



DOES TORTURE PREVENTION WORK?

**OUTLINE OF A 3 YEAR RESEARCH PROJECT
COMMISSIONED BY THE APT TO:**

Richard Carver

**Senior Lecturer in Human Rights and Governance
Centre for Development and Emergency Practice
Oxford Brookes University**

2012-2014

Table of Contents

1. Introduction.....	3
2. The research question and some obstacles to answering it.....	5
2.1. Human rights impact evaluation	5
2.2. Evaluating prevention.....	6
2.3. Quantifying torture	6
2.4. Measuring risk.....	7
3. Research Design	7
3.1. Stage 1: Exploratory study.....	7
3.2. Stage 2: Pilot study.....	8
3.3. Stage 3: Refinement of methodology.....	8
3.3.1. <i>Selection of countries</i>	9
3.3.2. <i>Identification of research partners</i>	9
3.3.3. <i>Development of research methodology</i>	9
3.3.4. <i>Finalization of country selection</i>	10
3.4. Stage 4: Main research period	10
3.5. Stage 5: Reporting	10
3.5.1. <i>Outputs</i>	10
4. Methodological issues	11
4.1. Qualitative Comparative Analysis	11
4.2. Assessing decline in risk of torture	12
4.2.1. <i>Multiple systems estimation</i>	13
4.2.2. <i>Official reported data</i>	13
4.2.3. <i>Informal reports and key informants</i>	13
4.3. Evaluating national level impact.....	14

1. Introduction

This concept paper has been prepared at the request of the Association for the Prevention of Torture to outline the design of a research project evaluating the effectiveness of torture prevention.

Although APT is commissioning this research, and will provide practical support for it throughout, the organization has made it clear from the outset that the conduct of the research should be independent of the APT and conducted by reputable scholars and other experienced and credible researchers. Parts of the research will be published in peer-reviewed publications in order to reinforce the scholarly credentials of the exercise.

The scope of the project is ambitious, in that it aims to identify the key factors leading to a reduction in the risk of torture and other ill-treatment in a substantial number of different countries. At the same time, in-depth research into the proximate effects of various torture prevention activities will give greater insight into the impact of these activities and is intended to develop effective evaluation tools for future use by preventive mechanisms.

The overall comparative study will be conducted using a method called Qualitative Comparative Analysis, which is highly suitable both for human rights impact assessment and for detailed comparative analysis of situations where causality may be extremely complex and unclear, where the question of necessary and sufficient conditions is important, where there is a relatively small number of case studies, and where available data may be of different types and not directly comparable by other methods (such as multiple linear regression). Other aspects of the in-depth country studies will use a variety of methods, some of which are tentatively proposed in this document and all of which will be agreed and finalized in the third stage of the project.

The table below summarizes the main stages of the project, with approximate timings, the main activities, and expected outputs. These are developed and explained in the narrative that follows.

	Description	Duration	Activities	Outputs
Stage 1	Exploratory research, gathering initial data on countries where torture assumed to be reduced in order to develop hypotheses	6 months	Desk study of 5+ countries	Initial report and proposals on methodology
Stage 2	Pilot study, testing robustness of comparative methodology	6 months	Desk study and limited field research on up to five countries, with refinement of causal conditions and testing of QCA	Research report
Stage 3	Selection of countries and development of methodology in conjunction with research partners	6 months	Meeting of participants in project Drafting of methodology	Methodology document
Stage 4	Main research period	12 months	Detailed case study research on 12+ countries, using quantitative and qualitative methods. Use QCA to create Boolean comparative analysis identifying key factors in torture prevention Detailed evaluation of specific preventive interventions	Peer-reviewed journal articles Interim reports on website
Stage 5	Writing of final report(s)	6 months	Writing/editing of publications	Final publication (multi-author edited volume) available in various languages freely online Peer-reviewed journal articles Evaluation methodology for NPMs

2. The research question and some obstacles to answering it

The research question is: “What interventions contribute to reducing the risk of torture and other cruel, inhuman, or degrading treatment or punishment?” Or, to put it more simply and crudely: “Does torture prevention work?”

Before considering detailed questions of research design or methodology, there are several problems with approaching this question that will have to be addressed by this project, if not exactly resolved. In descending order of hierarchy they are as follows:

- The difficulties inherent in evaluating the impact of any human rights interventions, both because of lack of clarity and agreement over how to measure human rights and because of the problem of disentangling causality in situations where there will be multiple factors that bring about change.
- The difficulties entailed in evaluating *prevention*. By definition we are interested in the non-occurrence of certain events and the reasons for this.
- The difficulties inherent in the fact that the issue being investigated is torture – the nature of the violation is that there are no generally accepted figures for its incidence and the secrecy in which it occurs makes reliable data almost impossible to obtain.
- The problem of measuring risk. The main danger here is that many of the indicators often proposed to suggest reduced risk are the very same preventive interventions that are the subject of the inquiry.

2.1. Human rights impact evaluation

There are two underlying problems connected with human rights impact assessment (referring here to what Todd Landman calls direct impact assessment – that is the effect of deliberate interventions intended to improve human rights¹). First, there is little doubt that large scale changes in human rights result from a variety and combination of factors, making it impossible to attribute change with any certainty and difficult to assess different contributions. There is a second obstacle, which is how it is possible to measure the degree of respect for human rights, whether at any given moment or over time. This is possibly more difficult in relation to torture than any other human rights issue, because of the absence or unreliability of data (see point 2.3 below).

In principle it might be possible to use quantitative techniques – specifically time series regression analysis – to measure the effects of specific interventions on the incidence of torture, but the impossibility of assembling reliable data on incidence of torture makes this impractical as a general method. (This does not exclude the possibility of doing this at country level if the data are available.)

However, the approach suggested in this paper combines a general qualitative comparative analysis, across a series of country case studies, combined with more localized evaluation of interventions at the country level. The latter could be qualitative or quantitative.

¹ Todd Landman, *Studying Human Rights* (London: Routledge, 2006).

Human rights impact assessment can be either *ex ante* or *ex post* – criteria and indicators can be established before the planned activity is carried out or devised subsequently. The first priority of this project will be a credible *ex post* evaluation, but one of the intended outputs will be a set of practical evaluation techniques (including *ex ante*) for National Preventive Mechanisms and others engaged in torture prevention.

2.2. Evaluating prevention

Strictly speaking, determining the effectiveness of preventive activities is establishing whether events that would otherwise have happened have not happened, as a result of the activities. While any impact assessment has a counterfactual element to it – addressing the question of whether events would (or would not) have occurred given different conditions – the attempt to establish a negative in this instance poses almost insoluble problems.² It is certainly possible to envisage a hypothetical situation in which preventive activities were deemed effective in mitigating torture, even when the actual incidence of torture had increased.

However, for the purposes of this study, decline in the reported incidence of torture will have to serve as a proxy for the effectiveness of prevention.

2.3. Quantifying torture

Underlying the methodological challenges for this project is, of course, the immense difficulty in determining whether torture has taken place and quantifying its incidence. The very nature of torture is usually an event that takes place secretly without any formal record.

Many global comparative indices use general measures of the presence or absence of torture, usually derived from international NGO reporting or sources such as the US State Department Human Rights reports. These are usually unconvincing as a precise index of the occurrence of torture and next to useless when the intention is to measure change over time.

Another approach is to use reports of torture, usually through official mechanisms, as an indicator of the incidence of torture. The problem with this is that the increase in reports may be primarily an indication of the existence of an effective reporting mechanism – a development that may actually be associated with a decline in the incidence of torture at the very moment when its reporting increases. Hence, for example, the indicators developed by the UN Office of the High Commissioner for Human Rights may be useful and relevant³ – but not necessarily as indicators of the incidence of torture.

An additional methodological issue is the spatial one. Certain activities may have an impact, positive or negative, on the incidence of torture outside the borders of the

² However, for the counterfactual element in comparative case studies, see Charles C. Ragin and John Sonnett, 'Between Complexity and Parsimony: Limited Diversity, Counterfactual Cases and Comparative Analysis', in Sabine Kropp and Michael Minkenberg (eds.), *Vergleichen in Der Politikwissenschaft* (Wiesbaden, 2004).

³ *Report on indicators for promoting and monitoring the implementation of human rights*, HRI/MC/2008/3, 6 June 2008.

country under consideration. How, for example, is the impact of the prohibition on refoulement to be factored in?

None of this is to say that torture cannot be quantified. It may, for example, be possible to use a technique such as multiple systems estimation to infer the incidence of torture, although this is dependent on the presence of multiple agencies reporting cases of torture. This is unlikely to be feasible in all countries included in the case study. Additionally, it will be harder to use for the purpose of this study, which is essentially a longitudinal one, where it will be important to measure change in the frequency of torture over time.

These practical problems do not mean that it is impossible to arrive at a credible estimation of whether or not the risk of torture has decreased. It does mean, however, that it is not likely to be possible to use the same technique across all the case studies. Hence the need for a method of comparative analysis that will allow the use of data of different types.

2.4. Measuring risk

Finally, the real measure of the effectiveness of torture prevention is how far the *risk* of torture has been reduced. Yet the indicators that are usually selected for reduction of risk – for example by the OHCHR in its indicators project – are to a large extent the very same preventive measures, the effectiveness of which the research is intended to evaluate. Although the general approach will be to take incidence of torture as a proxy for risk, this may not be adequate in all situations and consideration will have to be given to devising further indicators of risk that are independent of the preventive interventions being evaluated.

3. Research Design

The proposed framework for the research is a comparative analysis of several case studies. The method that will be used for this comparative analysis is Qualitative Comparative Analysis (QCA). This method, described in greater detail in section 4 below, is appropriate for comparative research in which the precise data for each case study may not be directly comparable. Various other methodological issues about the collection of data for each separate case study will need to be resolved later – a step that is built into the research design.

3.1. Stage 1: Exploratory study

The first stage of the research is to gather initial data primarily on several countries where *it is assumed* that the risk of torture has been reduced. Note that this hypothesis will be tested at a later point. The purpose of this initial research is to identify those factors that are considered by well-informed observers to have been important in achieving this risk reduction. In other words, the purpose is to identify those torture prevention interventions that will then be tested for their effectiveness. This stage will also identify other factors – such as political and social change, or indirect preventive steps, such as investigation and prosecution of torturers – that are regarded as having been influential. Although the focus in this exploratory phase is on countries where the risk of torture is believed to have decreased, the converse

situation will also have to be explored, identifying factors that lead to the continuation or even increase in the risk of torture.

Another purpose of this phase will be to determine what are reliable methods for estimating decreases (or not) in the incidence of torture. So, although they will be selected on the basis of a *prima facie* reduction in torture, an attempt will be made to test this assumption, or at least to determine the methods by which it might be tested.

A number of countries have been identified as possible case studies for this phase of the research, although not all will be selected: Turkey, Argentina, Maldives, Georgia, Philippines, Ghana, Togo, Uganda, Spain, and South Africa. Countries that are not studied here may be included in the broader comparative case studies later in the research. The countries will be selected for geo-political variety, as well as possible differences in the factors that have been important. Given that the sampling is purposive rather than random, and that it is only being used to construct a hypothesis, not to test it, there would be no problem with adding countries to the initial ones selected in this phase, subject to time constraints.

Research in the phase would be desk-based and conducted by the lead researcher. The phase would be completed within six months, with the output a research report of 20-30 pages.

3.2. Stage 2: Pilot study

In the second stage, the impact of the various causal condition assembled in Stage 1 will be tested in a pilot study of a maximum of five countries. In this stage countries will be selected to reflect a variety of cases in which the risk of torture has or has not been reduced. The primary purpose of this phase is to test the robustness of the comparative methodology.

It is normal in conducting QCA, which is an essentially qualitative methodology, to refine the definition of the causal conditions identified in order to develop a workable comparison. This pilot phase will be essential to do this.

It will be of particular importance to determine if the proposed comparative framework is able to cope with the complexity of the number of causal conditions that are likely to be identified. It is possible that the method will have to be refined to take account of remote and proximate causal conditions,⁴ or to sequence the causal conditions over time.⁵

This stage will be conducted by the lead researcher and may involve some limited field research in the countries selected, in addition to desk research. The output will be a further research report of 20-30 pages.

3.3. Stage 3: Refinement of methodology

⁴ C.Q. Schneider and Claudius Wagemann, 'Reducing Complexity in Qualitative Comparative Analysis (QCA): Remote and Proximate Factors and the Consolidation of Democracy', *European Journal of Political Research*, 45/5 (2006 2006), 751-86.

⁵ Neal Caren and Aaron Panofsky, 'TQCA: A Technique for Adding Temporality to Qualitative Comparative Analysis', *Sociological Methods & Research*, 34 (2005), 147-72.

The third stage of the research involves several tasks:

- Preliminary selection of countries for case studies.
- Identification of research partners.
- Development of research methodology.
- Finalization of country selection.

3.3.1. Selection of countries

Clearly, to test the hypothesis that certain preventive interventions have contributed to a reduction in the risk of torture, it will be necessary to include case studies that go beyond those explored in stage 1 (which were primarily those where torture has been reduced, apparently as a consequence of certain preventive interventions). As far as possible these should include: countries where torture has been reduced without these interventions; countries where the interventions have had no apparent, or indeterminate, effect. Of course, it is possible that countries in the second category cannot be found, but the point is that a range of scenarios should be included. Sampling in this phase is still purposive rather than random and case studies should be selected on the basis of geopolitical spread, as well as the likely availability of suitable research partners.

3.3.2. Identification of research partners

The aim is to identify a mix of academic and non-academic partners, preferably those who will be able to continue with monitoring of torture prevention after the end of the project – in other words, that the research will play a capacity-building role. There are an increasing number of university human rights centres that straddle these roles and would make ideal partners. It will be important that teams are multi-disciplinary, with a minimum of social scientific and legal backgrounds, but also in some instances other areas of expertise, such as medical/psychological.

Partner institutions will be selected through an open call for proposals at a relatively early point in the process (when a provisional list of countries is available). The aim will be to identify institutions that can demonstrate both a commitment to the issue being researched and the requisite skills and/or potential to participate in the study.

3.3.3. Development of research methodology

The research will operate at two distinct levels.

- A set of comparative case studies;
- In-depth studies of the impact of certain measures at a national level.

Clearly the essence of a comparative approach is that the same data must be gathered in all instances, in order that it can be compared. However, one of the strengths of the Qualitative Comparative Analysis (QCA), proposed as the key comparative tool, is that different types of data can be used to contribute to the various conditions and outcomes that are included in the Boolean analysis (even, despite the name, quantitative data). QCA can take account of complexity of

causality, as well as unevenness of data. Each country study will at a minimum have to generate all the necessary data to feed into the comparative analysis. However, additionally it will be possible in many instances to carry out in-depth research into certain aspects of preventive work. This might include the impact of training, the outcome of monitoring visits to closed institutions, the impact of legal reforms, or other issues. Since this detailed research will not be possible in all case studies, the comparative possibilities are fewer. Nevertheless, it may prove possible to conduct a detailed comparative analysis of some of these issues across a small number of countries. Clearly general agreement on methodology will be a precondition for such comparative analysis.

3.3.4. Finalization of country selection

The list of countries included in the case study is likely to be modified on the basis of researchers/institutions participating in the project and the methodologies employed.

Stage 3 is expected to take six months and will include a meeting of approximately two days involving all countries, to agree terms of reference, develop methodologies and finalize selection of countries.

3.4. Stage 4: Main research period

Stage 4 is the conducting of the research and preparation of interim research reports. It will be important to maintain a high level of communication between researchers during this phase, to ensure that data are being gathered in appropriate formats, as well as to exchange lessons from the country research.

It will be important to have some mechanism of quality control to ensure that data are being properly gathered and to conduct further pilot analyses to ensure that the methodologies are working as intended. It would probably be appropriate to establish a quality control/methodology review committee.

It is anticipated that this stage will last 12 months.

3.5. Stage 5: Reporting

Stage 5 is the drafting of the final report, which is expected to take six months.

3.5.1. Outputs

It is intended that the final research report will be published independently of the APT, to underline that this is independent research and to increase its credibility for a scholarly audience. However, consideration will have to be given to how to ensure that this does not limit the availability of the report to a general audience. Although this may be difficult to achieve with an academic publisher, there are serious non-academic publishers who would probably be willing to produce a free online version (for example Pambazuka Press in Oxford). This final publication would be in the form of an edited volume that would include comparative analysis, thematic discussion, and country case studies. The publication will be made available in several

languages, and at minimum in national languages of all countries covered throughout the duration of the project.

It is also intended that there should be other publications in the course of the project. Researchers and participating institutions will be encouraged to publish the outcomes of their research in reputable, peer-reviewed journals. The idea is to generate a variety of serious scholarly studies on the impact of torture prevention.

Interim reports could also be made available – for example the reports of stages 1 and 2 – and non-academic summaries of the findings would also be available for APT's own dissemination.

One important output of the project will be a set of evaluation tools for torture prevention bodies, including National Preventive Mechanisms under the OPCAT. The methods developed in the course of the project for evaluating both micro-level and national impact of preventive measures will be available for the use of NPMs and other bodies working on torture prevention.

4. Methodological issues

Design of the methodology for this project will work at three distinct levels:

1. The overall comparative case study methodology, which will be Qualitative Comparative Analysis.
2. Design of national level methodologies to identify the causal conditions (preventive factors) and outcomes (level of risk of torture) to be included in the QCA.
3. Other national level methodologies to evaluate the impact of specific interventions, including legal changes, visiting programmes, training and education etc.

4.1. Qualitative Comparative Analysis

Qualitative Comparative Analysis (QCA) is a set-theoretic technique that has been developed for providing a rigorous approach to comparative case studies in instances where quantitative data may be difficult to come by and where one of the objects of study is the interplay of a variety of causal factors. Despite the name, QCA is an effective way of incorporating data that may be a combination of quantitative and qualitative.

QCA works by using Boolean logic, expressing outcomes and causal conditions⁶ in dichotomous terms (did ratify OPCAT v did not ratify OPCAT; risk of torture reduced v risk of torture not reduced).⁷ These are assembled as a “truth table” that includes all logically possible combinations of causal conditions and outcomes (whether or not

⁶ QCA practitioners use the terms “outcome” and “causal condition.” These can be thought of as equivalent to the dependent and independent variables used in quantitative social scientific research.

⁷ This is known as “crisp set” QCA, in which every variable is expressed in Boolean terms as 1 or 0. “Fuzzy set” QCA allows measurement of degrees of presence or absence of different factors. Hence it would be possible, for example, to express how far custodial officials have been trained at various intervals between 0 (no one trained) and 1 (everyone trained). In general, however, it may be that for this research fuzzy set QCA would create a spurious precision. Nevertheless, this is an option that should be considered further.

they actually exist). Algebraically these are then minimized to display the combinations that do (or do not) lead to the outcome – in other words, in this instance, it will display all combinations of factors that have led to reduction in the risk of torture.

Charles Ragin, the American sociologist who has pioneered the technique of QCA, argues that the Boolean approach meets the needs of synthetic comparative research in five particular ways (that make it well suited to this particular project):

- An ability to examine a large number of cases
- An ability to address complex causal conjunctures
- An ability to produce parsimonious explanations (if desired)
- An ability to investigate cases both as wholes and as parts
- An ability to evaluate competing explanations.⁸

QCA will make it possible to determine which factors, or combinations of factors, are necessary or sufficient to prevent torture.⁹ Hence it will be important to include not only relevant preventive interventions, such as treaty ratification, visits to closed institutions, training, and legal reforms, but also the types of activity labelled by APT as indirect prevention (such as investigation and prosecution of cases of torture), as well as changes in the broader political, social and legal context. A common scenario for improved torture prevention is one in which preventive interventions result from an improved political environment, such as the end of an authoritarian governmental system. Yet it will be important to determine whether improved torture prevention would result from the change in political environment alone, or whether the specific interventions are required. Conversely, it would be important to establish whether a similar portfolio of preventive interventions would be effective in the absence of these political preconditions. In other words, it should be possible to determine both the necessity and sufficiency of particular factors that tend to lead to a reduction in the risk of torture.

Ragin argues that this methodology transcends the distinction between qualitative and quantitative analysis.¹⁰ This will be important for this particular study. For the purposes of answering the question did the risk of torture decrease or not, for example, different approaches will be taken in different countries. In some instances multiple systems estimation may be possible (which is a quantitative method). In others, a variety of other approaches may be used, including qualitative comparison between reports gathered by different monitoring bodies, formally reported cases of torture, or structured interviews with key informants. It is important to have a comparative method that allows the inclusion of data gathered using a variety of different methods, providing of course that each of these is robust, valid and reliable.

In other words, QCA functions as a kind of “meta-method” for the entire project. There will be a variety of other national-level methods, to be developed in the second stage of the project.

4.2. Assessing decline in risk of torture

⁸ Charles C. Ragin, *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies* (Berkeley: University of California Press, 1987).

⁹ Bernard Grofman and Carsten Q. Schneider, 'An Introduction to Crisp Set QCA, with a Comparison to Binary Logistic Regression', *Political Research Quarterly*, 62 (2009).

¹⁰ Ragin, *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*.

The difficulty in assembling reliable data on incidence of torture has already been noted. The problem for the purposes of this research is that estimates of the incidence must be made for at least two (and preferably more) different points in time. It is simply unrealistic to imagine that the same method can be used in all countries to measure changes in the incidence of torture. Detailed methods will be elaborated in Stage 3 of the project. These are likely to fall into one of the following three categories: multiple systems estimation; use of reported cases; NGO/NHRI documentation combined with key informant data.

4.2.1. Multiple systems estimation

Multiple systems estimation is a statistical technique that is used for estimating unknown populations. It was originally devised to estimate fish populations from fish that are tagged and later recaptured. The same technique has been used in social research, including estimating those affected by certain human rights violations, for example by truth commissions in Guatemala, Peru and East Timor.¹¹ The technique is based upon the ratio between cases reported to two or more monitoring groups and the total number of violations. The margin of error decreases with the greater number of data sources.

MSE is in principle ideal for estimating a total population that is unknown – in this instance victims of torture – and has indeed been used by truth commissions for estimating numbers of torture victims. The problem, however, is likely to be that the small amount of data available will create an unacceptable margin of error. For the purposes of this research, which aims to track changes in the frequency of torture over time, MSE may prove to be less suitable.

4.2.2. Official reported data

In some instances there may be a considerable amount of data available through reports of torture to official institutions, whether investigations, complaints to institutions such as NHRIs, or judicial proceedings. Some of these data will not be reliable, given that numbers of complaints may not correspond in any predictable way to numbers of actual instances of torture. The larger issue, however, is that the availability of channels of complaint inevitably stimulates usage. So, when authorities introduce a means to complain as a preventive measure, this may encourage complaints at a time when the actual incidence of torture is declining.

These sorts of data are most likely to be useful in countries where the same procedure has been in place for a significant period of time, which makes it more likely that variations in reported cases will reflect the reality. Careful evaluation will be required in each instance.

4.2.3. Informal reports and key informants

In a number of countries it will be necessary to rely on qualitative and often essentially anecdotal data. In these instances the project will depend upon the

¹¹ Landman, *Studying Human Rights*.

expertise of key informants in organizations that regularly monitor torture at a domestic level, including national human rights institutions and non-governmental organizations. Although this method would be incapable of detecting subtle changes in the level of risk of torture, it is actually likely to be reliable for significant declines in the level of risk (or the opposite).

4.3. Evaluating national level impact

Much of the contribution of this research will be in detailed evaluation of specific interventions at the national (or even local) level. Usually it will not be possible to link specific interventions to the overall decline in torture, but it is nevertheless extremely useful to be able to determine whether preventive activities have the intended proximate effect.

For example, what is the effect, both short-term and long-term, of the criminalization of torture? The expectation, in a situation where torture is a serious problem, would be a short-term increase in prosecution of officials for torture followed by a long-term decline.

What is the impact of training of officials? Training is usually only evaluated in the short term. Longer term follow-up would indicate whether knowledge and skills acquired through training has been retained and applied, and also whether it has been generalized to other officials who were not trained directly.¹²

What is the level of compliance of authorities with the recommendations made by visiting mechanisms? If visiting is conducted on a systematic basis it should be possible to measure compliance through repeat visits. Systematic analysis and comparison of these data could provide important indications of the impact of visiting mechanisms.

These and other examples of detailed impact assessment are essential to the success of this research. They will provide the texture that will illustrate exactly *how* torture prevention works. They will also offer of a legacy of evaluation techniques that can be used in future by organizations engaged in torture prevention.

¹² Richard Carver, 'Assessing the Effectiveness of National Human Rights Institutions', (Versoix: International Council on Human Rights Policy, 2005).