

Global Forced Displacement Data: the Politics of Counting People on the Move

Sigrid Weber *University College London*

Abstract: Quantitative analysis of the causes and consequences of forced displacement has increased in political science in recent years. At the core of these analyses is time-series-cross-sectional data on global forced displacement produced by international and national policy actors. This research note discusses the inaccuracy and political bias in data on forcibly displaced populations by focusing on challenges for political scientists. With a replication of Blair, Grossman and Weinstein (2022), I demonstrate that the inferences we draw from comparative analyses of forced displacement data are sensitive to both the data generation process and the political nature of the data. The note concludes that more awareness for the politics behind counting people on the move is crucial.

Introduction

Research on forced displacement, not only in political science but also in the social sciences more broadly, has dramatically increased in the past decade. More than six times the number of papers have been published between 2010 to 2019 than in the decade before (Dionigi and Tabasso, 2020). Reasons for the increased interest in forced migration certainly are the growing global scale of displacement, the realisation that population movements have far-reaching consequences for politics and the improved availability of data on these mobile and vulnerable populations. Researchers increasingly use data on the numbers of refugees, internally displaced persons (IDPs) and other forcibly displaced populations in a comparative way, comparing the causes and consequences of human mobility across time and countries with sophisticated quantitative empirics and time-series-cross-sectional data (e.g., Blair, Grossman and Weinstein, 2022; Turkoglu, 2022; Devictor, Do and Levchenko, 2021).

In this letter, I ask the question whether we know enough about the quality but also the politics of forced displacement data to confidently draw conclusions for academic research. Can we truly compare refugee and IDP numbers across countries and time in quantitative analyses? I first highlight different sources of ambiguity and bias in forced displacement data and I indicate that the data used in current research may not always be as accurate, as comparable and as apolitical as we might want the data to be. I then replicate a prominent and recent quantitative study on forced displacement by Blair, Grossman and Weinstein (2022). In this replication, I take into account the underlying data generating process in

the replicated models. I demonstrate how differences in the production of forced displacement data change - albeit not fully alter - the inference we might draw from the analysis. I conclude the note with a summary why it is crucial that academic researchers are more conscious about the data generating process when using forced displacement data. I highlight how this data issue relates more broadly to vulnerable populations and I point out possible ways to move forward as a discipline.

Many forced migration scholars have criticised the centrality of statistics in refugee and migration studies as well as the reliability of data on forced displacement (e.g., Kibreab et al., 1985; Harrell-Bond, Voutira and Leopold, 1992; Cardona-Fox, 2020). Critical voices often conclude that forced migration researchers "have been remarkably inattentive to the issue of quantitative data" (Crisp, 1999). Given the recent explosion of studies using forced displacement data, a moment of reflection on the origin and politics behind data seems crucial, in particular for political scientists that publish and read outside of "traditional" journals for forced migration studies. This letter hence particularly articulates the problem for academic researchers in quantitative political science rather than policy makers or other audiences. Furthermore, forcibly displaced populations or populations at risk of displacement - such as stateless persons - are extremely vulnerable. From an ethical point of view, quantitative research on these populations should be careful to generate findings that can stand up to scrutiny in regards to the underlying politicised nature of the data.

The difficulty of counting forcibly displaced persons: sources of inaccuracy

Estimates of the number of people fleeing from conflict, violence and disasters are important data sources for quantitative political science on forced displacement. A lot of recent research would not have been possible without data on global forced displacement (e.g., Braithwaite, Salehyan and Savun, 2019; Zhou and Shaver, 2021), for example provided by the United Nations High Commissioner for Refugees (UNHCR).

These data are not produced with academics in mind but are produced predominantly to inform humanitarian and development planning and aid allocation. Governments, journalists, politicians and the public also use the data to frame the importance of human mobility, to discuss reactions to displacement and to advance their agenda. Despite this central role of forced displacement data, there are several sources of inaccuracy in relation to the number and composition of refugees, IDPs, and other mobile populations.

Inaccuracy in the definition of forced displacement data

A first problem arises from *definitional challenges*. Although the term *refugee* is a legal term from the 1951 UN Refugee Convention, different organizations and world regions have counted refugees based on different definitions (Crisp, 1999). Since 1953, the United Nations has been expressing a need to enhance the comparability of international migration data and has made significant advances. Most recently, in the 2016 *New York Declaration for Refugees and Migrants*, the UN has reiterated the importance of improved data collection. In 2018, the newly founded *Expert Group on Refugee and IDP Statistics* has published *International Recommendations on Refugee Statistics* to standardize how countries and international organizations count refugees in their official statistics (EGRISS, 2018). These efforts to provide guidelines on what constitutes a refugee for statistical purposes show that no unity is yet achieved in practice. As a result, an individual counted as a refugee in one country - for example because they fled violent events - might not qualify for that status in another part of the world where a stricter definition is applied. Another example for academics is the question whether the data they use to compare refugees across countries includes Palestinians and Venezuelans - as different organizations include these people on the move in their refugee statistics or exclude them depending on the agencies that carries the mandate to protect these populations. Hence, researchers using global data must ensure that the populations they are *theoretically* interested in are sufficiently covered by the data sources they are using and comparable definitions are used.

The problem of definitions becomes even more acute when considering the concept of *internally displaced persons*. In contrast to the refugee definition, being an *IDP* is not a legal status or term. It remains unclear how far a person has to move within their own country to be considered displaced, what to do if international borders change and when IDPs cease to be IDPs (Crisp, 1999). While refugees end their legal status by resettling to their home country, or by migrating or naturalising abroad, there is no standardized way to account for the end of internal displacement in official IDP numbers (Cardona-Fox, 2020). The 2020 *International Recommendations on IDP statistics* aim to provide a unified statistical definition for IDPs but fail to conceptualise an applicable measure of exits of IDPs from the statistical stock because they have overcome their displacement-related vulnerabilities (EU and EGRIS, 2020). Some political scientists acknowledge that definitions of forced migrants "may vary from country to country" (Turkoglu, 2022) but discussions of the implications of comparing IDPs across contexts and with refugee numbers are largely absent in the discipline.

Definitions are inherently political and are used in politicised ways, for example by national statistical

offices (Ustek-Spilda, 2020). Colombia, despite relatively advanced IDP tracking systems, has shown unwillingness to remove IDPs from the country's register (Cardona-Fox, 2020). Because there is no agreement when IDPs should be taken out of official statistics, ambiguous definitions mean that the stock number of IDPs in Colombia hardly declines. This can have implications for academic findings given that the case of Colombia is very commonly studied by political scientists and economists to understand the phenomenon of forced displacement (e.g., Ibáñez and Vélez, 2008; Ibanez, 2009; Saldarriaga and Hua, 2019; Engel and Ibáñez, 2007; Balcells and Steele, 2016). Inaccurate numbers do not invalidate these findings but raise questions whether the framing used in papers is accurate and whether generalisations are possible.

Inaccuracy in the production of forced displacement data

The difficulties to define consistent and comparable statistical definitions of refugees and IDPs across time and contexts also relates to operational realities (Crisp, 1999): it is actually quite challenging to collect reliable data on mobile populations. The data generating organizations, like UNHCR or the Internal Displacement Monitoring Centre (IDMC), face several *practical and operational challenges*: They collect data on IDP and refugee numbers in emergency situations, during violence and conflict, in coordination with low-capacity states and other actors that have various other priorities. Not all moving individuals and households want to be counted, double-counting or under-counting occurs, and mixed migration movements make precise numbers even more complicated.

In Western countries, the number of refugees can often be captured through the asylum system as individuals have to make a claim for asylum and there is a clear number of recognized refugees. In the Global South, refugee recognition is often based on *prima facie* recognition. *Prima facie* arrivals usually occur in situations of large-scale movements in which entire groups are considered refugees when crossing the border and requiring humanitarian assistance. In these instances, the number of forcibly displaced persons entering our datasets can be an estimate by field personnel or may be based on camp registration rather than granted asylum applications. A number of different approaches and combinations of approaches can be used to estimate in- and out-of-camp refugee and IDP populations from door-to-door censuses, field surveys, estimations based on the spatial area and population density, the use of administrative records such as vaccination data but also statistical modeling with satellite imagery or call detail records to track movements in urban contexts (Ratnayake, Abdelmagid and Dooley, 2022).

Whether the different ways to produce data on displaced populations can truly be comparable across vastly different contexts is a critical question. Schwartz (2022) highlights that even though states might

accept refugees in *prima facie* procedures, increasing our estimates of forced migrants in the country, governments also use a wide range of strategies to return displaced persons under the radar, for example by halting *prima facie* recognition, threatening to cease refugee status or creating hostile living conditions. I return to this question in the replication study to discuss whether refugee numbers through *prima facie* recognition can be compared to refugee numbers based on individual asylum applications.

Once refugees and IDPs settle in camps, at specific sites or enter assistance programs, it may become easier to collect reliable demographic data. However, with the establishment of aid programs, incentives to register as IDP or refugee may distort accurate statistics in contexts in which non-displaced households are also considerably vulnerable (Crisp, 1999). Some scholars would hence go as far as saying that there is a "cloud of uncertainty and unreliability surrounding African refugee statistics" (Kibreab et al., 1985).

Kibreab (2004) has documented reasons why refugee communities inflate the size of their families at pre-registration, withhold information on deaths in the family, or register at different sites to ensure sufficient aid. For examples, refugees from Ethiopia were walking back across the border into Ethiopia and returned repeatedly to Somalia in the mornings as "new arrivals" to Tog Wojalle refugee camp because it was only 7km away from the border and this ensured them to receive more support. A full enumeration of refugees in the camp revealed in 1987 that only 32,000 refugees lived in the camp rather than the assumed 87,000 (Kibreab, 2004; Christensen, 1982).

In the context of IDPs, Ratnayake, Abdelmagid and Dooley (2022) have emphasised that many humanitarian efforts to estimate IDP populations - with population statistics and spatial data - during recent large-scale displacements in Burkina Faso, Cameroon, Syria, and Yemen are inaccurate, for example because they over- or under-estimate population denominators (also see Diggle et al., 2017; Checchi and Besson, 2022). Operational challenges induce (un)systematic inaccuracies.

For researchers, one crucial question is then what it means for their findings if some data is produced under more difficult operational conditions than others. If access for data collection is more difficult in one context, what are the implications for comparative analysis? Very few research articles discuss potential over- or undercounting of the underlying displacement data that could lead to systematically biased results in their models.

Inaccuracy in the use of forced displacement data

These sources of inaccuracy culminate in what I call *usability challenges*. By this I mean the inappropriate use of forced displacement data in a way that the data were not intended to be use or cannot be used. A

common example is to combine and aggregate forced displacement data from different organizations and sources to enable country-level comparisons across sections and time when it remains unclear whether the data generating process is truly comparable. If researchers combine displacement data from various sources and organizations without knowing or inquiring about the difference in definitions or the comparability, the data itself may be somewhat inaccurate but the comparisons drawn by researchers add additional systematic uncertainty. Researchers have also not paid attention to the fact that the way data is collected over time changes for a given country unit and may imply temporal discontinuities in refugee numbers due to methods rather than substance.

Another challenge to use forced displacement data are cognitive biases by the user/researcher. One such cognitive bias is the idea that more recent data is considered "better" or "more accurate" without testing and probing this (Shaver et al., 2022). The policy sector, in particular humanitarian and developing organizations, have been under considerable pressure to quantify their missions, to develop dashboards presenting the scope of their operations, and to be held accountable with numbers and figures. This pressure to provide quantitative data, to visualise it, and to make it used by a wider audience through databases and dashboards does not necessarily mean that the *quality* of data has improved. The fact that researchers can now download datasets on forced displacement while this required access to the data producers in the past is a step forward to foster open and equal science. It does not imply that the data itself requires less explanation for the user through the producer and is automatically of higher quality.

New technological solutions are increasingly portrayed as the way to improve refugee protection and data - for example by using bio-metric registration data. However, Jacobsen and Sandvik (2018) amongst others discuss that these data improvements may improve accountability towards donors of UNHCR and other agencies but may not improve the protection and accountability towards refugees. In 2021, Human Rights Watch reported that UNHCR improperly collected and shared personal information from ethnic Rohingya refugees with Bangladesh, which was then shared with Myanmar (Human Rights Watch, 2021). Given the vulnerability of people on the move, academic researchers have no choice but to pay attention to the politicised nature of displacement data and the ethics of producing such data in the first place as it may contribute to the surveillance of vulnerable populations. This problem is more acute for fine-grained and spatially dis-aggregated analyses than for global comparisons. However, the current academic practice, including in this author's own work, is insufficiently discussing the ethical implications of collecting and using increasingly fine-grained data of vulnerable populations.

The politics of counting forcibly displaced persons: sources of bias

Difficulties in enumerating forcibly displaced populations have led some analysts to express strong scepticism towards counting refugees (e.g. Harrell-Bond, Voutira and Leopold, 1992), in particular because inaccuracies in the data not only come from definitions, wrong usage or operational issues. In the following sections, I discuss additional political challenges that may introduce systematic political bias.

Bias in forced displacement data by states

First, global forced displacement data suffers from major *political challenges*. The producers of forced displacement data have their own agenda. Crisp (1999) explains well that refugees are a "political failure" for countries of origin that may hence have the incentive to downplay numbers of forcibly displaced persons or emphasise that only civilians involved in the opposition, in criminal or rebel activity against the state are leaving the country. In contrast, countries of origin may inflate returnee numbers as citizens deciding to come back not only show support for the government but also provide a source of development assistance.

At the same time, destination countries, in particular in the developing world, may exaggerate refugee numbers in their territory to receive humanitarian assistance and improve their political standing vis-a-vis other states. An example from the IDP context is that countries like Azerbaijan, Cyprus, Ukraine, and Georgia have inflated estimates of IDPs in their own country by including children born in displacement in their IDP registries (Cardona-Fox, 2020). Crisp (1999) also highlights that some countries of asylum may not inflate but deflate refugee numbers to de-escalate the relationship with refugee-causing neighbour countries or in diplomatic efforts.

Donor states or refugee-receiving states in the Global North also have politicised incentives to change refugee data. Governments in developed countries often selectively present refugee data, for example to justify their moves towards more restrictive asylum policies. Such manipulation may mean that governments emphasise sharp increases in asylum applications while not reporting the proportion of refugees they take in from a global perspective or their often lower acceptance rates than in the Global South (Crisp, 1999). Donor states giving money for refugee assistance may want to use forced displacement data to demonstrate the effectiveness of their programs, or want to use the data to justify cuts to their programs and budget lines.

In the earlier mentioned example of over-counting refugees from Ethiopia in Somalia, we can find a

good example for purposeful bias in refugee numbers. While a full census of Tog Wajalle camp took place and revealed that less refugees had found refuge in the camp than originally recorded, additional enumerations in other refugee camps were opposed by the government of Somalia. "The Somalia authorities learned from the outset that the best way to secure the continuous and increasing flow of international aid was by dramatically inflating the number of refugees in the country." (Kibreab, 2004). Donor states like the US but also the implementing National Refugee Council likely had their own interest and did not correct the known and deliberate inflation of refugee numbers. Refugees in Somalia were starving despite the oversupply of food but the warehouses of the National Refugee Council were stocked. The overlapping interests by various organizations in this example also demonstrate why it can be slow and difficult to correct de- or inflated refugee numbers that end up in official reports and academic comparisons.

But even beyond the actions of governments, humanitarian organizations face their own political dilemmas: Unbiased and accurate displacement data certainly helps to improve planning and humanitarian responses. However, there are also incentives to adjust displacement numbers to the political needs of the organization. On the one hand, aid organizations have incentives to report high numbers to increase their chances to raise funds for their cause. On the other hand, they have incentives to demonstrate the effectiveness of their aid by seeing declines in refugee numbers over time. Blitz (2021) for example criticises the way UNHCR has dealt with data on stateless populations - a group of people that is often at high risk of displacement. In particular, UNHCR's *iBelong* campaign claiming to eradicate statelessness by 2024 may create the incentives to downplay and strategically shift the definition of who counts as a stateless person to achieve the proclaimed goal. These incentives do not translate into large-scale strategic distortions of refugee numbers, but the conflicting interests to over- or underreport at localised points within the organization persists.

Bias in forced displacement data by international institutions

Intertwined with these political challenges, there is another difficulty arising from *accountability and institutional challenges*. Within an institution that produces forced displacement data, different incentives to report high or low numbers prevail. For example, field offices providing numbers might have different incentives, driven by needs for aid supplies, than those reporting and generating the official statistics in regional and headquarter offices. Additional institutional challenges come from competing numbers across institutions. While UNHCR may have the clear mandate to report on refugees, UNHCR has become increasingly active in situations of internal displacement. The expansion of UNHCR's work to IDPs, state-

less populations and other populations on the move means that UNHCR increasingly produces operational estimates of IDPs that may differ from estimates by other agencies. For example, the estimates on IDPs by IDMC do not always overlap with UNHCR statistics. Competing estimates from the International Organization for Migration, in particular in mixed migrations settings, further complicate the picture. Given that multiple international agencies and national statistical offices are involved in the production of statistics on displacement, it remains the task of the researcher to identify the most reliable data sources.

Overall, the critical question for researchers using forced displacement data is whether they can creatively identify ways to mitigate the bias from accountability challenges and political incentives in their models. One solution could be to compare numbers of refugees from origin and destination countries with the estimates of humanitarians and development agents. While this may not be feasible for all countries in global analyses, this could be relevant for outliers in the data or for country contexts with known political controversy around displacement numbers.

Demonstrating the problem: a replication of Blair, Grossman and Weinstein (2022)

To demonstrate the aforementioned challenges, I replicate recent findings by Blair, Grossman and Weinstein (2022) in the *American Political Science Review*. Using gravity models, they show in many convincing robustness analyses that "liberal displacement policies attract forced migrants in the Global South" through diffusion of policy knowledge and transnational ethnic kin. The paper provides an important contribution to the literature as very few researchers focus on displacement policies in the Global South.

The authors estimate the rate of refugees (i.e. new arriving refugees proportional to the local population) per year for dyads of origin and destination countries. To construct the rate of refugees, the supplemental material specifies that they use a combination of two data sources. First, the UNHCR Population Statistics Database provides the directed dyadic number of asylum-seeker applications. This number is based on individuals filing for asylum and could be prone to under-counting as not all people on the move, in particular in the Global South, ask for asylum in the countries they cross. Second, UNHCR released data on the directed dyadic number of *prima facie* arrivals to Fearon, Shaver et al. (2021). These numbers are particularly relevant in neighbouring states to the displacement-causing state in which the scale of displacement emergencies often lead to group-based *prima facie* recognition of refugees at scale. Blair, Grossman and Weinstein (2022) combine these two data sources as the main dependent variable. The authors are aware of the problem that refugee *recognition* via individual status determination is endogenous to liberal policies and circumvent this problem by focusing on applications and not recognized

refugees. However, they do not acknowledge that the two data sources for refugee numbers suggest a profoundly different data generating process.

The crucial empirical question is whether refugee numbers in countries with largely *prima facie* recognition can truly be compared to refugee numbers based on individual asylum applications. Are 10,000 applications through the "regular" asylum system the same as 10,000 refugees recognised in a *prima facie* decision? Imagine two countries A and B that are both exposed to 10,000 refugees arriving from a neighbouring conflict-ridden country. In country A, refugees have to go through a system of individual asylum seeking while country B does not have the capacity to process these asylum applications and a group-based *prima facie* recognition kicks in. Although both countries are equally exposed to displacement, are we likely to record 10,000 refugees in both countries? On the one hand, we could see under-counting in country A as not everyone will ask formally for asylum. On the other hand, an emergency recognition of these refugees in country B could lead to higher numbers than 10,000 (as more (double-)register in camps or are counted multiple times) or lower rates (as still not everyone registers). Importantly, the data generating process likely has an impact on the measurement of their dependent variable.

In my replication, I am hence interested to see if the results by Blair, Grossman and Weinstein (2022) hold if I control for *where* the data is coming from. They fit Poisson pseudo-maximum likelihood (PPML) gravity models in which the dependent variable is the refugee rate and the main independent variable is an interaction between their score of liberal policies towards displacement and a facilitating variable (e.g. information or ethnic kinship). Conditional on the facilitator, liberal displacement policies should increase refugee numbers. The models include origin and destination fixed effects, year fixed effects, and controls for region, intercapital distance, contiguity, shared language, bilateral stock of nonforced migrants, GDP/capita ratio and its squared term, population, unemployment¹, civil war occurrence, and repression at origin and destination. I fit the same models but I add two binary variables that describe whether any of the refugees in the dyad were recorded in individual asylum applications (1) or not (0) and whether any of the refugees in the dyad were recorded in the *prima facie* data source (1) or not (2). Table 1 provides this adjusted replication of the main results in Blair, Grossman and Weinstein (2022).²

I do not find a consistent interaction between policy liberality and the facilitator variables that are used by Blair, Grossman and Weinstein (2022). Information or the presence of ethnic kinship groups in countries with more liberal policies towards refugees does not consistently increase the rate of refugees in a given dyad. I still find, in line with Blair, Grossman and Weinstein (2022), that more ethnic kinship groups in

¹Since this variable is not included in the replication data, I exclude it from the model.

²For the original replication without model adjustments, see appendix.

Table 1: Replication of table 1 in Blair, Grossman and Weinstein (2022): Effect of policy liberality and a facilitator on refugee rates with a control for whether refugees were captured through asylum or *prima facie*. Red text shows by what percentage the coefficient has declined when controlling for data sources.

	ICT		TEK presence		TEK numbers	
	(1)	(2)	(3)	(4)	(5)	(6)
Policy liberality index (5-year MA) x facilitator	-0.068 (1.506) ▼-102.07%	0.008 (1.119) ▼-99.73%	1.625 (1.208) ▼-51.23%	0.816 (0.860) ▼-68.64%	1.903*** (0.469) ▼-34.51%	1.374*** (0.417) ▼-41.08%
Facilitator	0.003 (0.547)	0.002 (0.526)	0.426 (0.308)	0.566** (0.281)	-0.092 (0.136)	-0.042 (0.143)
Policy liberality index (5-year MA)	-0.019 (1.555)	0.631 (1.201)	0.238 (1.475)	0.881 (1.232)	-0.789 (1.479)	0.014 (1.197)
Asylum applications	0.708*** (0.193)	0.712*** (0.193)	0.609*** (0.200)	0.610*** (0.201)	0.729*** (0.234)	0.725*** (0.234)
Arrival by prima facie	3.194*** (0.398)	3.200*** (0.397)	3.163*** (0.367)	3.170*** (0.368)	3.111*** (0.362)	3.114*** (0.361)
Constant	-8.066 (16.693)	-5.120 (16.516)	-8.511 (16.612)	-8.357 (15.947)	-11.963 (16.091)	-9.759 (15.950)
Observations	113,252	113,252	113,703	113,703	113,703	113,703
Pseudo-R2	0.880	0.880	0.890	0.890	0.892	0.892
AIC	178626	178483	193129	192942	189283	189394
Summary index weighting	ICW	EW	ICW	EW	ICW	EW

*** p<0.01, ** p<0.05, * p<0.1

Robust, dyad-clustered standard errors are in parentheses. In each column, the header denotes the respective facilitator variable. The models include origin and destination fixed effects, year fixed effects, and controls for region, intercapital distance, contiguity, shared language, bilateral stock of nonforced migrants, GDP/capita ratio and its squared term, population, civil war occurrence, and repression at origin and destination. **In contrast to Blair, Grossman and Weinstein (2022), I control for whether the displacement data comes from asylum application and/or prima facie recognition.**

The policy summary indices are constructed using inverse covariance weighting (ICW) or equal weighting (EW).

more liberal regimes increase refugees. For two of the three facilitators in the original study, I do not find the postulated link. Regarding the two binary variables I have included in the models, one can see that refugee arrival in *prima facie* procedures increases the recorded rate much more than individual asylum procedures. The table also reports in red the percentage decrease in the effect size of the interaction when I include my two binary variables in the models compared to the original models. I find that the effect size for the ICT facilitator is reduced by 99.73 to 102.07% and in one model even flips the sign. The effect size for the number of ethnic kinship groups is only reduced by around 34.51 to 41.08%.

Next, I analyse whether I find a robust relationship between the facilitators and policy liberality if I do not focus on a combination of refugee numbers but if I only focus on asylum applications. Following Schwartz (2022), I assume that accepting refugees through *prima facie* mechanisms often occurs because governments do not have much choice in emergency situation, because *prima facie* recognition offers only a minimum level of protection compared to the protection outlined in the Refugee Convention and the recognition is often temporally precarious as hosting governments 'voluntary' return refugees. I hence make the assumption that if there is a link between policy liberality and increased refugee movements towards liberal states, this relationship should be much stronger regarding formally seeking asylum than

regarding *prima facie* recognition. If potential refugees hear information, for example by their ethnic kin, about liberal policies towards displaced persons in a specific country, I expect that this will increase individual asylum applications much more than *prima facie* recognition because the latter is more ad hoc and not necessarily provides refugees with a secure status.

In Figure 1, I hence report the average marginal effect of an increase in the facilitator variables at different levels of liberality on refugee rates for asylum applications only. I do not change the model by Blair, Grossman and Weinstein (2022) or include additional variables. I only limit the dependent variable to data from one source: asylum applications. One can see in this replication that only in the case of the number of kinship groups one can observe an effect of policy liberality on refugee numbers.³

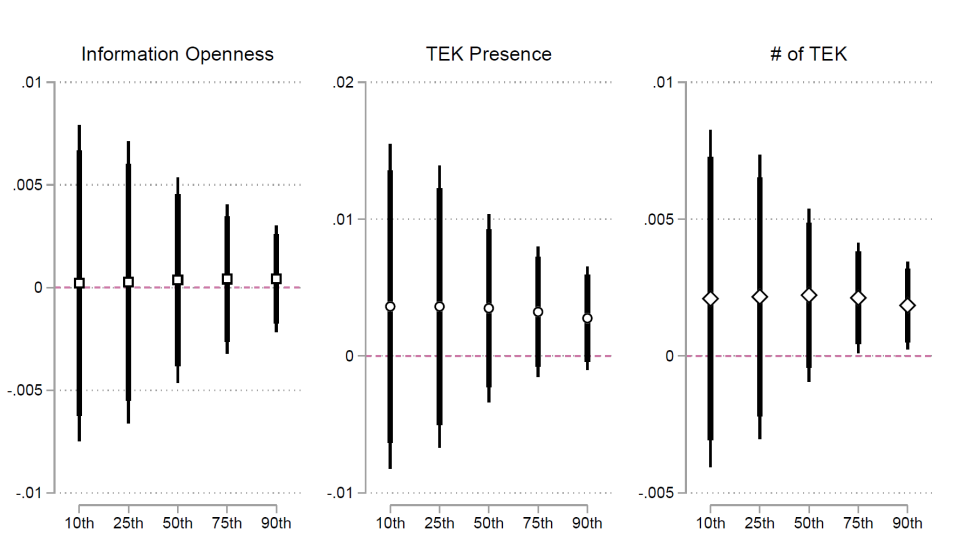


Figure 1: The conditional effect of displacement policy liberality on FDP flows. The plot shows the average marginal effect of the respective facilitator (information openness, presence of TEK and number of TEK) at different levels of displacement policy liberality. In contrast to Blair, Grossman and Weinstein (2022), the dependent variable is only refugee flows from refugee applications only. The 90% and a 95% confidence interval are marked.

Overall, the replication suggests that when I control for whether these refugees were recorded *prima facie* or through asylum applications, I still find that more kinship ties and a liberal displacement policy are somewhat associated with increased refugee numbers but the results are substantially weaker and inconsistent. This is an important implication. With its direct title, this letter in the *American Political Science Review* claims a clear connection between liberal displacement policies in the Global South and higher refugee attractiveness. Although many of the robustness tests are very convincing, this finding is easily politicised: If liberal displacement policies attract refugees, one can imagine many policy mak-

³If I focus on *prima facie* recognitions only in the same models, they do not converge.

ers claiming the need for even more restrictive policies towards people on the move. Given that a lot of countries recognise refugees in the Global South through *prima facie* recognition because they face emergency scenarios with low-capacity states and high refugee movements, my replication shows that there are some doubts whether we can compare those refugee rates with those from more stable and formalised systems through individual asylum applications that are also present in some of the sampled countries. It is important to understand this criticism in the context of other researchers that illustrate that refugee movements and asylum systems - and in particular the strategic manipulation of de facto policies towards refugees - vary profoundly in the Global South and make simple conclusions difficult (Schwartz, 2022).

Conclusion: why should we particularly care about forced migration data?

This letter discusses accuracy challenges in forced displacement data due to definitional, operational and usability challenges. I also discuss the politics of producing forced displacement data, highlighting sources of political bias through political, institutional and accountability problems. For these reasons, studies using forced displacement data in a comparative way should make sure that their use of data provides robust evidence for policymakers.

This is particularly important because forcibly displaced populations are a vulnerable population group. The quantification and tracking of such groups involves key ethical questions (Breslin, Shareck and Fuller, 2019; Taylor, 2016). Common under- and overcounting of vulnerable displaced populations is not a mere problem for statisticians - it is relevant to academics because it is a politicised matter (Harrell-Bond, Voutira and Leopold, 1992). Researchers that aim to shape policies towards migrants and refugees face the responsibility to be particularly sensitive towards potential biases that undermine the robustness of conclusions.

It is hence crucial that authors and reviewers in political science discuss sources of displacement data, in particular in global over-time comparisons and localised analysis. Demonstrating awareness for the data-generating process (e.g. changes in data collection strategies over time, comparability across units, etc.) seems crucial. Authors and reviewers should also enter a more serious dialogue about the ethical implication of using forced displacement data, in particular in fine-grained and dis-aggregated studies.

References

- Balcells, Laia and Abbey Steele. 2016. "Warfare, political identities, and displacement in Spain and Colombia." *Political Geography* 51:15–29.
- Blair, Christopher W, Guy Grossman and Jeremy M Weinstein. 2022. "Liberal Displacement Policies Attract Forced Migrants in the Global South." *American Political Science Review* 116(1):351–358.
- Blitz, Brad K. 2021. "Protection through Revisionism?".
- Braithwaite, Alex, Idean Salehyan and Burcu Savun. 2019. "Refugees, forced migration, and conflict: Introduction to the special issue".
- Breslin, Samantha, Martine Shareck and Daniel Fuller. 2019. "Research ethics for mobile sensing device use by vulnerable populations." *Social Science & Medicine* 232:50–57.
- Cardona-Fox, Gabriel. 2020. "The politics of IDP data." *Refugee Survey Quarterly* 39(4):620–633.
- Cecchi, Francesco and Emilie Sabine Koum Besson. 2022. "Reconstructing subdistrict-level population denominators in Yemen after six years of armed conflict and forced displacement." *Journal of Migration and Health* 5:100105.
- Christensen, Hanne. 1982. "Survival strategies for and by camp refugees : report on a six-week exploratory sociological field study into the food situation of refugees in camps in Somalia."
URL: <https://digitallibrary.un.org/record/47217?ln=en>
- Crisp, Jeff. 1999. *Who has Counted the Refugees?: UNHCR and the Politics of Numbers*. Centre for Documentation and Research, United Nations High Commissioner for Refugees.
URL: <https://www.unhcr.org/research/working/3ae6aoc22/counted-refugees-unhcr-politics-numbers-jeff-crisp.html>
- Devictor, Xavier, Quy-Toan Do and Andrei A Levchenko. 2021. "The globalization of refugee flows." *Journal of Development Economics* 150.
- Diggle, Emma, Wilhelmina Welsch, Richard Sullivan, Gerbrand Alkema, Abdihamid Warsame, Mais Wafai, Mohammed Jasem, Abdulkarim Ekzayez, Rachael Cummings and Preeti Patel. 2017. "The role of public health information in assistance to populations living in opposition and contested areas of Syria, 2012–2014." *Conflict and health* 11(1):1–12.

Dionigi, Filippo and Domenico Tabasso. 2020. "Academic Trends in Forced Displacement."

URL: https://www.unhcr.org/people-forced-to-flee-book/wp-content/uploads/sites/137/2021/10/Filippo-Dionigi-and-Domenico-Tabasso_Academic-Trends-in-Forced-Displacement.pdf

EGRIS. 2018. "International recommendations on refugee statistics."

Engel, Stefanie and Ana María Ibáñez. 2007. "Displacement due to violence in Colombia: A household-level analysis." *Economic Development and Cultural Change* 55(2):335–365.

EU and EGRIS. 2020. "International recommendations on IDP statistics."

Fearon, James D, Andrew Shaver et al. 2021. Civil war violence and refugee outflows. Technical report Empirical Studies of Conflict Project.

Harrell-Bond, Barbara, Eftihia Voutira and Mark Leopold. 1992. "Counting the refugees: gifts, givers, patrons and clients." *Journal of Refugee Studies* 5(3-4):205–225.

Human Rights Watch. 2021. "UN shared Rohingya data without informed consent: Bangladesh provided Myanmar information that refugee agency collected."

URL: <https://www.hrw.org/news/2021/06/15/un-shared-rohingya-data-without-informed-consent>

Ibanez, Ana Maria. 2009. "Forced displacement in Colombia: Magnitude and causes." *The Economics of Peace and Security Journal* 4(1).

Ibáñez, Ana María and Carlos Eduardo Vélez. 2008. "Civil Conflict and Forced Migration: The Micro Determinants and Welfare Losses of Displacement in Colombia." *World Development* 36(4):659–676.

Jacobsen, Katja Lindskov and Kristin Bergtora Sandvik. 2018. "UNHCR and the pursuit of international protection: accountability through technology?" *Third World Quarterly* 39(8):1508–1524.

Kibreab, Gaim. 2004. "Pulling the wool over the eyes of the strangers: refugee deceit and trickery in institutionalized settings." *Journal of Refugee Studies* 17(1):1–26.

Kibreab, Gaim et al. 1985. *African refugees: reflections on the African refugee problem*. Africa World Press.

Ratnayake, Ruwan, Nada Abdelmagid and Claire Dooley. 2022. "What we do know (and could know) about estimating population sizes of internally displaced people." *Journal of Migration and Health* p. 100120.

Saldarriaga, Juan Francisco and Yuan Hua. 2019. "A gravity model analysis of forced displacement in Colombia." *Cities* 95:102407.

Schwartz, Stephanie. 2022. "Refugee return without refoulement." *Working Paper* .

Shaver, Andrew, Benjamin Krick, Judy Blancaflor, Sarah Yein Ku, Xavier Liu et al. 2022. The Causes and Consequences of Refugee Flows: A Contemporary Re-Analysis. Technical report Empirical Studies of Conflict Project.

Taylor, Linnet. 2016. "No place to hide? The ethics and analytics of tracking mobility using mobile phone data." *Environment and Planning D: Society and Space* 34(2):319–336.

Turkoglu, Oguzhan. 2022. "Look who perpetrates violence and where: Explaining variation in forced migration." *Political Geography* 94.

Ustek-Spilda, Funda. 2020. "Statisticians as back-office policy-makers: Counting asylum-seekers and refugees in Europe." *Science, technology, & human values* 45(2):289–316.

Zhou, Yang-Yang and Andrew Shaver. 2021. "Reexamining the effect of refugees on civil conflict: a global subnational analysis." *American Political Science Review* 115(4):1175–1196.

Appendix

Table 2: Replication of table 1 in Blair, Grossman and Weinstein (2022): Effect of policy liberality and a facilitator on refugee numbers. Full replication with the available replication material (two variables are missing)

	ICT		TEK presence		TEK numbers	
	(1)	(2)	(3)	(4)	(5)	(6)
Policy liberality index (5-year MA) x facilitator	3.282* (1.892)	2.979** (1.488)	3.332** (1.376)	2.602*** (0.933)	2.906*** (0.547)	2.332*** (0.454)
Facilitator	-1.123 (1.043)	-1.232 (1.035)	-0.043 (0.350)	-0.023 (0.310)	-0.376** (0.148)	-0.371** (0.148)
Policy liberality index (5-year MA)	-1.979 (1.377)	-0.753 (1.208)	-1.577 (1.224)	-0.790 (1.123)	-2.736** (1.204)	-1.629 (1.072)
Constant	-19.182 (25.301)	-16.366 (25.400)	-17.739 (22.393)	-16.416 (21.580)	-23.257 (21.313)	-19.755 (21.374)
Observations	113,252	113,252	113,703	113,703	113,703	113,703
Pseudo-R2	0.818	0.818	0.834	0.834	0.838	0.837
AIC	270462	270527	290162	289815	284388	284583
Summary index weighting	ICW	EW	ICW	EW	ICW	EW

*** p<0.01, ** p<0.05, * p<0.1

Robust, dyad-clustered standard errors are in parentheses. In each column, the header denotes the respective facilitator variable. The models include origin and destination fixed effects, year fixed effects, and controls for region, intercapital distance, contiguity, shared language, bilateral stock of nonforced migrants, GDP/capita ratio and its squared term, population, civil war occurrence, and repression at origin and destination. The policy summary indices are constructed using inverse covariance weighting (ICW) or equal weighting (EW).