March 2016

Evaluating Transitional Justice: The Role of Multi-Level Mixed Methods Datasets and the Colombia Reparation Program for War Victims

Phuong Ngoc Pham
Harvard Medical School and Harvard T.H. Chan School of Public Health, ppham@hsph.harvard.edu

Patrick Vinck
Harvard Medical School and Harvard T.H. Chan School of Public Health, pvinck@hsph.harvard.edu

Bridget Marchesi
University of Minnesota, Bridget_Marchesi@hks.harvard.edu

Doug Johnson
Harvard University, douglas_johnson@hks.harvard.edu

Peter J. Dixon
University of California - Berkeley, pdixon@berkeley.edu

See next page for additional authors

Follow this and additional works at: https://ir.lib.uwo.ca/tjreview
http://dx.doi.org/10.5206/tjr.2016.1.4.3

Evaluating Transitional Justice: The Role of Multi-Level Mixed Methods Datasets and the Colombia Reparation Program for War Victims

Authors

Phuong Ngoc Pham, Patrick Vinck, Bridget Marchesi, Doug Johnson, Peter J. Dixon, and Kathryn Sikkink

Evaluating Transitional Justice: The Role of Multi-Level Mixed Methods Datasets and the Colombia Reparation Program for War Victims

Phuong Ngoc Pham, Harvard Medical School
and Harvard T.H. Chan School of Public Health
ppham@hsph.harvard.edu

Patrick Vinck, Harvard Medical School
and Harvard T.H. Chan School of Public Health
pvinck@hsph.harvard.edu

Bridget Marchesi, University of Minnesota
Bridget_Marchesi@hks.harvard.edu

Doug Johnson, Harvard University
douglas_johnson@hks.harvard.edu

Peter Dixon, University of California, Berkeley
pdixon@berkeley.edu

Kathryn Sikkink, Harvard University
kathryn_sikkink@hks.harvard.edu

P. Pham, P. Vinck, B. Marchesi, D. Johnson, P. Dixon, K. Sikkink 61

Abstract

This paper examines the role of mixed and multi-level methods datasets used to inform evaluations of transitional justice mechanisms. The Colombia reparation program for victims of war is used to illustrate how a convergent design involving multiple datasets can be used to evaluate the effectiveness of a complex transitional justice mechanism. This was achieved through a unique combination of (1) macro-level analysis enabled by a global dataset of transitional justice mechanisms, in this case the reparations data gathered by the Transitional Justice Research Collaborative, (2) meso-level data gathered at the organizational level on the Unidad para las Victimas (Victims Unit), the organization in charge of implementing the reparations program and overseeing the domestic database of victims registered in the reparations program, and (3) micro-level population-based perception datasets on the Colombian reparations program collected in the Peacebuilding Data database. The methods used to define measures, access existing data, and assemble new datasets are discussed, as are some of the challenges faced by the inter-disciplinary team. The results illustrate how the use of global, domestic, and micro-level datasets together yields high quality data, with multiple perspectives permitting the use of innovative evaluation methods and the development of important findings and recommendations for transitional justice mechanisms.

Introduction

The last decade has seen an intense debate among scholars and practitioners about the impact of transitional justice (TJ) mechanisms on societies making the transition from authoritarianism to democracy and/or conflict to peace. In these debates, transitional justice is often portrayed as sparse, ineffective, counterproductive, insensitive to local customs and sensibilities, or problematic in other ways.[1] Other

---

[1] For only a very partial list of this literature, see, for example, Christine Bell, "Transitional Justice, Interdisciplinarity and the State of the Field or Non-Field," *The*

62 Evaluating Transitional Justice

scholars have found positive effects accruing from TJ mechanisms, alone or in combination with others.[2] The debate about the impact of transitional justice has been intense, in part because scholars work within different academic traditions and draw on different methods of research and levels of analysis, leading them to arrive at different conclusions. In particular, there are differences between scholars and practitioners who rely on macro-level databases, qualitative case studies, population-based surveys, and cross-national studies. All these approaches are empirical and evidence-based, but each method of inquiry has its own inherent strengths and limitations with regard to drawing causal conclusions.

Empirical research using population-based surveys, for example, is often focused on providing "assessments of communities' needs and perceptions of and attitudes toward peace and justice, as well as systematic and rigorous measurement of the potential and actual impacts of TJ mechanisms."[3] 'Impact' in these studies concerns primarily the impact on victims and, more broadly, the population (e.g., measuring changes in perception of government, justice, and social

---

International Journal of Transitional Justice, 3.1 (2009): 5-27; Wendy Lambourne, "Transitional Justice and Peacebuilding after Mass Violence," *The International Journal of Transitional Justice* 3.1 (2009): 28-48, 30; Laurel E. Fletcher, Harvey M. Weinstein, with Jamie Rowen, "Context, Timing, and the Dynamics of Transitional Justice: A Historical Perspective," *Human Rights Quarterly* 31.1 (2008): 163-220; Jelena Subotic, *Hijacked Justice: Dealing with the Past in the Balkans* (Ithaca: Cornell University Press, 2009); Jack Goldsmith and Stephen Krasner, "The Limits of Idealism," *Daedalus* 132.1 (2003): 47-63; and Adam Branch, "Uganda's Civil War and the Politics of ICC Intervention," *Ethics & International Affairs* 21.2 (2007): 179-198, 189, 192.

[2] See, for example, Tricia D. Olsen, Leigh A. Payne, and Andrew G. Reiter. "The Justice Balance: When Transitional Justice Improves Human Rights and Democracy," *Human Rights Quarterly* 32.4 (2010): 980-1007. See also Hunjoon Kim and Kathryn Sikkink, "Explaining the Deterrence Effect of Human Rights Prosecutions for Transitional Countries," *International Studies Quarterly* 54.4 (2010): 939-963.

[3] Phuong N. Pham and Patrick Vinck, "Empirical Research and the Development and Assessment of Transitional Justice Mechanisms," *The International Journal of Transitional Justice* 1.2 (2007): 231-248.

Transitional Justice Review, Vol.1, Iss.4, 2016, 60-94

P. Pham, P. Vinck, B. Marchesi, D. Johnson, P. Dixon, K. Sikkink 63

cohesion). Cross-national research, on the other hand, has been more concerned with the macro-effects of TJ mechanisms at the national level, for example, on human rights practices, democratization, and conflict, or at the international level.[4] But impact in these studies tends to ignore local effects. Cross-national research and where possible longitudinal survey research can also explore long-term effects, such as the impact of TJ mechanisms over a decade or more, as people's needs and priorities change, institutions develop, and local understanding of international norms shifts. Cross-sectional and longitudinal evaluations could foreseeably yield results that conflict with one another: a TJ mechanism may be found to respond to the needs of communities and be well-received by them, and yet not appear to have effects at the macro-national level, such as building peace or improving human rights practices. There is an additional, "meso-level" of research, which looks at the functioning of organizations within the state, for example an institutional assessment of truth commissions, courts, or reparations units. Such institutional assessments of TJ mechanisms are rare, however, as is research that incorporates two or more of these macro-, micro- or meso-approaches.

Geoff Dancy makes a similar point, contrasting what he calls two general groups of perspectives on transitional justice, one consisting of those who highlight "top-down" or state-level goals, such as the consolidation of democracy or conflict resolution, and another which employs a "bottom-up" approach, including the individual's perception of justice. In particular, Dancy concludes that while one key purpose of macro-level impact evaluation is to explore causal claims and disputes, TJ evaluations "can and should produce knowledge that

---

[4] See, for example, Kathryn Sikkink, *The Justice Cascade: How Human Rights Prosecutions Are Changing World Politics* (New York: W.W. Norton and Co., 2011), which explores how human rights prosecutions are changing world politics.

Transitional Justice Review, Vol.1, Iss.4, 2016, 60-94

64 Evaluating Transitional Justice

is not exclusively causal in nature."[5] Indeed, the same tension exists in survey research between methods concerned with measuring impact and those concerned with relational and interpretive questions. One of our claims below is that the use of mixed-method and multi-level research allows us to simultaneously pursue causal, relational, and interpretive knowledge about transitional justice.

The application of bottom-up and top-down research on human rights prosecutions shows that each approach yields findings that are both contradictory and complementary. Some cross-national research on human rights prosecutions has found them to be associated with improvements in core human rights practices, either alone or in combination with other TJ mechanisms.[6] But survey research with victims, as well as ethnographic research in transitional contexts, has been much more ambiguous about the positive effects of prosecutions and other TJ mechanisms.[7] For example, population-based research by Pham and Vinck shows that respondents typically expect courts to bring about peace, arrest perpetrators and provide reparations, and thus that "unmet expectations and disenchantment will ultimately undermine the legitimacy of the courts in the eyes of the survivor."[8] The findings of both methods of inquiry reveal multiple truths that need to be taken into account when assessing the impact of

[5] Geoff Dancy, "Impact Assessment, Not Evaluation: Defining a Limited Role for Positivism in the Study of Transitional Justice," *The International Journal of Transitional Justice* 4.3 (2010): 355-376.

[6] See, for example, Kim and Sikkink 2010; and Olsen, Payne, and Reiter 2010.

[7] For more critical accounts of transitional justice from an ethnographic perspective, see, for example, Kimberly Theidon, *Intimate Enemies: Violence and Reconciliation in Peru* (Philadelphia: University of Pennsylvania Press, 2013); Rosalind Shaw, Lars Waldorf, and Pierre Hazan, eds. *Localizing Transitional Justice: Interventions and Priorities after Mass Violence* (Stanford: Stanford University Press, 2010); Rebekah Park, *The Reappeared: Argentine Former Political Prisoners* (New Brunswick: Rutgers University Press, 2014).

[8] Patrick Vinck and Phuong N. Pham, "Consulting Survivors: Evidence from Cambodia, Northern Uganda, and Other Countries Affected by Mass Violence," in *The Human Rights Paradox: Universality and Its Discontents*, eds. Steve J. Stern and Scott Straus, 107-124 (Madison: University of Wisconsin Press, 2014).

P. Pham, P. Vinck, B. Marchesi, D. Johnson, P. Dixon, K. Sikkink 65

transitional justice. TJ mechanisms are simultaneously legal, political, social, economic, cultural, and technical.

At the core of our approach, we believe that different research methods are useful depending on the context and type of research questions. Moreover, different research methods should be seen as complementary rather than competitive. To illustrate the usefulness of this complementary approach, but also some of its challenges, this paper examines the role of mixed research methods and multi-level datasets to inform the evaluation of one specific TJ mechanism: the reparations program for victims of the Colombian civil conflict. The evaluation illustrates how a convergent design using multiple datasets was used to evaluate the effectiveness of a complex TJ mechanism. This was achieved through a unique combination of (1) macro-level analysis enabled by global datasets of TJ mechanisms, in this case the reparations data gathered by the Transitional Justice Research Collaborative; (2) meso-level data gathered at the organizational level by Unit for Comprehensive Attention and Reparation of Victims, or Victims Unit (Unidad para la Atención y Reparación Integral a las Víctimas, VU) the organization in charge of implementing the reparations program and overseeing the domestic database of victims registered in the reparations program; and (3) micro-level datasets on population perceptions of the Colombian reparations program, collected using the mixed-methodology of the PeacebuildingData database. These three components were conducted in parallel over a 10-month period in 2014-2015. Together these methods provided insight into the legal, political, social, economic, cultural, and technical aspects of reparations. The research, however, did not integrate all possible methods that can yield rich insight into human behaviors. Notably, our research did not include extensive ethnographic research, limiting some the ability to capture nuances in understandings of local expectations and experiences. At the same time, we acknowledge that this kind of multi-level mixed-methods research (MMR) design is resource intensive and might be difficult for individual researchers to replicate. However, we also find evidence that research collaborations

66 Evaluating Transitional Justice

are a more fruitful way to implement multi-method research. The Harvard evaluation shows that some of the challenges of MMR, especially when implemented by an individual researcher,[9] can be overcome through the collaborative team model.

The methods used to define measures, access existing data, and assemble new datasets are discussed, as well as the benefits and challenges of conducting such "phased" research. The results illustrate how the use of global, regional, and local datasets together yields high quality data, with multiple perspectives permitting the development of important evaluation findings and recommendations for TJ mechanisms.

Background of the Reparations Program in Colombia

The Colombian government established an ambitious reparations program in 2011. The reparations Law 1448 provides for "a set of judicial, administrative, social, and economic measures, both individual and collective, to benefit the victims who individually or collectively have suffered harm for events that occurred as of January 1, 1985, as a result of violations of international humanitarian law or of grave and manifest violations of international human rights law provisions, which occurred on occasion of the internal armed conflict within a framework of transitional justice, that make possible the effective enjoyment of their rights to truth, justice, and reparation with guarantees of non-repetition." The law, and subsequent presidential decrees and rulings by the Constitutional Court, called for comprehensive reparations measures including compensation, rehabilitation, restitution, satisfaction, and guarantees of non-repetition. It took an expansive view of its temporal coverage,

---

[9] For a discussion on the shortcomings of MMR when implemented by individual researchers and the benefits of MMR when implemented by multiple researchers working in collaboration see, Amel Ahmed and Rudra Sil, "When Multi-Method Research Subverts Methodological Pluralism—or, Why We Still Need Single-Method Research," *Perspectives on Politics* 10:4 (2012): 935-953.

P. Pham, P. Vinck, B. Marchesi, D. Johnson, P. Dixon, K. Sikkink 67

coordination of benefits, eligibility criteria for victims, and forms of reparation. It defined a broad set of beneficiaries including all possible injury types (e.g., physical, emotional, economic, and fundamental rights), a requirement of finding and registering those victims, and a set of benefits to which they were entitled, not just to repair tangible harms but to restore victims to full citizenship.

To coordinate and implement this comprehensive reparation program, the law created the VU to formulate, implement, and coordinate public policies at national and regional levels with respect to victims' reparations. Part of its mandate is to coordinate the National System for Attention and Comprehensive Reparation of Victims (Sistema Nacional de Atención y Reparación Integral a las Víctimas, SNARIV), an intra-governmental coalition of 39 national ministries, agencies, and units, and 13 governmental organizations charged with the formulation and implementation of plans, programs, and actions to promote the reparation of victims. The VU further provides advice and counsel to the Government on how reparations might best serve the interests of victims and protect them from exploitation. Furthermore, the VU endeavors to establish individualized "reparation plans" for each person registered as a victim of the armed conflict, known as a Plan for Comprehensive Attention, Assistance and Reparation (Plan de Atención, Asistencia y Reparación Integral a las Víctimas, PAARI). The PAARI results from a consultation with a VU staff person on the individual needs to be matched with the available services in their area that would help them move from a state of vulnerability to independence and self-reliance.

A multi-level, mixed-method evaluation was implemented between late 2014 and mid-2015 to evaluate the efforts of the VU to implement the comprehensive reparations measures called for in Law 1448: compensation, rehabilitation, restitution, satisfaction, and guarantees of non-repetition. This evaluation was requested by the VU and supported by USAID through a subcontract from Management

68 Evaluating Transitional Justice

Systems International (MSI) with our Harvard-based team.[10] The evaluation included three levels of analysis: (1) a macro-global benchmarking study comparing the Colombian program to other reparations programs around the world (known as Component 1); (2) a meso-institutional analysis of the VU's reparation and coordination functions (Component 2); and (3) a micro-examination of the implementation of reparation measures by the VU from the perspective of its beneficiaries, and more broadly, victims of the armed conflict (Component 3). The evaluation focused on the reparations function of the VU, with the goal of better understanding its implementation and highlighting opportunities for improvement.[11]

The use of this three-level research approach was designed to comprehensively assess the work of the VU in order to identify strengths, gaps, lessons learned and timely recommendations for the VU's continued work. Each of the components had an important place in the study in its own right. The international comparative study in Component 1 involved a macro-level analysis that examined the design of the Colombian reparations model in light of other historical cases worldwide and drew lessons from this comparison to strengthen the Colombian model for the future. Component 2's institutional assessment provided a meso-level analysis that diagnosed the VU's institutional capacity for (a) direct provision of individual and collective reparation measures, in accordance with its normative mandate and historical context, and for (b) inter-institutional coordination for reparations delivered to victims through the

---

[10] The evaluation's funding was under the project title "Evaluation and Analysis for Learning" to the Carr Center for Human Rights Policy at the Harvard Kennedy School and the Harvard Humanitarian Initiative at the Harvard T.H. Chan School of Public Health.

[11] Because the provision of reparations has not been randomized and because no baseline evaluation was conducted, this is not an impact evaluation of the VU's reparations function. We do, however, seek here to draw conclusions about associations between the VU's work and the observable variables captured in our quantitative and qualitative data.

SNARIV. The third level of analysis (Component 3) examined the implementation of the reparations measures to date from the perspectives of (a) the general Colombian population, (b) Colombians who have registered as victims of the armed conflict but have not yet received reparations, and (c) registered victims who have received reparations payments. For Component 3, the team used both qualitative and quantitative methods of data collection. This research was designed to utilize a phased approach where each component would be implemented over the course of overlapping research phases, but we adapted this to more of a parallel approach in response to realities encountered in the field.

Research Methods for Each Perspective

Component 1: Macro-Level Comparative Analysis
For the first phase of the Colombian Victims Unit evaluation, the macro-level perspective (Component 1) was undertaken through a process of "benchmarking," which compared Colombia's laws, institutions, and the results to date of the reparations program to other reparations programs around the world. For Component 1 we conducted a broad comparison of the Colombian reparations program with 45 other reparations policies in 31 other transitional countries in the world. We then undertook a more in-depth comparison of the Colombian program with policies in a reference group of reparations policies in five other comparable countries: Guatemala, Indonesia, Peru, South Africa, and Morocco. While no country matches Colombia in terms of the duration of its conflict, size of its victim population, and size of its reparations program, we narrowed down this list based on several key criteria: (1) transition type, (2) conflict attributes, and (3) reparation policy attributes. The final list of comparative cases was negotiated with the VU. The VU requested that we select cases that would provide "lessons learned" in terms of policy design, institutional features, and implementation. While this phase took approximately two months, we drew heavily on data that members of our team had already gathered, as we describe below.

70 Evaluating Transitional Justice

Component 1 data was generated from three sources. The first source was the reparations data from the Transitional Justice Research Collaborative (TJRC). The TJRC is an international collaboration to produce a comprehensive database on TJ policies around the world, and to produce analysis of these policies using data from the database.[12] Since 2010, the TJRC has collected data on a variety of legal and quasi-legal TJ processes designed to reckon with past human rights violations following periods of political turmoil, state repression or armed conflict. The TJRC collects data on international, foreign and domestic criminal prosecutions; amnesty policies; truth commissions; reparation policies; vetting policies; civil trials for damages and customary forms of justice. The data cover all regions of the world from 1970 to 2013.

The first criterion for inclusion in the TJRC database is that the reparation policy is meant to address past human rights violations, defined as physical integrity harms limited to political and other extrajudicial killings, arbitrary or unlawful deprivation of life, disappearances, torture, and political imprisonment. The second criterion for inclusion is that a policy has to have broad coverage and coordinated benefits. Policies meant to address one-time events such as a single massacre, or civil trials for granting damages to specific individuals are not included in the reparations database. The database is focused on administrative reparations, where state policies seek to compensate, through a variety of measures, for specific types of violations and for various classes of victims, but without formal judicial proceedings to determine responsibility. The third criterion for inclusion is that the reparations policy must be official and domestic.

---

[12] TJRC operates from the Harvard University Kennedy School, Oxford University, and the University of Minnesota, and involves scholars from leading research universities in the United States and United Kingdom. Its research is supported by the National Science Foundation (Grant No. 0961226) and the Arts and Humanities Research Council (Grant No. 0AH/I500030/1) relating to the project titled "The impact of transitional justice on human rights and democracy," and "Alternative accountabilities for human rights violations." For an overview of the dataset, see www.transitionaljusticedata.com.

P. Pham, P. Vinck, B. Marchesi, D. Johnson, P. Dixon, K. Sikkink 71

Official policies originate through official state channels such as state-sponsored truth commissions, legislation or laws, and executive orders or decrees. The database excludes cases where policies are recommended by official state bodies but never actualized through legislation, decree or other policy instrument. Domestic reparations policies are state policies that aim to compensate victims for harms perpetrated within its territorial borders. Thus, the database also excludes international reparations polices that originate in international organizations such as the United Nations (e.g., Iraq UN resolution 687 of 1991) and the International Criminal Court (e.g., Lubanga judgment of 2012).

Notably, Colombia was at first not included in the TJRC database, [13] which was originally designed to capture only those countries that have undergone transitions from authoritarianism—Colombia has remained democratic despite its long history of internal armed conflict. [14] This speaks to the challenges of narrowing the definition of TJ mechanisms while recognizing that the term has come to refer both to political transitions as well as the transition from war. [15] Since that time, however, the TJRC database has been expanding to include all transitional countries both from authoritarianism to democracy and from war to peace.

In addition to the resources of the TJRC database, the second source for Component 1 was a review of primary documents such as laws, decrees and other official policy documents. A third set of sources included secondary documents such as academic books and

---

[13] The TJRC database was expanded in 2014-2015 to include all countries of the world from 1970-2014. The database uses a filtering procedure which allows users to filter cases by regime type, civil war status, and transitional justice mechanism.

[14] Although Colombia differs from most transitional cases due to its level of democratic development and other state-level attributes, Colombia has much in common with other states transitioning from civil war to peace.

[15] Ruti Teitel refers to this conceptual expansion as the "normalization" of transitional justice. See Ruti G. Teitel, *Globalizing Transitional Justice: Contemporary Essays* (New York: Oxford University Press, 2014).

72 Evaluating Transitional Justice

articles, reports and evaluations published by international and domestic non-governmental organizations, and evaluations published by international organizations.

Strengths and Weaknesses of the Macro Method

The strength of this research method was that it allowed the evaluation to appreciate unique features of the Colombian program that might not have been apparent without such a global comparison. The evaluation found, for example, that the number of victims the Colombian reparations program aims to serve is far broader and larger than any other reparations program, in both absolute terms and relative to population size. The VU uses a larger list of victimizing acts than any other country in the TJRC database. No other case in the database comes anywhere close to the number of victims already registered by the Colombian registry. The Colombian registry now includes more than 15% of the current population of Colombia; none of the other programs have registered or compensated more than 1% of their populations. In sum, Component 1 showed that there really is no genuine reference group for the reparations program in Colombia in terms of the absolute size of the number of registered victims. It also showed that the differences between the Colombian program and the other reparations programs are largely the result of the decision in Colombia to include displaced people in the reparations program and the huge size of the displaced population in the country. If displaced people were not included in the Colombian program, the number of the registered victims would be approximately 2% of the population, still twice the size of other large reparations programs, but somewhat more in line with the other large and complete programs in the database. Understanding these unique and challenging aspects of the Colombian program helped the evaluators to appreciate the dilemmas faced by the VU as it scaled up for an unprecedented task.

The weakness of this method was that it tended to focus on the formal and legal characteristics of the Colombian program, since this was easier to compare with the data on other reparations programs

P. Pham, P. Vinck, B. Marchesi, D. Johnson, P. Dixon, K. Sikkink 73

in the TJRC database. The data in the TJRC database are comprehensive but also relatively thin. This comparison led us to a generally favorable evaluation because the Colombian program in law and in its conception was so ambitious and comprehensive. However, this research method yielded relatively little understanding about how the reparations programs were actually being implemented, either in Colombia or in the other 44 programs in the database. To help compensate for these shortcomings, we later selected a reference group of five of the most comparable reparations programs in the database for a more in-depth comparison. Based in part on the request of the Colombian VU, the comparison group included other large-scale programs in post-conflict countries, based on the criteria noted above: Guatemala, Indonesia, Peru, South Africa, and Morocco. Furthermore, Guatemala and Peru were selected based on similarities with Colombia, especially regional comparability. Indonesia and Morocco were selected to show different collective reparations models and enable a comparison between Colombia and states outside of the Latin American region. South Africa was chosen for its prominence in TJ and some similarities in level of democratic development before and during transition. The cases were also chosen due to high data availability. For this part of Component 1, we moved beyond the data in the TJRC database and conducted additional qualitative research on the five countries. This part of Component 1 was valuable in that it involved much more detailed and in-depth research using both primary and secondary written sources. We were limited, however, by the existing sources available to us; those for some countries were better than those for other countries. Nor did we have access to any interviews or survey data from the other five reference group countries.

Furthermore, a weakness of the macro method is in the criteria according to which variables and comparators are included or excluded from databases. As noted, Colombia has maintained a relatively strong democratic history compared to most other countries where TJ mechanisms have been used and for that reason was not originally

74 Evaluating Transitional Justice

included as a "TJ country." More generally, a weakness of any large-scale TJ database is that it must necessarily simplify the coding in categories of mechanisms that can vary widely.

Nevertheless, the comparison with the reference group also clarified some of the successes and challenges of the Colombian program. For example, we learned the VU had been very efficient in delivering compensation to victims compared to other cases in our database. None of the other reference group countries have compensated so many individuals. Only Indonesia approaches the number of people receiving reparations in Colombia, but it has done so only through collective reparations that have both direct and indirect individual beneficiaries. Not only has the VU compensated more victims, but it has done so in a relatively short period of time. At the same time when we compared the number of individuals already compensated to the total size of the pool of victims, Colombia still faces a huge task to provide reparations for the individuals in its registry.

Ultimately, while the macro-level comparison confirmed the remarkable ambition of the Colombian program, the meso- and micro-level analyses were necessary to clarify what such ambition has meant in practice—both in terms of the capacity of the VU to live up to its mandates as well as the consequences for the program's reception by the public and victimized populations.

Component 2: Meso-Level Perspective: Institutional Analyses
Few TJ studies cross multiple levels of analysis. Even when they do, the meso-level is still often ignored. Yet the meso-level is home to the very institutions and interactions that we are interested in studying, in this particular case, the VU and the state institutional structure within which it functioned. For our purposes, understanding the institutional arrangements, normative systems, and interactions among various actors and institutions at this level of analysis was necessary to theorize about current and future policy implementation.

P. Pham, P. Vinck, B. Marchesi, D. Johnson, P. Dixon, K. Sikkink 75

Over the course of five months, Component 2 examined the VU's performance with regard to its two main institutional mandates: 1) the delivery of a set of reparation measures, specifically compensation of individuals and community level reparations, while also providing guidance and support to the reparation projects of other Federal agencies; and 2) the coordination of SNARIV, the interagency task force of 39 governmental units and entities that have a role in the provision of reparation measures, as set out by national legislation and presidential decrees.

For this evaluation, one reason that the meso-level was so difficult to analyze is that the VU is embedded in a complicated social system comprising victims, political entities and processes, and citizens. The preferences of these constituencies are often in conflict, putting the VU in the challenging and undesirable position of having to expend significant resources on seeking acceptable compromises.

We can think of this complicated social system in terms of both vertical and horizontal interactions that have to be analyzed and theorized. The vertical landscape stretches from individual victims interacting with local or regional officials—such as when individual victims make a declaration at the municipal level to begin the process of registration—up to the VU, which in turn, operates in a vertical position when interacting with the Executive. The vertical landscape also includes the VU's interacting with Colombia's departments and municipalities, as well as the donor community. The horizontal landscape primarily comprises the 39 government entities and 13 NGO allied organizations that make up the SNARIV. It was initially a challenge just to map the vertical and horizontal landscapes and understand how different entities in the network interact. Once mapped, the analyst then needs to understand how vertical or horizontal institutional arrangements constrain or facilitate the VU's implementation of the reparations policy.

We were able to take on this challenging meso-evaluation because of the mandate of the evaluation and the structure of the team largely made of Colombians with deep institutional knowledge of the

76 Evaluating Transitional Justice

Colombian government and of the political and legal environments, and many contacts in the central government, human rights community, and international donor community.[16] Although they have a variety of expertise, such as strategic management or government budgeting, all are trained in evaluation methodology and generally follow the guidelines and standards typical of, for example, the Development Assistance Committee (DAC) at the Organization for Economic Co-operation and Development (OECD).

The composition of the C2 team strengthened the evaluation for several reasons. Across components, the Harvard team was seen as independent and autonomous. This increased the credibility of the analysis and recommendations. The VU and the Government's broader transitional efforts are affected by wide-scale mistrust of the general population toward the State. Furthermore, independence was especially important for C2 because we analyzed and eventually made recommendations about politically sensitive subjects that dealt with political and economic resources. Whereas Harvard brought greater independence, the MSI consulting team brought a closeness to the subject matter that enhanced the validity and usefulness of the analysis and recommendations. This was achieved through a variety of mechanisms. The MSI consulting team allowed us to regularly involve key stakeholders in our analysis including VU leadership, VU management and various stakeholders in the territories. Due to their proximity to the VU headquarters in Bogota and their strong relationships with VU leadership, the MSI consulting team was able to observe and sometimes participate in VU operations and meetings as well as have quick access to important primary documents. The importance of having access to a diverse set of primary documents, such as human resources information on organizational arrangements and staffing, national planning documents, internal budgets and

---

[16] The consultants, hired by MSI, included in particular Francisco Osuna and Ana María Rivera, who led the Colombia-based part of the team for the Component 2 organizational diagnosis.

P. Pham, P. Vinck, B. Marchesi, D. Johnson, P. Dixon, K. Sikkink 77

forecasts, programmatic information, and registry data cannot be overstressed. Without these data, our analysis would have been incomplete and most likely erroneous. Finally, the MSI consulting team also served as *interlocuteur,* helping the Harvard team contextualize data and develop more useful insights and recommendations. In some cases this kind of intermediary might not be necessary, but in current-day Colombia, where the political and legal landscape is ever changing due in part to the ongoing peace negotiations, we benefited from weekly briefings. At the same time, since the Harvard team had more distance, we were able to ask certain questions and see particular issues that were not always obvious to researchers on site.

As evaluation specialists, the MSI team worked mainly from the perspective of evaluation research. Evaluation research is a form of social science research, but it differs from research where the main goal is causal inference and theory-building. The main goal of evaluation research is the systematic gathering, processing, and assessment of information to provide useful feedback about, in this case, a policy. Evaluation utilizes many of the same methodologies as other more traditional social science research. However, a key difference is that an evaluation takes place within a political and organizational context and it often requires group skills. The diverse group skills, abilities, management experience and political "savvy" allowed us to implement a complex multi-methods research design (MMRD) using multi-level combinations (Figure 1):
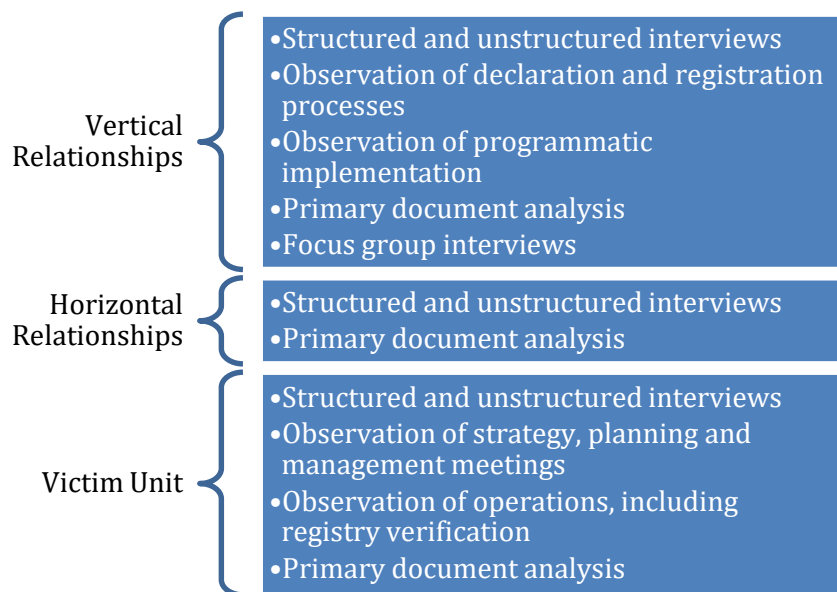
78 Evaluating Transitional Justice

Figure 1: MMRD and Multi-Level Combinations

Overall, component 2 yielded unique insight into the strategies and functioning of the reparation program. One of the main findings is that the VU needs to develop more effective strategies and tools to help prioritize their work. We found strong evidence that suggests the VU does not have a consistent, transparent, or actionable prioritization strategy. For example, the VU used *hundreds* of key performance indicators (KPIs), developed by the central government, to evaluate its progress and ask for resources through various planning instruments. The multitude of performance indicators made it difficult not only for the evaluators, but also for middle- and high-level managers within the VU to develop programs and allocate resources due to the number of goals and lack of within-unit guidance on which goals to prioritize. This also highlights the need for increased engagement by the central government through the SNARIV Executive Committee—the intergovernmental advisory committee tasked with helping the VU set

P. Pham, P. Vinck, B. Marchesi, D. Johnson, P. Dixon, K. Sikkink 79

priorities. We make a strong recommendation that in order to "scale-up" the reparations program, the Executive Committee needs to meet regularly and frequently in order to set and oversee priorities in collaboration with the VU. Once priorities are set, the VU should make public the rationale for these prioritizations. We argue that increased engagement at the highest level of government; increased public communication explaining the rationale for who and what has been prioritized; and increased engagement between the VU and victims to *ex-ante* inform the prioritization strategy and *ex-post* reduce the risks of dissatisfaction will enable more efficient and effective policy implementation.

Strengths and Weaknesses of the Meso-Method
Collecting multi-level combinations of data strengthened C2. We were able to collect a wide variety of data from different sources in varied contexts. When data converged, we could be confident that our findings were not biased, for example, by our data sample or our instruments (e.g., interview guide or survey). When data did not converge, we had to either collect more data or qualify our reporting.

While data collection was one of the strengths of C2, data processing and analysis were more challenging. From a data processing perspective, the quantity of data produced was paralyzing. Almost all of the primary documents and collected data were in Spanish, which added another layer of complexity for a largely English-speaking research team. We also lacked an efficient data management system where raw and processed data were efficiently catalogued and shared within the C2 team and across the other component teams. Such a large research team as we used in C2 was also expensive and would be difficult for most researchers to replicate. Since the VU had requested the evaluation, we had unprecedented access to VU staff and reports. Such access is often not available to TJ field researchers. Also, there was sometimes a mismatch between the number of researchers providing primary research and those doing analysis. In the classic sense of too many trees making it hard to see the forest, the large

80 Evaluating Transitional Justice

amount of data generated by the C2 research team sometimes made it difficult to see the big picture of VU institutional functioning. It is also important to recognize that Colombia is not like other conflict cases in that it is not difficult to find highly skilled, well-educated local partners. Colombia has a strong university system and there are very smart, knowledgeable people available to work with as local partners. This is likely not the case in a host of conflict states.

From an analytic perspective, it was difficult to gain analytical traction without a theoretical footing. For example, when we identified critical gaps in territorial implementation, we were reluctant to bring in the theoretical work of political scientists focused on institutional weakness in transitional democracies. Yet, that literature would have helped us make sense of spatial variation in policy implementation. We were reluctant to bring in such a theoretical perspective because data collection, processing, and analysis were focused on identifying the critical gaps between inputs, outputs and desired outcomes rather than necessarily developing explanations for why or how those gaps emerged and were sustained.[17] In retrospect, we could have benefited from more explicitly processing and analyzing data using the relevant theoretical lens. It is also possible that data from C2 could have helped develop hypotheses or a theory of change that would be beneficial to measure at the end of the program.

Component 3: Micro-Level Perception Datasets
Component 3 (C3) of the evaluation examined the VU's implementation of reparation measures from the perspective of the general population, victims of the armed conflict and the program's beneficiaries. To achieve its objectives, the study used a convergent mixed-methods design, gaining a comprehensive understanding of the

---

[17] The C2 evaluation strategy was based on a management-oriented systems model. In particular, the MSI team used the Logical Framework to guide data collection, processing, and analysis. The Logical Framework focuses on identifying critical gaps between inputs, outputs, and desired outcomes.

P. Pham, P. Vinck, B. Marchesi, D. Johnson, P. Dixon, K. Sikkink   81

effects of reparations through the combination of (1) qualitative data from key informant interviews and focus groups and (2) quantitative data from surveys conducted among randomly selected individuals. While consultation with the VU and stakeholders on the research methods and instruments was conducted in the beginning of the project, data collection and analysis was conducted over the course of five months toward the end.

The qualitative component used interviews and focus groups with selected groups of victims to describe and explain the perceived effects of reparations, the mechanisms behind those results, and the nature of barriers to success. With focus groups, the evaluation team gathered perspectives, opinions and experiences with the declaration and reparation processes. Qualitative methods also provided important insights into Colombia's unique situation of compensating victims while the conflict is ongoing. Where focus groups were not possible, in-depth interviews were used. Field team members were instructed to cover the themes in the interview and focus group instruments completely, but also to probe for understanding the themes from various perspectives.

For our quantitative research, in the absence of baseline data to compare individuals' perceptions and experiences before and after the receipt of reparations, we undertook a comparative assessment among three groups: the general population, people registered as victims of the armed conflict, and people registered as victims of the armed conflict who have received compensation.

General population—random multi-stage cluster sampling procedure: At the first stage, a random sample of 40 municipalities was selected from a list of all municipalities in the country, proportionate to population size.[18] At the second stage, eight blocks (in urban areas) or rural zones were

---

[18] According to 2014 National Statistical Department (DANE, Departamento Administrativo Nacional de Estadística) population estimates http://www.dane.gov.co/index.php/poblacion-y-demografia/series-de-poblacion.

Transitional Justice Review, Vol.1, Iss.4, 2016, 60-94

82 Evaluating Transitional Justice

randomly selected in each municipality. In each selected block or zone, interviewers used a random geographic method to select a dwelling,[19] and a random alphabetical method to select one resident of the dwelling to respond to the questionnaire. Interviewers interviewed an even number of men and women per block so as to preserve gender parity in the sample. The sample size for Bogota was 150; the sample size for all other 39 municipalities was 30, for a total target of 1,320 interviews.

Registered Victims—multi-stage cluster sampling procedure: At the first stage, 30 municipalities were randomly selected from a list of all municipalities, proportionate to the population of registered victims residing in each municipality, using an anonymized database of victims likely to be contactable[20] provided by the VU. At the second stage, unique identifiers (anonymized) were selected using a simple random selection procedure. The anonymized identifiers were provided to the National Consulting Center (CNC, Centro Nacional de Consultoría), a private research firm contracted by our funder for the project, which contacted the selected individuals for interviews. The minimum sample size was 30 registered victims per municipality (60 in Bogota), for a total goal of 930 interviews.

---

[19] Interviewers used a standard method to select a corner of the block to start, from which point they selected one household per side of the block to conduct the interview (on blocks with four sides, which are the majority). On blocks with three sides, two interviews were conducted on the longest of the three sides. On very small blocks (fewer than 16 households) or where the quota of interviews could not be completed, interviewers proceeded to the block diagonal to the selected block. To select a household, interviewers started at the first residence on shorter sides and the third residence on longer sides. Once a residence was selected, interviewers made a list of all eligible residents of the dwelling (18 and older), regardless of whether they were home or not and selected the first name according to alphabetical order.

[20] Victims were considered contactable if the VU had confirmed or updated their contact information within the last year using other available databases.

P. Pham, P. Vinck, B. Marchesi, D. Johnson, P. Dixon, K. Sikkink 83

Repaired victims—multi-stage cluster sampling procedure: Sampling for the repaired victims followed the same procedure as the registered victims. The VU provided an additional database of registered victims who had received compensation payments and thus were considered partially "repaired". For this sample, only those who had received compensation payments directly from the VU were eligible. The minimum sample size was 30 repaired victims per municipality, for a total goal of 900 interviews. The most common experience of violence among the population of registered victims of the armed conflict was displacement (84%). However, relatives of victims of homicide or disappearance were prioritized by the VU for compensation. As a result, victims of these specific crimes are over-represented in this sample, in comparison with the overall experience of violence among conflict victims in Colombia.

Component 3 (C3) sought to answer a number of research questions from beneficiaries' perspectives. In general, we sought to capture victims' self-reported perceptions of the program's various reparations measures and to identify what have been the successes and difficulties in gaining information about, accessing and participating in the program. More specifically, we sought to gauge the program's successes and challenges in terms of social cohesion and what the VU defines as "transformative" reparations: overcoming moral, emotional, and physical damages and contributing to socioeconomic stabilization. Because this is a state-based program, and because attitudes toward the state and its role in the conflict vary quite dramatically in Colombia, we also sought to measure victims' and citizens' confidence in the state and perceptions of the rule of law in Colombia, and to analyze their relationship to the reparations program. Finally, because Law 1448 outlines both individual and collective reparations measures, we included a qualitative component in C3 to identify the effects that "collective victims" (including ethnic groups or institutions such as schools) report with respect to participation in the diagnosis of damages, the formulation of a collective reparations plan, the strategy

84 Evaluating Transitional Justice

for rebuilding social fabric, and the implementation of collective reparations plans.

With such in-depth questions, and without the ability to measure causation through randomized treatment, the C3 team designed a multi-methods approach that would let us (a) draw relational (i.e., non-causal) comparisons among beneficiaries of reparations, victims in general and the general Colombian population and (b) interpret these trends in conjunction with in-depth, qualitative data. In addition, there are subsets of questions relevant to (a) members of the general population who self-identified as victims, (b) registered victims, and (c) repaired victims, which allowed us to look more closely at these subgroups in the analysis.

Strengths and Weaknesses of the Micro-Method
The clear strength of the micro-level analysis, by examining both the perceptions and the impact of the reparations program from the beneficiary's perspective, is that it yields data on indicators of the ultimate effectiveness of the program: whether it is reaching its intended beneficiaries and having the desired effects of providing them with a sense that the State is committed to repairing the harms they experienced and, ultimately, of affecting their lives. And while surveys are constrained by the limitations of self-reported data, as noted below, they do allow for both direct and indirect measurement of complex concepts. To gauge victims' confidence in the state, for example, we were able to ask directly about how the reception of compensation had made respondents feel about the state as well as how they feel about specific state-based institutions, unrelated to the receipt of reparations. Our comparative design, in turn, then enabled us to look for variation between those who had received reparations and those who had not.

Furthermore, studies of TJ are rarely able to collect fully representative data on the victim population, especially on a scale like that of Colombia's. Even rarer are studies that can simultaneously collect nationally representative data to which these data can be compared. Only through large-scale quantitative research can

P. Pham, P. Vinck, B. Marchesi, D. Johnson, P. Dixon, K. Sikkink 85

researchers capture the trends, magnitudes, and relative proportions across entire populations and sub-populations. This is increasingly important for the study of TJ, as countries develop large-scale reparations programs for large swaths of their populations.

Finally, another key strength of this component is that it has collected both qualitative and quantitative data. While the quantitative element allows for understanding of trends, magnitudes, and relative proportions, the qualitative data provides in-depth contextual data. For instance, in the qualitative interviews respondents noted that they thought the state's current system of prioritizing victims for reparation was random or based on luck. In the quantitative survey, two-thirds or more of the three sampled populations said that those most in need should be prioritized (Population (Pop.) 68%, Registered (Reg.) 70%, Repaired (Rep.) 66%). There was also support among the three populations, especially repaired victims, for a "first-in-first-out" scenario (Pop. 10%, Reg. 12%, and Rep. 19%). There was some support for a system of prioritization based on the victimizing act (Pop. 16%, Reg. 17%, and Rep. 13%). Still, the clear preference was for a system of prioritization based on need (defined in the survey as "vulnerability"), and the qualitative results suggested that beneficiaries do not perceive the current system as taking this approach. Indeed, this is not the approach currently followed by the VU.

One clear limitation of the micro-level perspective is that it relies on self-reported data, the reliability and validity of which may have been affected by a number of factors, including inaccurate recall of past events, misunderstanding of questions or concepts, reactions to the sensitive nature of the questions, and intentional misreporting (e.g., for socially unacceptable answers). Though we minimized such risks through careful development and piloting of the questionnaire to make the questions sufficiently clear and to reduce potential bias, it is not possible to account for how respondents were influenced by these factors.

Furthermore, while a population-based survey often uses random sampling and thus arrives at a more representative sample,

86 Evaluating Transitional Justice

surveys of specific populations that require a list of names, such as the list of victims who had registered and who had been repaired, depend on two factors. First, they are dependent on the quality and completeness of these lists. The Colombian Victims Registry, however, openly acknowledges that its lists are incomplete—on the one hand because there are in fact a number of different lists that have been collected over the years and are now being consolidated, and on the other hand because of the extremely high percentage of displaced victims, whose contact information may change often. Second, surveys of listed populations are dependent on the ability to find those who are randomly selected. This was hard to do for similar reasons: victims' contact info can be out of date, especially for displaced individuals and families who may be forced to move frequently. As with the C2 team, however, we worked with a very capable team of Colombian survey administrators from the CNC. The CNC compiled a list of randomly selected respondents, made phone calls, and was able to complete almost all the interviews we needed for our sample.

Ultimately, the collection of micro-level data in conjunction with the meso- and macro-analyses described above was very beneficial not only in interpreting our data, but also in drafting the initial survey instrument and in crafting tailored policy recommendations to the VU. One clear finding from the comparative analysis of C1, for instance, was that there can be an ambiguous relationship between *administrative* reparations programs, which do not include judicial proceedings, and justice, particularly in contexts where the state bears some responsibility for violence. In Guatemala, for instance, the administration of reparations payments without any judicial proceedings, according to one case study of a Mayan community, seems to have exacerbated victims' sense of injustice.[21] Noting this, the C1 team thought it was important to target victims'

---

[21] Lieselotte Viaene, "Dealing with the Legacy of Gross Human Rights Violations in Guatemala: Grasping the Mismatch between Macro Level Policies and Micro Level Processes," *The International Journal of Human Rights* 15.7 (2011): 1160-1181.

P. Pham, P. Vinck, B. Marchesi, D. Johnson, P. Dixon, K. Sikkink 87

perceptions of justice vis-à-vis the reparations payments they were receiving, especially since such payments are the VU's primary obiligation in terms of providing reparations (in contrast to its coordination role). We found that among a broad range of both positive and negative emotions, victims most often said they felt sad upon receiving their compensation checks, particularly those who had lost loved ones to the conflict. Moreover, two-thirds of repaired victims responded that the payment had *not* delivered justice when asked about their sense of justice.

In terms of our policy recommendations, we therefore stressed the importance of providing victims with the psychological and social support they likely needed in addition to compensation. We also stressed the importance of strategizing about how the VU might at least help victims arrive at a partial sense of justice by, for example, coordinating with Colombia's National Center of Historical Memory and some of the truth-seeking work they are pursuing. We also used the C2 findings related to institutional capacity and coordination to stress the importance of the VU's helping victims fully understand the reparations program and reparative process. Partly as a result of poor coordination, we concluded, beneficiaries of reparations are conflating reparations with both humanitarian assistance measures and other state-based subsidies for victimized populations (for displaced families, for example). Better coordination and messaging is needed, therefore, to ensure that victims not only receive compensation but understand their specific significance as reparations and not just another state subsidy.

Conclusion: Benefits and Weaknesses of Multi-Methods Research on Transitional Justice

The evaluation of the reparation program in Colombia provided a unique opportunity to implement a multi-level analysis at the macro (international policies), meso- (institutional), and micro- (population) levels. Each of the research methods employed for the Colombian Victims Unit evaluation had its own strengths and weaknesses, but

88 Evaluating Transitional Justice

undertaken together, simultaneously and by a coordinated team, they served to assemble a comprehensive picture of the Colombian reparations program from the perspective of its place in the global TJ context, its institutional functioning, and its ultimate responsiveness to the population it was set up to serve.

Component 1 focused primarily on the formal and legal characteristics of the Colombian reparations programs in comparison to others around the world; it did not result in an understanding of how the reparations program in Colombia was actually being implemented. Component 2 provided an incredible wealth of detailed information about the institutional framework, which was still quite formal and legal. The organizational charts and the flow charts describing the formal processes did not always capture the nature of how power and influence worked in the Colombian bureaucracy. Component 2 began to reveal the political issues and economic constraints the VU faced in trying to carry out its mandate. Some C2 research also went deeply into the local level, as members of the research team visited collective reparations programs and saw how it was being implemented in practice. Component 3, finally, complemented the first two by providing insight on the beneficiaries' and the general population's perceptions about the reparation programs and how it has impacted them from a self-reported perspective, according to a variety of complex benchmarks: social cohesion, attitudes toward the state, moral and physical wellbeing and socioeconomic stabilization. Taken together, the evaluation results provided invaluable insights and information to the government and the VU on how to revise their program activities to achieve greater efficiency and efficacy, as well as achieve greater impact with regards to alleviating the physical and psychological harm.

The need for a comprehensive approach renders pointless any debate about the most appropriate research methods by which TJ

P. Pham, P. Vinck, B. Marchesi, D. Johnson, P. Dixon, K. Sikkink 89

measures should be evaluated or their impact assessed.[22] As noted, we believe this project demonstrates that different research methodologies can—and should—be complementary rather than competitive. As Straus and Finkel point out in relation to studies of genocide, "Because genocide is usually a top-down process, in which national states are initiators and authorizers of mass violence, theorizing causation at the macro level is a natural fit for genocide studies. At the same time, the dynamism at the macro level should not overshadow important research agendas at the meso and micro levels." [23] In our experience, studies of transitional justice equally benefit from the application of multiple perspectives and approaches. Both the violence that Colombia has experienced, as well as the responses it has undertaken to address the aftermath of this violence, entail macro-, meso- and micro-level processes that only such multi-methodological research can comprehensively unravel.

We originally conceived of data collection, processing, and analysis in a sequential or chronological format, from the macro-level to the meso- and micro-levels, where the results from earlier analyses inform later analyses. However, we changed course due to time and other resource constraints. Instead, we collected, processed and analyzed data using a parallel format. The parallel format involves two or more datasets that are analyzed separately and the findings are then combined or integrated. The advantage of a parallel format is that findings can be triangulated across different data sources using mixed methods; and findings from one analysis can be expanded or deepened using data from another part of the analysis.[24] This potential remains

---

[22] See Colleen Duggan, "Editorial Note, Special Issue: Transitional Justice on Trial - Evaluating Its Impact," *International Journal of Transitional Justice* 4.3 (2010): 315-328.

[23] Evgeny Finkel and Scott Straus, "Macro, Meso, and Micro Research on Genocide: Gains, Shortcomings, and Future Areas of Inquiry," *Genocide Studies and Prevention* 7.1 (2012): 56-67.

[24] For more information about evaluation design (e.g. parallel vs. sequential), see, "Evaluations," U.S. Agency for International Development https://www.usaid.gov/results-and-data/progress-data/evaluations.

Transitional Justice Review, Vol.1, Iss.4, 2016, 60-94

90 Evaluating Transitional Justice

underexplored at this time. Due to time and staffing constraints, we were unable to thoroughly analyze all of the collected data and look for all convergent and divergent findings. One possible shortcoming of our evaluation is that we focused much of our discussion and feedback on convergent findings rather than on divergent ones. Furthermore, the challenges of working across multiple disciplinary fields also presented themselves, but we think they made both the research process and the product richer.

The combination of multi-methods research design and the parallel format maximized the opportunity to triangulate findings. For example, in Component 1 we categorized the Colombian reparations policy as one of the most complete and complex in the world.[25] In Component 2, we found that different aspects of the VU's management practices were overly complex. In Component 3, we found that repaired victims were often confused about who provided reparations and what benefits and services counted as reparations, as assistance, or as subsidies. Although each component used different data sources and focused on a different level of analysis, a consistent finding across components was that the VU had to do more to simplify its interpretation and implementation of the policy

The three levels of data combined to guide the formulation of recommended changes in program implementation, both from the policy and the institutional levels, to achieve great efficiency and effectiveness in the ultimate delivery of reparations to the beneficiaries. The issue of prioritization of victims provides a clear example of how the three research streams were each important in formulating the evaluation's final recommendations. Component 1 highlighted the ambitious nature of the reparations program in Colombia with its commendably broad eligibility criteria and resulting large number of registrations. However, through the meso-level, institutional analysis

---

[25] Following de Greiff, a *complete* policy covers the full range of beneficiaries and a *complex* policy offers beneficiaries a range of benefits and services; see: Pablo De Greiff, ed., *The Handbook on Reparations* (New York: Oxford University Press, 2006).

P. Pham, P. Vinck, B. Marchesi, D. Johnson, P. Dixon, K. Sikkink 91

undertaken under Component 2, there was clear evidence to suggest it was essential that the VU prioritize its tasks more clearly since its institutional constraints—resources and capacity—would not allow it to fulfill the full scope of the mandate assigned to it and respond equally to the large number of claims received. Component 3, however, found that the population's expectations of the program were high, and that any attempt to undertake large-scale changes to the program could have serious negative consequences, particularly if they involved removing certain groups of victims' eligibility for benefits. Based on these three types of data, the evaluation team was able to propose recommendations to the VU that although prioritization is necessary, its criteria must be clearly conveyed to the population through a strong communication strategy.

Ultimately, the combination of the three components yielded valuable information for a variety of important TJ considerations, including the design and implementation of administrative reparations programs, their relationship to the judicial process and their impact on victims' sense of recognition and wellbeing. These are all key issues in the TJ literature that demand both rigorous data collection and access to institutions and victimized populations. This latter point is an important one. Researchers may be more comfortable working in one context or the other and simultaneous access to both can be difficult to arrange. Yet, TJ concerns both worlds. It is simultaneously an institutional response—a "toolkit"—and a set of social and political processes that people experience.[26] TJ scholarship is often concerned with the gap between these two components of the field, but less often are TJ scholars able to access both at the same time.

As we have noted, the resources and coordination to accomplish such integrated research, however, are not without their challenges. Perhaps most fundamentally, the difficulty of combining

---

[26] Pablo de Greiff, "Theorizing Transitional Justice," in *Transitional Justice: NOMOS L1*, eds. Melissa S. Williams, Rosemary Nagy, and Jon Elster, 31-77 (New York: New York University Press, 2012).

92 Evaluating Transitional Justice

macro-, meso- and micro-analyses comes with the inevitable differences that exist between academic fields. As we have noted, TJ problems are simultaneously legal, political, social, economic, cultural, and technical. Our team combined researchers with backgrounds in public health, sociology, political science, business administration, and evaluation. Agreeing on common definitions and approaches, therefore, was always challenging. This, however, also forced those involved to confront their assumptions and approaches to make the research methods fully attuned to the problems under investigation and not, as can happen, vice-versa. Moreover, while our team and methods were multi-disciplinary, we were not able to utilize all methods that are useful for research in transitional contexts. Notably, we did not incorporate ethnographic methods, which are valuable for uncovering the complex and often hidden cultural layers of transitional justice and, as Kimberly Theidon writes, "moving beyond the black and white of statistics to explore the grey zone that characterizes the complex realities of a fratricidal war."[27]

Furthermore, this was a resource-heavy project, with significant resources dedicated to travel and staff hours at both Harvard and at MSI in Colombia. Our first component drew on data that had already been gathered, cleaned, and analyzed by the TJRC team with funds from a separate grant. Both the funds required for such an ambitious project as well as the in-country capacity on which we drew raise relevant questions about the project's replicability. Through MSI, we were able to work with staff with expert knowledge of the VU and Colombia's legal context. For Component 3, the VU provided our team with a database of all recipients of reparations, which CNC then spent significant time calling and tracking down to interview. Other countries undergoing "transitional" processes may likely not have similar internal capacity to support what was a relatively

---

[27] Kimberly Theidon, "Transitional Subjects: The Disarmament, Demobilization and Reintegration of Former Combatants in Colombia," *International Journal of Transitional Justice* 1.1 (2007): 66-90, 74.

P. Pham, P. Vinck, B. Marchesi, D. Johnson, P. Dixon, K. Sikkink 93

small core team for a project of this scale.[28] At the same time, the importance of the research to enhance the effectiveness and efficiency of the reparation program should not be understated. From this perspective, and considering the huge cost of the reparation program itself, the resource allocated to the research were not out of proportion. The integrated research approach was more valuable and cost-effective than a series of independent research, and its cost was minimal compared to the cost of the policy under consideration. We found that working as a collaborative team of individual researchers with different but complementary methodological expertise can yield new insights that would be nearly impossible for individual researchers to achieve even with a very sophisticated MMR design. We also note that future work of this kind must be more attentive to triangulating different kinds of evidence, synthesizing findings across different levels of analysis, and investigating contradictory findings. Ultimately, we suggest that there is room for further effort within the TJ research community for collaboration on similar large-scale endeavors in addition to the pursuit of scholars' individual research agendas.

In conclusion, this evaluation of the Colombian reparations process demonstrates the value of a multi-level research approach that proved highly effective in assembling a comprehensive, multi-perspective picture of the effectiveness of this particular TJ measure. Having multiple research approaches looking at different aspects of the program allows for triangulation of findings, which leads to programmatic recommendations that reflect more than one perspective or type of expertise. In a highly complex setting such as that of the Colombian peace and TJ processes, the multiple research approaches also contribute to the validity of the assessment, since they

---

[28] See Morten Jerven, *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It* (Ithaca: Cornell University Press, 2013). Our Component 3 team has shown that, at least for the general population, similarly ambitious micro-level surveys can be implemented in partnership with local institutions. See www.peacebuildingdata.org for more details on surveys carried out in Liberia, the Democratic Republic of Congo, Uganda, and elsewhere.

Transitional Justice Review, Vol.1, Iss.4, 2016, 60-94

94 Evaluating Transitional Justice

are the result of a study led by a multi-disciplinary team and implemented by several actors working in coordination. We believe that the multi-level approach merits future adoption in relation to other TJ processes and measures. This would not only strengthen those individual processes and measures by providing evaluative data and enhancing learning about the effectiveness of transitional justice, it would also serve to improve and refine each research method— whether macro, meso or micro—and the means of using all three simultaneously.