Journal of Behavioral Economics for Policy, Vol. 5, Issue 1, 13-17, 2021

# Bundling violence: How do people trade-off combinations of violent acts?

Magda Osman[1]*, Isabelle Mareschal[1], Emily Hannon[1]

**Abstract**

Violence risk assessments are used in a variety of settings (e.g. forensic, psychiatric, public) to determine risk of violence of a given individual. The accumulation of data generated from risk assessments can be used by practitioners and policy makers to determine aggregate levels of violence so as to determine the demand of future services, particularly in the domain of violence reduction or prevention. For instance, in prisons risk assessments can reveal the extent to which, between prisons, and across time, violence is on the rise, so as to determine strategies to reduce violence. Often violent outbreaks occur in which different combinations of violence acts within each outbreak are observed. In order to determine future demands on prison services, what approach should be taken to assess if an outbreak in one prison is less than, equal to, or more violent overall than an outbreak in another prison? This is a particularly challenging question to answer because the aggregate score (total violence score) will significantly vary depending on how the severity of violent acts is taken into account; some risk assessments treat all violent acts as equally violent, some rank violent acts according to severity, and some use a weighted sum. In the present study we present a non-expert sample with putative violent bundles (combinations of violent acts), having taken place in a prison setting, in order to determine the most common way in which people intuitively aggregate violent acts.

**JEL Classification:** I30; C91; D81

**Keywords**

violence — public protection policy — prison violence — risk assessment — aggregate violence

[1] *Biological and Experimental Psychology, School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London, E14NS*

*__Corresponding author__: m.osman@qmul.ac.uk

Supplementary materials are available here:
docs.google.com/document/d/1Kp_uaZPGr705arhrUTHLj7Ux9v7-7fjof_Ix_uiA71k/edit?usp=sharing

## Introduction

Individual risk assessment of violence is essential to practitioners (e.g., health care professionals, clinicians) and in turn serves a useful function for a number of organisations (e.g., criminal justice agencies; Global Peace Index [GPI]; Vera Institute of Justice; World Health Organization [WHO]); in the latter case aggregating the records of violent acts from the assessments conducted can be used to quantify levels of violence in order to gauge current rates of violence, as well as predict future level of violence in a number of settings (e.g. forensic, psychiatric, public) (Fazel & Wolf, 2018; Warren, et al., 2018). From the risk assessments that are carried out, risk managers are then able to formulate critical details (e.g. care, treatment, parole, sentencing) based on aggregating the details from risk assessments performed (e.g., UK National Offender Management Service [NOMS]; U.S. FBI's Uniform Crime Reports (UCR)) to communicate rates of violence. Furthermore, aggregating individual assessments of violence can be used to determining levels of violence in a given population, to then develop policy interventions for violence prevention work (e.g. Lee, 2016).

Thus, how different violent acts (e.g. assault, rape, homicide), that vary in their severity (e.g., low, medium, high), are aggregated will depend on whether or not assessment tools differentiate violent acts by level of severity in the first place, and if they do, whether the different violent acts are weighted in some way by their level of severity. For instance, an individual could commit a combination of violent acts in one episode, or several individuals each commit a different violent act in one episode (e.g., outbreaks of violence in prison); the latter of which is an example used in this study. This raises the question, how are different violent acts aggregated? How this is answered matters with respect to determining total violence in a given location, or context (Osman et al., 2017). As a result, this has fundamental implications for behavioural economic analysis of the cost of violence (Institute for Economics & Peace, 2018), and practical implications for economic analyses determining the success of policy interventions, such as violence prevention strategies (Shiffman, 2020).

In the present study we focus on a population of non-experts sampled from the general population, and examine how they typically rank combinations of violent acts, as an index of how to aggregate different types of violent acts oc-

curring in a violent episode. We exclusively focus on violent acts that are likely to produce physical harm, which most commonly appears in violent risk assessments, though we acknowledge that psychological and sexual harm are also highly prevalent, but not common to all risk assessments. In addition, we exclusively focus on a sample of non-experts for the reason that this provides an insight into lay people's intuitive understanding of how they approach aggregating violence, which can in turn inform policy makers as to how to potentially improve communication of aggregated violence to lay audiences, from which policy decisions are made.

## Violence risk assessments

There are a wide range of tools used to assess violence (see Supplementary Materials for examples of risk assessments of violence (physical, sexual, psychological) and aggression). The range reflects the fact that they are often updated to ensure they are the most reliable and the most valid they can be (e.g. Johnson, et al., 2019; Sedgwick, et al., 2016) and because they are applied in a variety of contexts. Given the range, there are two broad factors that can be used to differentiate between different tools of assessment, the latter of which is of relevance to the main focus of this study. First, they vary as to whether or not they are actuarial (also called statistical - e.g., COMPAS, VRAG, LSI-R, OGRS – see supplementary materials for details) or used for structured professional judgement (SPJ). It is then possible to determine the predictive validity, for instance, of actuarial assessments for individual cases (e.g., Hart et al., 2007; Chapman, 2017) and relative to each other through meta-analyses (Fazel et al., 2012; Yang et al., 2010).

## Bundling violence

The second factor concerns how violence is conceptualised with respect to the way in which different violent acts are latter aggregated. As mentioned earlier, the way in which violent acts are aggregated have implications for a number of behavioural, economic and social policy matters. Aggregated violent acts to determine total violence then can be used to estimate cost of violence which can include welfare loss based on morality rate (Soares, 2006) as well as calculating direct costs and expenditures on criminal justice and crime prevention (Soares, 2015). In addressing aggregation of violent acts based on the ways in which violent acts are conceptualised in violence risk assessment, one needs to consider how combinations of violent acts ought to be treated. It is best to help illustrate this with an example. If we imagine that an outbreak of violence has occurred in two prisons, then we would expect that typically that would include a range of acts of low severity, such as pushes and shoves, to more violent acts such as punches and attacks with weapons. In fact, previous work shows that, regardless of gender, education, religiosity, and in cross country comparisons, people intuitively rank a range of violent acts such as the ones referred to here in much the same way (e.g., Osman & Pupic, 2019; Osman et al., 2017). Comparing violent outbreaks between two prisons likely consisting of combinations of violent acts of more or less severity, necessarily invites some basic assumption about how violent acts ought to be differentiated, to then be able to aggregate the acts in some systematic way to compare total violence between the two prisons. In fact, analyses of this kind feature in reports such as those published by the U.K.'s Government agency HM Prison and Probation service that present statistics on levels of violence in prisons (e.g. National Offender Management Service (NOMS) Annual Report and Accounts, 2016-2017).

Returning to our example, in our illustration of outbreaks in two prisons the following (2, 8) denotes a violence bundle of 2 low violent acts and 8 acts of high violence – outbreak 1, and a second violence bundle (7, 4) of 7 low violent acts and 4 high violence acts – outbreak 2. There are three main ways in which professionals could rank violent events. Firstly, ranking could be done according to the number of acts of the highest level of violence. This would rank outbreak 1 (N = 8) as more violent than outbreak 2 (N= 4). For instance, the WHO adopts a procedure like this, such as assessing total interpersonal violence by homicide by country, and over time. The highest violent acts are implied by the fact that they resulted in death, and are aggregated to determine changes in rates of violence over time and by country.

Returning to our illustration, another way to rank the two violent outbreaks in order to determine which of the two was more violent is by simply summing the number of violent acts per outbreak, in which case outbreak 2 (N = 11) is more violent than outbreak 1 (N = 10). As has been noted by others (e.g., Fazel & Wolf, 2018), many risk assessment tools adopt a scoring method like this, meaning that they do not differentiate the type of violent act by its severity (e.g., CARDS, CVS, Gunn Roberts scale, LHA, M55, VRS – see supplementary materials for details). In addition, many prominent agencies in turn adopt the same scoring method (e.g., Center for Disease Control and Prevention U.S. [CDC]); though it should be noted that scoring methods of this kind are used later to determine severity of violence, or severity of danger, but not by initially distinguishing violent acts by severity.

The third method that can be used to make comparisons between the violent outbreaks in our prison example is through a weighted sum ranking system. For example, to discriminate between low and high violent acts, the low violence acts could be multiplied by 1, and the high violent acts by 5, then each combination of weights is summed to produce an aggregate score of violence for later comparison. In our example outbreak 1 (N = 42) is more violent than outbreak 2 (N = 27). The weighted sum ranking system varies by risk assessment tool (e.g., Attacks, MCVS, MOAS, NVA, QOVS, VRAG – see supplementary materials for details), and by agency (e.g., GPI), because how the weights are derived is subject to expert judgment, often through the Delphi Method, which involves consultation with multiple professionals who provide either qualitative and or quantitative feedback on their characterisation of violence.

To recap, the first method (High Violence Sum, hereafter HVS) neglects low violent acts in the aggregating scoring of violence in the two different outbreaks, the second method (Simple Sum, hereafter SS) neglects the differentiation by severity of violence, and the third method (Weighted Sum Rank, hereafter WSR) takes into account severity through the assignment of weights. In answer to the question which outbreak is more violent? The first and last method would answer outbreak 1, and the second would answer outbreak 2. While it is clear that experts differ with respect to how to rank combinations of violent acts, what is not yet known is how lay people approach the issue of ranking combinations of violent acts which also indicates how they aggregate violent acts.

## Present studies

The objective of this study was to investigate the typical approach that non-experts take to aggregating combinations of violent acts, either by taking severity into account (HVS/WRS) or not (SS). To achieve this required piloting the materials we would later use in our experiments, via three separate pilot studies. This was due to the fact that the types of materials used were new, and so it was important to assess the materials by comprehensibility and reliability.

After refining our materials, our instructions, and the number of trials needed for participants to familiarise themselves with the experimental set up, we devised 4 experiments. All four experiments included two types of tasks: Forced-choice outbreak task and Missing Value outbreak task (see supplementary materials). The four experiments systematically varied the exemplars of the two tasks according to whether they were odd [Exp 2 & 4] or even values [Exp 1 & 3], and varied the instructional details to aid responding to the tasks so that they were either minimal [Exp 1 & 2] or extensive [Exp 3 & 4]. We considered this a minor manipulation to determine the extent to which the pattern of responding was robust, and therefore immune to superficial properties of the bundles (psychological research suggests that cognitive fluency can be affected by whether numbers are odd or even). The second manipulation examined whether the comprehensiveness of the task instructions would impact the way participants responded, so we varied the level of detail we presented in the task instructions (Less – Experiment 1 and 2, More – Experiment 3 and 4).

Thus, this is an exploratory study using a variety of materials that are highly innovative in order to empirically address an important question that has significant policy implications, particularly with respect to how the public conceptualise combinations of violent acts, and the extent to which that aligns with experts (incl. practitioners, risk managers, policy makers). Our purpose is to examine how lay people approach the aggregation of violent acts, so as to provide insights to policy makers that may help improve communication of social and economic policies that depend on estimates of total violence.

## Methods

Each experiment comprised four main sections (Demographic details, Provision of Instructions and Scene setting, Forced-choice outbreak task, Missing-values outbreak task). The first involved responding to a series of questions gathering basic demographic details after which participants were then presented with the general instructions, and then the two main tasks (Forced choice outbreak task, Missing values outbreak task) which were designed to assess the extent to which non-experts aggregate combinations of violent acts either by taking severity into account (HVS/WRS) or not (SS). For full details on the methods of the experiments see supplementary materials section.

## Results

Only the top line analyses are presented in this section, the remaining analyses that were conducted are presented in the supplementary materials.

### Forced choice outbreak task

**Response consistency overall:** We coded responses based simply on whether the selections corresponded with High Violence Sum/Weighted Sum [HV/WSR] or Simple Sum [SS], or were inconsistent, that is with no obvious strategy. At a gross level, we are simply looking at the total number of consistent responses [in line with HV/WSR or SS] or inconsistent responses, irrespective of experiment (Exp 1 to 4) and trial type (Trial 1 to 4). Approximately 60% of responses were consistent (N = 1083/1639), and there is a significant difference between the proportion responding consistently from inconsistently, chi-squared is $\chi^2$ (df =2; consistent, inconsistent) = 169.45, $p_{14} < .01$, Fisher's Z–$r$ = .33; here application of Fisher's Z–$r$ is as a test of significant of the difference between the correlation coefficients entered into the chi-squared analysis.

Looking at the consistency of responses across tasks, 60% of participants either 75% or 100% of the time selected HVS/WSR options, and 16%[1] of participants selected the SS option either 75% or 100% of the time. This suggests that the majority of participants were consistently selecting responses that aggregate combinations of violent acts by taking into account severity, but there is a non-negligible proportion of participants that consistently respond in line with aggregating combinations of violent acts regardless of severity.

### Missing value task

**Response consistency overall:** We coded responses based simply on whether the values that were entered did in fact correspond with the instructions. At a gross level, simply looking at the total number of consistent responses, by which we mean consistent with the instructions presented on each trial (e.g.,

[1] 40% of participants consistently selected the HVS/WSR response across all 4 forced choice tasks, and 8% consistently selected the SS option for all 4 forced choice tasks.

generate a value that makes the comparator bundle more/less violent than the target bundle) and inconsistent responses, irrespective of experiment (Exp 1 to 4) and trial type (Trial 1 to 4), approximately 50% of responses were consistent (N responses = 779/1548), and a chi-squared analysis, confirms that there is no significant difference between the proportion responding consistent with instructions from inconsistent with instructions, chi-squared is $\chi^2$ (df =2; consistent, inconsistent) = .07, $p > .79$.

However, looking more closely at the pattern of responses people gave, we re-classified responses according to whether participants gave values in line with the instructions (i.e. consistent), and then classified the inconsistent responding via three sub-categories. First, those that gave values that were clearly in the opposite direction of what was instructed (e.g. providing values that increased the overall violent bundle, rather than decreasing it as instructed) which is explicitly incorrect. Second, those that simply entered the value that matched that of the exemplar in the trial (e.g., enter a value for high violent acts X, where the value entered was 4, and the example of the high violent act in the target bundle was 4), which we term "matching". Third, those that generated values that ignored the instructions so that they made the comparator bundle sum to the same total number of violent acts as the target bundle (e.g. enter a value for high violent acts X, where the value entered was 6 and the low violent act is 6 [total summing to 12], and the example was high violent acts 8, low violent acts 4), which were in line with SS approach.

In order to determine consistent patterns of responses, we classified those as consistent if they responded with either a consistent [HVS/WSR], Matching, or SS strategy 75% of the time as well as 100% of the time. 7%[2] of participants generated the "matching" value consistently, 2% of participants generated SS responses, and 35% of participants generated HVS/WSR responses 75% or 100% of the time, and 2% consistently generated values that were explicitly incorrect[3].

## Implications and conclusions

There are a variety of ways of conceptualising violence in risk assessments, which has implications for how violent acts are aggregated. In the present study we identify three broad categories (High Violence Sum [HVS], Weighted Sum Ranking [WSR], Simple Sum [SS]). Two of these take into account the severity of the violent acts (e.g., HVS, WSR) and one does not (SS).

The present study is the first of its kind to examine which of the two approaches taken in violent risk assessments is most frequently expressed in a non-expert audience. We investigated this in two tasks set in prison contexts, assessing the way lay people make relative comparisons between different combinations of violent acts. In four experiments we found that people most commonly take into account the severity of violent acts. What this finding suggests is that, of the many violence risk assessments that exist, the conceptualisation of violence most compatible with a lay audience is one which discriminates violent acts by their severity. Moreover, how that information is presented to a lay audience, especially if they are required to think about making relative comparisons, can alter how they make those comparisons as well (e.g., how they might compare combinations of low and high violent acts).

Why might these findings be of importance? Many organizations (e.g., Global Peace Index [GPI]; U.S. Justice Dept.; U.K. Ministry of Justice; World Health Organization [WHO]; Vera Institute of Justice) generate reports that report on aggregate (total) levels of violence. How violence is conceptualised has direct implications for the estimates of total violence, and the economic analyses conducted from them (e.g. calculating welfare loss), and the policies designed around (e.g. success of violence prevention). These details are communicated to lay audiences, to inform them about, for instance, the risk of violent exposure in their neighbourhoods, and the policies that are designed to reduce exposure. How these details are interpreted against personal beliefs and assumptions about violence and levels of violence may differ from how they are intended by policy makers. For example, one might imagine a case where parents might consider moving from where they live to another location to raise their children and want to determine how safe it is. Crime statistics might report the rates of different types of violent acts (e.g., stabbings [e.g., knife crime], punches, kicks [e.g., violent assaults]). How this data is aggregated to communicate how violent crime has increased or decreased, or varies by different locations may or may not take into account the severity of violent acts (e.g., Bureau of Crime statistics – U.S. Justice Dept.). Our findings show that this appears to be at odds with the way in which lay people aggregate combinations (bundles) of violent acts in order to make comparisons. When they do these comparisons our findings reveal that they take severity of violent acts into account more often than not. Therefore, it is important to recognise that there might well be a misalignment in the way people intuitively make comparisons between bundles of violent acts, and how experts communicate their findings on which lay people utilise this information.

[2] 21% of participant consistently generated the HVS/WSR across all 4 forced choice tasks, and 4% consistently generating the SS value for all 4 missing value tasks.

[3] These were respondents that generated values that were inconsistent with any of the three strategies being examined.

## Acknowledgments

We would like to thank Nick Baigent for his significant contributions throughout this project, and being the inspiration behind the core ideas that set the foundations for what we have tried to realise in experimental form.

## References

Almvik R, Woods P, & Rasmussen K. (2000). The Brøset Violence Checklist, sensitivity, specificity and inter-

rater reliability. *Journal of Interpersonal Violence, 15*: 1284–96.

Chapman, M. (2017). A review of violence risk assessment for the general clinician. *Psychiatric Annals, 47*(9), 449–453.

Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24 827 people: systematic review and meta-analysis. *Bmj, 345*, e4692.

Fazel, S., & Wolf, A. (2018). Selecting a risk assessment tool to use in practice: a 10-point guide. *Evidence-based mental health, 21*(2), 41-43.

Grann, M., Belfrage, H., & Tengström, A. (2000). Actuarial assessment of risk for violence: Predictive validity of the VRAG and the historical part of the HCR-20. *Criminal Justice and Behavior, 27*(1), 97-114.

Hart, S. D., Michie, C., & Cooke, D. J. (2007). Precision of actuarial risk assessment instruments: Evaluating the 'margins of error' of group v. individual predictions of violence. *The British Journal of Psychiatry, 190*(S49), s60–s65.

Institute for Economics & Peace (2018). The Economic Value of Peace 2018: Measuring the Global Economic Impact of Violence and Conflict, Sydney, October 2018. Available from: http://visionofhumanity.org/ reports (accessed, 2020).

Johnson, K. L., Desmarais, S. L., Tueller, S. J., & Van Dorn, R. A. (2019). Methodological limitations in the measurement and statistical modeling of violence among adults with mental illness. *International journal of methods in psychiatric research*, e1776.

Lee, B. X. (2016). Causes and cures VI: The political science and economics of violence. *Aggression and Violent Behavior, 28*, 103–108.

Osman, M., Pupic, D., & Baigent, N. (2017). How many slaps is equivalent to one punch? New approaches to assessing the relative severity of violent acts. *Psychology of violence, 7*(1), 69–81.

Osman, M., & Pupic, D. (2019). Judging Levels of Violence: Factors That Influence Assessment of Severity of Violent Acts and Their Consequences. *Psychology and Cognition, 4*, 1–11.

Sedgwick, O., Young, S., Das, M., & Kumari, V. (2016). Objective predictors of outcome in forensic mental health services—a systematic review. *CNS spectrums, 21*(6), 430–444.

Shiffman, G. M. (2020). *The Economics of Violence*. Cambridge Books.

Soares, R. R. (2006). The welfare cost of violence across countries. *Journal of health economics, 25*(5), 821-846.

Soares, R. R. (2015). Welfare costs of crime and common violence. *Journal of Economic Studies, 42*(1), 117-137.

Warren, J. I., Wellbeloved-Stone, J. M., Dietz, P. E., & Millspaugh, S. B. (2018). Gender and violence risk assessment in prisons. *Psychological services, 15*(4), 543-552.

Yang, M., Wong, S. C., & Coid, J. (2010). The efficacy of violence prediction: a meta-analytic comparison of nine risk assessment tools. *Psychological bulletin, 136*(5), 740-767.