Mawdsley J.


Copyright:


Date deposited:

12/05/2016

Embargo release date:

12/11/2017

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence
Comparing Militaries: The Challenges of Datasets and Process-Tracing

Jocelyn Mawdsley (Newcastle University)

Introduction

Why compare militaries? A great deal of the seminal social science literature on the military has been focused on nationally specific topics, and the nuances of national strategic cultures are generally held to be important in understanding why states and their militaries vary on questions of defence. However, it is undeniable that a reliance on what Williams (2007: 100) calls ‘insular case studies’ may mean that wider trends are missed. For some the appeal of comparative research lies in the sense that militaries are changing rapidly: post-Cold War changes in the nature of warfare, the globalisation of security challenges, the increasing multinational nature of missions and a growing homogeneity amongst advanced militaries mean that comparative work is needed to understand the changes and the implications they might have. Cross-national comparisons can help us to better understand the processes and mechanisms underlying change at the national level and thus identify what is important and what less so. Caforio (2007), for instance, claims that the extent of the recent changes to the functions and roles of the military, particularly in highly-developed states, makes cross-national military research increasingly vital. For others, the desire to generate generalizable theories makes hypothesis testing on multiple cases necessary.

The type of research questions that inform comparative social science research on the military can vary considerably. Researchers might wish to see how similar pressures for change have impacted on militaries of a similar type. Forster (2005) for example examined the extent of change in civil-military relations in European states following the end of the Cold War. Researchers might want to discover the best way to demobilise armed forces in a post conflict society, so a comparative study of states that have undergone such a process might be instructive; as in the previous case this might be best achieved by a series of detailed case studies. Or they might want to discover how particular variables affect the way in which militaries are managed in comparable states. Born et al (2003) for example carried out a comparative study of democratic control of the armed forces, stressing the need for variable-driven comparative research to understand the issue properly rather than relying on highly nationally specific case studies.
In short, military researchers, like other comparative researchers, may find that their research questions drive the choice between what Della Porta (2008) calls case-oriented and variable-oriented research. Other factors that may influence this choice are researchers’ epistemological preferences, their methodological skillsets e.g. statistical or linguistic abilities and practical considerations such as data availability or the feasibility of fieldwork. While mixed methods studies are growing in popularity, many military researchers wanting to do comparative work will find themselves choosing between large-N and small-N comparative research designs. Both pose particular challenges to the military researcher, beyond the more general difficulties of comparative research. This chapter aims to give an overview of some of the military-specific issues that large-N and small-N comparative research pose.

**Large-N Comparisons**

Quantitative methods have long played an important role in military research. The influence of the RAND Corporation’s quantitative research on US defence policy during the Cold War has had a substantial impact on how social scientists engage with military research (Barnes, 2008). While the RAND-style of statistical modelling has been criticised for divorcing military analysis from its historical and social context and thus lacking substance (Gray 2002), it is nevertheless quantitative analysis that often appears to give the clear answers to research questions that the military establishment is seeking. Statistical analysis is therefore a popular form of research methodology for those carrying out research for the military or hoping to have impact on military thinking. This institutional preference, which as Müller-Wille (2014) points out is often based on only a rudimentary methodological understanding, can lead to the military commissioning and using poorly designed studies, which particularly in conflict environments, where reliable baseline data are absent, have involved methodological compromises in for example survey sampling techniques that render the findings unreliable. The clear answers that quantitative research appears to offer may not in fact be the case.

As Barnes (2008) argues, the bias towards quantitative work in military-commissioned research, has been matched by a general trend towards the quantification of the social sciences, particularly in the United States. Indeed some fields of military research such as civil war research have become predominantly quantitative in orientation (Florea, 2012). For comparative researchers, carrying out large-N cross-national studies on military topics entails
an early initial choice as to whether the researcher will collect data themselves or draw on existing datasets. The feasibility of data collection is obviously driven by the research question(s) being investigated. It is much easier for example to devise a survey instrument, and to expect reasonable response rates, if like Born et al (2003) you are studying military democratic accountability issues in advanced democracies, than if you are studying causes of internal conflicts in failed states. Data collection may also involve the gatekeeper and access problems common to all military research, with the additional challenge for comparativists that unless the study is officially sanctioned / commissioned, it requires multiple national gatekeepers to grant the same access to the requisite number of people, and the infrastructure being in place to support effective data-collection. This is a particularly challenging hurdle in under-developed or conflict / post-conflict regions (Müller-Wille, 2014). Given this, it is perhaps unsurprising that many studies draw on existing datasets. Alongside official data compiled by international and national authorities, there are longstanding and freely available datasets compiled by researchers such as the Uppsala Conflict Data Program, which records instances of armed conflict. Other popular sources include the Correlates of War project, now based at Penn State University, which compiles datasets on variables that may influence the outbreak of war, or the Stockholm International Peace Research Institute’s (SIPRI) data on military expenditure and arms exports.

Using existing datasets for comparative research is not however entirely straightforward. Schrodt (2014: 291), for example, questions the value of continual ‘reanalysis of a small number of canonical datasets, even when those have well-known problems’. In particular, Schrodt (2014) questions the value of multiple analyses of a dataset, when the researcher only makes minor specification, operationalization or methodological variations on the original. Moreover, reusing existing datasets means that you are tacitly accepting the data collection techniques, the coding decisions and the accuracy of the data. In a similar vein to Schrodt, Florea (2012) suggests that uncritical use of datasets can lead researchers into a conceptual morass. He gives the example of civil war datasets, which pinpoint the start and endpoint of civil wars by a casualty-threshold metric, which he argues leads to arbitrary and problematic coding and the conflation of civil war with violence. For the rest of this section, the chapter will consider the military expenditure (MILEX) data to show the difficulties facing a comparative researcher, who wishes to use this data in a large-N comparative study. MILEX data matters as it is not only used directly, but also is used as an indicator in other key military-related datasets such as the Correlates of War dataset.
Military Expenditure (MILEX) Data

There are some basic questions about data that any comparative researcher needs to ask. How reliable are the data? Can the states in question be reasonably compared using this measure or has it been subject to what Sartori (1970) described as conceptual stretching? Do the data actually answer my research question? These are all issues that arise with MILEX data but which are complicated further by the military nature.

Firstly, let us turn to data reliability. MILEX data are compiled by a number of national, international and independent bodies. The International Monetary Fund (IMF) and the United Nations’ (UN) Office for Disarmament Affairs collect this data. The United States’ Bureau of Arms Control, Verification and Compliance issue annual ‘World Military Expenditures and Arms Transfers’ reports. For those states that are members, the Organisation for Economic Co-operation and Development (OECD), the North Atlantic Treaty Organisation (NATO) and the European Defence Agency (EDA) also issue annual reports. Finally, the independent research institutes the International Institute for Strategic Studies (IISS) and SIPRI compile The Military Balance and the SIPRI Yearbook respectively on an annual basis. For the researcher, who may be at first excited by this plethora of publically available data, it is confusing to note that the figures vary from body to body.

Brzoska (1981) argued that this variance is due to two separate problems; data origin and data preparation. Firstly, definitions of military expenditure vary and all sources are reliant to some degree on governments reporting in good faith their expenditure. There are multiple reasons why a government may choose not to report or to exaggerate or under report spending. A government may want to heighten or alleviate regional geopolitical tensions, it might be responding to domestic concerns about levels of defence spending or it may just consider such information best kept secret. Secondly, if these data are to be presented in a comparative fashion, there will need to be adjustments to account for different budgetary cycles, adjustments for inflation and the conversion into a single currency (usually the US dollar). Particularly when overall military expenditure is broken down into sub-categories such as defence procurement or military research spending, differences between the way governments report accumulate. Lebovic (1999) additionally points out that MILEX reports are heavily reliant on estimates, where governments have failed to or only partially reported (or are assumed to be unreliable), which may be revised subsequently as new data become available and that this practice leads to substantial divergence between the various sources.
Lebovic (1999) argues that the level of error in the data may be responsible for the degree of disagreement between researchers on a variety of hypotheses linked to MILEX data such as arms race models or the hypothesis that military spending inhibits economic growth.

Let us examine one case as an example. As has been argued already, governments may deliberately choose to under- or over-report military expenditure for political reasons, which can also skew analysis, where identification of deviant cases or outliers is the researcher’s aim in a comparative study. Here perhaps the example of West Germany during the 1980s is instructive. In general German defence data from this period were thought to be of low quality (Cowen, 1986). Brzoska (1981) claimed that West Germany produced different figures on defence spending for domestic and international (NATO) consumption to suit national distaste for any hint of militarism and the foreign policy need to be seen by its allies as a reliable NATO member. But the issue was particularly problematic when it came to analysis of the role that expenditure on defence research played in creating economic growth. Germany, like Japan, had often been used by economists as an example to support the ‘crowding out’ hypothesis, (i.e. that government investment in defence R&D crowds out civilian R&D, which would have resulted in higher rates of growth), given that until the early 1990s, despite a much lower level of defence R&D spending than comparable states, it enjoyed considerably higher rates of growth (Kaldor et al, 1986). This challenged the widespread view that spending on defence research correlated positively with economic growth.

German peace activists like Rilling (1988) pointed out that this picture of German research suited the actors involved and was therefore rarely challenged even if it seemed unlikely – the research community prefer their civilian image, while it had suited successive German governments to claim that defence production is in private hands and has little to do with them. Rilling also claimed that by 1988 ‘the military research budget in the FRG requires a far larger percentage of the national science resources than is officially declared’ (Rilling, 1988: 317). His argument was that the visibility of such research was minimised by a combination of statistical secrecy, the large percentage of such work being carried out in the private sector rather than in national research facilities and a conscious effort to marginalise the work by subsuming it in non-military spending categories. Rilling was making a political point. However, parliamentary questions from the politician Edelgard Bulmahn led to an admission by the German government that the figures on military research supplied to the OECD were too low; for 1990, the year in question, revised figures were estimated to be
almost double the amount that stood in the statistics (Liebert, 1998). Had the economists been aware of the inaccuracy of the reported data, would Germany have been such a deviant case? As it happens, subsequent meta-analyses of multiple studies on military research expenditure and economic growth suggest that the positive correlation was much weaker than was thought at the time, and that the crowding out hypothesis may well be correct (Dunne and Braddon, 2008). Nevertheless, the misleading German data may have contributed to what Lebovic (1999) sees as the inconsistent results in these studies. While all sources relying on government reporting are subject to similar problems, the particular sensitivity of military data makes it likely that these problems occur more frequently.

Our next problem is whether states can be reasonably compared by such a measure or whether it is subject to conceptual stretching (Sartori, 1971). It should be noted that SIPRI, at least, is of the opinion that despite both governments and researchers using MILEX data to closely compare states, they are in fact more appropriately used for ‘comparisons over time and as an approximate measure of the economic resources devoted to military activities’ (Omitoogun and Sköns, 2006: 270-71). They point to one well-known case during the Cold War as an example of how this can be misleading. There was a lack of credible official statistics on the Soviet Union’s military spending and so estimates were generated by the ‘building-brick’ method. As Omitoogun and Sköns (2006) point out, this was critiqued at the time as being methodologically unsound, as it used US costs and relative prices to estimate costs in the Soviet Union. Despite the fact that the data were known to be problematic, this approach led politicians and media in the West to uncritically accept that the Soviet Union had higher military expenditure than the United States and used this to argue for defence spending to increase in the West. A similar worry might apply to the way in which Chinese MILEX data are being used politically at present, given they too are largely based on estimates in the absence of reliable official data.

The final question is whether the data actually answer the research question. Here MILEX data offer a good example of what Florea (2012) sees as a current problem with quantitative civil war research; namely that data continue to be collected, under classifications developed during the Cold War, but are these classifications still relevant to the way we conceptually now understand civil wars? Omitoogun and Sköns (2006) identify two potential problems with the use of MILEX data for contemporary security research: the impact of the war on terrorism and the broadening of the concept of security. The blurring of the dividing line between internal and external security means that increasingly governments are including
items in MILEX reporting that they would not have done previously. Unless MILEX reporting is sufficiently disaggregated, this may mean a confusing picture for those seeking to use MILEX data as an indicator of increased hostility in a region. Moreover, MILEX data do not give a picture of the impact of non-military security concerns on government spending e.g. rises in climate change protection spending. As military force may not always be the response to newer security concerns, its use as an indicator of the likelihood of conflict may diminish. In short, while large-N comparative studies offer the tempting possibility of testing and yielding generalizable hypotheses, and for military researchers there are multiple existing datasets, care needs to be taken not to misuse the data.

**Small-N Comparisons**

Small-n comparative research generally relies on the development of small number of detailed case studies. Vennesson (2008) argues that there are three types: the descriptive case study, the interpretive case study and the hypothesis generating or refining case study. Case study research on military matters has often been of a historical nature, but has been used to go beyond description, and to test theories on subjects like military doctrine. This type of research became more prevalent in Security Studies from the 1980s onwards, when more archival material was made available, which led to various studies questioning established interpretations of security events and the theories that had been largely unquestioned during the Cold War, in particular challenging deterrence theory (Walt, 1991).

Vennesson (2008) argues that researchers working with case-studies can take a positivist or an interpretivist perspective. For a positivist the main purpose of process-tracing in a case study is to establish or evaluate whether the causal process of the proposed theoretical framework can be observed. The case study allows the investigation of links between different factors and the evaluation of the relative importance of potential causal factors. For an interpretivist, Vennesson (2008) argues that the focus is not just on what happened but on how it happened. In other words ‘interpretive approaches to political science focus on the meanings that shape actions and institutions, and the ways in which they do so’ (Bevir & Rhodes, 2003: 17). Here the importance of the case study is to investigate the relationship between actors’ beliefs and their behaviour. This type of approach for example might be valuable for a researcher interested in change in the armed forces. These different perspectives will influence how the researcher designs and carries out case study research. Military researchers come from both perspectives.
Case study research generally employs a process-tracing methodology. Usually this would entail an appropriate combination of open-ended interviewing, participant observation and document analysis (Venesson, 2007). As most military researchers are unlikely to be able to engage in participant observation particularly in a comparative context, this section will concentrate on the particular issues for comparative researchers in document analysis and interviewing.

**Process-Tracing-Documentary Analysis**

For any military researcher working on recent or contemporary topics, access to relevant government documents is likely to be heavily restricted. Military documents are more likely to be classified and for longer than other government documentation. The problem is magnified for a researcher carrying out a cross-national study. Not only does the researcher have to learn to navigate different national archives (each with their own peculiarities in terms of procedures, cataloguing and regulations), but the researcher also has to accept that decisions to classify / declassify documents are unlikely to be the same in each state. If one of the states being researched has strict rules about what may be classified in the first place and a robust freedom of information system, a detailed case study can be developed, only to find that none of the equivalent documentation is available in the other cases. Perhaps even more dispiritingly, but not unexpectedly, Deschaux-Beaumes (2012) reports lengthy freedom of information procedures with eventual rejections from both her case study states (France and Germany).

This problem is also prevalent in international military organisations’ archives, which might be expected to be a helpful source for the comparative researcher as their documents are applicable to more than one state. Mastny (2002) for example bemoans the lack of a comprehensive history of NATO due to the declassification problem. Although in theory NATO documents are released after thirty years, objection by any member state can block their release. Deschaux-Beaume (2012) also complains about the lack of availability of EU documents relating to its Common Security and Defence Policy (CSDP). Political events of course may mean that archival material becomes available unexpectedly in one state that reveals information on its allies. For example, Heuser (1993) was able to draw on material from the opened East German military archives to trace the evolution of Soviet military doctrine through Warsaw Pact training documents.
The comparative military researcher needs to be resourceful and flexible in seeking documentation to help build up their case studies. There are various sources that can be helpful. Military researchers have the good fortune that military matters are the subject of multiple well-informed specialist magazines both in paper format and online. While the Jane’s Information Group with titles such as Jane’s Defence Weekly and Jane’s Intelligence Review is perhaps the best known publisher (and has been publishing in the field since the 19th century), other useful publications include the online Defense News and Flight International. Most states also have journals that cut across the worlds of academia, practitioners and think tanks such as the British Royal United Services Institute (RUSI) Journal or the French Revue Défense Nationale. Military training establishments may also have their own publication series, which can offer insider insight. Depending on the states studied, if there is an active and institutionalised peace movement, their publications and collections of documents can also be helpful. Finally, of course, there are the defence-related think tanks and research institutes, which whether independent or closely associated to a national Ministry of Defence offer both publications but also libraries, with often well-catalogued and extensive grey literature collections including official documentation, which can be easier to navigate than official archives. The most helpful sources are likely to vary between the case studies, so it is worth investing time in conversations with more experienced researchers in each state to get a sense of where the most useful material is likely to be found, so that fieldwork can be as efficient as possible. Clearly as well, linguistic proficiency or otherwise will limit how much access a researcher can get, and varying language skills may lead to imbalance between case studies. In sum though, documentary analysis on its own is unlikely to produce a full picture of the topic investigated, and so many comparative military researchers make interviewing a key part of their research design.

*Process-Tracing: Interviewing*

Interviewing can be a highly valuable information gathering resource. As Deschaux-Beaumes (2012) argues, it is perhaps of particular value for researchers without existing connections to the military, as it brings them into contact with a highly specific and distinct social field. The degree of difficulty in setting up series of interviews is highly dependent on who you wish to interview, the sensitivity of the topic and the states that are the case studies. While there are always access issues, the gatekeeper problem is reduced if, for instance, you seek to interview people working within ministries (even if they are serving officers), than if for instance, you seek to interview multiple junior members of a regiment on their base. A basic problem for
comparative military researchers is equality of access between states. A topic might be more sensitive in one state than another. Or the time between fieldwork trips might mean that a topic was not sensitive in the first country, but a subsequent security incident has made it sensitive before the fieldwork is carried out in the other states. The gatekeeper in one system might be more supportive of researchers than in another. This might involve methodological compromises, whereby fewer interviews are carried out in one country than another. Again, while organigrammes of MoDs vary substantially, if you are interviewing in ministries asking an official in state A, who their opposite number in state B is, enables a form of snowball sampling to take place. Deschaux-Beaume (2012) also makes the point that personal relationships between officers from different militaries can be a significant enabler for comparative research. Finally, some states are more accustomed to researchers carrying out military research than others – this may be beneficial or disadvantageous to the researcher, depending on whether they are viewed as a novelty or a threat.

Interviewing soldiers or veterans in a conflict or post-conflict setting clearly increases the access and sensitivity problems. This may mean that some methodological flexibility is required. For example, Eriksson Baaz and Stern (2009) report that in their research on sexual violence carried out by soldiers in the Democratic Republic of the Congo, that semi-structured group interviews were more fruitful than their initial intent to use individual interviews, as they were less intimidating for the interviewees. In conflict zones, the principle of informed consent from interview partners is particularly relevant, and the researcher needs to be highly sensitive to cross-cultural differences and power dynamics.

Carrying out interviews across several nation states also requires cross-cultural sensitivity in the research design, concerning the conditions under which serving military personnel may agree to be interviewed. In particular, when interviewing military and defence ministry personnel, their willingness to be interviewed on the record may vary between countries. It is important to recognise that the patterns emerging from fieldwork in one country may not be duplicated in another. Deschaux-Beaume (2012) offers the example of the different legal protection for freedom of speech for French and German soldiers, and the problems this caused for her own comparative study of the French and German militaries’ experiences of the European Union’s security institutions and networks. The French military statutes state that they may only give their own opinions when off duty and then only with the reserve expected of military personnel. The German armed forces in contrast are officially ‘citizens in uniform’ and therefore protected by the right to freedom of speech contained in the
German Basic Law, if they criticise German defence policy on the record in an interview. Practically for Deschaux-Beaume’s (2012) research, this meant that some of her interviewees would not allow interviews to be recorded or to have their names or positions identified. Such factors need consideration in the initial research design for as Deschaux-Beaume (2012: 110) points out there will need to be trade-offs between ‘research deontology and methodological rigour’, and an inability to record a substantial number of interviews obviously means that some forms of analysis of interview data are consequently ruled out.

The question of the need for reflexivity on the part of the military researcher is also more complex when the researcher is carrying out comparative work. Higate and Cameron (2006) rightly point out that a reflexive military researcher needs to be aware of the potential of co-option, or being militarised, of the issues raised by interviewing in highly masculine environments, of the power relations between the researcher and the interviewee, and the question of whether or not the researcher is viewed as an insider or not. In addition to this already daunting list, the comparative military researcher needs to be aware of their own relations to each military organisation. Does shared citizenship, language or experience for example lead to unconscious bias in favour of one set of interviewees over those without that sense of familiarity? How much has the researcher’s or interviewee’s linguistic abilities helped or hindered effective communication?

Carrying out case-study based comparative research on militaries can be a very fruitful enterprise. It is however likely to involve a degree of methodological flexibility and an acceptance that the ways in which the case studies are developed, will depend on local factors and will vary. The researcher has to be reflexive about the decisions being made, and also ready to be resourceful and resilient.

**Conclusion**

Caforio (2007) is correct to argue that our understanding of the nature and extent of the changes to militaries, following the end of the Cold War, can be greatly enhanced by comparative studies. Both large-N and small-N comparative studies can help to provide a richer and more detailed picture, and help researchers to generate generalisable theories, which may enhance our understanding of international security. There is a need to know what remains nationally specific and what broader trends are emerging.
Nevertheless, it is necessary to accept that the difficulties in carrying out military research in general are increased in much comparative work. This chapter has reviewed some of the challenges involved in conducting large-N and small-N comparative research in this difficult research terrain. It is certainly far from an exhaustive account of the potential issues involved, which will vary substantially between projects.

Questions of access, gatekeeping, reliability and availability of information and negotiating an unusual environment pose methodological challenges to military researchers working on a nationally-focussed study. The additional needs for cross-cultural awareness and sensitivity, the difficulties of consistently comparing different national militaries, and in many cases multiple periods of fieldwork, mean that a comparative military researcher may need to be even more reflexive and flexible about the methodological choices made during the period of research design and subsequently. This should however not deter the researcher: comparative military research can make both a substantial contribution to knowledge and be enriching for the researcher.

**Bibliography**


Deschaux-Beaume, Delphine (2012) Investigating the military field: Qualitative research strategy and interviewing in the defence networks, Current Sociology, 60(1): 101–117


Forster, Anthony (2005), Armed Forces and Society in Europe, Basingstoke, Palgrave Macmillan


