplinary field, differences in assumptions, strategies and methods complicate any assessment of quality. Our approach to answer the question on what constitutes high quality research is to link criteria related to the methodological choice of the comparative strategy with criteria nominated by an interdisciplinary group of experts who are routinely involved in comparisons on health care systems and policies, the EHPG.

Based on this input we have identified five criteria which we think will help assess the quality of comparisons of health care systems and health policies: (1) adequate use of theory, (2) explicit selection of comparator countries, (3) rigour of the comparative research design, (4) an appreciation of the complexity of comparison at the country level, and (5) contribution to knowledge.

This list draws together quality criteria already established in social research, including those that reflect the principles of social research in general, and combines them with aspects that are pertinent to comparative studies in general and cross-country comparisons in particular.

We think these are particularly pertinent in relation to health care systems and policy research, given the complexity of the social phenomenon, the ‘macro-social’ units, it is concerned with.
Publication and discussion of the list of criteria should advance the development of a greater degree of consensus in this field on what constitutes high quality research. We also hope that others will build on and amend this list through using it to shape their own research and to judge the quality of others’ research.
References


Brannen, Julia (2005), Mixing Methods: The entry of qualitative and quantitative approaches into the research process, International Journal of Social Research Methodology, 8:3, 173-184.


Rothgang, Heinz; Cacace, Mirella; Frisina, Lorraine; Grimmeisen, Simone; Schmid, Achim; Wendt, Claus (2010), The State and Healthcare: Comparing OECD Countries. Houndmills/Basingstoke: Palgrave.


Schoen, Cathy; Osborn, Robin; Squires, David; Doty, Michelle; Pierson, Roz; Applebaum, Sandra (2011), New 2011 survey of patients with complex care needs in eleven countries finds that care is often poorly coordinated. Health Affairs 2011; 12.


Appendix 1: EHPG survey questionnaire

What is a ‘high quality’ cross-country comparison?

There has been a long-standing interest in cross-national comparisons of health systems and policies among policy analysts and policy makers alike. Cross-national comparisons are a broad field and there is much variation with regard to the subject of comparison, underlying rationale, the disciplinary perspectives of the research and consequent methodological approach. The background that motivates us to further exploring that field is the project “An ‘On-call’ Facility for International Healthcare Comparisons”, which has, since October 2005, been providing policy makers at the Department of Health (England) with reviews of health policies in other countries. This work was commissioned based on an expectation that learning from other countries would assist health policymaking by giving proposed developments an international perspective and adding to the range of policy ideas and options emerging from domestic policy processes.¹

One question that concerns us in particular is that of the quality of cross-national comparative work. We know that context and case selection matter, and attention should be given to detail. Approaches to analysis are important, with preference given to systematically organised, framework- or theory-led approachesii; preferences also relate to the subject matter, the type of methods employed and the purpose and perceived relevance of the work. Yet, on the whole, there seems to be little (if any) consensus on the criteria that characterise ‘high quality’ cross-national comparisons.

So, who better to ask than the members of the European Health Policy Group? We would like to invite you to help us by listing up to five cross-country comparisons you consider as ‘high quality’ and briefly give reasons for your choice. We will then reflect on and try to crystallise these characteristics. We aim to present your (anonymous) choices at one of the next EHPG meetings and to publish the analysis in a peer-reviewed journal.

Please list up to five cross-country comparisons you consider as ‘high quality’ and briefly give reasons for your choice. What, in your view, makes the publications you have selected qualify as ‘high quality’ cross-national comparative work? We appreciate if you consider as many of the above mentioned criteria (case selection, framework, data comparability etc.) as possible in your answer.

1. …
5.

Finally, we kindly ask you to give us information on your disciplinary background _________ and your nationality_______________.

# Appendix 2: Nominations for high quality cross-country comparisons and criteria

<table>
<thead>
<tr>
<th>Short Citation (see App 3 for full reference)</th>
<th>Ability to explain</th>
<th>Practical relevance &amp; health policy learning</th>
<th>Use of theory</th>
<th>Comparative research design</th>
<th>Case selection</th>
<th>Use of data</th>
<th>Cultural and contextual sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Tuohy 1999</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Marmor et al. 2009</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Autier et al. 2010</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>4 Bevan 2010</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Connolly et al. 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>6 CMWF IHP Survey</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Evans et al. 1991</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Flood 2003</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 France 2008</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Freeman 2000</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>11 Glaser 1991</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 Gottweis 1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>13 Green 2001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 Hajer 1995</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>15 Halffman 2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>16 Immergut 1992</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 Immergut 1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 Jasanoff 2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>19 JHPPL 2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>20 Knorr-Cetina 1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Koechlin et al. 2010 (OECD)</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Kudrle/Marmor 1982</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Maioni 1998</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Moran 1999</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Moran 2000</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Marmor et al. 2005</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Nolte/McKee 2008</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Okma/Crivelli 2010</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Payer 1988</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Redelmeier/Fuchs 1993</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Rothgang et al. 2010</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Sen (2011)</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>The Joint Can/US Survey</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Van Doorslaer et al. 1999</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Wendt 2009</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>Wendt/Kohl 2010</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>White 1995</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 3: References for selected high quality cross-country comparisons

2. Marmor, Theodore R.; Freeman, Richard; Okma, K.G.H. (2009); Comparative Studies and the Politics of Modern Medical Care
5. Connolly, S.; Bevan, G.; Mays, Nicholas (2010), Funding and performance of healthcare systems in the four countries of the UK before and after devolution. London: The Nuffield Trust (4)
6. Publications by The Commonwealth Fund basing on data from the International Health Policy Survey
24. Moran, Michael (1999), Governing the Healthcare State: a comparative study of the United Kingdom, the United States and Germany, Manchester: Manchester University Press
30. Redelmeier, Donald; Fuchs, Victor (1993), Hospital Expenditures in the United States and Canada, New England Journal of Medicine 328 (11) 772-78. (5)
32. Publications basing on data from The Joint Canada/United States Survey of Health
33. van Doorslaer, Eddy; Wagstaff, Adam; van der Burg, Hattem; Christiansen, Terkel; Citoni, Guido; Di Biase, Rita; Gerdtham, Ulf-G.; Gerfin, Mike; Gross, Lorna; Hakinnen, Unto; John, Jürgen; Johnson, Paul; Klavus, Jan; Lachaud, Claire; Lauritsen, Jørgen; Leu, Robert; Nolan, Brian; Pereira, Joao; Propper, Carol; Puffer, Frank; Rochaix, Lise; Schellhorn, Martin; Sundberg, Gun; Winkelhake, Olaf (1999), The redistributive effect of health care finance in twelve OECD countries. Journal of Health Economics 18 (3) 291-313. (5)
36. White, Joe (1995), Competing Solutions