# Who Reviews Whom, Where and Why? Evidence from the Peer Review Process in the OECD Development Assistance Committee

Alice Iannantuoni* Simone Dietrich† Bernhard Reinsberg‡

June 2022§

**Abstract**

The study of international organizations' peer review systems has focused largely on their efficacy in disseminating best practices, with mixed results. This paper informs the debate from a new angle: we evaluate the extent to which decisions about who reviews whom and where result from bureaucratic guidelines; or whether these decisions are shaped by the particularistic interests of member states, thus affecting efficacy evaluations. Our empirical case is the long-standing OECD Development Assistance Committee, where foreign aid donors have their practices reviewed by two peer examiners every few years. Using quantitative and qualitative methods, we study (i) the assignment of peer examiners (1962–2020), and (ii) the selection of recipient countries visited for assessment during the review (1996–2020). Our analyses show that the choice of peer examiners is driven by the IO's bureaucratic process, while the preferences of reviewed donors shape the selection of recipient countries for field visits.

*Department of Political Science and International Relations, University of Geneva, alice.iannantuoni@unige.ch.

†Department of Political Science and International Relations, University of Geneva, simone.dietrich@unige.ch.

‡School of Social and Political Sciences, University of Glasgow, bernhard.reinsberg@glasgow.ac.uk.

§Draft prepared for the PEIO 2022 annual meeting. Please do not circulate without permission. The authors thank Katharina Fleiner and Tomas Spahini for research assistance and Çağlayan Başer, Ekrem Başer, Nuole Chen, Daniel Nielson, and participants of the ECPR Joint Sessions Workshop "The Performance of Global Governance Institutions" for helpful comments on a previous draft.

# 1 Introduction

Increasing interdependence has led to the creation of different types of international regimes to solve transnational problems and promote compliance with agreements among states. One prominent regime is the peer review system hosted by international organizations (IOs). Peer reviews can be found across issue areas as diverse as environment, finance, development, security, and human rights. While they can take different forms, such reviews always entail the evaluation of practice and performance of a state by other states. Contrary to coercive modes of governance, peer reviews are "soft governance" instruments that seek to induce states to comply with the standards touted by IOs (e.g., Pagani 2002; Liverani and Lundgren 2007; Groenendijk 2012; Lehtonen 2007; Carraro, Conzelmann, and Jongen 2019).

The existing literature on peer reviews in world politics has focused on their effectiveness, suggesting that they can influence state behavior, through a number of possible mechanisms. Some have argued that peer reviews affect change through policy learning (Lehtonen 2007; Mahon and McBride 2009; Marcussen 2003). Others suggest that reviews incentivize compliance by informing an international audience about the extent to which states conform with agreed-upon standards (Terman and Voeten 2018). Negative assessments can hurt the reputation of states and this, in turn, mobilizes domestic actors to pressure governments to change their ways (De Ruiter 2013; Meyer 2004; Carraro 2019). Others again understand states as social creatures that care about the approval of their peers, and thus work to receive positive reviews and avoid reputational damage (Carraro, Conzelmann, and Jongen 2019). Akin to global performance indicators, peer reviews establish a framework for comparability among states that reveal shortcomings and thus serve as instruments of social pressure: because laggards of performance can easily be identified, these rankings induce states to improve their practices (e.g., Kelley and Simmons 2015; Koliev, Sommerer, Tallberg 2020).

If negative assessments can be costly for states, it is plausible that states might act to influence aspects of the peer review process that could increase the chances of a positive assessment. Indeed, recent work by Terman and Voeten (2018) shows that strategic relations among states affect peer review assessments in the United Nations Universal Periodic Review: examiner countries that evaluate partners tend to be less critical in evaluating their human rights records than they are vis-à-vis countries with whom they do not share strategic relations. Such a pattern provides evidence of politics infiltrating the peer review process and raises the stakes of assignment decisions about who gets reviewed by whom, how, and where—a decision which, thus far, scholars have implicitly assumed to be largely process-driven and impartial. If states themselves were indeed able to influence assignment decisions, then politics might enter the peer review process well before more or less favorable reviews are written. Such a possibility would have implications for inferences that scholars draw when assessing state compliance with peer review recommendations.

In this paper, we assess the extent to which review assignment decisions are indeed impartial and independent, or whether they are politicized. Although peer review tools are often portrayed as examples of relational governance, with states evaluating other states, their credibility rests on the consistent deployment of a comparable methodology (Jongen 2018). Typically, this includes explicit IO guidance on how to conduct the peer review process, including self-assessments, reporting, and publication requirements. While member states need to broadly agree on the mechanics of such a process, it is in the purview of IO secretariats to devise standard procedures and subsequently ensure that they are followed. Conversely, the absence or inconsistent implementation of bureaucratic procedures invites political biases that, if present, should be accounted for in research on peer review efficacy. For example, if states were to advocate with IO staff to be paired with peers with whom they share strategic relations and which, according to Terman and Voeten (2018), may be expected to provide less critical assessments, then compliance records would not be easily comparable across countries. The politicization of peer review assignments might also undermine the credibility of the peer review process, making it less likely that states remain open to learning from each other and comply with suggested behavior changes (Carraro and Jongen 2018; Carraro 2019).

We tackle this question by using the Organization for Economic Cooperation (OECD) Donor Assistance Committee's (DAC) peer review system as our empirical case. We consider the OECD DAC a relevant laboratory for three reasons: First, the DAC has the longest-standing system of peer evaluation within the OECD, with a total of over 500 reviews between 1962 and 2020. Second, while previous research has recognized the role of the DAC peer review in advancing best practices among donor countries in international development cooperation (Ashoff 2013; Eyben 2013; Paulo and Reisen 2010; Liverani and Lundgren 2007), little attention has been paid to how this system works and how it shapes states' aid policies and practices. Third, the foreign aid literature has produced robust evidence documenting that donor strategic interests matters in foreign aid decision-making (e.g., Alesina and Dollar 2000; Bermeo 2017; Berthélemy 2006; Heinrich 2013; Schraeder, Hook, and Taylor 1998), and we believe that strategic thinking may also infiltrate the foreign aid peer review process, Further, there is a mounting body of evidence suggesting that donor domestic politics matters. For example, Dietrich (2021) shows that variation in domestic ideological structures of DAC donors not only yields divergent preferences about how donors seek to promote development about. It also produces different preferences about which standards and best practices to prioritize and champion in the OECD DAC review process. If states can expect more favorable reviews from peers that are strategically or ideologically aligned with them, they might lobby the Secretariat to deviate from the protocol and grant them conditions that make a more favorable review outcome more likely.[1]

[1]For research that documents collusion between IO secretariats and states see for example Dijkstra (2017)

Drawing on author interviews with officials from the OECD Secretariat and designated DAC peer review officials as well as archival research, we develop a theoretical framework that focuses on the role of the IO secretariats as stewards of the peer review methodology. Given the rational-legal authority, IO secretariats are in a unique position to assume this role (Barnett and Finnemore 2004). Designated secretariats have analytical expertise and are expected to develop and monitor comparative peer evaluation frameworks that assess how member states compare against the IO's agreed-upon standards and practices. At the same time, secretariats are responsible for developing and implementing bureaucratic protocols that provide guidance about how the peer review process is implemented including rules about who gets reviewed by whom, how, when, and where. As we will elaborate in more detail below, DAC peer review assignment protocols are expected to maximize efficiency and learning including, for example, that states be matched with peers of similar size, language, geography, or that they not review the same member twice in a row (OECD 2020).

We develop a bureaucratic process model that features mandate-driven criteria and that, if implemented correctly and consistently, should ensure that assignment decisions are impartial and independent of state interests. We then augment the bureaucratic process model with factors that reflect particularistic donor interests and that could be expected to affect assignment decisions made by the OECD Secretariat. In a first step, we use this approach to explain decisions about who reviews whom, i.e., which two DAC member states will serve as examiners for a particular review. In a second step, we explain the choice of recipient country where field visits take place that allow peers to examine the reviewed donor's aid practices in in-depth.

We empirically assess this bureaucratic process model on originally collected data that exploit information on examiner pairings from 516 OECD DAC peer reviews from 1962 to 2020 as well as originally collected data on field visits described in review documents from 1996 onwards. Our evidence suggests that the two assignment decisions under study are largely impartial: states are not more likely to be examined by peers with whom they are strategically or ideologically aligned. Instead, the choice of "who reviews whom" seems driven by the procedural criteria laid out in the guidelines. The results on which recipients are visited also dovetail with the notion that the review process largely follows the IO's rule book. However, we find some evidence that donor strategic interest play a role in decisions about where to assess donor performance. By allowing input from reviewed donors on the choice of where to evaluate donor performance, the secretariat does respond to particularistic requests from states, while broadly implementing a bureaucratic protocol that ensure decisions about where conduct field visits remain largely impartial and independent Taken together, our findings lend support to a bureaucratic process explanation, suggesting that the OECD Secretariat largely manages to keep key components of the aid peer review architecture from getting politicized.

By shedding light on what drives peer assignment decisions, our paper makes contributions to a burgeoning literature on peer reviews. First, this paper is the first to systematically examine decisions about who

gets to assess whom, when, and where in peer review. Second, although our paper does not examine state compliance with review recommendations, it informs the debate indirectly. Our findings suggest that the OECD Secretariat largely succeeds in keeping the choice of peer examiners impartial and free of politics. We do find evidence that reviewed donors exert some influence on where the evidence on peers gets collected, and argue that this knowledge should be factored in by scholars when designing research on the effectiveness of the DAC peer review. Third, from an empirical perspective, our study casts light on the relatively understudied DAC peer review system. With foreign aid from DAC donors reaching about 180 billion US dollars annually, this lack of attention is surprising. That said, we would welcome future research on the extent to which our findings about review assignment decisions hold across a broader range of peer review systems in other areas like human rights, trade, or the environment.

Further, our study contributes to the literature in international political economy that examines the degree to which international bureaucracies are independent as opposed to influenced by member state pressure (Dijkstra 2017; Dreher, Sturm, and Vreeland 2009; Lang and Presbitero 2019; Stone 2011). While input from reviewed donors plays a role in decisions about where evidence on gets collected, our findings on the determinants of peer examiners suggest that, at least in the OECD DAC, the Secretariat fulfills its role as stewards of bureaucratic protocol and manages to insulate the review process from political interference. With its focus on the politics of peer review in foreign aid, our study extends the empirical scope of this line of research which, to date, has largely focused on aid implementation decisions.

We also contribute to the literature on foreign aid. By focusing on institutional processes and the politics of the DAC peer review system, we complement empirical analyses of donor behaviors that take these norms as given and that scrutinize donor practices such as foreign aid tying, aid fragmentation, and aid quality (Ganga and Girod 2021; Knack and Eubank 2011; Palagashvili and Williamson 2021; Steinwand 2015). Our results about donor influence in assignment decisions about where evidence gets collected is consistent with findings in the aid allocation literature that document, for example, that donors favor former colonies (Alesina and Dollar 2000). Finally, our findings contribute to studies on foreign aid effectiveness (Bearce and Tirone 2010; Doucouliagos Paldam 2009; Dreher, Minasyan, and Nunnenkamp 2015). If peer reviews were biased because of politically-motivated choices on who reviews whom or on where evidence gets collected, they would produce less accurate assessments and, in turn, be a less useful tool for monitoring and encouraging best practices in aid giving. With its focus on the DAC peer review, our study thus provides a segue into aid effectiveness research, emphasizing the possibility of inappropriate policy recommendations (in addition to selective implementation of peer recommendations) as a potential source of aid ineffective aid giving.

The rest of the paper is organized as follows: we first present background information on the OECD DAC and its peer review system. We then develop theoretical expectations for the bureaucratic process

model. We present a first set of empirical analyses on explanatory factors related to the assignment of peer examiners from 1962 to 2020, finding that the bureaucratic process model receives more empirical support than the possibility of donor interest politicizing examiner assignment. The second set of empirical analyses examines the selection of recipient countries for in-depth field visits when the review process is ongoing for reviews from 1996 to 2020. We find that, while still largely within the bureaucracy-driven guidelines for the peer review process, donor interest plays a role in this decision. We conclude with a discussion of our findings and point to future research.

## 2 The Peer Review System in the OECD DAC

The Organisation for Economic Co-operation and Development's (OECD) Development Assistance Committee (DAC) is a forum in which member states come together to "promote coordinated, innovative international action to accelerate progress towards the Sustainable Development Goals (SDGs) in developing countries and improve their financing" (OECD 2022). Originally established in 1960 as the Development Assistance Group (DAG) within the Organisation for European Economic Co-operation (OEEC), the predecessor of OECD, the DAC today includes 29 member countries[2]. Officially classified as partner, the European Union acts as full member alongside seven "participant" countries;[3] and six "observer" organizations.[4] The DAC is the premier IO where traditional foreign aid donors come together to discuss and harmonize their development and cooperation efforts. It is a consensus-based organisation that works to accomplish three objectives: (i) to set standards for foreign aid giving, (ii) to facilitate learning among its members as well as observers, and (iii) to collect and publish data on development cooperation. It also produces soft law instruments, including official "Recommendations," "Declarations," and donor peer reviews, that seek to enshrine standards for best practices in aid giving, to spread these best practices, and to facilitate their "enforcement" across its member states.

Since the early 1960s, peer reviews of DAC members have taken place in regular intervals. While in annual intervals in the early years, reviews were held every two to three years throughout the 1980s. Since the 1990s, reviews of DAC members take place every four to six years. For each peer review, two DAC member states (referred to as "examiners" throughout the paper) serve as peers that examine national aid bureaucracies and practices of the state under review (referred to as the "reviewed donor"). The Secretariat organizes and facilitates the peer review process in different ways. Notably, the Secretariat proposes schedules that

[2]Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, the Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, United Kingdom, and United States.

[3]Azerbaijan, Bulgaria, Kuwait, Qatar, Romania, Saudi Arabia, and United Arab Emirates.

[4]Asian Development Bank, African Development Bank, Inter-American Development Bank, International Monetary Fund, United Nations Development Programme, and World Bank.

determine who gets reviewed by whom, when and where. Because the DAC is a consensus-based organization, the review schedules, who reviews whom, when, and where are subject to discussions with member state delegates to the DAC. These delegates can object to the Secretariat's process proposals and demand changes. It is during these discussions where states may attempt to influence assignment decisions, rendering them less impartial and more likely to further their particularistic interests.

Once initiated, the peer review process typically lasts several months, beginning with the reviewed donor producing a Memorandum of Understanding (MoU) where it presents the state of its development cooperation efforts to the Secretariat and the organization's staff in charge of the review process. The peer review team then travels to the reviewed donor's capital to visit relevant development actors, as well as to one or more of the donor's recipient countries to closely examine aid activities and practices in the field.[5] The assessments conducted in the field, both in the reviewed donor's capital and in recipient countries, are an integral part of the review process. Peer examiners pass on their notes from these assessments to the Secretariat and DAC staff who are in charge of drafting the review report including feedback from examiners as necessary.

Once drafted, the review report is presented and discussed in a formal DAC meeting. Importantly, the report includes a section titled "DAC's Main Findings and Recommendations," or "Summary and Conclusions" in older documents, where an explicit set of recommendations are outlined for how the reviewed donors should proceed to improve its aid giving practices. In earlier peer reviews, recommendations used to be fewer in number and less detailed in terms of precise metrics and expectations. Over time, they have become more specific and thorough. During the formal peer review meeting, the reviewed donor has the chance to question and possibly even request changes to the report document before a final draft gets distributed to all member states and made public. In this paper, we focus on two key aspects of the peer review mechanism: the choice of examiner countries and the choice of the aid recipient(s) to feature in an in-depth assessment. We now turn to developing a theoretical framework that explains these two decisions.

# 3 A Process-Driven Model of Peer Review Assignment

What determines which OECD DAC member states will serve as peer examiners? And, how are one or more of the reviewed donor's field operations chosen for field visits and special examination? It may be that assignment decisions may be the result of institutional protocols implemented by the Secretariat that maximize the efficiency and credibility of the process. Or, it may be that assignment decisions are politicized insofar as particularistic interests of member states affect assignment.

[5] On occasion, the reviewed donor's field operations are examined through meetings or calls with relevant stakeholders, without a physical visit to the recipient country.

Our in-depth study of the OECD DAC peer review process reveals that review assignment decision-making, although in the purview of the Secretariat, contains opportunities for member states to influence assignment choices. As designated stewards of peer review methodology, the Secretariat is responsible for developing and implementing institutional procedures that provide guidance for the implementation of peer reviews including guidelines that determine who gets reviewed by whom, how, when, and where. Insights from interviews with secretariat staff reveal that these guidelines are set up to achieve the following objectives: they seek to facilitate the practical business of conducting reviews and they ensure that the review process remains impartial and credible for all members of the organization.

According to interviewed staff of the Secretariat and archival research, several factors matter for review efficiency and credibility and thus become part of our process-driven model of peer review assignment. When assigning examiners, the Secretariat should select examiners that are not currently engaged in another review. The peer examiner should be able to conduct the review efficiently (e.g., newer DAC members might serve as examiners alongside more experienced members or reviews should be conducted by peers that share the same language). And, there should not be an obvious conflict of interest (e.g., EU member states may not serve as examiners for a peer review of EU's aid giving) or repeated pairings of the same examiners for a given donor, among other factors.

However, because the OECD DAC is a consensus-based organization, the Secretariat's guidelines are not only subject to approval of member states. They also undergo regular and extensive consultations with member states, providing members with opportunities to react to the Secretariat's assignment proposals. It is during these consultations when we would expect donor interests to influence assignment; or when the reviewed donor might ask to be paired with peers that are ideologically or strategically aligned with them. As Terman and Voeten (2018) have shown, politically motivated pairings produce biased reviews and recommendations insofar as examiners have incentives to portray an aligned peer's aid practices in a better light, or produce recommendations that are easy to implement or that might overlook malpractice.

Concerning the choice of where to conduct field visits to assess the reviewed donor's aid program, we also learn that efficiency and credibility matter. According to official OECD peer review methodology,[6] the Secretariat solicits direct input from donors at this stage: i.e. the reviewed member is asked to provide suggestions of possible recipients to consider for field visits to the review team. The guideline states that, "in the interests of transparency and objectivity," members are asked to provide multiple options[7] that satisfy a set of process-oriented criteria, which the review team will also use in making the final selection:

[6] See for example document DCD-DAC(2010)19-FINAL dated March 20, 2017, and titled, "Guidance for Selecting Partner Countries to Visit as Part of the Peer Review Process."

[7] "The review team should be offered a choice of three countries in cases where there is only one visit to a partner country and a choice of six countries in cases where the Peer Review includes two visits."

> "In selecting options for partner country visits, the reviewed member should consider the following criteria. (i) The country should be a significant partner for the reviewed member in terms of relationship, and level and scope of engagement; (ii) the country should be representative of the reviewed member's programme for accountability purposes; (iii) it should provide an opportunity for the review team to reflect on the reviewed member's application of its development co-operation policies and operational guidance; (iv) the programme should also provide for other learning opportunities (such as those related to global issues, exit strategies and aid management issues); and (v) the country should not have been visited by a peer review team in the same or the previous year."

Consultations and exchanges between the Secretariat and member states on recipient country field visits allows members to influence this choice: the reviewed donor is explicitly asked to name possible recipients to the Secretariat; or to challenge Secretariat decisions after an initial decision about field visits is made. In both scenarios it is possible that the reviewed donor, who we expect to have incentives to show its aid program at its best, will advocate for aid recipients who are especially important to the donor, such as those receiving a large proportion of the donor's aid budget, or those who used to be former colonies of the donor.

Drawing on these insights we develop two parsimonious bureaucratic process models that explain decisions about who reviews whom when and where. These models feature mandate-driven criteria that are referenced by archival documents and in author interviews with staff from the DAC and the Secretariat. We then augment these bureaucratic process models with factors that reflect particularistic donor interests and that we expect to influence assignment decisions during consultations between the Secretariat and delegates of member states.

In our empirical analyses, we interpret statistically significant associations between variables that capture the official peer review methodology and assignment choices as evidence that the review assignment is in line with the DAC's rules and guidelines. We interpret statistically significant associations between politically salient factors and assignment choices as evidence that the review process might be politicized by reviewed donors who are able to influence the Secretariat and obtain a favorable set up. If politically salient factors are not systematic predictors of either choice, we would interpret this pattern as indicative that the review process is process-driven and not politicized and that the peer review process remains Secretariat-driven throughout. We first analyze the determinants of peer examiners for the full universe of peer reviews including 516 reviews from 1962 to 2020. Second, we assess the choice of recipients who receive an in-depth assessment for the 116 recent reviews for which field visit information is available, from 1996 to 2020.[8]

[8]See the Appendix for descriptive information about the peer review data.

# 4 Determinants of Peer Examiner Choice

The first key decision in the DAC peer review process is to choose which member states will serve as peer examiners. In this section, we analyse how reviewed donors and examiners are paired to assess whether the assignments are more consistent with the idea of a standardized, bureaucracy-driven process or with the possibility that this decision is influenced by DAC member states' interests.

## 4.1 Determinants of Peer Examiner Choice: Data and Methods

We construct a panel data set covering the universe of 516 DAC peer reviews from 1962 to 2020, in which the unit of analysis is the pair of reviewed donor and peer examiners nested within each review. This approach ensures that we account for the potential interdependence of examiner choices in any given DAC review. We construct this data set in three steps. First, we identify the potential set of peer examiners—i.e., countries who are DAC members at the time the review is undertaken. Second, we compute all unique pairs of potential examiners. The set of potential examiners (and potential examiner pairs) varies over time as the organization's membership grows: in our sample, the number of potential examiners ranges from 10 countries in 1962 to 28 countries since 2016. Our panel data set is therefore unbalanced in the number of observations for any given review. Lastly, we add all relevant covariates, i.e., information on the reviewed donor, on each peer examiner, on the reviewed-donor-examiner-pair dyad, and the examiner dyad.

Our dependent variable is a binary indicator equal to one for those potential examiner pairs that actually served as examiners for a given review. We rely on linear probability models and include fixed effects at the review level, controlling for any idiosyncratic effects of each specific review. We are therefore left with examiner effects, reviewed-donor-examiner-pair effects, and examiner-pair effects to explain the choice of peer examiners.

Our key independent variables are organized into two sets and follow our main theoretical considerations for the choice of peer examiners. The first model is the "bureaucratic process" model, which captures the official criteria for examiner selection as developed in DAC guidance. Although this guidance has been updated over the years, long-standing criteria include that two DAC members act as peers for each review. These examiners are designated by the Secretariat and are matched, where possible, with "reviews covering a member of similar size and complexity, and language and geography" (OECD 2021). Furthermore, examiners should not serve as an examiner of the same donor twice in a row.

We operationalize these criteria at the reviewed-donor-examiner-pair level as follows. To account for the criteria related to "size" and "complexity" we construct an indicator of being a major donor—France, Germany, Japan, United Kingdom, and United States—and include two indicators for whether (i) the review

team includes at least one peer examiner in the same category as the reviewed donor, and (ii) both examiners are of the same type. We include a similar set of relational indicators for membership in the European Union (EU). These EU indicators are time-varying across reviews because countries joined the EU at different points in time. As further proxy for complexity, we add two relational indicators for whether at least one or both peer examiners have the same organizational model as the reviewed donor (DAC 2009). In addition, we include indicators for common language. Specifically, in line with DAC peer review guidance, we measure whether at least one or both examiners match on an indicator for French language.[9] Similarly, we capture whether at least one or both examiners are contiguous to the reviewed country. Finally, we include an indicator for whether the same examiner pair reviewed the same donor in its previous review; as well as three indicators capturing whether at least one examiner or both, individually and as a pair, have reviewed the same donor before.

The second model is where we capture the potential interest of the reviewed donor. As discussed, DAC members may have preferences for being reviewed by examiners who they are aligned with or from whom they might otherwise expect a more favorable review. We proxy donor interests through relational variables that have either been commonly used as proxies for donor interest in existing aid allocation literature or that we develop specifically in the context of the DAC peer reviews. Our first relational variable measures foreign policy alignment between the reviewed donor and the potential examiner pair. Consistent with numerous previous studies, we proxy foreign policy alignment as the ideal-point distance of any of the two donor-examiner pairs based on their voting patterns in the UN General Assembly (Bailey, Strezhnev, and Voeten 2017). If DAC members have a preference for reviewing countries that are closely (or not closely) aligned with their foreign policy and lobby the Secretariat to be paired with them, we would expect this measure to be a statistically significant predictor of which examiner pair is ultimately chosen. Since there are two examiners per review, we construct both the average and the range of the pairwise relational variable.

A second set of relational variables draws on similarities in donor political economies or institutional rule books and their underlying beliefs about the role of markets in public service delivery that have been shown to shape donors' aid delivery tactics (Dietrich 2021). On the one hand, Anglo-Saxon donor bureaucracies are pressured to demonstrate short-term results and therefore avoid engaging directly with poorly-governed recipient governments. On the other hand, donor bureaucracies in more statist economies are less constrained to engage with recipients in a long-term fashion. DAC members with similar political economy structures may be more ideologically aligned and therefore less likely to question each other's aid practices. We create two indicator variables which capture whether at least one or both donor-examiner pairs are of the same

[9]Reviews of francophone donors may be carried out in French, before being translated into English (OECD 2021).

political economy type.[10]

Turning to domestic politics, we posit that donors may prefer to be paired with examiners that have similar domestic orientations. For example, they may prefer to be reviewed by like-minded examiners, or "sympathetic interlocutors," in order to reduce the risk of unpredictable reviews. If DAC members have a systematic preference for reviewing and being reviewed by peers that are institutionally and ideologically similar, they could lobby the Secretariat to be paired with them. We consider two sets of domestic politics alignment variables. The first is the average difference of partisan ideology between donor-examiner pairs, as well as the range of these differences (Fuchs and Richert 2018). The second set contains the average and range over the differences in the index of deliberative democracy between donor-examiner pairs. The deliberative democracy index is a pertinent measure of the quality of democracy, which serves to address the limited variation of conventional democracy measures such as the polity index in the sample of advanced donor countries (Teorell et al. 2020).[11]

For our main analysis, we opt for parsimonious models that only include review fixed effects. This serves to capture any idiosyncratic effects of the review, the reviewed donor, and any year effects. In robustness tests, we remove fixed effects and instead estimate random-effects models, which assume a normally-distributed systematic error component across reviews and which allow us to include additional features of the review, such as the involvement of a third examiner and the number of available donors in the peer examiner pool. We cluster standard errors on reviews, considering the interdependent choices of examiner pairs within reviews. Unless otherwise stated, our sample excludes reviews in which the EU institutions are being reviewed, given that many covariates are undefined.

## 4.2 Determinants of Peer Examiner Choice: Results

Table 1 juxtaposes the two models that include variables that help explain the process of examiner assignment and the extent to which this process is driven by process-driven factors or donor interest. We proceed by first introducing each model separately, and then estimating a joint model.

In Table 1, Model 1, we probe the "process-driven" model. We find that DAC peer review pairs are chosen such that at least one examiner is comparable to the review donor in size, while it tends to be unlikely that both examiners are ($p<0.1$). Substantively, if at least one examiner within a potential examiner pair is of similar size to the reviewed donor, it is over 55% more likely for that pair to be chosen, compared to pairs where both examiners are of different size than the donor. Similarly, if one of the examiners shares the language of the reviewed donor, that potential examiner pair is over 17% more likely to be chosen. We do

[10]Following Dietrich (2021), the list of liberal market economies includes Australia, Canada, Ireland, New Zealand, the United Kingdom, and the United States from 1985–2020, and Denmark, Finland, Norway, and Sweden from 1995–2020.

[11]Table A1 in the Appendix presents descriptive statistics of all variables in the analysis.

not find significant effects for common membership in the EU, nor with respect to the independence of the aid agency. Geographic proximity is also not statistically significant.

**Table 1:** Determinants of Peer Examiner Choice

| | (1) | | (2) | | (3) | |
|---|---|---|---|---|---|---|
| At least one similar size | 0.555*** | (0.111) | | | 0.645*** | (0.133) |
| Both similar size | -0.126° | (0.066) | | | -0.154 | (0.106) |
| At least one similar EU | 0.072 | (0.069) | | | 0.086 | (0.089) |
| Both similar EU | -0.065 | (0.066) | | | -0.123 | (0.082) |
| At least one common language | 0.177* | (0.089) | | | 0.203° | (0.114) |
| Both common language | -0.090 | (0.062) | | | -0.080 | (0.081) |
| At least one contiguous | -0.014 | (0.076) | | | -0.123 | (0.093) |
| Both contiguous | -0.128 | (0.177) | | | -0.018 | (0.246) |
| At least one same model | -0.011 | (0.056) | | | -0.036 | (0.073) |
| Both same model | 0.021 | (0.096) | | | 0.091 | (0.126) |
| Review twice in a row | -0.075 | (0.076) | | | -0.105 | (0.096) |
| Both before as pair | 0.143 | (0.134) | | | 0.101 | (0.162) |
| At least one before | -0.005 | (0.085) | | | -0.032 | (0.116) |
| Both before | 0.119° | (0.065) | | | 0.157° | (0.086) |
| Average UNGA distance | | | -0.206° | (0.115) | -0.073 | (0.136) |
| Dispersion of UNGA distances | | | -0.078 | (0.084) | -0.137 | (0.091) |
| At least one similar CPE | | | -0.011 | (0.111) | 0.031 | (0.111) |
| Both similar CPE | | | -0.086 | (0.096) | -0.081 | (0.100) |
| Average partisan distance | | | 0.090 | (0.140) | 0.091 | (0.139) |
| Dispersion of partisan distances | | | 0.011 | (0.114) | 0.030 | (0.114) |
| Average difference in democracy | | | -0.101 | (0.563) | -0.042 | (0.571) |
| Dispersion democracy differences | | | -0.870* | (0.376) | -0.721° | (0.384) |
| Observations | 83646 | | 54425 | | 54425 | |
| Within R2 | 0.001 | | 0.000 | | 0.001 | |

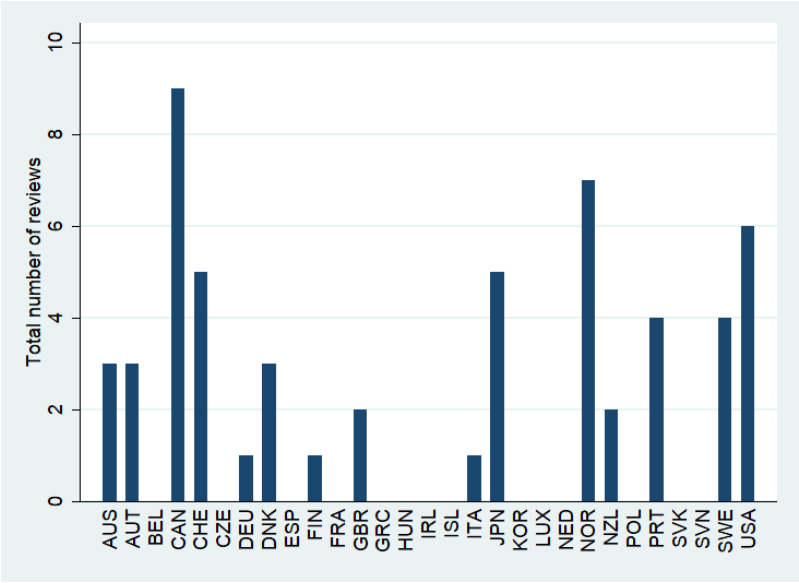
Linear regressions with review fixed effects and standard errors clustered on reviews in parentheses.
Significance levels: ° <.1 * <.05 ** <.01 *** <.001

Model 2 tests the "donor interest" model. Across a battery of operationalizations, we do not find much support for the proposition that reviewed donors get disproportionately paired with examiners who are similar to them or strategically aligned with them. We only see that highly unequal pairings of donors with respect to deliberative democracy are less common. The remaining donor alignment variables, such as the dispersion in pairwise ideal-point distances, partisan ideology, and donor political economy are not systematically related to the choice of examiner pairs.

Model 3 pits the two sets of explanations against each other. The key result remains that pairs where at least one examiner is of similar size as the donor are significantly more likely to be chosen. No other variable in this model has a statistically significant coefficient at the traditional level ($p<0.05$). Taken together, the results across the three models do not suggest that the choice of peer examiners is made on the basis of political or strategic considerations. Instead, we interpret these pairings chosen by the Secretariat to indicate that the assignment follows peer review guidelines.

While our main models exclude reviews of EU institutions as a donor, we can use this special case to probe the extent to which examiner pair choice follows a bureaucratic process. DAC peer review guidelines

**Figure 1:** Number of reviews of the EU institutions by examiner country




state that no EU member country should review aid giving by EU institutions. This limits the examiner pool considerably, which is why we expect our bureaucratic process variables to matter less systematically than for ordinary DAC member states. In further analysis, we indeed verify that the EU institutions were almost never reviewed by a EU member state, with only two exceptions among 516 reviews (Figure 1). Table 2 corroborates this result when controlling for other determinants of examiner choice.

**Table 2:** Determinants of Examiner Peer Choice for Reviews of the European Union

| | (4) | | (5) | |
|---|---|---|---|---|
| | EU is reviewed | | Other reviewees | |
| At least one major donor | 0.079 | (0.266) | 0.183** | (0.065) |
| Both major donors | -0.397° | (0.223) | -0.458** | (0.144) |
| At least one EU member | -2.596*** | (0.214) | 0.112 | (0.070) |
| Both EU members | -0.216° | (0.125) | 0.069 | (0.067) |
| At least one same model | -0.327 | (0.250) | -0.011 | (0.055) |
| Both same model | -0.023 | (0.379) | 0.029 | (0.096) |
| Review twice in a row | -0.112 | (0.385) | -0.056 | (0.076) |
| Both before as pair | -0.188 | (0.668) | 0.166 | (0.134) |
| At least one before | 0.195 | (0.311) | 0.009 | (0.084) |
| Both before | -0.671° | (0.370) | 0.129° | (0.065) |
| Observations | 4295 | | 83646 | |
| Within R2 | 0.016 | | 0.000 | |

Linear regressions with review fixed effects. Standard errors clustered on reviews in parentheses. Significance levels: ° <.1 * <.05 ** <.01 *** <.001

In summary, we have found that the choice of peer examiners is broadly in line with a process-driven model whereby examiner teams tend to include at least one donor of similar size and examiners are chosen to

avoid obvious conflicts of interest (such as EU institutions being reviewed by EU member states), as per DAC peer review guidelines. In contrast, the empirical patterns are inconsistent with donor interest arguments as we find no evidence that examiner choice is driven by similar foreign policy preferences, political institutions, partisan alignment, or donor political economies.

In a series of robustness checks, we establish further support for these findings in a number of ways. First, by expanding the set of variables that proxy for donor politics to include arms trade, military alliances, and diplomatic representation (Table A2). Second, by using a time-varying measure of major donorship by considering the top five donors in every year in terms of their gross domestic product (Table A3); and an alternative time-varying measure for relative donor size based on aid budgets (Table A4). We further show results without the review fixed effects (Table A5), clustering standard errors on years (Table A6), and including additional examiner characteristics (Table A7).

# 5 Determinants of Recipient Country Visits

In the second set of analyses, we investigate the choice of recipient countries that are visited for in-depth assessments by the peer review team. These visits are an integral element of the peer review process and yield important insights on the reviewed donor's programming in the field. These insights are then featured in the DAC peer review document. Of the 121 peer reviews (1996–2020) for which we have access to the full document, 116 (about 96%) report at least one recipient-specific assessment; 55 include at least two recipient-specific assessments; five have at least three; and one has information on four recipient country assessments. Overall, this results in 177 recipient-specific assessments across 116 reviews.

Recipient assessments typically consist of a field visit and are undertaken by the peer review team. In fact, 162 of the 177 recipient-specific assessments in our sample (about 92%) takes the form of a field visit, which is summarized in a report that serves as an annex to the final peer review document. When a physical field visit is not possible, the peer review commission evaluates the donor's operations in that recipient through phone interviews with relevant actors (in four instances in our sample) or by inviting relevant actors to answer questions about recipient country programs during a meeting in the reviewed donor's capital (in seven instances), or by meetings and calls from other locations (three instances).[12]

What determines which recipient countries are chosen for such in-depth assessments? As described in Section 3, the Secretariat asks reviewed donors to suggest a few potential options for visits that satisfy the process-oriented criteria peer review, including, for example, that visits should take place in recipients that are significant development partners of the reviewed donor, that broadly represent the reviewed donor's aid

[12]In one case, a field visit to the recipient country took place after the completion date of the peer review document.

program, and that were not visited for another peer review in the same or the previous year. Descriptively, we see that this last criterion is not always followed: rather, it is not uncommon for the same recipient to receive visits for reviews in back-to-back years; and in six instances, the same recipient was visited for more than one review in the same year, although this has not happened since 2003.[13]

Aside for privileging recipients who have not been the subject of in-depth assessments in the same or previous year, peer review teams are asked to visit recipients who are important to the reviewed donor's aid program and thus make for a good setting from which to learn about its aid giving practices more broadly. With this in mind, while the reviewed donor does suggest a list of possible recipients for field visits, it is the peer review team that ultimately makes the decision of where to go.

On the one hand, we expect reviewed donors to want to showcase their programs in recipients that they favor, with whom they have long-standing relationships, or which are strategically important. On the other hand, we expect that the peer review team might prefer to perform field assessments in recipients that are easy to navigate and with infrastructure in place that facilitates evaluation. This may include recipients that are relatively better developed, that score more highly on governance indicators, and that are more democratic. The Secretariat's interests might also favor recipients who are particularly important not to the reviewed donor in particular, but to the group of DAC donors as a whole. We test these expectations on the same sample of 116 peer reviews for which we have information on their recipient-specific assessments.

## 5.1 Determinants of Recipient Country Visits: Data and Methods

We construct a review-donor-recipient data set covering 116 DAC peer reviews from 1996 to 2020. We identify the potential set of recipients that could be assessed for any review as any country that has appeared as a recipient of the reviewed donor's ODA flows in any of the three years prior and up to the year of the peer review, according to the OECD DAC Creditor Reporting System (CRS) data. Each observation in the data is therefore a triad composed of (i) the reviewed donor, (ii) the review year, and (iii) a recipient of that donor's aid that could potentially be chosen for a recipient-specific assessment.

We construct our dependent variable as an indicator equal to one for recipients that were the subject of a recipient-specific assessment for a donor-year review, and zero for recipients who were not chosen for assessment in that review. Note that this indicator may be equal to one for up to four observations per review, because some reviews have more than one recipient-specific assessment.[14] The five reviews in our sample which have no recipient-specific assessments are excluded from the analyses.

To we operationalize the key DAC criteria for the choice of recipient countries to be visited in two ways:

[13] See Figure A3 in the Appendix.

[14] As discussed, 116 reviews from 1996 to 2020 include at least one recipient-specific assessment; 55 include at least two recipient-specific assessments; five have at least three; and one has information on four such assessments.

(i) the importance of the recipient to the reviewed donor's aid program; and (ii) whether the recipient was visited for any other review in the previews year. The latter is a simple binary indicator. The former is measured as the percentage of the reviewed donor's bilateral ODA commitments to that recipient in the year before the review. In other words, this captures how important a given recipient is in the reviewed donor's aid portfolio, by looking at how much of that donor's aid it receives.

In addition to privileging recipients who are important to the reviewed donor, the review team might be interested in favoring recipients who are important to the group of DAC donors as a whole. We measure this as the amount of aid committed to a given recipient in the year before a review by all DAC donors, as a percentage of the total amount committed by all DAC donors to all recipients in that year. In other words, this captures how important a given recipient is to DAC donors in general, by looking at how much of their collective aid it receives.

The peer review team might also have a preference to visit recipients with characteristics that facilitate the conducting of their evaluation. Thus, we include a measure of democracy (Polity V); government effectiveness, rule of law, and political stability and absence of violence/terrorism (all from the World Bank's Governance Indicators); and life expectancy and GDP per capita (from the World Bank's World Development Indicators). These variables allow us to assess whether recipients who are more politically developed—and perhaps easier to conduct a visit in—get chosen more frequently. We use a measure of international tourism (from the World Tourism Organization) to proxy for the ease of access of a given recipient; and the the World Bank's Statistical Capacity Indicator to capture the extent to which a given recipient can be expected to keep good records and thus facilitate operations and learning during a peer review. All of these recipient characteristics are included in our models for the year prior to the peer review.[15]

On the other hand, the reviewed donor might want to steer the review team toward its most favored recipients—beyond the ones who are most important in terms of how much of its aid their receive. In particular, the aid allocation literature has long established robust evidence of a colonial bias, whereby donors favor recipients that used to be their colonies. Thus, we include an indicator of colonial history, equal to one for recipients who were primarily a former colony of the reviewed donor and zero otherwise. If we found that field visits tend to take place in recipients with which the reviewed donor shares a colonial history, we would interpret this as evidence of donor interests influencing decisions about where to visit.

In some of our model specifications, we control for the size of the pool of potential recipients, given that it depends on how many recipients each donor has in its portfolio at a given time. We also use document-format fixed effects, which allows us to account for the fact that donor peer reviews change formats over time; and recipient fixed effects, to account for other unmodeled recipient-specific idiosyncrasies.

[15]See the Appendix for a detailed description of all data sources.

## 5.2 Determinants of Recipient Country Visits: Results

Similar to the previous set of analyses, we estimate linear regression models with a binary outcome variable. Table 3 shows the results of the main specifications in our analyses. The even-numbered specifications (Models 6, 8, and 10), include as predictors only the two explicit criteria outlined in the peer review methodology: a recipient's importance to the reviewed donor's aid program, and whether it was already visited in a review in the previous year. These models are, respectively, a random-effects model without controls (Model 6), a format fixed-effects model controlling for the size of the pool of potential recipients (Model 8), and a format-and-recipient fixed-effects model also controlling for pool size (Model 10).

Conversely, the odd-numbered specifications in Table 3 also include a recipient's importance to the DAC donor community as a whole as well as all other recipient characteristics, as discussed in Section 5.3. Parallel to the even-numbered specifications, these three models are, respectively, a random-effects model without controls (Model 7), a format fixed-effects model controlling for the size of the pool of potential recipients (Model 9), and a format-and-recipient fixed-effects model also controlling for pool size (Model 11).

**Table 3:** Determinants of Recipient Choice

| | (6) | (7) | (8) | (9) | (10) | (11) |
|---|---|---|---|---|---|---|
| Aid from Donor, % Donor Tot | 0.83*** | 0.94*** | 0.83*** | 0.94*** | 0.77*** | 0.90*** |
| | (0.035) | (0.067) | (0.035) | (0.068) | (0.037) | (0.069) |
| Visited in Previous Year | 0.0049 | −0.018** | 0.0049 | −0.018** | −0.021*** | −0.035*** |
| | (0.0042) | (0.0063) | (0.0042) | (0.0063) | (0.0045) | (0.0066) |
| Aid from DAC Donors, % DAC Tot | | −0.32 | | −0.32 | | −0.66* |
| | | (0.18) | | (0.18) | | (0.28) |
| Former Colony of Donor | | 0.041*** | | 0.041*** | | 0.041*** |
| | | (0.0087) | | (0.0087) | | (0.0088) |
| Democracy | | 0.000 32 | | 0.000 32 | | 0.000 21 |
| | | (0.000 33) | | (0.000 33) | | (0.000 98) |
| Government Effectiveness | | −0.0028 | | −0.0030 | | −0.0036 |
| | | (0.0070) | | (0.0071) | | (0.013) |
| Rule of Law | | 0.010 | | 0.010 | | −0.0028 |
| | | (0.0066) | | (0.0067) | | (0.015) |
| Political Stability | | −0.0039 | | −0.0039 | | −0.0071 |
| | | (0.0029) | | (0.0029) | | (0.0057) |
| Life Expectancy | | −0.000 16 | | −0.000 13 | | 0.000 69 |
| | | (0.000 25) | | (0.000 26) | | (0.0011) |
| GDP Per Capita | | −0.000 003 3*** | | −0.000 003 3*** | | 0.000 000 63 |
| | | (0.000 000 70) | | (0.000 000 70) | | (0.000 003 6) |
| International Tourism | | −0.000 000 00 | | −0.000 000 00 | | −0.000 000 00 |
| | | (0.000 000 00) | | (0.000 000 00) | | (0.000 000 00) |
| Statistical Capacity | | 0.000 38* | | 0.000 38* | | 0.000 083 |
| | | (0.000 15) | | (0.000 15) | | (0.000 32) |
| Effects | Random | Random | Format | Format | Format and Recipient | Format and Recipient |
| Controls | None | None | Pool Size | Pool Size | Pool Size | Pool Size |
| Num.Obs. | 13.790 | 5.593 | 13.790 | 5.593 | 13.790 | 5.593 |
| R2 | 0.039 | 0.051 | 0.039 | 0.051 | 0.062 | 0.067 |
| R2 Adj. | 0.039 | 0.049 | 0.039 | 0.048 | 0.049 | 0.048 |

Linear regressions, standard errors in parentheses. Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001.

First, note that the coefficient for a recipient's importance to the reviewed donor's aid program is positive and significant across all specifications. This suggests that receiving a larger proportion of the reviewed donor's total bilateral commitments makes countries more likely to be chosen for a visit by the peer review

team. This is in line with official guidelines for the peer review process. Also in line with those guidelines are the results for the indicator of whether a recipient was visited in the previous year: its coefficients are negative and significant in most specifications.

Moving beyond what is in the official peer review guidelines, we find strong evidence that former colonies of the reviewed donors are significantly more likely to be chosen for an in-depth assessment during peer reviews. Given how a pro-colony bias is well documented in the aid allocation literature, we are not surprised to find results consistent with the idea that donors seek to maintain a special relationships with their former colonies through their aid giving activities. Note that this is a significant predictor even as we are controlling for a recipient's importance to the reviewed donor's aid program—in other words, beyond following DAC guidelines to propose recipients who are important to their aid program, reviewed donors steer the choice towards important recipients who are also former colonies.

Interestingly, we do not find support for the idea that the importance of a recipient to the DAC as a whole matters in the choice of recipient visits: the peer review team does not seem to favor visiting recipients who get a lot of aid from all DAC donors. Results are also null for nearly all other recipient characteristics included in the odd-numbered specifications. Notable exceptions are GDP per capita, which is a negative and significant predictors in Models 7 and 9—indicating that less economically developed recipients are more likely to be chosen; and statistical capacity, which is a positive and significant predictors in Models 7 and 9—indicating that recipients who might keep better records and thus facilitate the review team's job are more likely to be chosen.

It is also important to get a sense of the magnitude of these effects. Substantively, a one-percent increase in the amount of the reviewed donor's aid that the recipient gets makes such recipient between 8 and 9 percent more likely to be chosen for a visit than those who are not—all else equal; and recipient countries who are former colonies of the reviewed donors are about 4 percent more likely to be chosen. The effect of a recipient's GDP per capita and of its statistical capacity, however, are much smaller.

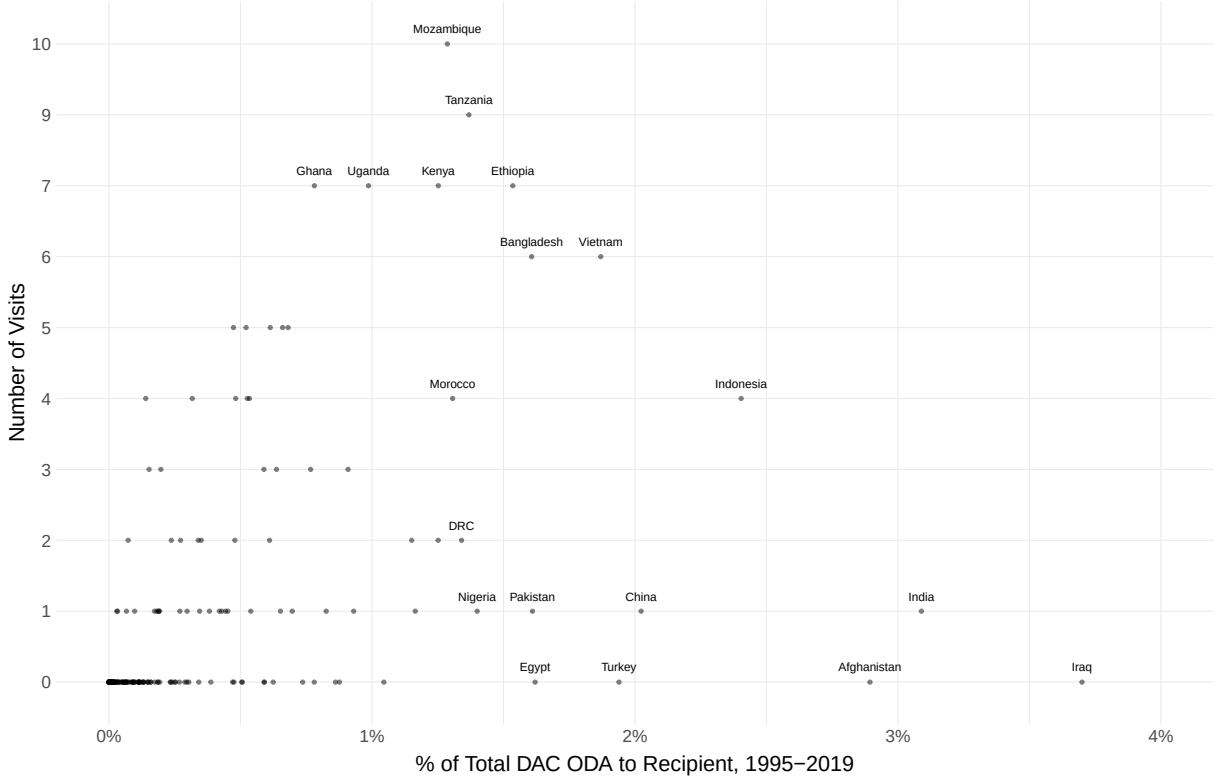
Taken together, these results indicate, first, a general compliance with DAC peer review methodology guidelines: recipients who are important to the reviewed donor's aid program are more likely to be visited, while recipients who were just subject of a visit for another review in the previous year are less likely to be chosen again. However, within the scope of those guidelines, reviewed donors seem to be able to steer the review team toward visiting recipients who are their former colonies—yet another way in which they seek to maintain a privileged relationship with those recipients.

Third, we find some evidence that DAC peer review teams favor less economically developed recipients, but also those recipients with greater statistical capacity. In order to further understand this, consider the correlation plotted in Figure 2 below: on the $x$-axis, the percentage of aid commitments from all DAC

donors to a given recipient from 1995 to 2019, as a percentage of the those donors' total commitments to all recipients over that same time period; on the $y$-axis, the number of total visits to that recipient in our sample of peer reviews.

Note how some recipients are never or rarely visited, despite getting the greatest proportions of DAC aid: e.g., Afghanistan, India, and Iraq. Conversely, other recipients are visited very frequently, in spite of receiving comparatively much less DAC aid: e.g., Ghana, Mozambique, Tanzania. This suggests that DAC peer review teams tend to visit recipients who are economically underdeveloped and targets of sustained development programs; as opposed to recipients who received a lot of DAC aid in the context of specific crises or for strategic reasons.

**Figure 2:** Correlation Between Recipient Importance to DAC Donors and Number of Visits




We interpret these findings to suggest that, overall, the official process and guidelines set in place by the Secretariat drive this step of the peer reviews as well. Recipients are largely chosen with an eye to what would be most informative for the peer review team to be able to draw conclusions about the reviewed donor's development assistance program as a whole: i.e., they visit recipients who receive a larger share of the reviewed donor's aid, who are economically underdeveloped, and whose statistical capacity may facilitate the review process. Yet it is also evident that, within the scope of the guidelines, reviewed donors are also

able to stir this choice toward recipients who they care to maintain a special relationship with their former colonies. In future research, it will be important to assess whether this influences the content of reviews: as visiting former colonies could result in more favorable, or otherwise biased, reviews and recommendations.

# 6 Conclusion

In this study, we systematically examined factors that shape important assignment decisions in the implementation of the OECD DAC's peer review system. Peer reviews have become a widespread "soft" mechanism for international organizations to assess and promote their members' compliance with agreed-upon standards and best practices. Yet, while most scholarship examines peer reviews with a focus on questions related to its efficacy, only limited knowledge exists about the process through which peer review assessments are constructed: whether key assignment decisions taken by the OECD Secretariat about who reviews whom, how, and where are impartial and independent; or whether the very states that are subject to peer review might be able to influence and politicize the process in ways that make favorable reviews more likely.

In the context of the DAC, we know from existing research that donors differ in their understanding of what constitutes best aid giving practices and which aid standards should be prioritized. For instance, we know that donor countries with domestic economies that privilege market-oriented solutions for public service delivery also favor aid delivery tactics that maximize efficiency criteria in the short-run and yield to relatively more bypass of the recipient public sector in implementation. On the other hand, more statist donors privilege aid delivery tactics that engage recipient country governments and gear their development assistance toward building capacity in recipient institutions. In light of these differences, we investigated the extent to which political economy characteristics, among other factors, affect assignment decisions. Across a number of metrics, we do not find evidence that donors tend to be reviewed by their most similar peers in the DAC: neither by similarities in domestic political economy structures, nor by measures of policy alignment. Nor do we find evidence that assignment decisions reflect strategic interests. Instead, we find evidence that assignment is largely process-driven as tested by our model; and that, as a result, the Secretariat succeeds in insulating peer review assignment from politics. This finding contributes to the empirical literature in international political economy that evaluates IO independence and, which has often found evidence indicating collusion or impartiality on the part of the Secretariat.

We also assess the choice of where the DAC peer review teams evaluate a donor's aid program through in-depth field visits. Here, too, we find evidence that the Secretariat's guidelines to select recipients are largely followed and which aim to maximize insights about donor aid programs and their practices: the choice favors recipient countries that are important to the reviewed donor's aid program and that have not

already been visited in the previous year. It privileges recipients who are economically underdeveloped, but whose statistical capacity may facilitate the review process. At the same time, we also find evidence of particularistic interests insofar as former colonies of the reviewed donors are more likely to be chosen for a visit, even while holding other characteristics equal. This finding contributes to the foreign aid allocation literature as it shows that key factors that we know affect donor decisions about how much aid to commit to recipient countries, like former colonial status, also shape decisions in the DAC peer review process. In future research, it will be important to assess whether this influences the content of reviews insofar as visiting former colonies might result in more favorable, or otherwise biased, reviews and recommendations.

Taken together, our results indicate that in the case of the OECD DAC peer review modalities and assignment mechanisms are largely explained by process-driven considerations, and not by donor interests. For the study of peer review compliance and aid effectiveness, our findings suggest that the assignment mechanism is not tainted by politically motivated selection effects. For future research, we would welcome scholars to assess the extent to which our results on review assignment decisions hold across a broader range of peer review systems in other, more contentious areas like human rights or the environment. It is possible that IO secretariats face more pressure in review assignment decision-making when standards and best practices championed by the organization are more heavily contested across members. Further, it is possible that assignment decisions are more susceptible to member interests and when the IO's membership is more heterogeneous. After all, the OECD DAC is composed of a relatively homogeneous group of democratic and developed member states.

# References


1. Abbott, K, & Snidal, D. (1998). "Why States Act through Formal International Organizations." *Journal of Conflict Resolution* 42(1): 3–32.
2. Abbott, K, & Snidal, D. (2013). "Taking Responsive Regulation Transnational: Strategies for International Organizations." *Regulation and Governance* 7(1): 95–113.
3. Albaret, M. (2016). "Multilateralism under Transformation." *Routledge Handbook of International Organization* (10511). http://www.routledgehandbooks.com/doi/10.4324/9780203405345.ch38.
4. Alter, K., and Meunier. (2009). "The Politics of International Regime Complexity." *Perspectives on Politics* 7(1): 13–24.
5. Ashoff, G. (2013). 50 Years of Peer Reviews by the OECD's Development Assistance Committee: An Instrument of Quality Assurance and Mutual Learning. German Development Institute.
6. Bailey, M. A., Strezhnev, A., & Voeten, E. (2017). Estimating dynamic state preferences from United Nations voting data. Journal of Conflict Resolution, 61(2), 430-456.
7. Barnett, M. N., & Finnemore, M. (1999). The Politics, Power, and Pathologies of International Organizations. International Organization, 53(4), 699–732.
8. Bayer, R. (2006). Diplomatic Exchange Data set, v2006-1. `http://correlatesofwar.org` (Accessed 15 June 2021).
9. Broome, A., Homolar, A., & Kranke, M. (2018). Bad science: International organizations and the indirect power of global benchmarking. European Journal of International Relations, 24(3), 514–539.
10. Brown, R. (2010). "Measuring Delegation." *Review of International Organizations* 5(2): 141–75.
11. Carcelli, S. (2021). "Bureaucratic Structure and Treaty Compliance." Working Paper.
12. Carraro, V., Conzelmann, T., & Jongen, H. (2019). Fears of peers? Explaining peer and public shaming in global governance. Cooperation and Conflict.
13. Carraro, V. (2019b). "Electing the experts: Expertise and independence in the UN human rights treaty bodies." *European Journal of International Relations*, 25(3), 826-851.
14. Conzelmann, T. (2014). "The Politics of Peer Reviewing: Comparing the OECD and the EU." In *The Politics of Informantion. The Case of the European Union,* ed. T et al. Blom. , 49–62.
15. Dahlberg, Stefan, Aksel Sundström, Sören Holmberg, Bo Rothstein, Natalia Alvarado Pachon, and Cem Mert Dalli. 2021. The Quality of Government Basic Dataset, version Jan21. University of Gothenburg: The Quality of Government Institute.
16. Dietrich, S. (2021). *States, Markets, and Foreign Aid*. Cambridge: Cambridge University Press.
17. Dijkstra, Hylke. "Shadow bureaucracies and the unilateral control of international secretariats: Insights from UN peacekeeping." *The Review of International Organizations* 10, no. 1 (2015): 23-41.
18. Dingwerth, K., Schmidtke, H., & Weise, T. (2011). "Speaking Democracy: Why International Organizations Adopt a Democratic Rhetoric." In *Legitimacy and Legitimation in Global Governance*, eds. Jonas Tallberg and Michael Zürn.
19. Doshi, R., Kelley, J. G., & Simmons, B. A. (2019). The power of ranking: The ease of doing business indicator and global regulatory behavior. International Organization, 73(3), 611–643.
20. Dreher, Axel, Valentin Lang, B. Peter Rosendorff, James Raymond Vreeland, 2018, Buying Votes and International Organizations: The Dirty Work-Hypothesis, CEPR Discussion Paper No. 13290.

21. Dreher, Axel, Jan-Egbert Sturm, and James Vreeland, 2009, Development Aid and International Politics: Does membership on the UN Security Council influence World Bank decisions? Journal of Development Economics 88: 1-18.

22. Eyben, R. (2013). Struggles in Paris: The DAC and the purposes of development aid. European Journal of Development Research, 25(1), 78–91.

23. Farrell, H., & Newman, A. (2014). "Domestic Institutions beyond the Nation-State: Charting the New Interdependence Approach." *World Politics* 66(2): 331–63.

24. Flüglister, K. (2012). Where Does Diffusion Take Place? The Role of Intergovernmental Cooperation for Policy Diffusion in Switzerland. European Journal of International Relations, 51(1), 316–349.

25. Freyburg, T., Krizic, I., Lavenex, S., & O'Flynn, C. (2017). *Democratic Governance and Transgovernmental Networks*.

26. Fuchs, A., & Richert, K. (2018). Development minister characteristics and aid giving. European Journal of Political Economy, 53, 186–204.

27. Groenendijk, N. (2009). EU and OECD benchmarking and peer review compared. Third EUCE Annual Conference: The EU in a Comparative Perspective, 1–18.

28. Hale, T., Held, D., & Young, K. (2013). "Gridlock: From Self-Reinforcing Interdependence to Second-Order Cooperation Problems." *Global Policy* 4(3): 223–35.

29. Honig, D., & Weaver, C. (2019). "A Race to the Top? The Aid Transparency Index and the Social Power of Global Performance Indicators." *International Organization*, 73(3), 579–610.

30. Hooghe, L., & Marks, G. (2015). "Delegation and Pooling in International Organizations." *Review of International Organizations* 10(3): 305–28.

31. Johnson, Tana. "Organizational Progeny: Why Governments Are Losing Control Over the Proliferating Structures of Global Governance." *Transformations in Governance*, 2014.

32. Jongen, H. (2018). "The authority of peer reviews among states in the global governance of corruption." *Review of International Political Economy*, 25(6), 909-935.

33. Karlas, J., Parízek, M. (2019). "The Process Performance of the WTO Trade Policy Review Mechanism: Peer-Reviewing Reconsidered." *Global Policy*, 10(3), 376-384.

34. Kelley, J. G., & Simmons, B. A. (2015). "Politics by number: Indicators as social pressure in international relations." *American Journal of Political Science*, 59(1), 55–70.

35. Kelley, J. G., Simmons, B. A. (2021). "Governance by other Means: rankings as regulatory systems." *International Theory*, 13(1), 169-178.

36. Keohane, R., & Nye, J. (1977). *Power and Interdependence: World Politics in Transition*. New York: Longman.

37. Leeds, B., Ritter, J., Mitchell, S., & Long, A. (2002). "Alliance Treaty Obligations and Provisions, 1815–1944." International Interactions, 28(3), 237–260.

38. Lehtonen, M. (2007). "Environmental policy integration through OECD peer reviews: Integrating the economy with the environment or the environment with the economy?" *Environmental Politics*, 16(1), 15–35.

39. Lehtonen, M. (2020). "Harder governance built on soft foundations: experience from OECD peer reviews." *Journal of Environmental Policy  Planning*, 22(6), 814-829.

40. Liverani, A., & Lundgren, H. E. (2007). "Evaluation Systems in Development Aid Agencies: An Analysis of DAC Peer Reviews 1996–2004." *Evaluation*, 13(2), 241–256.

41. Mahon, R., & McBride, S. (2009). "Standardizing and disseminating knowledge: The role of the OECD in global governance." *European Political Science Review*, 1(1), 83–101.

42. Marcussen, M. (2003) "Multilateral surveillance and the OECD: playing the idea game," in K. Armingeon M. Beyeler (eds.), *The OECD and European Welfare States*, pp. 13–31 (Cheltenham, UK and Northampton, MA, USA: Edward Elgar).

43. Mayer, T., & Zignago, S. (2011). "Notes on CEPII's distances measures: The GeoDist database."

44. McVoy, E. C. (1940). "Patterns of diffusion in the United States." *American Sociological Review*, 5(2), 219–227.

45. Meyer, J. W., Drori, G. S., & Hwang, H. (2006). "World Society and the Proliferation of Formal Organization." *Globalization and Organization: World Society and Organizational Change*, pp. 25–49.

46. Moravcsik, A. (2004). "Is There a 'Democratic Deficit' in World Politics? A Framework for Analysis." *Government and Opposition* 39(2): 19–36.

47. Niemann, D., & Martens, K. (2018). "Soft governance by hard fact? The OECD as a knowledge broker in education policy." *Global Social Policy*, 18(3), 267–283.

48. Nielson, D., & Tierney, M. (2003). "Delegation to International Organizations: Agency Theory and World Bank Environmental Reform." *International Organization* 57(2): 241–76.

49. Organisation for Economic Co-operation and Development's (OECD). 2021. *DAC Peer Review Methodology, Updated 2021.*

50. Pagani, F. (2002). "Peer Review As a Tool for Co-Operation and Change." *African Security Review* 11(4): 15–24.

51. Paulo, S., & Reisen, H. (2010). "Eastern Donors and Western Soft Law: Towards a DAC Donor Peer Review of China and India?" *Development Policy Review* 28(5): 535–52.

52. Porter, T., & Webb, M. (2007). The Role of the OECD in the Orchestration of Global Knowledge Networks. Canadian Political Science Association Annual Meetings.

53. Prakash, A., & Potoski, M. 2016. "Dysfunctional Institutions? Toward a New Agenda in Governance Studies." *Regulation and Governance* 10(2): 115–25.

54. Raustiala, K., & Victor, D. (2004). "The Regime Complex for Plant Genetic Resources." *International Organization* 58(2): 277–309.

55. Singer, J. D. (1972). "The "Correlates of War" project: Interim report and rationale." *World Politics*, 24(2), 243-270.

56. SIPRI (2021). SIPRI Arms Transfers Database (TIV of arms exports from all). `http://www.sipri.org/databases/armstransfers/sources-and-methods/` (accessed 15 June 2021).

57. Teorell, J., Kumlin, S., Sundström, A., Holmberg, S., Rothstein, B., Alvarado Pachon, N., & Dalli, C. M. (2020). The Quality of Government Dataset, version Jan20. University of Gothenburg: The Quality of Government Institute.

58. Vabulas, F., & Snidal, D. (2013). "Organization without Delegation: Informal Intergovernmental Organizations (IIGOs) and the Spectrum of Intergovernmental Arrangements." *Review of International Organizations* 8(2): 193–220.

59. Weaver, C. E. (2008). The World Bank's Hypocrisy and the Poverty of Reform. Ithaca: Cornell University Press.

60. Webb, M. C. (2004). Defining the boundaries of legitimate state practice: Norms, transnational actors and the OECD's project on harmful tax competition. Review of International Political Economy, 11(4), 787–827.

61. Westerwinter, O. (2019). *Transnational Public-Private Governance Initiatives in World Politics: Introducing a New Dataset*. St. Gallen.

62. Woodward, R. (2009). *The organisation for economic co-operation and development (OECD)*. Routledge.

# Appendix

## A1 OECD DAC Peer Reviews: The Data

The analyses presented in this paper rely on metadata about the OECD DAC donor peer reviews. We collected information about the entire universe of existing peer reviews—516 reviews conducted 1962–2020—through documents available at the OECD library.[16] For all 516 reviews, we know the DAC member who was reviewed, the date of the review, and the peer examiner countries, as we are able to retrieve this information without the need to access the full peer review report. These are the data we use in the analyses of the choice of peer examiners.

For 121 of the most recent reviews, we also have access to the full peer review report, which are available online through the OECD library.[17] This includes most (but not all) reviews from the late 1990s, and all reviews starting from the year 2000. The full peer review documents contain information about the in-depth assessments of one or more recipients of the reviewed donor's aid, which was collected manually by research assistants. They also allow us to classify reviews by document format, an indication of the different "eras" in which reviews were conducted under different operating guidelines. These are the data we use in the analyses of the choice of recipients for in-depth assessments.

Figure A1 illustrates the universe of peer reviews, highlighting: (i) the reviews for which we have access to the full report document; (ii) those for which, in spite of having access to the full document, we cannot find information on in-depth recipient assessments; and (iii) the early reviews which featured three peer examiner countries rather than two. Figure A2 focuses on the 121 reviews for which we have full documentation and illustrates the different review formats.

[16] Specifically, information about review dates and examiners for older review was retrieved from the "Chronological List of Development Assistance Committee (DAC) Aid Review (Peer Review) Meetings (by year) - 1962 to 1994.pdf" document.

[17] The available peer reviews are stored under slightly different classifications over time, e.g., "Development Co-operation Reviews," "OECD Development Assistance Peer Reviews," and "OECD Development Co-operation Peer Reviews."

**Figure A1:** Universe of OECD DAC Peer Reviews

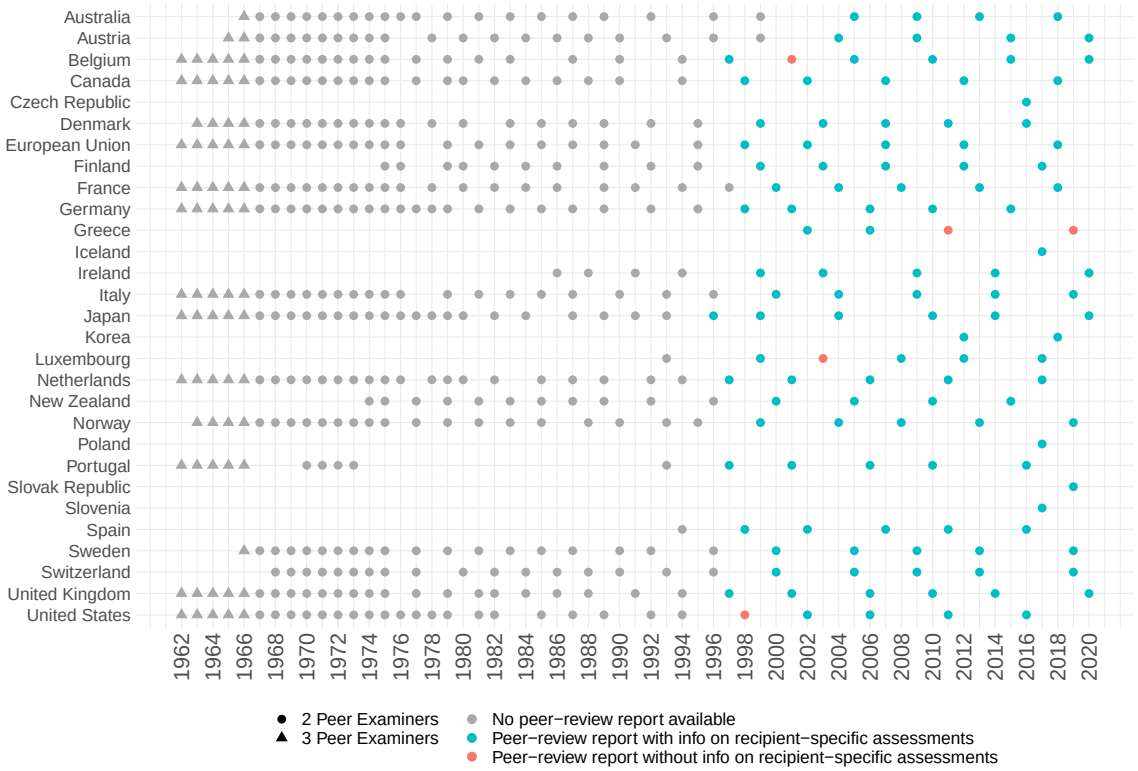
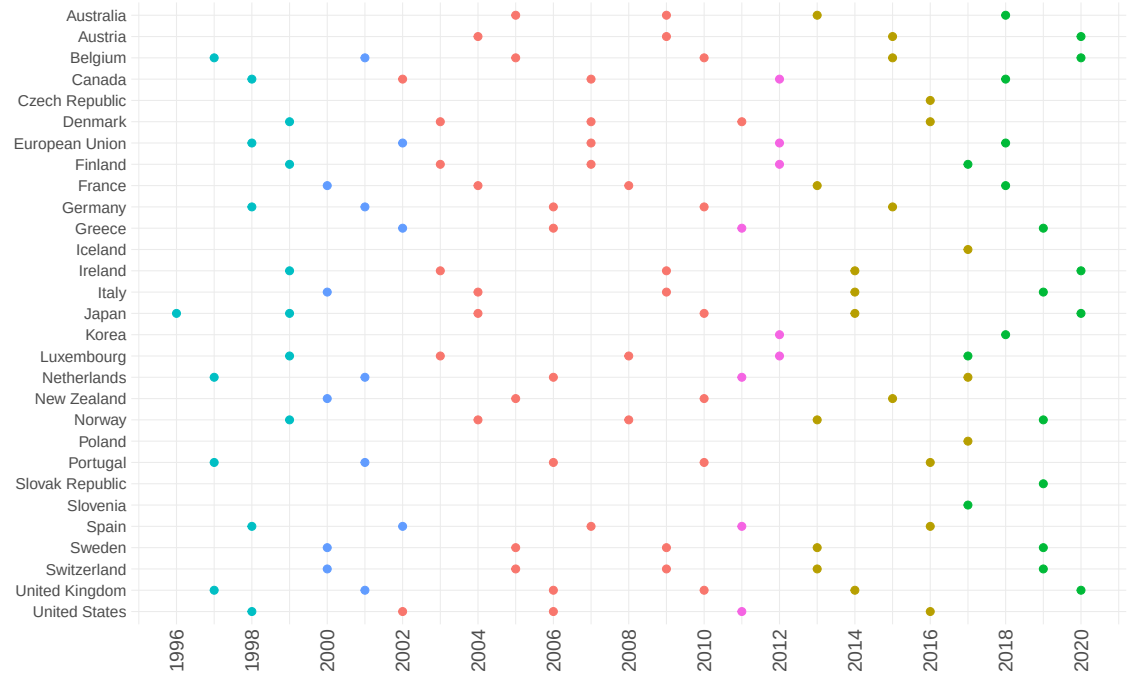



**Figure A2:** Distribution of Formats Across Available Peer-Review Documents



# A2 Determinants of Peer Examiner Choice: The Data

**Table A1:** Definitions and Descriptive Statistics for Data on Peer Examiner Choice

| Variable Name | Description | Count | Mean | SD | Min | Max |
|---|---|---|---|---|---|---|
| Is examiner | Binary indicator for whether a given pair of examiners conducted a given review (OECD 2020) | 87941 | 0.058 | 0.076 | 0 | 1 |
| At least one similar size | Binary indicator for whether at least one examiner is of the same size (where size is measured as being a major donor—United States, United Kingdom, Germany, France, Japan) | 87941 | 0.826 | 0.379 | 0 | 1 |
| Both similar size | Binary indicator for whether both examiners are of the same size (where size is measured as above) | 87941 | 0.501 | 0.5 | 0 | 1 |
| At least one similar EU | Binary indicator for whether at least one reviewee–examiner pair has the same status with respect to being a EU member in the year of review | 87491 | 0.701 | 0.458 | 0 | 1 |
| Both similar EU | Binary indicator for whether both examiners have the same status as the reviewee with respect to being a EU member in the year of review | 87491 | 0.247 | 0.431 | 0 | 1 |
| At least one common language | Binary indicator for whether at least one reviewer matches on an indicator of French language (in which the review may be carried out). French-speaking members are Belgium, Canada, the EU institutions, France, Luxembourg, and Switzerland. Note that a match also exists if both are English-speaking countries | 87941 | 0.796 | 0.403 | 0 | 1 |
| Both common language | Binary indicator for whether both reviewers match on the French language indicator with the reviewee | 87941 | 0.396 | 0.489 | 0 | 1 |
| At least one contiguous | Binary indicator for whether at least one reviewee–examiner pair entails states that share a border (Mayer and Zignago 2011) | 83646 | 0.191 | 0.393 | 0 | 1 |
| Both contiguous | Binary indicator for whether both reviewee–examiner pairs are contiguous | 83646 | 0.026 | 0.158 | 0 | 1 |
| Review twice in a row | Binary indicator for whether both examiners also reviewed the reviewee at its immediately prior review | 87941 | 0.189 | 0.391 | 0 | 1 |
| Both before as pair | Binary indicator for whether both examiners as a pair reviewed the reviewee before at any point in time | 87941 | 0.065 | 0.247 | 0 | 1 |
| At least one before | Binary indicator for whether at least one examiner was allocated this reviewee ever before (OECD 2020) | 87941 | 0.797 | 0.402 | 0 | 1 |
| Both before | Binary indicator for whether both examiners were each (individually) allocated this reviewee ever before (OECD 2020); the theoretically interesting effect can be obtained by adding the coefficient for "one examiner reviewed reviewee" (which includes cases of two examiners) | 87941 | 0.408 | 0.491 | 0 | 1 |
| Average UNGA distance | Average of the pairwise idealpoint distances between each examiner and the reviewee, based on UN General Assembly voting behavior (Bailey, Strezhnev, and Voeten 2017) | 71445 | 0.021 | 0.576 | -2.444 | 2.216 |
| Dispersion of UNGA distances | Dispersion of UNGA distances, defined as the difference between the larger distance and the smaller distance across both pairs; distances are absolute distances | 71445 | 0.347 | 0.431 | 0 | 2.857 |
| At least one same CPE | Binary indicator for whether at least one reviewee–examiner pair had the same "donor political economy" (or CPE), that is, whether they were both LMEs or non-LMEs (Dietrich 2021) | 87941 | 0.861 | 0.346 | 0 | 1 |
| Both same CPE | Binary indicator for whether both reviewee–examiner pairs had the same CPE (Dietrich 2021) | 87941 | 0.593 | 0.491 | 0 | 1 |

| Variable Name | Description | Count | Mean | SD | Min | Max |
|---|---|---|---|---|---|---|
| Average partisan distance | Average of the partisan ideology distance between each examiner and the reviewee, based on a quasi-continuous partisan ideology measure (Bjornskov and Potrafke 2013) | 66640 | 0.423 | 0.297 | 0 | 1.5 |
| Dispersion of partisan distances | Dispersion of partisan ideology distances, defined as the difference between the larger distance and the smaller distance across both pairs; distances are in absolute terms | 66640 | 0.289 | 0.324 | 0 | 1.5 |
| Average difference in deliberative democracy | Average of the differences in the V-Dem deliberative democracy index (Coppedge et al. 2016) between each examiner and the reviewee; direction of difference is preserved; | 73284 | 0.009 | 0.136 | -0.832 | 0.832 |
| Dispersion of differences in deliberative democracy | Dispersion of differences in deliberative democracy, defined as the difference between the higher difference and the lower difference in deliberative democracy across both pairs | 73284 | 0.072 | 0.122 | 0 | 0.832 |
| Arms trade between at least one | Binary indicator for whether at least one reviewee–examiner pair had any arms trade in the year of review (SIPRI 2021) | 87941 | 0.449 | 0.497 | 0 | 1 |
| Arms trade between both | Binary indicator for whether both reviewee–examiner pairs had any arms trade in the year of review (SIPRI 2021) | 87941 | 0.116 | 0.32 | 0 | 1 |
| Alliance between at least one | Binary indicator for whether at least one reviewee–examiner pair had any alliance obligation, which includes bilateral alliances and joint membership in multilateral alliances—drawn from ATOP database (Leeds et al. 2002) | 87941 | 0.405 | 0.491 | 0 | 1 |
| Alliance between both | Binary indicator for whether both reviewee–examiner pairs had alliance obligations, which includes bilateral alliances and joint membership in multilateral alliances (Leeds et al. 2002) | 87941 | 0.145 | 0.352 | 0 | 1 |
| Diplomatic ties between at least one | Binary indicator for whether at least one reviewee–examiner pair had any diplomatic representation (Bayer 2006). Since the data end in 2005, we carried them forward based on the last available year | 87941 | 0.695 | 0.46 | 0 | 1 |
| Diplomatic ties between both | Binary indicator for whether both reviewee–examiner pairs had any diplomatic representation (Bayer 2006). Since the data end in 2005, we carried them forward based on the last available year | 87941 | 0.625 | 0.484 | 0 | 1 |
| Average ODA budget ratio | Geometric mean of the ratio of (logged) ODA budgets between the reviewee and examiners; data on ODA commitments available from DAC1 table (OECD 2019) | 80497 | 1.043 | 0.233 | 0.41 | 2.476 |
| Dispersion of ODA budget ratio | Ratio of the larger ODA budget ratio and the smaller ODA budget ratio across both reviewee–examiner pairs; underlying data available from DAC1 table (OECD 2019) | 84203 | 1.247 | 0.23 | 1 | 2.757 |
| First pair role-swopped | Binary indicator for whether the current reviewee was a examiner at the most recent (past) review in which the first examiner was being reviewed (OECD 2020) | 87941 | 0.66 | 0.474 | 0 | 1 |
| Second pair role-swopped | Binary indicator for whether the current reviewee was a examiner at the most recent (past) review in which the second examiner was being reviewed (OECD 2020) | 87941 | 0.604 | 0.489 | 0 | 1 |
| Average ODA/GNI (%) | Average ODA/GNI (%) quota across both examiners, available from DAC1 table (OECD 2019) | 85734 | 0.416 | 0.187 | 0.029 | 1.405 |
| (Logged) average ODA | Average ODA budget of both examiners, available from DAC1 table (OECD 2019) | 85810 | 7.502 | 0.972 | 3.928 | 10.348 |
| Number of available examiners | Number of potential examiners available for review, which is determined by DAC membership only (OECD 2020) | 87941 | 19.557 | 4.738 | 10 | 28 |
| Three examiners | Binary indicator for whether the review involved three examiners (OECD 2020) | 87941 | 0.06 | 0.237 | 0 | 1 |
| Review year | Year of the review | 87941 | 1991.39 | 17.609 | 1962 | 2020 |

# A3 Determinants of Peer Examiner Choice: Robustness Tests

**Table A2:** Determinants of Peer Examiner Choice: Additional Donor Politics Variables

| | (A1) | | (A2) | | (A3) | |
|---|---|---|---|---|---|---|
| Arms trade between at least one | 0.097 | (0.061) | 0.042 | (0.087) | 0.045 | (0.085) |
| Arms trade between both | 0.179° | (0.105) | 0.172 | (0.126) | 0.146 | (0.126) |
| Alliance between at least one | -0.013 | (0.081) | -0.018 | (0.105) | 0.011 | (0.106) |
| Alliance between both | 0.019 | (0.089) | 0.128 | (0.116) | 0.190 | (0.117) |
| Diplomatic ties between at least one | -0.515 | (0.384) | -0.285 | (0.660) | -0.316 | (0.662) |
| Diplomatic ties between both | -0.098 | (0.113) | -0.093 | (0.129) | -0.041 | (0.131) |
| Average UNGA distance | | | -0.133 | (0.136) | -0.013 | (0.151) |
| Dispersion of UNGA distances | | | -0.099 | (0.102) | -0.167 | (0.105) |
| Average power difference | | | 0.753 | (2.177) | -0.793 | (2.195) |
| Dispersion in power difference | | | 1.011 | (1.534) | 0.898 | (1.557) |
| Average partisan distance | | | 0.104 | (0.139) | 0.106 | (0.138) |
| Dispersion of partisan distances | | | 0.007 | (0.117) | 0.002 | (0.116) |
| Average difference in democracy | | | 0.138 | (0.597) | 0.040 | (0.602) |
| Dispersion democracy differences | | | -0.964* | (0.399) | -0.731° | (0.403) |
| At least one similar size | | | | | 0.630*** | (0.139) |
| Both similar size | | | | | -0.095 | (0.106) |
| At least one similar EU | | | | | 0.090 | (0.092) |
| Both similar EU | | | | | -0.144* | (0.083) |
| At least one common language | | | | | 0.204° | (0.106) |
| Both common language | | | | | 0.313 | (0.251) |
| At least one contiguous | | | | | -0.223* | (0.105) |
| Both contiguous | | | | | -0.074 | (0.269) |
| At least one same model | | | | | -0.074 | (0.075) |
| Both same model | | | | | 0.079 | (0.126) |
| Review twice in a row | | | | | -0.196* | (0.097) |
| Both before as pair | | | | | 0.105 | (0.164) |
| At least one before | | | | | 0.004 | (0.117) |
| Both before | | | | | 0.124 | (0.088) |
| Observations | 83646 | | 52387 | | 52387 | |
| Within R2 | 0.000 | | 0.000 | | 0.001 | |

Linear regressions with review fixed effects and standard errors clustered on reviews in parentheses.
Significance levels: ° <.1 * <.05 ** <.01 *** <.001

**Table A3:** Determinants of Peer Examiner Choice: Alternative Measure of Donor Size

| | (A4) | | (A5) | | (A6) | |
|---|---|---|---|---|---|---|
| At least one similar size (GDP-5) | 0.512*** | (0.087) | | | 0.528*** | (0.108) |
| Both similar size (GDP-5) | -0.055 | (0.064) | | | -0.019 | (0.096) |
| At least one similar EU | 0.080 | (0.069) | | | 0.107 | (0.091) |
| Both similar EU | -0.070 | (0.067) | | | -0.130 | (0.083) |
| At least one common language | 0.106 | (0.082) | | | 0.211* | (0.106) |
| Both common language | 0.260 | (0.198) | | | 0.315 | (0.250) |
| At least one contiguous | -0.056 | (0.084) | | | -0.216* | (0.104) |
| Both contiguous | -0.141 | (0.182) | | | -0.080 | (0.268) |
| At least one same model | -0.043 | (0.056) | | | -0.100 | (0.075) |
| Both same model | 0.025 | (0.096) | | | 0.084 | (0.126) |
| Review twice in a row | -0.064 | (0.074) | | | -0.180° | (0.098) |
| Both before as pair | 0.162 | (0.133) | | | 0.128 | (0.163) |
| At least one before | -0.007 | (0.085) | | | 0.005 | (0.117) |
| Both before | 0.121° | (0.064) | | | 0.122 | (0.087) |
| Average UNGA distance | | | -0.157 | (0.134) | -0.084 | (0.139) |
| Dispersion of UNGA distances | | | -0.102 | (0.101) | -0.103 | (0.103) |
| Average power difference | | | 0.696 | (2.141) | -0.728 | (2.180) |
| Dispersion in power difference | | | 1.025 | (1.516) | 0.362 | (1.570) |
| Average partisan distance | | | 0.106 | (0.140) | 0.106 | (0.139) |
| Dispersion of partisan distances | | | 0.012 | (0.116) | 0.022 | (0.116) |
| Average difference in democracy | | | 0.042 | (0.575) | -0.119 | (0.584) |
| Dispersion democracy differences | | | -0.925* | (0.388) | -0.693° | (0.394) |
| Observations | 83646 | | 52387 | | 52387 | |
| Within R2 | 0.001 | | 0.000 | | 0.001 | |

Linear regression with review fixed effects and standard errors clustered on reviews in parentheses.
Significance levels: ° <.1 * <.05 ** <.01 *** <.001

**Table A4:** Determinants of Peer Examiner Choice: Alternative Measure of Donor Size

| | (A7) | | (A8) | | (A9) | |
|---|---|---|---|---|---|---|
| Average ODA budget ratio | -1.121*** | (0.196) | | | -1.916*** | (0.379) |
| Dispersion of ODA budget ratio | 0.073 | (0.112) | | | 0.411* | (0.203) |
| At least one similar EU | 0.107 | (0.069) | | | 0.142 | (0.091) |
| Both similar EU | -0.013 | (0.068) | | | -0.096 | (0.085) |
| At least one common language | 0.097 | (0.085) | | | 0.254* | (0.108) |
| Both common language | 0.220 | (0.197) | | | 0.302 | (0.249) |
| At least one contiguous | -0.111 | (0.087) | | | -0.276* | (0.108) |
| Both contiguous | -0.269 | (0.176) | | | -0.305 | (0.237) |
| At least one same model | 0.005 | (0.058) | | | -0.096 | (0.076) |
| Both same model | 0.037 | (0.101) | | | 0.105 | (0.131) |
| Review twice in a row | -0.089 | (0.078) | | | -0.132 | (0.104) |
| Both before as pair | 0.177 | (0.137) | | | 0.128 | (0.165) |
| At least one before | -0.040 | (0.093) | | | -0.034 | (0.128) |
| Both before | 0.013 | (0.073) | | | 0.015 | (0.096) |
| Average UNGA distance | | | -0.157 | (0.134) | 0.007 | (0.156) |
| Dispersion of UNGA distances | | | -0.102 | (0.101) | -0.152 | (0.111) |
| Average power difference | | | 0.696 | (2.141) | 3.928 | (2.386) |
| Dispersion in power difference | | | 1.025 | (1.516) | 1.151 | (1.659) |
| Average partisan distance | | | 0.106 | (0.140) | 0.129 | (0.142) |
| Dispersion of partisan distances | | | 0.012 | (0.116) | 0.049 | (0.118) |
| Average difference in democracy | | | 0.042 | (0.575) | 0.341 | (0.840) |
| Dispersion democracy differences | | | -0.925* | (0.388) | -0.638 | (0.730) |
| Observations | 77007 | | 52387 | | 49879 | |
| Within R2 | 0.001 | | 0.000 | | 0.001 | |

Linear regression with review fixed effects and standard errors clustered on reviews in parentheses.
Significance levels: ° <.1 * <.05 ** <.01 *** <.001

**Table A5:** Determinants of Peer Examiner Choice: Random Effects and Additional Controls

| | (A10) | | (A11) | | (A12) | |
|---|---|---|---|---|---|---|
| At least one similar size | 0.360*** | (0.075) | | | 0.457*** | (0.113) |
| Both similar size | -0.167** | (0.064) | | | -0.192* | (0.092) |
| At least one similar EU | 0.063 | (0.064) | | | 0.082 | (0.086) |
| Both similar EU | -0.071 | (0.065) | | | -0.144° | (0.081) |
| At least one common language | 0.064 | (0.064) | | | 0.140 | (0.087) |
| Both common language | 0.204 | (0.194) | | | 0.232 | (0.245) |
| At least one contiguous | -0.037 | (0.070) | | | -0.153° | (0.089) |
| Both contiguous | -0.120 | (0.179) | | | -0.025 | (0.267) |
| At least one same model | -0.009 | (0.054) | | | -0.053 | (0.072) |
| Both same model | 0.030 | (0.095) | | | 0.092 | (0.124) |
| Review twice in a row | -0.082 | (0.074) | | | -0.159° | (0.095) |
| Both before as pair | 0.172 | (0.133) | | | 0.136 | (0.162) |
| At least one before | -0.002 | (0.065) | | | -0.009 | (0.094) |
| Both before | 0.106° | (0.060) | | | 0.109 | (0.080) |
| Available examiners | -0.040 | (0.082) | -0.023 | (0.152) | -0.102 | (0.230) |
| Three examiners | 0.868 | (1.420) | 1.447 | (1.745) | 0.505 | (2.635) |
| Average UNGA distance | | | -0.052 | (0.054) | 0.023 | (0.062) |
| Dispersion of UNGA distances | | | -0.098 | (0.100) | -0.177° | (0.104) |
| Average power difference | | | 0.315 | (0.582) | 0.468 | (0.639) |
| Dispersion in power difference | | | 1.249 | (1.226) | 1.437 | (1.317) |
| Average partisan distance | | | 0.104 | (0.104) | 0.110 | (0.103) |
| Dispersion of partisan distances | | | 0.011 | (0.113) | 0.007 | (0.113) |
| Average difference in democracy | | | -0.010 | (0.337) | -0.020 | (0.346) |
| Dispersion democracy differences | | | -0.870** | (0.328) | -0.690* | (0.330) |
| Observations | 83646 | | 52387 | | 52387 | |
| Within R2 | 0.001 | | 0.000 | | 0.001 | |

Linear regression with random effects, year effects, and standard errors clustered on reviews in parentheses.
Significance levels: ° <.1 * <.05 ** <.01 *** <.001

**Table A6:** Determinants of Peer Examiner Choice: Review Interdependence

| | (A13) | | (A14) | | (A15) | |
|---|---|---|---|---|---|---|
| At least one similar size | 0.549*** | (0.135) | | | 0.625*** | (0.172) |
| Both similar size | -0.114 | (0.070) | | | -0.111 | (0.096) |
| At least one similar EU | 0.065 | (0.070) | | | 0.092 | (0.093) |
| Both similar EU | -0.070 | (0.066) | | | -0.137° | (0.077) |
| At least one common language | 0.089 | (0.083) | | | 0.192° | (0.096) |
| Both common language | 0.252 | (0.155) | | | 0.313 | (0.203) |
| At least one contiguous | -0.057 | (0.086) | | | -0.209* | (0.099) |
| Both contiguous | -0.155 | (0.168) | | | -0.090 | (0.279) |
| At least one same model | -0.020 | (0.064) | | | -0.072 | (0.081) |
| Both same model | 0.025 | (0.081) | | | 0.083 | (0.114) |
| Review twice in a row | -0.085 | (0.088) | | | -0.190° | (0.101) |
| Both before as pair | 0.145 | (0.130) | | | 0.108 | (0.169) |
| At least one before | 0.001 | (0.079) | | | 0.006 | (0.108) |
| Both before | 0.125° | (0.068) | | | 0.128 | (0.086) |
| Average UNGA distance | | | -0.157 | (0.131) | -0.026 | (0.146) |
| Dispersion of UNGA distances | | | -0.102 | (0.087) | -0.177° | (0.097) |
| Average power difference | | | 0.696 | (2.036) | -0.650 | (2.079) |
| Dispersion in power difference | | | 1.025 | (1.018) | 0.961 | (1.142) |
| Average partisan distance | | | 0.106 | (0.144) | 0.109 | (0.143) |
| Dispersion of partisan distances | | | 0.012 | (0.120) | 0.013 | (0.117) |
| Average difference in democracy | | | 0.042 | (0.358) | -0.075 | (0.370) |
| Dispersion democracy differences | | | -0.925*** | (0.248) | -0.651* | (0.243) |
| Observations | 83646 | | 52387 | | 52387 | |
| Within R2 | 0.001 | | 0.000 | | 0.001 | |

Linear regression with review fixed effects and standard errors clustered on years in parentheses.
Significance levels: ° <.1 * <.05 ** <.01 *** <.001

**Table A7:** Determinants of Peer Examiner Choice: Added Review Characteristics

| | (A16) | | (A17) | | (A18) | |
|---|---|---|---|---|---|---|
| Highest unemployment | -0.002 | (0.003) | -0.003 | (0.005) | -0.001 | (0.005) |
| Lowest growth | -0.017 | (0.017) | -0.014 | (0.023) | -0.010 | (0.023) |
| Highest corruption | -0.297 | (0.263) | -0.149 | (0.433) | -0.063 | (0.443) |
| At least one similar size | 0.367*** | (0.140) | | | 0.504** | (0.169) |
| Both similar size | -0.088 | (0.072) | | | -0.067 | (0.114) |
| At least one similar EU | 0.197** | (0.076) | | | 0.210* | (0.098) |
| Both similar EU | 0.006 | (0.073) | | | -0.015 | (0.096) |
| At least one common language | 0.025 | (0.093) | | | 0.113 | (0.119) |
| Both common language | 0.089 | (0.210) | | | 0.204 | (0.265) |
| At least one contiguous | -0.103 | (0.090) | | | -0.227* | (0.113) |
| Both contiguous | -0.020 | (0.221) | | | -0.070 | (0.282) |
| At least one same model | 0.066 | (0.061) | | | -0.043 | (0.083) |
| Both same model | -0.064 | (0.104) | | | -0.068 | (0.129) |
| Review twice in a row | -0.004 | (0.089) | | | -0.197° | (0.119) |
| Both before as pair | 0.186 | (0.154) | | | 0.084 | (0.179) |
| At least one before | 0.055 | (0.081) | | | 0.162 | (0.112) |
| Both before | 0.138° | (0.074) | | | 0.127 | (0.099) |
| Average UNGA distance | | | -0.161 | (0.175) | -0.024 | (0.195) |
| Dispersion of UNGA distances | | | -0.215° | (0.118) | -0.288* | (0.123) |
| Average power difference | | | -0.319 | (2.606) | -2.684 | (2.628) |
| Dispersion in power difference | | | 1.973 | (1.595) | 2.317 | (1.632) |
| Average partisan distance | | | 0.111 | (0.156) | 0.123 | (0.153) |
| Dispersion of partisan distances | | | 0.095 | (0.126) | 0.091 | (0.125) |
| Average difference in democracy | | | 0.095 | (0.921) | -0.194 | (0.940) |
| Dispersion democracy differences | | | -0.665 | (0.657) | -0.352 | (0.663) |
| Observations | 58139 | | 38005 | | 38005 | |
| Within-R2 | 0.001 | | 0.001 | | 0.001 | |

Linear regression with review fixed effects and standard errors clustered on reviews in parentheses.
Significance levels: ° <.1 * <.05 ** <.01 *** <.001

**Table A8:** Determinants of Peer Examiner Choice: Additional Potential Determinants

| | (A19) | | (A20) | | (A21) | |
|---|---|---|---|---|---|---|
| At least one same CPE | 0.014 | (0.081) | -0.007 | (0.085) | -0.052 | (0.124) |
| Both same CPE | -0.020 | (0.067) | 0.000 | (0.069) | -0.027 | (0.102) |
| First pair role-swopped | 0.212*** | (0.058) | 0.189** | (0.062) | 0.151° | (0.082) |
| Second pair role-swopped | 0.289*** | (0.055) | 0.262*** | (0.061) | 0.341*** | (0.081) |
| At least one similar size | | | 0.525*** | (0.111) | 0.606*** | (0.139) |
| Both similar size | | | -0.112° | (0.066) | -0.116 | (0.106) |
| At least one similar EU | | | 0.059 | (0.070) | 0.079 | (0.092) |
| Both similar EU | | | -0.063 | (0.067) | -0.134 | (0.082) |
| At least one common language | | | 0.093 | (0.083) | 0.209° | (0.107) |
| Both common language | | | 0.256 | (0.199) | 0.327 | (0.252) |
| At least one contiguous | | | -0.056 | (0.083) | -0.187° | (0.103) |
| Both contiguous | | | -0.150 | (0.182) | -0.075 | (0.268) |
| At least one same model | | | -0.025 | (0.056) | -0.077 | (0.076) |
| Both same model | | | 0.020 | (0.097) | 0.075 | (0.129) |
| Review twice in a row | | | -0.074 | (0.075) | -0.179° | (0.097) |
| Both before as pair | | | 0.137 | (0.134) | 0.099 | (0.164) |
| At least one before | | | -0.056 | (0.087) | -0.038 | (0.120) |
| Both before | | | 0.024 | (0.069) | 0.036 | (0.091) |
| Average UNGA distance | | | | | 0.019 | (0.152) |
| Dispersion of UNGA distances | | | | | -0.157 | (0.106) |
| Average power difference | | | | | -0.656 | (2.152) |
| Dispersion in power difference | | | | | 0.833 | (1.545) |
| Average partisan distance | | | | | 0.097 | (0.138) |
| Dispersion of partisan distances | | | | | 0.022 | (0.117) |
| Average difference in democracy | | | | | -0.003 | (0.583) |
| Dispersion democracy differences | | | | | -0.620 | (0.394) |
| Observations | 83646 | | 83646 | | 52387 | |
| Within R2 | 0.000 | | 0.001 | | 0.001 | |

Linear regression with review-fixed effects and standard errors clustered on reviews in parentheses.
Significance levels: ° <.1 * <.05 ** <.01 *** <.001

# A4 Determinants of Recipient Choice: The Data

## A4.1 Data on Official Development Assistance

The data on foreign aid flows used in this paper come from the OECD DAC Creditor Reporting System (CRS). The version of these data used in this draft was updated on the OECD.Stat website on April 28 and downloaded by us on May 6, 2021.[18] We aggregate the CRS's project-level information into yearly figures by donor, recipient, and donor-recipient pairs as necessary. In our aggregation, we only consider flows marked as "ODA Grants" or "ODA Loans"—thus excluding flows marked as "Equity Investment," "Private Development Finance," and "Other Official Flows (non Export Credit)" from our analyses. The resulting data includes information on 47 years, from 1973 to 2019; 95 donors; and 203 recipients (note that 21 of the recipients are "regional" and two are "unspecified").

We use commitment data in million US dollars, 2019 constant amounts (variable `USD_Commitment_Defl`, which for 2019 data is the same as variable `USD_Commitment`). Note that, out of the 3,986,771 project-level observations from 1990 to 2019, 265 have a negative commitment amount (0.0066%); 802,917 have a commitment amount equal to zero (20%); and 1,013,029 are listed as not available (25%). As we aggregate amounts

[18] `https://stats.oecd.org/DownloadFiles.aspx?DatasetCode=CRS1`

from the project level to construct donor-year, recipient-year, and donor-recipient-year totals, we sum across all projects—including those with zero or negative amounts—while the projects with no commitment amount information are dropped.

## A4.2 Data on Recipient Characteristics

### A4.2.1 Quality of Government (QoG) Data

Many of the recipient characteristics included in our models are data from the Quality of Government (QoG) Basic Dataset 2021.[19] Specifically, we use the tenfold classification of colonial origin from Wahman, Teorell and Hadenius (Hadenius and Teorell, 2007; Teorell and Wahman, 2018; Wahman et al., 2013)—variable `ht_colonial` in the QoG Basic Dataset 2021—to construct our indicator of colonial history, which equal to one for recipients who were primarily a former colony of the donor, and zero otherwise.

We measure democracy with the Revised Combined Polity Score (`p_polity2`) from the Polity project (Marshall and Gurr, 2020); and government effectiveness (`wbgi_gee`), rule of law (`wbgi_rle`), and political stability and absence of violence/terrorism (`wbgi_pve`) with estimates from the World Bank's Governance Indicators. We also use measures of life expectancy at birth (`wdi_lifexp`) and GDP per capita in constant 2010 US dollars (`wdi_gdpcapcon2010`) from the Bank's World Development Indicators.

Note that, when including these covariates, we lose observations from the 22 recipients who are not countries in the QoG data.[20] This is a problem for Kosovo, which was visited once by Austria in 2020; and for the West Bank and Gaza Strip, which was visited by Italy in 2000. Since Italy also visited Ethiopia in 2000, that review remains in the sample. However, Austria only visited Kosovo in 2020—this means that lines from the Austria 2020 review remain in the sample for our models, but no recipient is selected with a value of one for the dependent variable once the Austria-2020-Kosovo observation is dropped.

### A4.2.2 UN Security Council Data

We include an indicator for whether a recipient country holds one of the rotating positions in the UN Security Council in the year of a review. These data are from Dreher, Sturm, and Vreeland (2009) and were updated by Dreher, Lang, Rosendorff, and Vreeland (2018).[21]

Note that the following recipients do not appear in these data: Anguilla, Aruba, Bermuda, British Virgin Islands, Cayman Islands, China (People's Republic of), Cook Islands, French Polynesia, Gibraltar,

[19] Retrieved on May 3, 2021, from `http://www.qog.pol.gu.sedoi:10.18157/qogbasjan21`.

[20] Anguilla, Aruba, Bermuda, British Virgin Islands, Cayman Islands, Cook Islands, French Polynesia, Gibraltar, Hong Kong (China), Kosovo, Macau (China), Mayotte, Montserrat, Netherlands Antilles, New Caledonia, Niue, Northern Mariana Islands, Saint Helena, Tokelau, Turks and Caicos Islands, Wallis and Futuna, West Bank and Gaza Strip.

[21] The data were last updated on August 4, 2020, and retrieved by us on May 10, 2021, from `https://www.uni-heidelberg.de/fakultaeten/wiso/awi/professuren/intwipol/datasets_en.html`.

Hong Kong (China), Kosovo, Macau (China), Mayotte, Montenegro, Montserrat, Netherlands Antilles, New Caledonia, Niue, Northern Mariana Islands, Saint Helena, Serbia, South Sudan, Timor-Leste, Tokelau, Turks and Caicos Islands, Tuvalu, Wallis and Futuna, West Bank and Gaza Strip. We manually code this indicator to be equal to one for China in all years—since it is a permanent member of the UNSC.

This means that, when including this indicator, we lose relevant observations for recipients who were visited in the cases of Kosovo and West Bank and Gaza Strip (already discussed); but also Montenegro and Timor-Leste.

#### A4.2.3 World Bank's World Development Indicators Data

We include information on recipients' tourism and statistical capacity from the World Bank's World Development Indicators (WDIs). To measure a recipient's level of tourism, we use the number of tourists who travel to a country other than that in which they have their usual residence, but outside their usual environment, for a period not exceeding 12 months and whose main purpose in visiting is other than an activity remunerated from within the country visited ("International inbound tourists (overnight visitors)" indicator). These data are originally from the *Yearbook of Tourism Statistics* of the World Tourism Organization, and they were last updated on June 30, 2021.

To measure statistical capacity in the recipient country, we use the overall average Statistical Capacity score from the World Bank's *Bulletin Board on Statistical Capacity*, last updated on July 30, 2021. This indicator is a composite score assessing the capacity of a country's statistical system. It is based on a diagnostic framework assessing the following areas: methodology; data sources; and periodicity and timeliness. Countries are scored against 25 criteria in these areas, using publicly available information and/or country input. The overall Statistical Capacity score is then calculated as a simple average of all three area scores on a scale of 0–100.

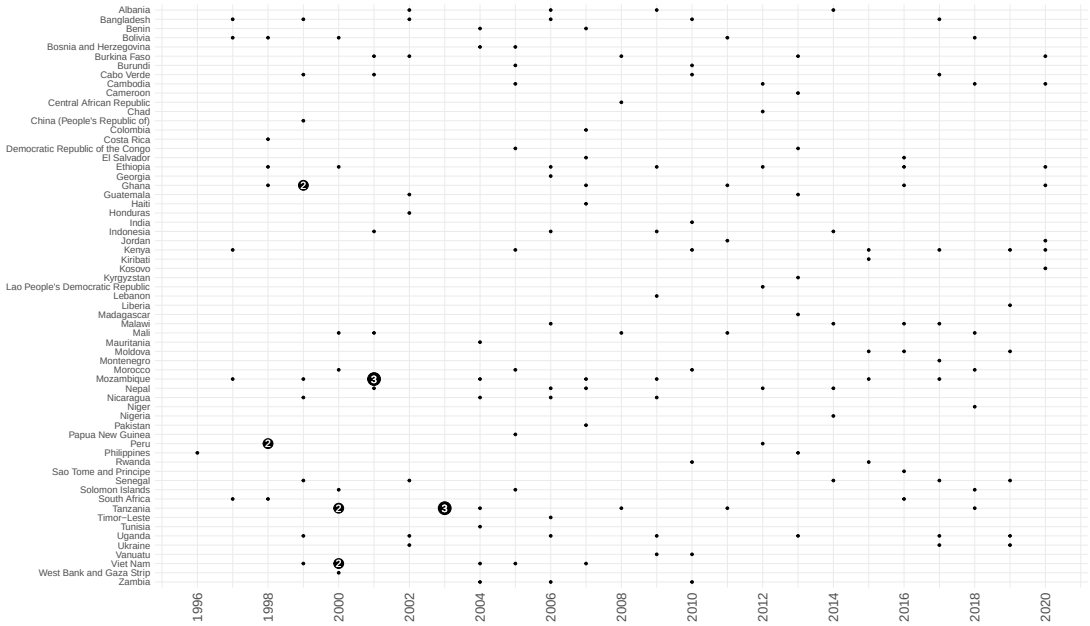
### A4.3 Data on Recipient Assessments

We retrieved information on recipient-specific assessments of the donor's development assistance program from the peer review reports. As depicted in Figure A1, we have access to the full report document for 121 peer reviews, 1996–2020; and 116 of them mention one or more recipient-specific assessment. The following figures provide additional descriptive information about these data.

Figure A3 lists all recipients subject of at least one field visit in the sample of 116 reviews for which we have visit information. Contrary to what outlined in the peer review methodology, it is not uncommon for the same recipient to receive visits for reviews in back-to-back years; and in six instances, the same recipient

was visited for more than one review in the same year, although this has not happened since 2003.

**Figure A3: Recipient Visit Count by Year.** Universe of 177 recipient-specific assessments conducted across our sample of 116 reviews, 1996–2020. Each data points represent one recipient-specific assessment, with dot size and label highlighting recipients who were visited for more than one review in the same year.



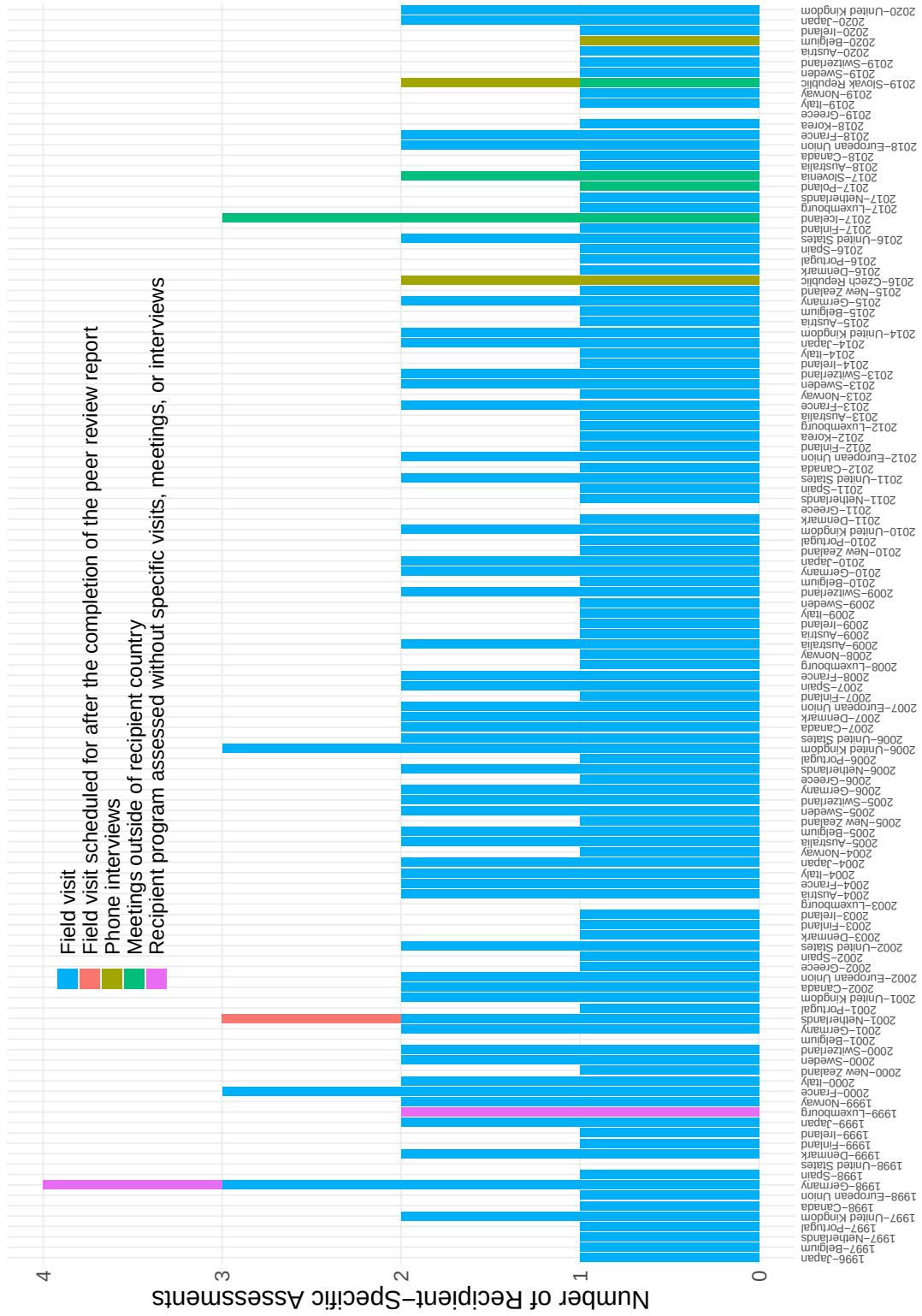



**Figure A4:** Number of Recipient-Specific Assessments per Review, by Type

**Figure A5:** Universe of Recipient-Specific Assessments

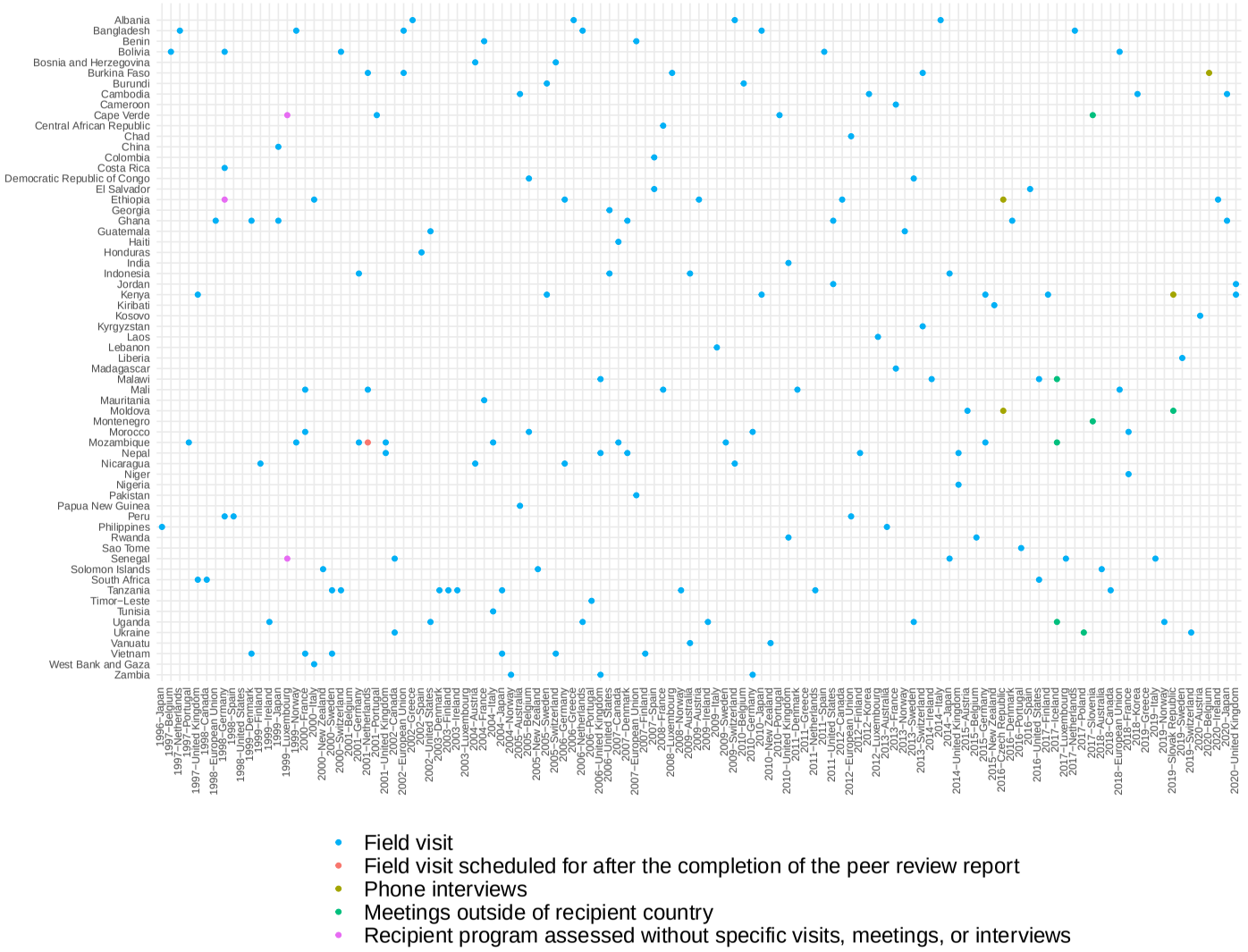


**Figure A6:** Number of Recipient-Specific Assessments per Donor

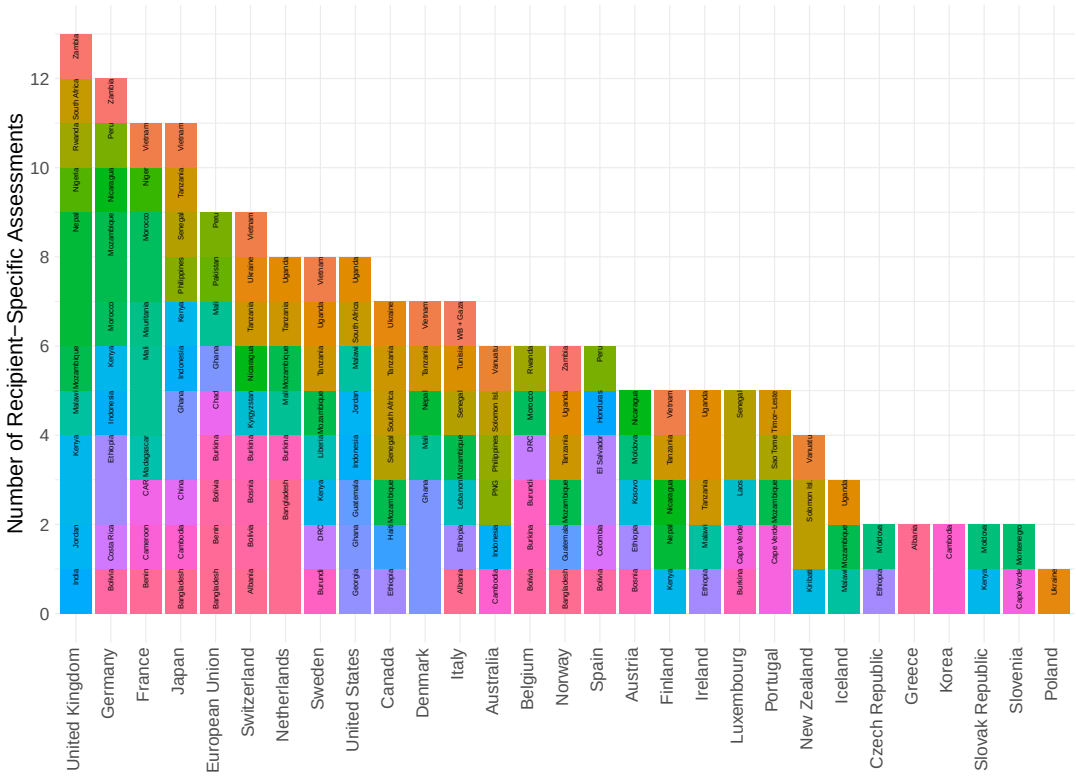



**Figure A7:** Number of Recipient-Specific Assessments per Recipient