

tive studies on welfare regimes (see Table 1) indicates that the number and coverage of cases varies considerably between studies, not to mention the differences in the applied methods and indicators used. Thus not all countries fit neatly into the ideal-type categories, and some cases have often been ignored.

More generally, the essay discussed the problems of generalizing from a selection of advanced welfare states (OECD or EU) in comparative welfare state research. The very process of welfare state development is closely related to the macro-historical and political processes leading to economic growth and political integration that are preconditions for membership in these international organizations. Therefore comparative welfare state analysis faces a selection problem: The causal factors explaining welfare state development are also related to the factors determining the available cases for analysis. Therefore configurational comparative methods seem better suited than cross-national quantitative methods such as linear regression analysis. Configuration analysis can be used to analyze subsets of causal factors explaining specific typologies. Although multi-value QCA can be used to code welfare state typologies based on qualitative and quantitative indicators, it is advisable to use fuzzy set QCA to allow for a more open conceptualization of welfare regime typologies. Typologies can be used then as membership set relations not unlike the conceptualization of a comparison of ideal-type and real-type observations as advocated by Max Weber.

References

- Arts, Wil and John Gelissen. 2002. "Three Worlds of Welfare Capitalism or More? A State of the Art Report." *Journal of European Social Policy* 12:2, 137–158.
- Arts, Wil and John Gelissen. 2010. "Models of the Welfare State." In *The Oxford Handbook of the Welfare State*. Francis G. Castles, Stephan Leibfried, Jane Lewis, Herbert Obinger, and Christopher Pierson, eds. (Oxford: Oxford University Press), 569–583.
- Castles, Francis G. and Deborah Mitchell. 1993. "Worlds of Welfare and Families of Nations." In *Families of Nations: Patterns of Public Policy in Western Democracies*. Francis G. Castles, ed. (Aldershot: Dartmouth), 93–128.
- Castles, Francis G. and Herbert Obinger. 2008. "Worlds, Families, Regimes: Country Clusters in European and OECD Area Public Policy." *West European Politics* 31:1/2, 321–344.
- Collier, David. 1995. "Translating Quantitative Methods for Qualitative Researchers: The Case of Selection Bias." *American Political Science Review* 89:2, 461–466.
- Ebbinghaus, Bernhard. 2005. "When Less is More: Selection Problems in Large-N and Small-N Cross-National Comparison." *International Sociology* 20:2, 133–152.
- Ebbinghaus, Bernhard and Philip Manow, eds. 2001. *Comparing Welfare Capitalism: Social Policy and Political Economy in Europe, Japan and the USA*. London: Routledge.
- Emmenegger, Patrick. 2011. "Job Security Regulations in Western Democracies: A Fuzzy Set Analysis." *European Journal of Political Research* 50:3, 336–364.
- Esping-Andersen, Gøsta. 1990. *Three Worlds of Welfare Capitalism*. Princeton, NJ: Princeton University Press.
- Esping-Andersen, Gøsta. 1993. "The Comparative Macro-Sociology of Welfare States." In *Social Exchange and Welfare Development*. Luis Moreno, ed. (Madrid: Csic), 123–136.
- Esping-Andersen, Gøsta. 1999. *Social Foundations of Postindustrial Economies*. Oxford: Oxford University Press.
- Ferrera, Maurizio. 1996. "The 'Southern Model' of Welfare in Social Europe." *Journal of European Social Policy* 6:1, 17–37.
- Geddes, Barbara. 1990. "How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics." In *Political Analysis*, James A. Stimson, ed. (Ann Arbor: University of Michigan), 131–150.
- Hall, Peter A. and David Soskice, eds. 2001. *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*. New York: Oxford University Press.
- Katzenstein, Peter J. 1985. *Small States in World Markets. Industrial Policy in Europe*. Ithaca, NY: Cornell University.
- Kvist, Jon. 1999. "Welfare Reform in the Nordic Countries in the 1990s: Using Fuzzy-Set Theory to Assess Conformity to Ideal Types." *Journal of European Social Policy* 9:3, 231–252.
- Lange, Peter and Hudson Meadwell. 1991. "Typologies of Democratic Systems: From Political Inputs to Political Economy." In *New Directions in Comparative Politics*. Howard J. Wiarda, ed. (Boulder, CO: Westview), 82–117.
- Leibfried, Stephan. 1992. "Towards a European Welfare State? On Integrating Poverty Regimes into the European Community." In *Social Policy in a Changing Europe*. Zsuzsa Ferge and Jon Eivind Kolberg, eds. (Frankfurt: Campus), 245–279.
- Pennings, Paul. 2005. "The Diversity and Causality of Welfare State Reforms Explored with Fuzzy Sets." *Quality & Quantity* 39, 317–339.
- Ragin, Charles C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.
- Ragin, Charles C. 1994. "A Qualitative Comparative Analysis of Pension Systems." In *The Comparative Political Economy of the Welfare State*. Thomas Janoski and Alexander M. Hicks, eds. (New York: Cambridge University Press), 320–345.
- Ragin, Charles C. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.
- Rihoux, Benoît and Charles C. Ragin, eds. 2009. *Configurational Comparative Analysis. Qualitative Comparative Analysis (QCA) and Related Techniques*. Los Angeles: Sage.
- Vis, Barbara. 2009. "The Importance of Socio-economic and Political Losses and Gains in Welfare State Reform." *Journal of European Social Policy* 19:5, 395–407.

One-Size-Fits-All? Measurement Issues in Medium-N Comparative Welfare State Analysis

Jon Kvist

University of Southern Denmark
kvi@sam.sdu.dk

Medium-N studies often aim to explore diversity, i.e., the simultaneous existence of similarities and differences across cases (Ragin 2008). Diversity-oriented analysis is vividly illustrated in comparative welfare state analysis, where the typology of welfare states advocates a configurational view of cases (Esping-Andersen 1990). For example welfare states are differentiated according to whether they are universal or selective on one dimension and generous or ungenerous on another dimension.

However, when conducting medium-N studies welfare state scholars often resort to measures that are not true to the meaning of configurations or to the theoretical concepts they intend to measure, and in some studies measures that are not comparable across cases or over time. Scholars appear to have too many cases to make informed cross-case comparisons and too few to make robust statistical analyses. The question thus arises as to whether medium-N studies sit uncomfortably between small-N and large-N studies when it comes to measurement issues.

What do scholars do when they have a total population of between 18 and 35 countries that meaningfully qualify as welfare states? Most scholars use one of two dominant approaches. One approach is for scholars to resort to comparing only a few countries, perhaps even single-case studies, and relating their empirical findings to theory building and other empirical findings in comparative welfare state research. As scholars using this approach are no longer in the field of medium-N studies, I will not consider their particular measurement challenges here.

Instead I will focus on the second approach, that of those who believe in “one size fits all.” This approach is namely to use the measures of large-N studies in medium-N comparative welfare state analyses. Typical large-N measures include aggregate measures like social expenditures, single-case indicators such as measures for a specific stylized type, averages for different groups of countries, and composite indexes covering different aspects. In such studies scholars use large-N measures and data as though these have been designed to accommodate a wide range of questions and research designs.

I argue that the practice of using large-N measures in medium-N studies is problematic with regard to the dual challenge of securing content validation and ensuring comparability (for more on these issues of measurement validity, see Adcock and Collier 2001). Securing content validation entails understanding how we can adequately capture the meaning of theoretical concepts and analytical constructs. Ensuring comparability entails making measures that are comparable with measures in other contexts. To illustrate the problems of using aggregate large-N measures in medium-N studies, I offer two examples. Each example focuses on a specific way of theoretically operationalizing the welfare state, namely welfare effort and benefit entitlements. Each example uses different measures of welfare states typically used in large-N studies. Together the two examples cover aggregate measures, single-case indicators, group averages and composite indexes that all have specific problems in meeting the challenges of content validation and comparability. But first I will sketch how different types of welfare states are conceptualized and measured.

Measuring Different Types of Welfare States

Comparative welfare state research is rich on theories and concepts. The conventional starting point is the three worlds of welfare capitalism as advanced by Gøsta Esping-Andersen (1990, 2009). Known as the Liberal, Conservative-Corporatist, and the Social Democratic types, these welfare state regimes are constituted by different factors that scholars have opera-

tionalized for inputs and outputs.

The dominant input aspect is welfare effort as measured by social expenditures as percentage share of gross domestic product (GDP). The dominant output factor is social rights, or benefit entitlements, which derive from the ways that the institutional design of policies allows access to benefits and stipulate benefit generosity.

In short, countries with a Liberal welfare regime provide welfare benefits only as a last resort, when the market and families have failed. We expect low social expenditures, with benefits that are selective and not generous.

Countries with the Conservative-Corporatist welfare regime provide good coverage to the labor market insiders and meager benefits to outsiders. We expect medium social expenditure levels, as well as generous benefits for insiders and ungenerous benefits for outsiders.

Countries with a Social Democratic welfare regime have universal coverage and generous benefits. We expect high social expenditures, with benefits that are particularly generous for low-income groups, generous for middle-income groups, and the least generous for high-income groups.

How do scholars capture these constitutive factors in their measurements? And what issues related to content validation and comparability do these measurements raise?

Social Expenditures

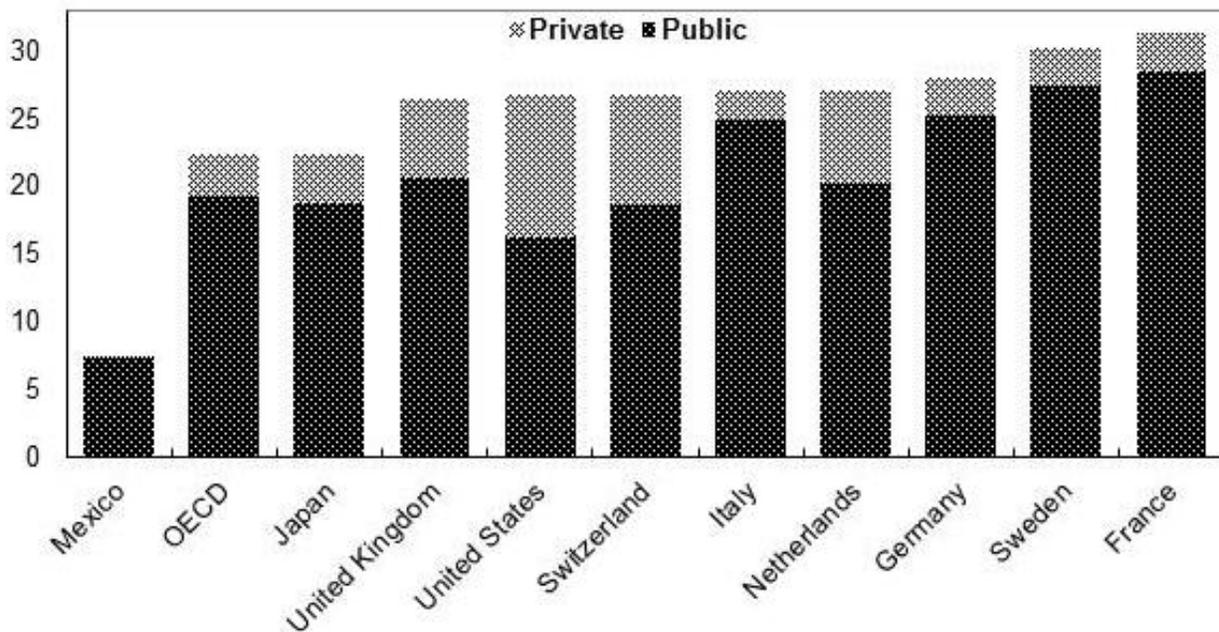
Public social expenditure as a percentage of GDP is by far the most frequent measure in comparative welfare state research. Arguably, the measure has proven effective in large-N studies for distinguishing between countries with a welfare state and those without a welfare state. Figure 1 shows the difference in social expenditure in ten countries and the average for countries that are members of the international Organisation for Economic Co-operation and Development (OECD). With 7.4 percent of GDP devoted to social expenditures, Mexico does not qualify for the group of countries with welfare states. Social expenditures in the countries with welfare states range between 22.0 percentage of GDP as in Japan (and the OECD average) and 31.3 percentage of GDP as in France.

However, social expenditures do not allow for accurate distinctions between countries with different types of welfare states, nor does social expenditure reflect the meaning of welfare effort as understood by researchers or people in general.

Public social expenditure cannot differentiate between countries with different types of welfare states. Figure 1 shows that Sweden with a Social Democratic welfare state type has high public social expenditures as expected. But at the same level we also find France, Germany, and Italy, which are associated with the Conservative-Corporate welfare state type. Sweden and Italy appear the same even though any scholar even slightly familiar with social policy in the two countries knows that they are very different: The Italian welfare state has lavish old-age pensions but rudimentary benefits in almost all other areas. The Swedish welfare state has extensive benefits across a wide range of areas. Social expenditures are not comparable across cases; a percentage point of GDP is seen as the same in

Figure 1: Public and Private Social Expenditure as Percentage of GDP, 2007

Source: OECD (2011)



the two countries although their allocation of expenditures on policy areas and programs differ markedly.

Public social expenditure does not encompass all relevant aspects of welfare state expenditures. As the OECD points out, social expenditures must take into account different tax systems and private social expenditures (Adema and Ladaïque 2009). Switching from gross public social expenditure to net public and private social expenditure has two important implications. First, social expenditure levels rise, sometimes dramatically, as in the U.S. Second, cross-national differences diminish. As Figure 1 shows, the Netherlands and Switzerland now have the same expenditure level as countries normally associated with the Conservative-Corporative welfare state-type—and so does the U.S.

Another weakness of social expenditure as a measure of welfare effort is that it aligns neither with scholarly interest in the politics of welfare reforms nor with people's perceptions of well-being. As Esping-Andersen (1990) notes, that actors have ever pressed for more social expenditure per se is hard to believe. When interpreted as an indicator of welfare effort, Finland had the biggest welfare state in the mid-1990s but less at the start and end of the decade, an observation that squared badly with public perception at the time. In the 1990s Finland hit an economic crisis in the first half of the decade, recovering in the second half. Public social expenditure varied from 24.1 percent of GDP in 1990, over 30.7 in 1995 to 24.2 in 2000 (OECD 2011). The measure depends as much on the development in the economy, GDP, as in social expenditure. Social expenditures do not pass the test of content validation.

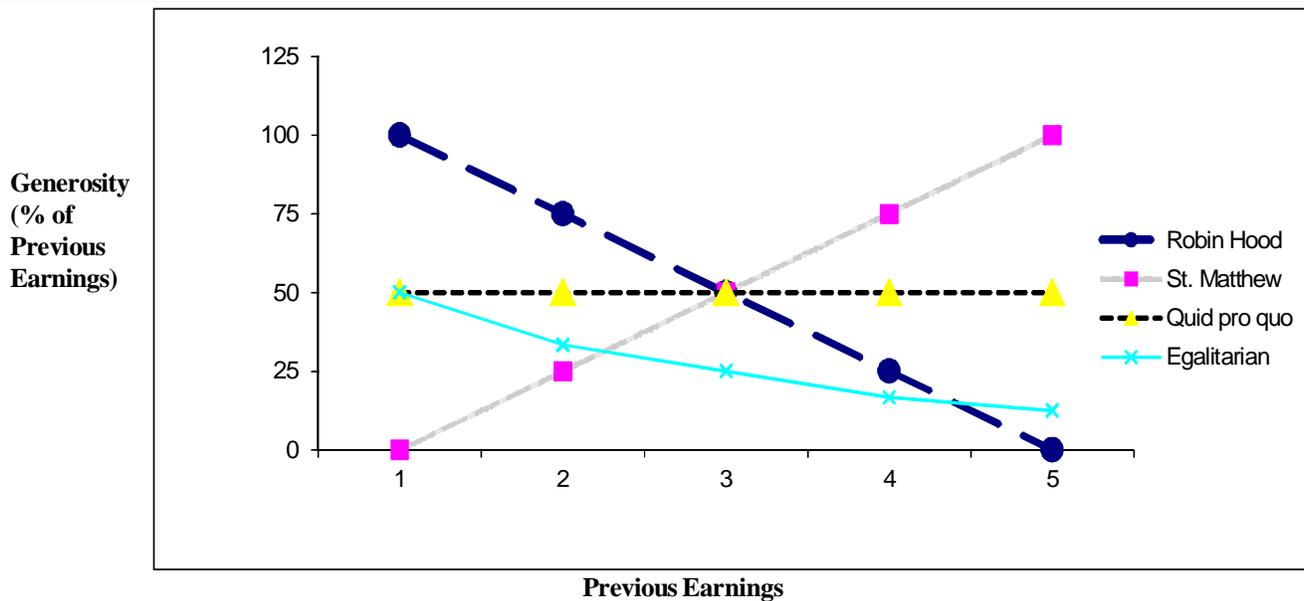
In sum, while social expenditure may have an explanatory power when used in large-N studies, as soon as we limit the analysis to a medium number of countries, the explanatory

power diminishes. To establish comparability, measures must take account of different tax treatment and provision of benefits in both the public and the private sector. After accounting for private expenditures and the tax system, we see how countries with welfare states range within a few percentage points, making it difficult to make clear distinctions between countries. Moreover, expenditure levels do not reflect the concern in many comparative welfare state analyses such as tapping into the total welfare effort or the redistributive strategies of who benefits. Finally, expenditure measures do not unmask differences in allocation means across different policy areas and programs. Scholars using expenditures as a measure for welfare state efforts must take account of all these concerns if their analyses are to yield meaningful results.

Benefit Entitlements

As described earlier, there are distinct theories about how the generosity of benefits varies for different socio-economic groups in different types of welfare states. I will illustrate the heterogeneous impact of the welfare state by referring to the expected profiles of benefit generosity according to four strategies of redistribution, as set out in Figure 2. The Robin Hood strategy is to take from the rich and give to the poor, resulting in a downward slope with benefits being very generous to people with low previous earnings and ungenerous benefits for high-income groups. The complete opposite is the case in the St. Matthew strategy, where more is given to those who have. The quid pro quo—or “something for something”—strategy implies that one receives according to what one pays in. Finally, the egalitarian strategy gives something to everybody, which translates into more generous benefits for the poor than for the rich.

Figure 2: Four Strategies of Redistribution Measured by the Generosity of Social Benefits as a Percentage of Previous Earnings



We can expect that European countries with a Liberal welfare state type have a profile following the egalitarian strategy (while this may not be the case for non-European countries with a Liberal welfare state type, we need not enter this debate now). Countries with a Conservative-Corporative welfare state type focus on status maintenance, thus following the quid pro quo strategy. Countries with a Social Democratic welfare state type that both is heavy on redistribution and gives generous benefits to middle-income groups have a profile that follows the Robin Hood strategy for low incomes and the quid pro quo strategy for higher incomes.

Single-Case Indicators

To measure benefit entitlements, scholars making medium-N studies use single-case indicators, composite indicators, or group indicators as large-N studies often do. The situation of the average production worker is the single-case indicator that is used in the two perhaps most important databases, i.e., the Social Citizenship Indicator Program (SCIP) and the Welfare State Entitlement Project (WSEP) (Korpi and Palme 2007, Scruggs 2004). As the average production worker becomes extinct, the comparability of measures over time becomes difficult. However, there is also a more general problem potentially associated with single-case indicators, namely that it is not possible to generalize from the situation of the single-case to other cases. To illustrate this, I have calculated net replacement rates for persons with previous earnings varying between 75% and 200% of the average production worker.

Figure 3 shows that cross-national differences increase with previous earnings. The gap between the countries with the lowest and highest net replacement rate increases from 8 to 39 percentage points from the lowest previous earnings level to the highest. At the level of the average production worker, the gap is 17 percentage points. Single-case indica-

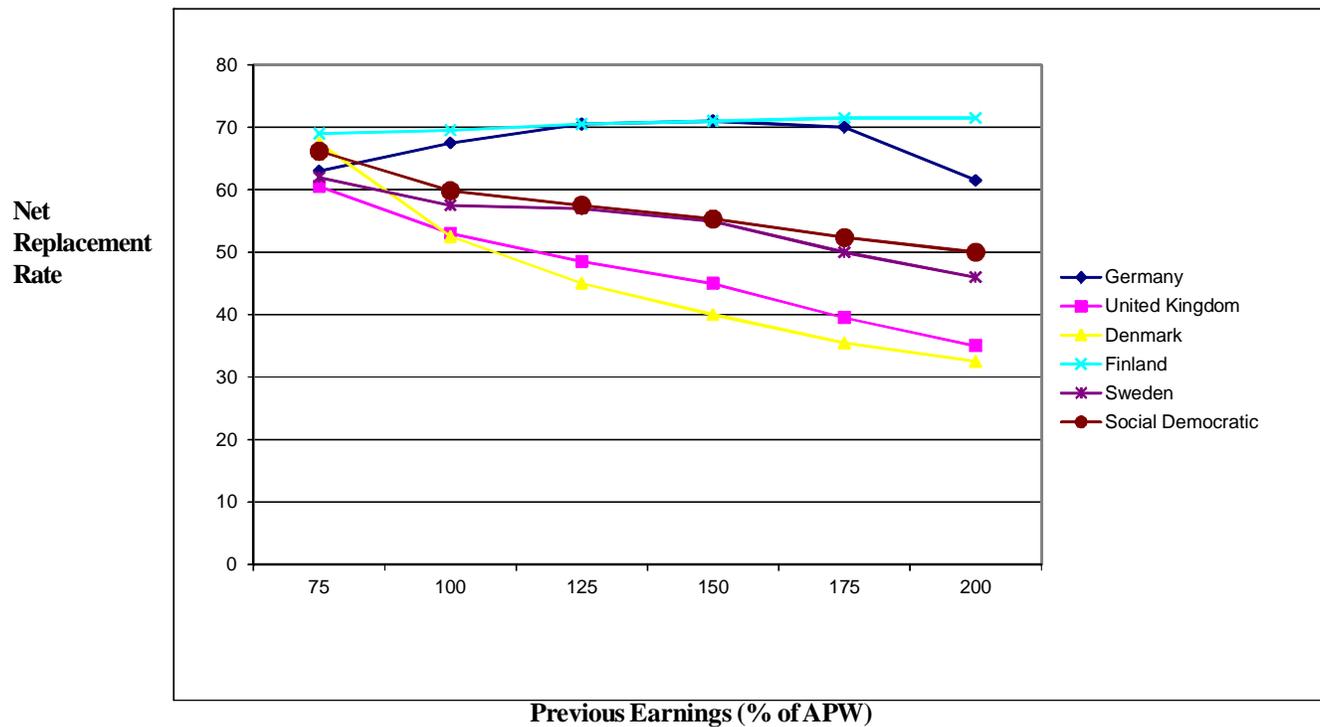
tors of social citizenship provide a partial picture of cross-national differences.

Nevertheless, single-case indicators may not provide the most relevant basis for advancing theory or advocacy of policy. If poverty among the elderly is the issue of concern, then the situation of persons with low previous earnings may be more relevant than that of the average production worker. If sustainability of pension systems is the issue of concern, then the situation of persons with high previous earnings may be the most appropriate to examine. For example, Danish full old-age pensions are the second highest for low-income groups, but the lowest for high-income groups. The Danish plan thereby looks effective in reducing poverty in old age as well as sustainable in economic terms. However, the measure only covers public benefits. Just as we saw that private social expenditures matter for the ranking of countries, such private programs also matter for benefit entitlements. In the Danish example, including private pensions would raise the level of benefits for middle- and high-income groups and modify the picture of a sustainable pension system.

The Finnish scheme comes close to the quid pro quo strategy. Because pension income is not taxed or is taxed more lightly than income from work, the German profile shows elements of the St. Matthew strategy, in that replacement rates increase until the plan is capped at an income level approximately 175% of that of the average production worker.

Making group averages masks intra-group differences. Comparative studies of social insurance tend to claim that countries belonging to the Conservative regime have the highest generosity and countries belonging to the Liberal regime the least generous. The Social Democratic countries are often said to occupy a middle position (see, e.g., Scruggs 2006). Figure 3 shows how group averages are problematic. The profile for the Social Democratic regime made up by the average for the three

Figure 3: Full Public Old-Age Pensions in Five European Countries as a Percentage of Previous Earnings, Single Person, 2006



Nordic countries in the study conforms nicely to the expectation across all income levels. However, in reality the Nordic countries are all over the map, with Finland being more generous than Germany, and Denmark being less generous than the UK for middle- and high-income earners.

Composite Indicators

Composite indicators come in two forms. Perhaps the most common composite indicator is an average of scores across two or more typical cases. Both databases on social rights and welfare entitlements (SCIP and WSEP) give the average net replacement rate for the standard and the minimum old age pensions. The other type of composite indicator is an indicator based on a variety of different types of information, typically put together in an index. Examples include the de-commodification indicator of Esping-Andersen (1990) and the generosity indicator of Scruggs (2006).

Compared with single-case indicators, composite indicators are more robust to differences across typical cases. An example is the OECD summary measure of benefit entitlement for unemployment benefits that is made up by the average of the gross unemployment benefit replacement rates for two earnings levels, three family situations, and three durations of unemployment (OECD 2007a). Distinct, but perhaps not representative, replacement rates are made less distinct in this average of 18 typical cases.

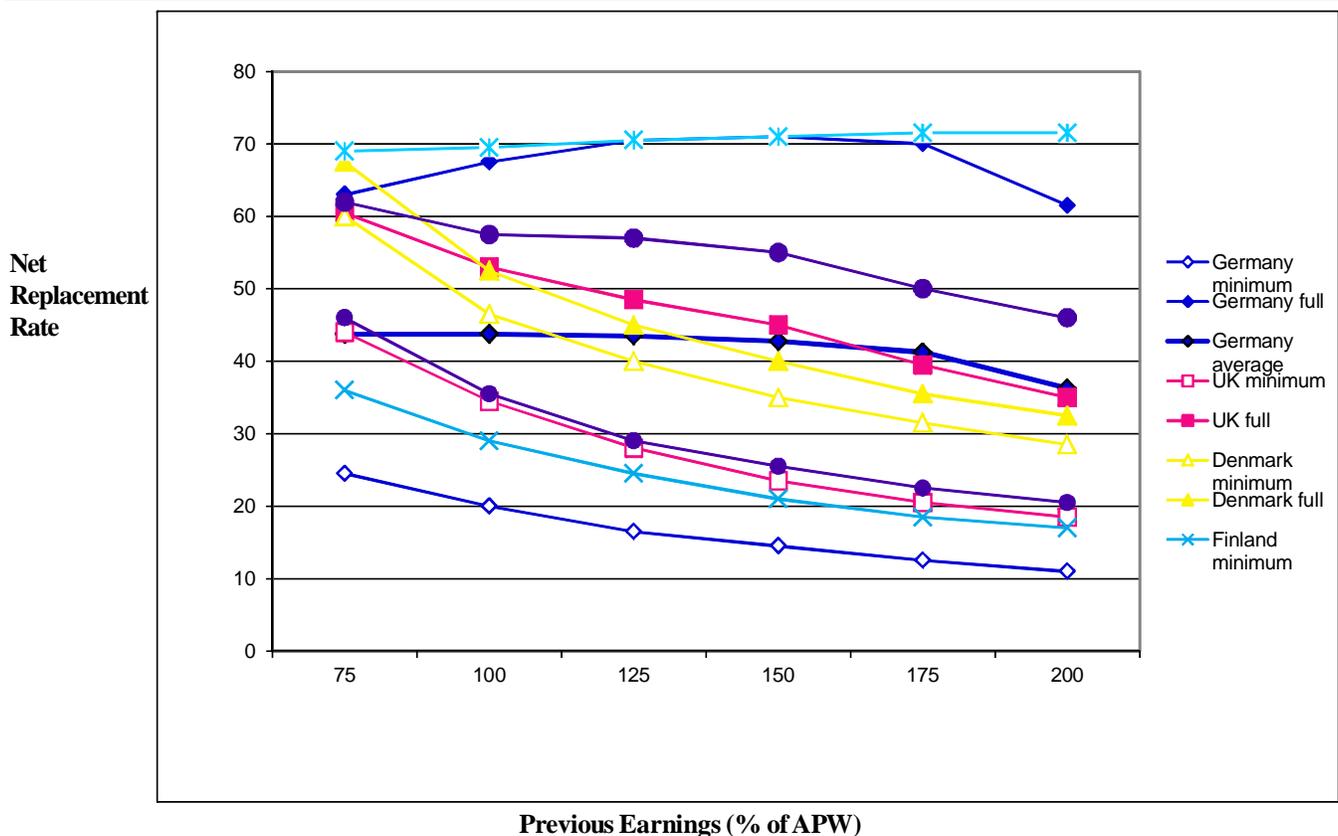
The risk, however, is that the composite indicators provide a mirror image of a fictive world, an image that is misleading for theoretical advances and policy advocacy. Figure 4 shows the situation of labor market insiders with full pensions

and outsiders with minimum pensions.

A good example is Germany at the earnings level of the average production worker. The German minimum pension is the smallest for all countries, whereas the size of the German full pension is only surpassed by a hair by the Finnish full pension. However, the average of the German minimum and standard pension at this earnings level is situated in the middle, i.e., above the level of the minimum pensions in all countries, (except Denmark) and below the full pensions. The measure of the average pension in Germany thus says little about the dual nature of the social rights and the pension system in Germany, and it misinforms us about the cross-national differences.

Indeed, comparing minimum and full pensions demonstrates important points about social rights and the different nature of public old-age pension systems in the five countries. The gap between minimum and full pensions is largest in Finland and Germany and smallest in Denmark. This result indicates how the public pension systems in Finland and Germany follow the *quid pro quo* strategy, providing the most generous pensions of all countries for labor market insiders and the least generous for outsiders. In contrast, the Danish public pension system is characterized by giving everybody nearly the same pension, almost independent of labor market or contribution record. Social rights in Finland and Germany are largely based on work history, in Denmark on residence. Perhaps surprisingly, the gaps between Swedish minimum and full pensions are not negligible, especially for middle- and high-income groups, and larger than the gaps between similar British pensions.

Figure 4: Full and Minimum Old-Age Pensions in Five European Countries as a Percentage of Previous Earnings, Single Person, 2006



Concluding Remarks

I have shown that the practice of taking aggregate large-N measures into medium N-studies has its problems. Aggregate measures like social expenditures have problems in establishing equivalence of meaning across cases when they do not include the tax system and private social benefits. Also, social expenditure measures fail to give clear distinctions of welfare state types and wrongly assume that countries with similar expenditure levels also have similar welfare state types. Single-case indicators risk flaws in making inferences from the specific case to other cases. Group averages risk masking intra-group differences. Composite indexes mix apples and oranges when they encompass information on benefit coverage and generosity and tell us little about either coverage or generosity.

To better capture the complex, configurational nature of welfare state types, scholars may benefit from rethinking their use of large-N measures in medium-N studies. Given the need for multiple measures, the practice of making aggregate measures should be replaced by a broad range of measures. Such a set of multiple measures should capture what is of central importance. For example, as family types are becoming more diverse and the male breadwinner model is less dominant, we need to investigate the situation of not only the dominant family type—increasingly dual-earner—but also the other family types. Similarly, when private social benefits are becoming

of more importance, scholars may be ill-advised to stick with the study of only public social benefits.

References

- Adcock, Robert and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95:3, 529–546.
- Adema, Willem and Maxime Ladaïque. 2009. "How Expensive is the Welfare State? Gross and Net Indicators in the OECD Social Expenditure Database (SOCX)." *OECD Social, Employment and Migration Working Papers*, No. 92, OECD Publishing.
- Esping-Andersen, Gøsta. 1990. *The Three Worlds of Welfare Capitalism*. Cambridge: Polity Press.
- Esping-Andersen, Gøsta. 2009. *The Incomplete Revolution. Adapting Welfare States to Women's New Roles*. Princeton: Princeton University Press.
- Korpi, Walter and Joakim Palme. 2007. *The Social Citizenship Indicator Program (SCIP)*. Swedish Institute for Social Research, Stockholm University.
- OECD. 2011. Social Expenditure Database via <http://www.oecd.org/els/social/expenditure>, retrieved 6 October 2011.
- Ragin, Charles. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: University of Chicago Press.
- Scruggs, Lyle. 2004. *Welfare State Entitlements Project Codebook*, version 1.1.
- Scruggs, Lyle. 2006. "The Generosity of Social Insurance, 1971–2002." *Oxford Review of Economic Policy* 22:3, 349–364.