

Factor interaction in prevention of torture

Reflections based on Carver and Handley's research

Hans Draminsky Petersen, MD.*

Key issues box:

- Preventive means should be seen as a complex of factors that interact in synergy and, together, they impact on practices in detention where torture is committed.
- Ranking the importance of factors can be problematic since it draws the focus of prevention to components that may only work fully in synergy with other components.
- The merging of many different observations concerning preventive actions of diverse quality may blur outstanding results. When designing prevention programmes inspiration should, inter alia, come from identified best practices. High quality work pays off.
- Assessing accurately the "burden" of torture (torture vs. severe torture) is difficult and the distinction is problematic from both a legal, strategic/political and practical/research point of view.
- The exclusion of intentional ill-treatment during interrogations in research on torture is problematic, given the blurred border to torture and given the frequent manipulation of definition or misclassification of torture by authorities.

Summary

"Does torture prevention work?" is a very comprehensive book based on commendably profound research in 16 countries. It contains a wealth of very important results concerning the relationship between a multitude of factors in the prevention and occurrence of torture. However, the results described may be interpreted in a manner different to how it was done in the book. The intention of this paper is to draw attention to some challenges in the research design and to give a broader view of the complexity of torture prevention.

The book: The authors have identified a host of preventive factors, organised them in clusters (detention law and practice; prosecution (of torturers) law and practice; complaint law and practice; and, monitoring law and practice), and scored them according to whether they fulfil international standards. A torture score comprising frequency, geographical spread and severity of torture was constructed (CHATS). Ill-treatment was excluded. Correlations between preventive factors, clusters and CHATS were calculated. However, the interrelationship between various factors and clusters was not analysed. The main findings included that detention practice had the strongest (negative) correlation to torture and that the torture scoring, pooled

*) ex-member and vice-chair of the UN Subcommittee on Prevention of Torture (SPT).

Correspondence to: hdp@dadlnet.dk

for all countries, declined during the study period (1985–2014). **Comment:** For several reasons, distinguishing ‘more severe torture’ from ‘torture’ is problematic. Excluding ill-treatment in the research is also problematic because the border between the two may be blurred and difficult to interpret and it may be manipulated by authorities, leading to falsely low torture incidence. The pooling of data of diverse quality may hide outstanding prevention results. Identification and implementation of best practices is recommended. Preventive means implemented with low quality may give legitimacy to practices in torturing detentions. The quality of preventive actions is key to efficiency. Factors and clusters of preventive means interact in synergy making each other fully effective. A new model for torture prevention is proposed, which emphasises that all preventive means interact together with transparency, lack of corruption and reprisals, forming the practices in detention where torture takes place. The political will to prevent torture is a key factor.

Keywords: Torture, ill-treatment, prevention, factors of prevention, interaction, quality, best practice, new model, political will

Introduction

Richard Carver and Lisa Handley have written a great and commendable book on prevention of torture based on original research undertaken in 16 countries covering the period 1985 - 2014,

“Does Torture Prevention Work” (Liverpool University Press 2016; 662 pages). It is a book definitely worth reading for its wealth of information relevant for torture prevention. It also invites reflection on how research on this subject should be designed and how results can be interpreted.

The amount of work invested in the research work is obviously immense. Apart from Carver and Handley (in addition to auxiliary staff, including statisticians), there

must have been researchers employed full-time in each country studied for up to two years, amounting to an estimate of at least 30 years of research. It is practically unthinkable that other researchers could find such time and money to do something similar. Therefore, this book will represent a key source of knowledge in the field of torture prevention and be a valued handbook on torture prevention for many years to come.

Given this inarguable status of the book, it is important that it is analysed thoroughly with respect to the design of the study as well as the interpretation of the results presented. The intention of this paper is therefore not to give a comprehensive overview of existing literature on prevention of torture, but to analyse and comment on the book on its own premise with the eyes of a practitioner. The overall goal is to encourage discussion among practitioners and academics on how torture prevention works and how its effectiveness can be improved. This paper will therefore give an overview of the findings and conclusions in the study, raise some questions regarding methodological approaches, and discuss the findings as well as the implications that may be drawn from this comprehensive study.

The present paper consists of four sections:

- (1) a section describing the content of the book (assuming that only a few professionals dealing with torture prevention have found time to read the whole book¹) emphasising some key issues concerning study design, results presented and the authors’ interpretation of the findings;
- (2) comments on issues mentioned in (1);
- (3) propose a model for the way factors important for torture prevention interact (Figure 2 and 3 of the original book).

¹ A summary of the book can be found at: http://www.apr.ch/content/files_res/apr-briefing-paper_yes-torture-prevention-works.pdf

- (4) a summary of conclusions and main considerations.

The content of the book

The design of the study

The study consists of:

- A section on quantitative analysis of the relationship between preventive measures and the incidence of torture for each year in the period 1985–2014 in the countries studied: The UK, Norway, Chile, Hungary, Indonesia, Israel, Peru, South Africa, Georgia, Tunisia, Turkey, Ethiopia, India, Kyrgyzstan, The Philippines and Argentina (approximately 100 pages).
- Qualitative descriptions of factors of importance for A from the 14 countries where torture was found. In Norway, torture did not exist and Argentina was excluded; nevertheless, some qualitative information from Argentina appears in the book.

The authors argue that qualitative studies succeed in describing complexity, but fail to convince that observations can be generalised; and that quantitative studies often have been unconvincing because they have relied on flawed measures, including failing to recognize that legal obligations must be translated into practice to have effect. They therefore combined quantitative and qualitative measures. They further identified multiple elements of prevention, separated legal obligations—*law*—, from their implementation—*practice*—, and developed a new measure of torture incidence.

Measures in torture prevention, the independent variables

The authors present a model to illustrate the relationship between various important factors in relation to torture prevention. According to this model (Figure 1), the

political environment, in interaction with law and training of relevant actors, determine the practice (implementation) of preventive means. This again determines the degree to which torture is practiced or prevented.

The authors have identified a number of means of prevention and organised them into clusters. The clusters of preventive measures are divided into two groups, one for relevant legislative measures, ‘law’ and one for the corresponding ‘practices’. The eight clusters are: ‘Detention law and practice’, e.g. existence of unofficial /secret detention centres; ‘Prosecution (of torturers) law and practice’, e.g. whether independent investigations were done; ‘Complaint law and practice’, e.g. whether cases were referred to investigating / prosecutorial authorities; ‘Monitoring law and practice’, e.g. conducting interviews with detainees. For more details, cf. below.

All elements of the clusters were scored from 0–2. The highest score of 2 was given if legal requirements or practice were in accordance with international standards. The sum of scorings of elements made up the index of the cluster. Higher scores indicate more complete fulfilment of international standards in law and practice.

All indices increased in value with time. Monitoring indices increased much more than the other indices and their values were by far greater than the values of all other clusters by the end of the study, i.e. by the end of the study, monitoring fulfilled international standards much better than the other groups of preventive means (clusters).

The authors found that law clusters scored better than practice clusters, i.e. law was more often in agreement with international standards than practice. In four graphs, the law and practice scorings are visualized over the study period. The gap between law and practice was greatest for the prosecution indices.

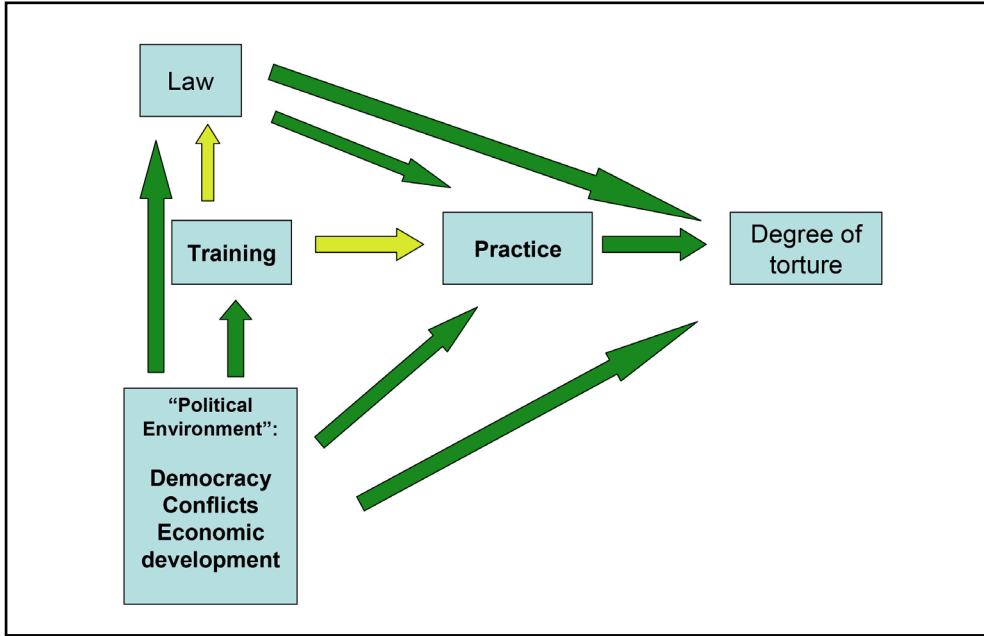


Figure 1: Various important factors in relation to torture prevention (p49)

In the multivariate analysis, four control variables—democracy, national conflicts, economic development and population size—were included. Level of democracy was measured using the Polity IV database; conflicts assessed using the Major Episodes of Political violence index; and for economic development Gross Domestic Product (GDP) was chosen.

Incidence of torture, the dependent variable

The authors assessed a number of existing torture scales and found *inter alia* that, based on one of these scales, torture is much more prevalent in the UK than in Ethiopia! Amnesty International (AI) reports, together with reports from the US State Department, formed the basis for these assessments. The authors emphasise that AI's reports are about known cases and not about incidence,

hence not apt for comparison over time or between countries. Countries with few cases and easy access to information (an aspect of democracy) may ring the torture bell more often than a country with many cases and difficult access to information and /or high risk of reprisals for reporting torture—the “human rights information paradox” (HRI paradox).

Torture was distinguished from other forms of ill-treatment where suffering is not inflicted deliberately (a criterion of torture in the UN Convention against Torture and Other Cruel, Inhuman or Degrading Treatment and Punishment (UNCAT)). The authors acknowledge that ill-treatment caused by poor conditions in prisons may have an impact on the general respect of detainees' rights, including freedom from torture, and forms a part of the mandate of monitoring mechanisms. However, the authors choose to investigate the impact of prevention on the incidence of torture and not on the incidence of ill-treatment. The severity criterion of the UNCAT was employed.

When assessing the incidence of torture, the authors used information from contemporaneous annual reports as a starting point and verified that information using other sources, i.e., reports from international and domestic human rights organisations, official and unofficial statistics and interviews with primary sources. All scores of torture produced by the country researcher were peer reviewed paying particular attention to years of changes.

The authors designed their own score for measuring torture, CHATS (the Carver-Handley Torture Score). Three measures are included in this:

- (1) *Frequency*: on a score 0-3, from absent to routine and pervasive.
- (2) *Geographical spread*: 1 for occurrence in particular regions or areas, 2 for the whole jurisdiction.
- (3) *Severity*: 1 for treatment sufficiently severe to qualify as torture, and 2 for more severe torture. The distinction is based on torture methods used, duration and repetition, and combination of methods.

A value was calculated for each country.

If torture exists (frequency score > 0) the geographical factor was only added to the scoring if torture was observed in the whole jurisdiction and then only with one point. Likewise, the severity factor only added one point to the score if the observed torture was severe.

The authors state that the distinction between torture and severe torture was included because prevention not always eliminates torture, but sometimes prompts changes of methods. This became apparent when dealing with Turkey where changes to less severe methods of torture preceded a decline in frequency of torture. The reason for introducing the distinction was to capture the intent of the state agents who commit torture.

The CHATS values, based on the mean values for all countries taken together,

decreased in the course of the study period. Turkey was mentioned as a success case.

Correlations between clusters of preventive measures and CHATS values

The scoring values of the eight cluster indices (detention law and practice; prosecution (of torturers) law and practice; complaint law and practice; monitoring law and practice), as well as the scoring values of elements of the clusters were correlated to CHATS values; results are given in tables provided in the book. The impact that a cluster or an element could have on other clusters or elements was not dealt with in the book.

The cluster detention practice came out as the best predictor (strongest negative correlation with CHATS) followed by prosecution practice. Correlations between the four law indices and CHATS were weak.

In the bivariate analysis, all eight clusters came out as significantly negatively correlated to CHATS (the better the law and practice, the less torture). Detention practice was by far the strongest predictor, followed by prosecution practice and monitoring practice. The remaining clusters only correlated weakly.

When the four control variables (democracy, national conflicts, economic development and population size) were included in multivariate analyses, detention practice was the only cluster that (nearly consistently) came out as significantly correlated negatively to CHATS.

Correlations between elements of preventive measures and CHATS

Correlations between elements of preventive measures and CHATS are, in the following, listed for each cluster according to strength of negative correlation.

Detention law: 'Recording of interrogations', 'family notification' and 'unlawfulness of unofficial detention' came out as significantly negatively correlated to CHATS.

Detention practice: ‘Unofficial detention not practiced’, ‘family notification of detainee’s whereabouts’, ‘detainee informed of right to lawyer’, ‘exercise right to lawyer’, ‘medical examination performed’, ‘lower reliance on confession’, ‘recordings of interrogations’, ‘prompt presentation before judge’, ‘video monitors used and available’, ‘training of police and custodial personnel’ were all correlated negatively and significantly to CHATS. All practice elements scored better than law elements.

Prosecution law: ‘Independent investigation’ was much stronger (negatively) correlated with CHATS than ‘existence of substantial penalties’ and ‘no statute of limitation’ and ‘criminalisation of torture’.

Nearly all the negative correlation between ‘prosecution practice’ elements and torture were far more convincing, ‘torture complaints lodged’ being the strongest, followed by ‘conviction rates’, ‘no amnesty for torturers’, ‘complaints investigated’ and ‘suspension of suspected torturers’.

Complaint law and practice: most issues are to a low or moderate degree negatively correlated with CHATS. That law ‘permits complaint mechanism to compel evidence’ and the practice to ‘refer cases to prosecution’ have the strongest correlations followed by ‘training of complaints personnel’, ‘effective investigation conducted and reports of activities published’.

Monitoring law: The best law measure was found to be the ‘power to conduct unannounced visits’, ‘power to conduct interviews’, and ‘requirements to produce reports’.

Monitoring practice: Best of all monitoring elements was that ‘monitors were not sanctioned’ for their monitoring activities followed by ‘conducting interviews’, ‘training of monitors’ and ‘publishing reports’.

Both international and national monitoring have a moderate and comparable negative correlation with torture.

Implementation and impact of recommendations based on monitoring were not analysed.

Immunity for inmates for communicating with monitors is mentioned as a very important factor for effective monitoring. Fear of reprisals is mentioned as a major obstacle to effective monitoring, but the issue of reprisals is not systematically analysed.

Law measures were (almost) consistently found to be more weakly correlated to CHATS than practice measures. This was not seen as if anti-torture laws were unimportant; but practice came out as a better *predictor* for occurrence of torture. Laws were seen as a necessary fundament for good practices and not as isolated measures. In this line there was a (weak) negative correlation between ratification of the UNCAT and its Optional Protocol and CHATS.

Training of personnel involved in arrests, interrogations and custody; prosecutors and judges; complaint personnel; police and prison facility monitors were significantly (but not strongly) correlated negatively to CHATS. Training of professional skills of police investigators and doctors was more effective than mere human rights training. The authors refer to anecdotes indicating that sometimes human rights training was regarded as a required ritual that had little bearing in the real world. When Soviet police investigators left Georgia and Kyrgyzstan after the collapse of the Soviet Union, the incidence of torture increased. The new police investigators were not capable of resolving crimes and resorted to torture.

Environmental /control factors: Democracy is supposed to be more prone to offer procedural guarantees, more likely to adhere to rule of law and hold perpetrators accountable. The observations from Georgia and Kyrgyzstan (cf. above) are noteworthy. India is a parliamentary democracy where the incidence of torture has been high during the whole study period.

It was found that democracy and existence of conflicts were (in some of the statistical models) correlated to CHATS in the expected manner, while Gross Domestic Product was not.

The book has a section on transitional justice. The main impression was that it does not work well, partly because of a great delay, but also because the public is “able to simultaneously believe that torture of middle class political activists (the minority of all cases) is heinous while torture of marginalised and allegedly criminal youth is unfortunately necessary” (p88) (e.g. Argentina).

Merging observations from different countries: The authors state that the record on monitoring and complaint is uneven in many of the countries studied, making it difficult to draw firm general conclusions on efficiency (when pooling very different results).

Comments and reflections

In the following I will present comments and reflections based on my reading of the study, and in particular, the points raised above.

Exclusion of ill-treatment from the analysis

The authors restricted their analysis to the concept of torture and excluded ill-treatment. Ill-treatment, in relation to torture and detention, may mean different things:

- (a) Deplorable conditions in cells, i.e., non-intentional suffering most often caused by lack of resources, which the authors excluded. This may be regarded as an aggravated form of “pain or suffering arising from, inherent in or incidental to lawful sanctions” (UNCAT § 1), which the Convention does not include under the definition of torture.
- (b) Beatings, threats, humiliations, deprivation and exhaustion procedures during interrogations and detention not reaching a level that qualifies it

as torture, i.e., *intentional* infliction of suffering by officials, for a purpose, but not meeting the severity criterion. This form of ill-treatment has similarities with torture; its infliction of pain and suffering is intentional, purposeful and committed by a public official.

The UN Convention against Torture and Other Cruel, Inhuman or Degrading Treatment and Punishment, deals with ill-treatment as well as torture. As an example, the fact that its §16 expresses that provisions in §§10-13 apply for other cruel, inhuman or degrading treatment and punishment as well as torture. Although only torture is mentioned explicitly in those articles, the connection between torture and ill-treatment is underlined.

A frequent and serious problem in the fight against torture in countries where torture is widespread is that the states do not acknowledge its existence. The threshold of suffering between ill-treatment and torture is set very high by authorities with the consequence that many cases that should have been classified as torture are only labelled as minor offences, e.g., abuse of power, or excessive use of force, or enhanced interrogation techniques. This problem is related to (c) below.

- (c) In countries with a certain level of access to information, torture consists of methods that do not leave physical marks e.g., suffocation with a plastic bag or with water; a combination of repetitive non-violent beatings, threats, sleep deprivation, physical and psychological exhaustion and confusion procedures (Spain). These are far more difficult to document than bruises and fractured bones. Nevertheless, these acts should be classified as torture.

It is not quite clear how the authors tackle these challenges, i.e., the manipulation of terminology by authorities and interpretation

of those cases on the border between torture and intentional ill-treatment. Moreover, in the chapter on Hungary, it appears that statistics on ill-treatment exist and that ill-treatment (supposedly in the intentional form related to detention and interrogation) apparently contributed to determine the CHATS value for Hungary.

If the intentional ill-treatment (above (b)) is excluded from the discourse it will, particularly in combination with (c), give states where torture is used in the fight against terrorism and crime, an incentive to shift from “classical” physical torture methods that leave marks and chronic physical conditions to more sophisticated methods that are easier to conceal, much more difficult to document and probably much more difficult to raise public awareness about; but equally effective for achieving the purpose (e.g., obtaining confession or information whether true or false, breaking up groups of “terrorists” or felons, or intimidating the general population) and detrimental for the long term psychological health of the victim—and of course still a serious crime. It could be feared that the success of torture prevention in Turkey found by the authors is a result of such a shift from classical very violent torture to more refined methods now labelled ill-treatment—or at best “less severe torture”; and labelled that way it does not add points to the CHATS value (cf. below). It is not clear how the criteria for distinguishing torture from ill-treatment and less severe from severe torture worked in practice. Is e.g. water boarding torture or severe torture? And how is the scoring of all three CHATS elements done if the two types of torture coexist?

A related problem is that the interpretation of the definition of torture may shift as mentioned in the book. British authorities used the “five techniques” (deprivation, exhaustion and confusing procedures) against

Irish Republican prisoners, which in 1978 was classified as “inhuman and degrading treatment” by the European Court of Human Rights (ECHR). However, in the 2000 case of *Selmouni v. France*,—the ECHR regarded a significantly less severe treatment as torture. This could make comparisons over time difficult and it underlines the difficulty that arises when only torture, and not ill-treatment, is considered in the analysis.

It could be argued that particularly because the distinction between torture and ill-treatment is blurred and often subject to manipulation, ill-treatment must be included in measurements of abuses in detention. Moreover, authorities who resort to ill-treatment described under (b) above can be expected to cross the ill-treatment/torture border if they find it necessary for their purpose.

The distinction between less severe and severe torture and the CHATS

The authors argue in favour of the distinction between less severe and severe torture by referring to observations that a shift towards less severe torture often precedes a decline in the incidence of torture and that they wanted to capture the intent of state agents who commit torture.

An alternative interpretation is that such a shift to less severe torture made it easier for authorities to misclassify cases of torture as e.g. abuse of power, hence achieving a better CHATS scoring. It could be that the intent of officers was to conceal the practice of torture using methods that are harder to detect, which can hardly be seen as a great success of prevention, particularly if such practice is accepted by superiors, politicians and the public as less controversial and thereby used more generally against more suspects.

CHATS is a measure of incidence (frequency and geographical spread) and

severity of torture. How observations on frequency were translated into scorings does not appear clearly with an example.

It appears that the way CHATS works—only severe torture adding points to the score—means that, for example, a few cases (1 point) of severe torture (1 point) spread over the whole country (1 point) will score as much as routine (3 points) less severe torture (0 point) in half of the country (0 point). A scenario that may occur in countries with internal conflict (Turkey?). Or: a country that shifts from more to less severe torture and at the same time spreads its practice from one province to the whole country will have an unchanged score. At first glance, this does not seem a fair picture of the practice of torture in the country. It would have facilitated the understanding of the appropriateness of the CHATS if the book had given examples of how the CHATS had been applied on observations from the individual countries.

A further consideration could be that subdividing torture into less severe (“soft” or “clean”) torture and severe (mainly physically mutilating?) could jeopardise the fundamental concept that *all* torture is extremely cruel and that it has severe consequences for the majority of victims—in its less severe form maybe “only” psychological and social. A country that shifts from severe to less severe torture will—all things being equal—be seen as having made progress, which is hardly in line with the spirit of the pertinent international conventions.

The case of Argentina illustrates another problem: (registered) incidence of torture fell when democracy was restored; but it persisted. It could be speculated that the initial fall in incidence was limited to political cases of high visibility and awareness, while torture of ordinary criminal suspects remained constant, and that (at least part of) a subsequent raise in

incidence was due to the HRI paradox: with more democracy, more cases of torture of ordinary criminal suspects (the majority of cases) became known due to more transparency in the administration of justice.

Environmental control factors included in the multivariate analysis

As the measure of economic development, GDP was chosen and it was found not to be related to CHATS. Some sort of average national economic potential could be thought to impact on incidence of torture, e.g. via availability of technical resources for police investigation and training of all relevant staffs. However, it could be argued that the level of inequality would be a much more relevant measure since inequality is correlated to political instability and crime, and torture could be a means to maintain power and wealth for the privileged, who feel a need to be tough on crime.

Correlations of preventive factors with CHATS

Components included in the clusters

- (a) ‘Unofficial detention not practised’ was found to have the “highest impact on the incidence of torture” (p68). This way of wording may be problematic since a cause-effect relationship is not fully documented. In the following the phrase ‘negative correlation to CHATS’ is used. However, the finding is not surprising since there is no outside scrutiny to protect the detainee’s rights. Unofficial detention has been practiced in countries in conflicts like Argentina in the seventies and later by the US as “extraordinary rendition.” In these cases it can be assumed authorities at the highest level were aware of this form of detention and used it deliberately as a means to, *inter alia*, extract information under torture. After unofficial or secret detention, many victims disappeared.

On the other hand, a ban on unofficial or secret detention may be seen as an indicator for governments' willingness to abolish or limit torture. If so, one could suppose that the government at the same time would introduce other measures against the use of torture, indicating that it may be problematic to look at clusters and elements separately because connections between components and clusters exist. However, unofficial detention could also be seen as an indicator that government has little control over this and potentially also other activities of the police. Again, this indicates that there are strong connections between various factors.

- (b) 'Training' of all relevant actors comes out as significantly (but not strongly) correlated negatively to CHATS. It could be supposed that the way training is organised and its quality is very important for its impact, i.e. implementation of changes of institutional practices. After two to three days of training a police officer may return to untrained colleagues, who may not be willing to listen to the wisdom learned, and a change of institutional culture cannot be expected. A huge number of training workshops carried out by a state may look impressive, but the impact may swiftly vanish if the management of the institution and colleagues are not prepared to implement changes. To be effective, training must be organised in a way that facilitates a change of the culture of the whole institution. The quality of the preventive means can be supposed to be key to impact.
- (c) The existence of 'medical examinations' is found to be among the elements that had the strongest negative correlation to

CHATS, which is a little surprising. In many countries, a medical examination is only built on an encounter between doctor and detainee of a few minutes (e.g. some places in Mexico). The examination can be regarded as a formality without practical significance and is sometimes conducted in the presence of the police. In some countries (e.g. Spain), police officers are often within hearing distance. Such factors will discourage many victims from relating exposure to torture. Documentation of torture is thereby impeded and the preventive effect of the medical examination is jeopardised.

Moreover, in some countries (e.g. Spain) a medical document mentioning allegations of torture and describing bruises very rarely leads to investigation and prosecution. To be effective in the prevention of torture, a medical examination and its record must be of a certain quality, and documentation of torture should have consequences in terms of reporting and investigation. Doctors working in places of detention are rarely fully independent and have had little or no training in how to identify a victim and document torture (although Turkey is mentioned as a case where training has taken place).

If a medical examination routinely and with good quality is carried out and the court subsequently acts in cases where torture is documented, this can be supposed to have impact. Such cases may indicate some commitment of the authorities to fight torture and it could be supposed that the government would implement other preventive actions, i.e. indicating connections between preventive means. If, on the other hand, the quality of the medical

examination is low and no action is taken if torture is documented, the medical examination setup can be regarded as a means to blur the facts. In the worst case, authorities will use it as an argument that torture does not exist. This may in particular be the case if the torture used in general is too sophisticated to be documented by overburdened non-specialist doctors working in places of detention.

When a judge considers an allegation of torture, the assessment is based on the statements of the alleged victim, those of police officers and a medical forensic report. In general, the credibility of officers is valued much higher than that of a suspect. Furthermore, the judge may not be fully independent and may not have a full understanding of medical procedures and terms and the potential for documenting facts. Medical reports are often of low quality due to lack of medical independence and skills, i.e. doctors lack training, instructions and supervision. In addition, they may be under time constraints and under pressure from police officers, who are present within sight and/or hearing distance.

- (d) 'Prompt presentation before a judge' is a component of the detention clusters, but the quality of the courts actions, inter alia, based on independence, dismissal of evidence obtained by way of torture and ill-treatment, freedom of corruption and scrupulous scrutiny of all material, including medical documents, is not dealt with in the study. The torture preventive effect of the court depends on other preventive factors, e.g. the work of the lawyer and the medical examination, underlining that some preventive factors within the same cluster (e.g. detention practice) interact.

It can be supposed that, if courts took action based e.g. on good quality medical documents and initiated investigations on possible torture (UNCAT obligation), it would impact on the practice in detention. (Some) officers might refrain from torturing due to the risk of prosecution.

- (e) Exercising 'the right to a lawyer' correlated negatively to CHATS.

The mere presence of a lawyer in the detention would deter officers from torturing. However, this right may be completely watered out as it has been e.g. in Spain where terrorist suspects, while in detention, may not talk with or even see the appointed lawyer, whose only role until the suspect appears in court is to witness the suspect's declaration to the police.

Preventive means may as such be used in ways that only give legitimacy to practices in torturing detentions, i.e., such means may be counterproductive when not being implemented with rigour and good quality.

- (f) 'Reliance on confession' as evidence is a historical and serious problem, particularly when the confession it is given to the investigating police. This problem still exists in many countries. The Italian criminologist Cesare Beccaria (1738-1794) commented: "Torture is the surest means to acquit the robust felon and condemn the weak innocent." Particularly in combination with requirements for the police to solve a certain number of crimes, the reliance on a confession as the only means of evidence may lead to more torture; e.g. the authors found that in Kyrgyzstan, torture incidence increased at the end of the month when the police struggled to fulfil the monthly demands.

Torture can be seen as a shortcut to “solving” crimes (i.e., getting a confession), particularly if police agents are badly trained in investigative techniques, are under undue stress, including being rewarded for “solving” crimes. Occurrence of torture may, as such, in addition to lack of training and undue incentives, also be related to lack of control of the practices of the police. With the collapse of the Soviet system, oversight and discipline among police officers vanished. Officers were vested with a lot of power, little training and low salaries, a classical prescription for corruption and an incentive for police officers to extort under torture or threats money from arrestees and their families—with or without existence of democratic elections.

- (g) ‘Corruption/extortion’ may be an incentive to use torture. Corruption was nevertheless not included in the research. If a police officer gets a salary that does not allow for making a living for himself and the family, there is a serious risk of becoming corrupt. Detainees may be forced by means of torture or ill-treatment to pay their way out of the police station or to exercise their rights. This practice has been documented by the UN Subcommittee on Prevention of Torture (SPT) in several country visits and this form of corruption is probably commonplace in many countries. It runs in parallel with torture as a way to disrespect detainees’ rights. Practices of corruption may be reduced via “bookkeeping” in detention; training, instruction, supervision and oversight of officers; laws; monitoring; complaints and prosecution—and fair salaries.

The practice of corruption may indicate that individual police officers operate in their own interest without

disciplinary action taken or any form of outside control; it may also be part of a pervasive institutional culture involving all those who have some power they can abuse, including strong prisoners. This underlines the connection between different means of prevention and corruption (see Figure 3).

- (h) ‘Complaint law and practice’ components were to a low or moderate degree negatively correlated with CHATS. That law permits complaint mechanisms to compel evidence and the practice to refer cases to prosecution had the strongest correlations. Power used by authorities to prevent torture, leading to legal consequences, seems to work to some extent.

However, the complaint system may be designed in a repressive manner. In Hungary, persons who have made allegations of ill-treatment against named officers may be charged with slander for making unfounded allegations. The basis for deeming an allegation “unfounded” could, *inter alia*, be a medical examination, the quality of which is crucial. The charge could be seen as a kind of reprisal.

Effective complaint mechanisms, which include investigation of well-founded complaints, require freedom from all forms of reprisals, including physical punishment in detention. If the risk of reprisals is considerable, the incentive for victims to complain will vanish. Effectiveness is related to other factors, such as medical documentation of torture and risk of reprisals, law, training and control means. Complaint practice and reprisals may influence each other, e.g. the more reprisals, the less complaints and prosecutions; and the more effective the complaint

and prosecution mechanisms, the less reprisals (see Figure 3).

- (i) The best ‘monitoring practice’ measure was that monitors are not sanctioned for monitoring activities. This is seen by the authors as the ability of monitors to do their work without threats and sanctions. This may be difficult to understand, since it cannot be assumed that it is the absence of sanctions, but rather the conducting of interviews and producing reports with recommendations for change that could cause the changes—together with the torture deterring effect of monitors’ presence in places of detention. It seems logical to regard absence of sanctions for monitors as a prerequisite for effective monitoring, like law is for practice. (It should be added that the deterring effect of monitors’ presence will normally require very frequent visits). It could be interesting to know where, how often and under which circumstances monitors were sanctioned in order to really understand the significance of this finding.

Conducting interviews, which is the main activity of monitors during visits, only came in as number two regarding strength of correlation to CHATS. The authors mention immunity for inmates for communicating with monitors as a very important factor for the efficiency of monitoring. Fear of reprisals is mentioned as a major obstacle to effective monitoring (for interrelations between reprisals and other factors, cf. above), but the issue of reprisals is not systematically analysed.

Publishing reports on monitoring and complaints was significantly and negatively correlated to CHATS. It would be an indication that *transparency* in administration of justice impact on public opinion and political will to implement

changes, leading to strengthening of preventive actions and consequently to reduced incidence of torture (Figure 3)

Quality of implementation of previous actions

The authors discuss another important factor in monitoring, namely the power, commitment and competences of actors, i.e. the way the work is done, or the quality of the preventive work. This could also be a valid point in relation to other practice measures, as mentioned under training, presentation before a judge, complaint and medical examinations. Are the resources of the monitoring body sufficient to allow for a reasonable number of visits with a reasonable number of staff? Are interviews conducted in private and not in the presence of police/prison guards (e.g. Ethiopia)? How are detainees selected for interviews; are interviewees e.g. the strong leading prisoners/capos, who may have common interests with staff in collecting bribes or informal payment for services that should be offered for free, thereby sharing an interests in concealing facts?

The authors state that the record on monitoring and complaint is uneven in many of the countries studied, making it difficult to draw firm general conclusions on efficiency when all results are pooled. One could argue the other way round: rigorous (i.e. weekly) monitoring of police stations in certain areas of Georgia reduced the incidence of police torture dramatically. Such intensive monitoring is not described in other countries. It could be inferred that intensive high quality monitoring is very effective (under certain conditions).

Georgia’s outstanding result was not reflected in the overall results. Monitoring in all countries taken together was found to be only weakly negatively correlated to torture. This could be interpreted as a dilution

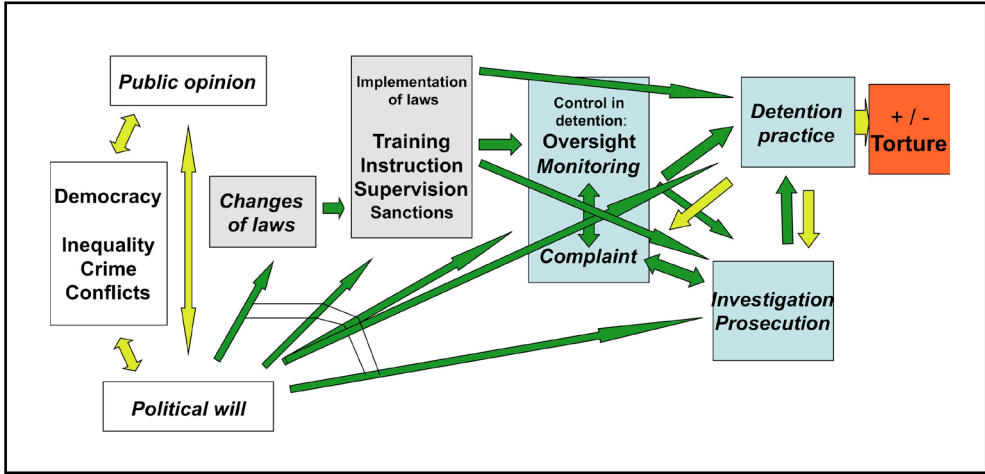


Figure 2: An alternative model for torture prevention—elements included in ‘Does torture prevention work?’

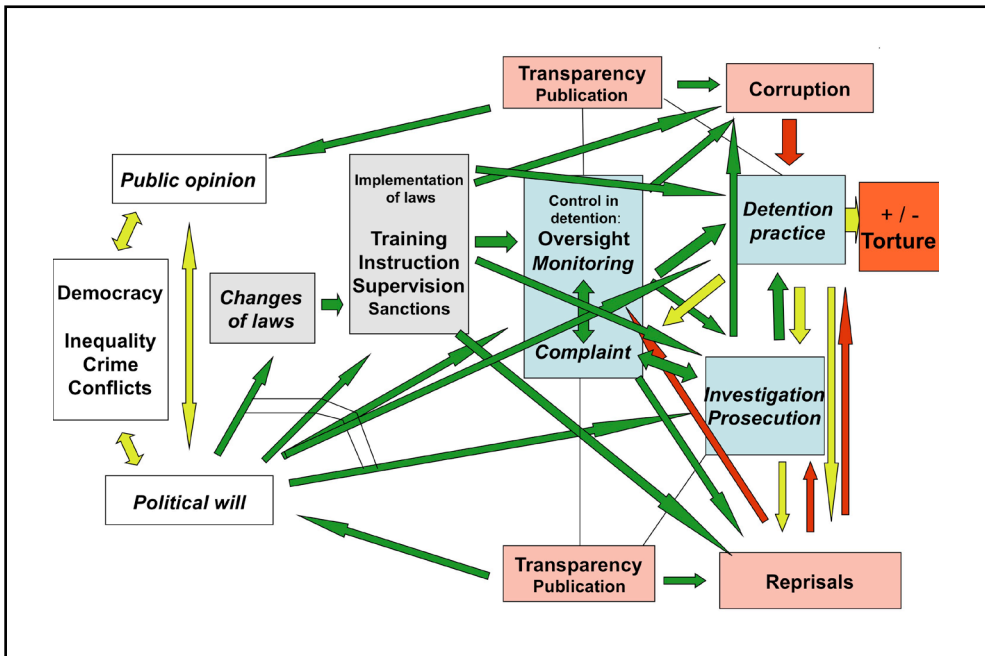


Figure 3: An alternative model for torture prevention with the addition of reprisals, corruption and transparency

of the remarkable results from Georgia when merged with the effect of ineffective (sporadic or low quality) monitoring in other countries. This points to a general key reflection: *The rigour and quality with which preventive measures in all categories are implemented is crucial for effectiveness. Identifying best practice could be key to develop better prevention.* However, it must be added that, simultaneously with the intensive monitoring and reporting, other actions were taken in Georgia: promoting treaty ratification and an anti-torture law, and educating the public and officials. This leads to suggesting an *interrelationship between some “first level elements” of prevention*” (cf. below), *i.e. public opinion and political will and different clusters—and a synergy when several interventions are combined* (Figure 2).

A proposed model for prevention of torture

The impact of monitors’ recommendations on other clusters e.g. detention practice was not analysed, *i.e.* the interrelation of effects of the different clusters and elements. The indices of the two monitoring clusters were those that increased by far the most during the study period—in parallel with the decline in CHATS values—and reached the highest scores, *i.e.* the highest compliance with international standards. A question here is: could it be that the practice of monitoring had influenced the detention practice, which in the authors’ statistic came out with the highest negative correlation to torture?

This leads to the general reflection already mentioned many times above: the clusters of preventive means work together. If some preventive means were hypothetically eliminated, it would most likely have a negative impact on the efficiency of other means. This also works the other way round, so, if one factor is

introduced or made effective, it would most likely have a positive impact on other factors.

The authors have discussed this concerning law measures that (almost) consistently were found to be more weakly correlated to torture than practice measures. They underline that this does not mean that anti-torture laws are unimportant, but only that practice is a better *predictor* for occurrence of torture. A question here is whether a predictor in itself is necessarily equal to preventive potential. A predictor may be heavily dependent on other preventive factors to exert its potential for prevention. Analysing separately the predictive value of a cluster or an element may lead to a misleading interpretation of cause-effect.

An alternative model for prevention emphasising the interaction between the elements of prevention can be seen at Figures 2 and 3.

Explanation of figures

The arrows indicate that the elements inside the boxes impact on elements in the boxes to which the arrow leads. The impact may be positive in terms of prevention, which is indicated by the arrows being coloured green, e.g. monitoring may lead to better detention practice.

- If the impact is negative, the colour of the arrow is red, e.g. reprisals will tend to lead to a reduced number of complaints and investigations, which may lead to more torture through bad practice in detention.
- If an element under some circumstances impacts torture negatively, and under other circumstances impacts it positively, the colour is yellow, e.g. a detention practice where torture is prevalent will lead to fewer complaints, while less prevalent torture in detention with smaller risk of reprisals will encourage more complaints to be lodged.

- The white boxes (first column) represent factors crucial for any change: public opinion, political will, democracy, the level of inequality and crime and conflicts in the state.
- The grey boxes represent the government's endeavours to make changes.
- The blue boxes represent the materialisation of the government's endeavours. They include the four practice clusters with which the authors worked.
- The pink boxes represent additional factors of importance for torture prevention.
- The connecting lines from transparency boxes indicate where transparency should work. Double lines indicate that the actions by the government to abolish torture could be supposed to work simultaneously.

Figure 2 only deals with the elements included in Carver's and Handley's model. The three blue boxes contain the practice clusters used by the authors. The box 'control in detention' is subdivided into (a) oversight / monitoring that interact with (b) complaint. Figure 3 adds corruption, reprisals and transparency. This figure is very complex and a reader may lose the overview. This was the reason for splitting the figure, which hopefully will facilitate the understanding of the complex pattern of interaction

Examples of interactions:

- (a) Public opinion and political will interact under the influence of the levels of existent democracy, crime, conflict and inequality. The arrows are yellow since all the elements may be in favour of torture prevention; but they may also give rise to more power to security forces and reduced rights of detainees, e.g. in a "war on crime and terrorism."
- (b) The political will is paramount to torture prevention. A government may undermine

all elements of torture prevention, in which case all five arrows going out from the box *political will* should be red. This will imply that most preventive means will not work well and consequently most of the positive impact (green arrows from the other boxes) would vanish or could change to be negative.

However, in the model proposed in this paper, it is supposed that a certain political will exist to prevent torture, consequently the arrows from the box *political will* are green. If the political will exists to fight torture, and the government is in control of its institutions, it can be supposed that preventive measures concerning all elements will be reinforced in parallel, which is indicated by the connecting double lines between arrows going out from the box *political will* in both figures. In future studies on torture prevention the political will to prevent torture could be a main issue.

- (c) Laws, including those on control measures (oversight, monitoring, complaints), as well as prosecution and detention practice, must be implemented effectively through training, instruction and supervision of all relevant personnel, as well as sanctions for not complying with laws and instructions. Effectively implemented laws would lead to a positive preventive effect.
- (d) Implemented control measures (oversight, monitoring and complaint practice) together with real consequences for torturer in case of infractions, i.e., prosecution practice, work together forming the practice in detention where torture is committed.
- (e) Good monitoring and complaint handling lead to more transparency, supposedly influencing public opinion and the political will to implement

changes—including the reduction of reprisals and corruption. Effective monitoring could also lead to more complaints and better handling of them, apart from the other effects that monitoring has on detention practice.

- (f) An effective prosecution practice and good detention practice that is not too unsafe (reprisals for complaining), would lead to more complaints. Effective handling of complaints—eventually via prosecution—would lead to a better detention practice.
- (g) Reprisals and corruption can be supposed to exist anywhere torture is used. They interact with monitoring, complaint mechanisms, transparency, prosecution and detention practice.

Torture prevention is here seen as a complex pattern of factors that impact on each other. In this model, good detention practice is the end objective of prevention and it is determined by and it interacts with the other elements.

Torture takes place in detention, which fits with the finding that detention practice is the most important predictor of torture. However, all preventive elements are necessary requirements for effective prevention and, together with additional factors like reprisals, corruption and transparency, they form the practices in detention, i.e. existence or absence of torture.

Summary of conclusion and main considerations

“Does Torture Prevention work?” is a very important book with a wealth of information from research in 16 countries and it is relevant for all who work with torture prevention. The book inspires the reader to reflect on the way elements of prevention work. The present paper comments on the

research design and analyses the results in an alternative way to the authors.

Instead of looking at prevention as components that work independently and in parallel, this paper recommends that prevention should be seen as a complex of factors that interact in synergy.

Detention practice had the strongest negative correlation to torture. This fits the proposed model (Figure 2), but, it gives an alternative interpretation: All factors are important. They interact directly or indirectly and, taken together, they impact on the practices in detention where torture is committed, in that way produces a preventive effect.

The scores of the monitoring clusters increased by far the most and reached the highest compliance with international standards over the study period. This does not necessarily undermine the finding that detention practice was found to be the best predictor of torture. It could be that monitoring exerted its preventive effect through its impact on other factors and first of all detention practice. The effect of the implementation of monitors’ recommendations were not analysed in the book.

Predictive values of preventive means for incidence of torture were not documented to be equal to preventive potentials; particularly considering the interaction between different means.

Ranking the importance of factors can be problematic since it draws the focus of prevention to components that may only work fully together with other components.

Merging of many different observations concerning preventive actions of diverse quality may blur outstanding results. When designing prevention programmes inspiration should, *inter alia*, come from identified best practice that should be analysed in order to find out why they worked. High quality work pays off. This would apply for all preventive means.

Training of all relevant actors is a prerequisite for high quality work. The same goes for implementation of learned knowledge and skills. The quality of implementation of such learning and control measures—supervision of and oversight with actors—was not assessed in the book.

The quality with which preventive means are implemented and maintained could be a research issue and it should be a priority for monitoring bodies.

Preventive means can be used in bad faith by authorities by converting them to mere formalities with no real value (e.g. the right to a lawyer); by not ensuring a minimum of quality (e.g., low quality medical examinations); or by being repressive (if a dismissal of a complaint leads to persecution for defamation). In this way, they may give legitimacy to a badly functioning administration of justice; and, in a worst case scenario, be counterproductive.

Assessing accurately the “burden” of torture (torture vs. severe torture) is difficult and the distinction is problematic from both a legal, strategic/political and practical/ research point of view.

The exclusion of intentional ill-treatment in research on torture is problematic, given the blurred border between ill-treatment and torture and given the frequent manipulation of the definition or misclassification of torture by authorities.

Transparency in administration of justice, reprisals against detainees who report torture and corruption (Figure 3) were not made subject to particular analysis. Transparency can be supposed to have a preventive effect through an effect on public awareness and decision makers.

Reprisals against detainees who report torture can be supposed to have a serious detrimental effect on torture prevention.

Like torture, corruption is an expression of a lack of respect for the rights of the detainee. Hence, torture and corruption can

be supposed to exist in parallel. Bribing with the aim of exercising one’s rights in detention may include the right not to be tortured.

A crucial condition for preventing torture is that authorities at all levels, particularly the government, are committed to do so. Such a commitment can be expected to influence many preventive means simultaneously.

With this paper, it is hoped that scholars and practitioners within the field of torture prevention will be encouraged to read the important book analysed here and take part in discussion on how best to design research and set priorities for torture preventive work. Such discussions and deliberations should encompass not only the existence of factors of importance for torture prevention, but also the identification of indicators for high quality torture prevention.

Does torture prevention work? Reply to Dr Hans Draminsky Petersen

Richard Carver*, Lisa Handley**

Introduction

We thank Dr Hans Draminsky Petersen for taking the time to engage so thoroughly with the arguments and findings of our book.² We also thank the editors of *Torture* journal for giving us the opportunity to respond. One of

*) Senior Lecturer in Human Rights and Governance, Oxford Brookes University

**) Visiting Research Academic, Oxford Brookes University

Correspondence to: rcarver@brookes.ac.uk

² Richard Carver and Lisa Handley, *Does Torture Prevention Work?*, Liverpool University Press, 2016.

the aims of our research was to open a serious debate on the impact of torture prevention measures and this is just such a discussion.

It is encouraging that there seem to be large areas of agreement between Dr Petersen and ourselves. Inevitably, in this response we will focus on the points where he disagrees with us, but we think it is important to emphasize the large amount of common ground.

We will start with a couple of general points about the methodology of our study, since these inform our response to Dr Petersen on a number of the issues he raises (and since we do not have space here to address every single point in his article). The first point is the very narrow scope of our research question. We addressed only the specific issue of whether the preventive mechanisms required in international law, or commonly recommended by treaty bodies, non-governmental organizations, and others, are actually effective in preventing torture. We did not, as Dr Petersen notes, address their impact on instances of ill-treatment falling short of torture. We also did not consider the possible impact of such factors as political will, corruption or other variables that fall within what we describe in our model as the “political environment.” We strongly maintain that the success of our study was in part a consequence of the narrowness of our research question: do preventive mechanisms reduce torture? While much writing on human rights in society emphasizes a “holistic” understanding of measures to protect rights, our feeling was that we would be better able to understand the impact of measures to prevent torture by isolating these factors (a “reductionist” approach, if you will). This is not, of course, to deny the importance of other factors, such as “political will,” which we, like Dr Petersen, consider important. However, these were tangential to the question that we sought to answer.

The second point about our methodology is that we used both quantitative and qualitative approaches. This may seem self-evident and, of course, Dr Petersen notes it in his article. The significance is that a number of his criticisms refer only to the quantitative findings, ignoring the fact that there is detailed discussion of many of the issues he raises in the qualitative narratives (some 80 per cent of the book), which flesh out and illuminate the statistical findings.

In the remainder of this response, we will address some of Dr Petersen’s specific criticisms of the book.

Severity of torture

One of Dr Petersen’s strongest criticisms is of our use of the severity of torture as an element of the scoring method that we devised. We embarked on the research hoping that we could use one of the existing indices that scholars have developed to record the incidence of torture. In the event, we quickly realized that these indices were often inaccurate both between countries and within any given country over time. Hence, we developed our own scores based upon a combination of contemporaneous reporting and primary research on the part of our team. There clearly needed to be a means to determine changes in the incidence of torture over time. Simply using the number of torture cases reported would not work for a number of reasons that Dr Petersen discussed. Our torture score (CHATS) therefore includes a number of elements: frequency of torture, geographical spread, and severity of the methods used. This approach could be applied equally in Norway and Ethiopia.

We reflected deeply on the severity question, with lengthy discussions within the research team and consultations with two past UN Special Rapporteurs. A number of points should be made clear. First, we understand that the severity of torture is always

situational and contextual. Our score does not make an assessment of any individual case. Secondly, our consideration of severity was absolutely *not* based on whether torture left visible marks. Stealth torture techniques are not, in principle, less severe than “classical” torture. The assessment is based upon a number of elements, including the duration of torture and the combination of different techniques. To answer Dr Petersen’s question, we would consider waterboarding, as normally employed, to constitute a more severe type of torture.

However, Dr Petersen’s core objection to our approach is that there is something inherently, perhaps ethically, wrong in distinguishing different levels of severity of torture. We disagree. A distinction is already made between different forms of ill-treatment, with some rising above the threshold required to be defined as torture. Once it is accepted that a severity threshold is possible, logically it makes no sense to say that all ill-treatment above that threshold is equally severe. Practically, we observed that the effect of prevention measures, in some instances, was a change in the methods used towards less severity (while still above the torture threshold). We certainly do not dissent from Dr Petersen’s observation that this is not “within the spirit” of the instruments prohibiting torture. But as a matter of social science, we considered the finding relevant.

The exclusion of ill-treatment from the study

As noted above, ill-treatment not constituting torture was not considered in this study, essentially in order to make the research question manageable. It is reasonable to assume that had our question been different, so our answer might also have been different.

We do not accept Dr Petersen’s suggestion that this lets governments off the hook because they redefine torture as

ill-treatment, since we did not use national definitions of torture but the one to be found in Article 1 of the UN Convention Against Torture.

A related point does deserve discussion. Where does purposive ill-treatment that falls short of the severity threshold fit into this? As we note in the book (p. 37), we are in agreement with leading scholars in the field (Rodley, Nowak, Evans) that the key distinction is between purposive and non-purposive ill-treatment and that torture could be simply defined as the former without departing from Article 1. At the outset, that was the approach that we took. The problem, of course, is that most of the secondary sources we drew upon did not share that definition. So, we found ourselves back with the severity threshold. We doubt that this greatly affected the scores that we assigned (just as the shifting legal definition that Dr Petersen alludes to had little impact). The reason is simply that our five-point scale was not fine-grained enough to capture these distinctions, and nor could it have been while maintaining any credible coder reliability.

Best practices

We were somewhat baffled by Dr Petersen’s contention that “The merging of many different observations concerning preventive actions of diverse quality may blur outstanding results.” The specific example that he cites on a couple of occasions is the intensive programme of monitoring visits to police detention in Georgia in the mid-2000s. First, we discuss this explicitly in the book (p. 414). Secondly, outstanding preventive interventions are not “blurred” in the quantitative analysis, but rather contribute to the overall finding. Each independent variable is coded on a three-point scale, so a less complete monitoring body will receive a lower score (contrary to what Dr Petersen seems to be arguing).

Finally, it is important to understand that the definition of “best practice” must be what works most effectively, not an *a priori* set of practices that are commonly believed to be good. For example, we concluded that independent complaints bodies (as distinct from independent prosecutors) had no detectable preventive impact. If that finding is correct, then any number of “best practices” in the complaints field would not result in a reduction in torture.

Preventive measures that fall short of “best practice”

Dr Petersen is surprised at our finding that medical examinations are an important measure in preventing torture. Given his long involvement with this issue, we are surprised at his surprise. We can only conclude that he has misunderstood the process by which we coded the presence or absence of preventive variables. In the example of medical examinations, if these were invariably compromised (for example by non-objective doctors or the presence of security officials during the exam) this would be coded as 0—in other words, the same as if there were no examination. Only if examinations were almost invariably independent would the maximum score of 2 be allocated. We completely concur with Dr Petersen’s observation on how such examinations are typically conducted—a point that is addressed in a number of the country chapters (and see also p. 72). Similar considerations apply to many of the preventive measures that he discusses. In our codebook, for example, if a lawyer’s presence is typically delayed, limited or otherwise compromised, this would not be allocated the maximum score of 2.

Interaction of preventive measures and environmental factors

The central argument of Dr Petersen’s piece is that different preventive measures interact with each other and with various political

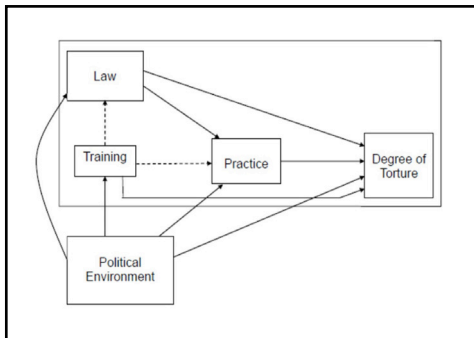
and social factors to result in an effective reduction in torture (and ill-treatment). Dr Petersen criticizes us for not taking account of such interaction.

Our multiple regression analysis did examine the interaction between different preventive measures. We also analysed those measures found to be closely related, such as law and practice (although there was considerable divergence for the most important variables) and the role of training in bringing practice closer to law. We also, as Dr Petersen notes, observed a high degree of correlation between the different sets of preventive measures, with governments interested in reform bringing in a variety of different prevention mechanisms.

Moreover, all the country studies consider the impact of all the sets of preventive measures in concert. As a general proposition, for example, we would accept that a monitoring body might make recommendations on detention safeguards, which are then put into practice resulting in a reduction in torture. However, according to our findings, it is still the detention safeguards that have the primary effect, not the monitoring body. The safeguards could equally have been put in place without any intervention from a monitoring body (the United Kingdom among our case studies offers a clear example, see p. 116 ff).

The aim of our model is modest. It is designed to illustrate those aspects of torture prevention that we addressed in our study. (Dr Petersen’s rendition of our graphic actually omits one element from the original version—a box indicating what falls within the scope of our research and what does not.) All the relationships in our model are testable and were indeed tested in our analysis. The box labelled Political Environment is important and no doubt interesting but was not the focus of our study for reasons explained above. In each country chapter there is some

discussion of the political environment, since this enables us to understand whether preventive measures were initiated and how effectively they were implemented. Of course, the answer is different in each case. To take one example, “political will” in Northern Ireland derived from a desire to end decades of conflict, in Turkey from a wish to join the European Union, in Chile from the effort to rebuild democratic governance, and so on. Our concern was simply to determine, if the political will *were* present to end torture, what methods would achieve this most effectively. Dr Petersen’s models unpack the Political



Environment box, introducing a series of factors that are doubtless relevant in a broad sense but do not help us to understand the simple question of whether preventive mechanisms work and which are most effective at reducing torture. They lead us into territory that is well-trodden, but which provides little practical guidance to the human rights community. Likewise, the claim that everything interacts with everything else is not, we think, a very helpful way of proceeding, leaving us back where we started—with a long list of preventive measures, each of equal and indemonstrable value.

Response by the author

Hans Draminsky Petersen, MD*

Thank you to Mr. Carver and Mrs. Handley (C&H) for their response to my paper and to the editor for the opportunity to respond.

I see discrepancies in expressed views as reflecting disagreements as well as misunderstandings. In the following I’ll respond to C&H’s reply.

The distinction between torture and severe torture: C&H argue as follows: a distinction already exists between different forms of ill-treatment, with some reaching above the threshold required to be defined as torture; logically it makes no sense to say that all ill-treatment above the torture threshold is equally severe, when seen as a matter of social science.

Seen with my medical eyes a key problem is that severity of torture and ill-treatment is very difficult to “measure.” Duration and application of different methods in combination are obviously relevant indicators. However, the problem could be illustrated by asking where exactly is the threshold drawn between ill-treatment and torture and between torture and severe torture? In my view an impossible question to answer in precise terms according to a formula within medical science. Hooding and forced standing for one hour may not be classified as torture, but if it goes on for weeks it certainly should. But drawing an exact “quantitative” line is in my view too difficult and may lead to exclusion of cases that assessed “qualitatively”/individually would be found relevant for the purpose of preventing torture and ill-treatment.

C&H's view is that torture is "situational and contextual." I don't know whether this includes personal factors: the victim's vulnerabilities, his/her own and those related to his/her network. It could be argued that assessing the severity of ill-treatment should include its impact on the victim, psychologically as well as physically. As an example, the impact is different if forced standing for a given period of time is applied to a fit young man, a frail or disabled person or a pregnant woman. In this example a cut-off value would not make sense. In my view classification of cases as torture would very often need an individual assessment.

In individual cases a court may have to draw a line between ill-treatment and torture since the distinction exists in international law. However, I disagree that this juridical distinction, based on profound deliberations in every individual case, can be used as an argument to apply the same distinction *generally* in medical or social science without appraising the individual cases or a sample.

Exclusion of ill-treatment from the study:

Some national governments and courts *misinterpret* the definition of torture in national and international law and *misclassify* acts that amount to torture as less serious crimes. Additionally, the national definition may be insufficient. These means are used by some national authorities to get themselves "off the hook"—and it leads to the perpetuation of impunity.

Whether or not the inclusion of purposive ill-treatment in the study would have affected research results is a matter of speculation. But it remains that it obviously is an objective to prevent ill-treatment as well as torture, including the cases that are

difficult to classify which fall into a grey zone between torture and ill-treatment.

Best practice:

I agree with C&H that best practices should be appraised by their results. The finding that intensive Georgian monitoring led to the dramatic reduction of police torture was among the very important ones. Maybe our disagreement stems from misunderstandings. An *outstanding* result should of course be analysed as to the reasons for achievements. One crucial characteristic, but probably not the only precondition for effectiveness, was that visits to problematic police stations were frequent, an observation that makes a lot of sense and could be applied in other settings.

Quite another issue is what happens when merging such an achievement with the results of "ordinary" monitoring in other countries. In C&H's overall results monitoring was not very strongly correlated to CHATS, while the Georgian monitoring supposedly was. I apologise in advance for using the following simplistic example to explain my view, but find that it provides a vivid picture: If we mix a cup of boiling water with 15 cups of water that are at an *average* temperature of 20 centigrade, the temperature in the mixture will be 25 degrees. If we only look at this we would not infer that one contribution to the whole was very hot. In the merging of research results an outstanding preventive potential may be overlooked/ blurred. Clearly, C&H's presentation of qualitative as well as quantitative data prevented this from happening. However, if merged results are used to recommend specific prioritisation of preventive means it becomes very problematic since an evident preventive potential is ignored.

C&H's example concerning complaints: if all identified attributes of an independent complaint mechanism are assessed to work well and the mechanism nevertheless is

*) ex-member and vice-chair of the UN Subcommittee on Prevention of Torture (SPT)

Correspondence to: hdp@dadlnet.dk

ineffective in preventing torture it could be that unidentified factors would have importance. According to my model such a factor could be the interaction with effective prosecution of torturers and/or reprisals.

The preventive effect of medical examinations: I appreciate the comments in response to my article. I am not sure that I have fully understood the way the scorings on medical examinations works, i.e. which attributes were used and how they were used, e.g. the quality of the medical examination; but I am somewhat reassured that the findings reflect that the medical examination in the 16 countries of the study on average were done in a more effective manner than what I have witnessed in some countries (which were incidentally not included in the study).

Interaction of preventive measures: Here we disagree. As argued in my paper I think the interactions of factors are crucial for effective prevention (cf. also above concerning complaints). Hence, assessing only the effect of individual factors and clusters may be misleading if we are looking for preventive *potentials*. For example, effective monitoring is one among various factors that may change practices in detention.

C&H's example from the UK is interesting. I read p. 117 as follows: In the study period "dramatic improvements in detention practice" were seen as corresponding to changes in legislation—in combination with better discipline among police officers, new police interviewing techniques, moving away from confession evidence, and better complaints and disciplinary procedures. In my view this points to the likelihood that preventive means interact and work in synergy.

I apologise for the omission of the box in C&H's figure illustrating the focus for

the study. However, as expressed in my own figures, I think that an analysis of how torture is prevented should include "political environment"/"political will" and that the example from the UK illustrates this well.

We disagree as to the helpfulness of including interaction of preventive means in the analysis of how torture prevention works and can be improved. C&H express that if everything interacts with everything it leaves us with a long list of preventive measures each of indemonstrable value. My view is different: understanding how factors interact may help us to make each of them effective, e.g. if complaint handling has no observable effect it could, *inter alia*, be an idea to look at the possibility to link it to prosecution of torturers and the existence of reprisals. When attempting to improve a complaint mechanism it is clearly helpful to have an updated assessment of its efficiency and the efficiency of the prosecution of torturers—as well as an analysis of the reasons for (in)efficiencies.

As appears from several of my remarks here, and from my paper, I regard the design of C&H concerning the elements of prevention and the grouping of them as a very useful starting point in the analysis of how torture prevention works and how it can be improved. It may be that a refinement based on experience/observations could improve the design further, e.g. introducing an item on the existence of, and police officers' compliance with, a protocol for interrogation in detention, cf. the UK example. That being so, as a conclusion I see an interesting perspective in combining this part of C&H's approach with an analysis of reasons for (lack of) effects of preventive means in the light of factor interaction.