City Research Online

City, University of London Institutional Repository

Politics or Performance? Leadership Accountability in UN Peacekeeping

Magnus Lundgren[1] , Kseniya Oksamytna[2], and Vincenzo Bove[3]

Abstract

International organizations face a trade-off between the need to replace poorly performing leaders and the imperative of preserving the loyalty of influential or pivotal member states. This performance-politics dilemma is particularly acute in UN peacekeeping. Leaders of peacekeeping operations are responsible for ensuring that peacekeepers implement mandates, maintain discipline, and stay safe. Yet, if leaders fail to do so, is the UN Secretariat able and willing to replace them? We investigate newly collected data on the tenure of 238 civilian and military leaders in thirty-eight peacekeeping operations, 1978 to 2017. We find that the tenures of civilian leaders are insensitive to performance, but that military leaders in poorly performing missions are more likely to be replaced. We also find evidence that political considerations complicate the UN's efforts at accountability. Holding mission performance constant, military leaders from countries that are powerful or contribute large numbers of troops stay longer in post.

[1]University of Gothenburg, Sweden
[2]King's College London, United Kingdom
[3]University of Warwick, Coventry, United Kingdom

Corresponding Author:
Magnus Lundgren, University of Gothenburg, Box 100, 405 30 Gothenburg, Sweden.
Email: magnus.lundgren@gu.se

Do international organizations (IOs) hold staff in key positions accountable for performance? In appointing and dismissing senior officials, IO secretariats are formally required to hire and fire on merit, but they frequently weigh considerations of merit against the anticipated reactions of the official's country of nationality. This presents IOs with a dilemma. If power and patronage of IO member states offset considerations of merit and can shield officials from accountability for poor performance, it may undercut organizational effectiveness and weaken legitimacy. At the same time, IOs need to preserve the loyalty of member states, whose support is crucial for their operational activities, by keeping their nationals in prestigious or lucrative positions for a significant amount of time.

The dilemma between politics and performance has been investigated in national bureaucracies but is rarely explicitly discussed, let alone systematically studied, in IOs. A key reason is that, at the international level, officials' performance is difficult to observe and member states' interest in specific positions difficult to gauge. The case of UN peacekeeping allows circumventing these barriers. In UN peacekeeping missions, top civilian and military leaders are expected to prevent violence, protect civilians, ensure peacekeepers' discipline, and keep peacekeepers safe. Mission performance can thus be evaluated using indicators such as battle deaths, civilian victimization, peacekeepers' misconduct, or casualties among peacekeepers. Furthermore, since countries contribute troops to specific missions voluntarily, we can estimate their ability and interest in exerting influence regarding particular leadership positions.

We investigate whether political influence or performance affects how long peacekeeping leaders stay in post. Using event history analysis, we examine newly collected data on 238 civilian and military leaders in thirty-eight UN peacekeeping missions between 1978 and 2017. We make three key findings. First, civilian and military leaders' tenures are varyingly sensitive to mission performance. Whereas Force Commanders in poorly performing missions face a higher likelihood of replacement, we observe no such association for Special Representatives of the Secretary-General. Second, Force Commanders are particularly unlikely to endure in post in missions that fail to stop armed violence, while sexual misconduct by peacekeepers does not affect the likelihood of replacement. Third, political considerations influence the tenure of peacekeeping leaders in several ways. Leaders from large troop contributors or permanent members of the Security Council are more likely to endure in post, and they may also be partly shielded from the effects of poor mission performance.

These findings enhance our understanding of IOs in several ways, further developed in the conclusion. To begin with, this is the first quantitative study of how accountability operates in UN peacekeeping. While the academic and policy literatures have discussed peacekeepers' misconduct and their alleged inaction in the face of violence, prior research has not provided a systematic understanding of how the UN Secretariat reacts to such incidents. Second, we demonstrate the existence of politics-performance trade-offs in IO senior staffing, extending this debate from the

national level to the international domain and from the discipline of comparative politics to International Relations. Our results indicate that IO secretariats face a balancing act between preserving the support of key member states by keeping their nationals in the job and increasing effectiveness by replacing leaders of poorly performing units, programs, or missions. The peacekeeping case suggests that this problem is particularly salient in IOs that depend on voluntary provision of resources by member states. Finally, our evidence points to specific forms of states' meddling in international bureaucracies, expanding our understanding of informal influence in IOs and suggesting how it may undermine accountability. An implication is that some prominent frameworks of accountability in global governance (Grant and Keohane 2005), which assume a separation between internal bureaucratic account-ability and external accountability vis-à-vis IO principals, may underestimate the degree to which the two are interconnected.

Performance and Accountability in International Bureaucracies

In recent decades, IOs have become increasingly autonomous, which is reflected in expanding regulatory powers (e.g., Hawkins et al. 2006) and the establishment of large supranational bureaucracies (e.g., Bauer and Ege 2016). These developments have triggered an interest in IOs' handling of their expanding autonomy and a growing literature on issues of accountability (Woods and Narlikar 2001; Verdirame 2011; Campbell 2018; Hirschmann 2020). Different forms of accountability exist in IOs. Chesterman (2008) distinguishes between legal accountability, emerging from compliance with rules, and political accountability, emerging from behavior con-sistent with the preferences of political principals. In an expansive inventory of international accountability structures, Grant and Keohane (2005, 36) identify "hierarchical" accountability as a characteristic of bureaucracies in which "[s]uper-iors can remove subordinates from office, constrain their tasks and room for discre-tion, and adjust their financial compensation." This is also known as "managerial accountability" (Wouters, Hachez, and Schmitt 2011; Kuyama and Fowler 2009). It differs from what Grant and Keohane call "supervisory" accountability, which is the relationship between IO's member states and its bureaucracy.

Much of the discussion of accountability in IOs has focused on supervisory and other forms of political accountability. For that reason, most of what we know about managerial accountability emerges from research on national bureaucracies. A key finding in that literature is that appointments of senior bureaucrats are subject to so-called politics-performance trade-offs (Egorov and Sonin 2011; Gallo and Lewis 2012; Hollibaugh, Horton, and Lewis 2014; Aaskoven and Nyrup 2019).[1] These trade-offs pit political considerations against considerations of merit. On the one hand, executives making bureaucratic appointments seek to reward political allies with lucrative and prestigious postings. On the other hand, executives need appoin-tees who are qualified to perform. This dilemma is mirrored in decisions on

bureaucrats' retention and dismissal. While research shows that performance explains the length of service of cabinet ministers (Berlinski, Dewan, and Dowding 2007), studies of turnover of ambassadors (Arias and Smith 2018) and senior national bureaucrats (Dahlström and Holmgren 2019) highlight political considerations.

There are many parallels between national and international bureaucracies, despite the fact that research on national and international public administration has evolved largely separately (Fleischer and Reiners 2021). Crucially, both national civil services and IOs purport to hire, retain, and promote staff on merit. However, since IOs are composed of individual states, there are inevitable collective action problems, and what is individually rational for one state may not be rational for the membership as a whole. In particular, member states lobby IO secretariats for staff and leadership appointments for their nationals (Dijkstra 2016; Kleine 2013), sometimes in situations where nationals of other countries might be more suitably qualified. As a result, citizens of powerful and rich countries are over-represented in international bureaucracies (Parízek 2017; Novosad and Werker 2019; Oksamytna, Bove, and Lundgren 2021). This is one way in which powerful states exercise what Stone (2011) calls informal influence.

These studies suggest that politics-performance trade-offs might be present in IOs, although they have not been explicitly conceptualized as such. While research has shown that international bureaucracies may anticipate member states' reactions (Martin 1993; Pollack 2003; Oksamytna and Lundgren 2021), to our knowledge, there exist no studies that investigate politics-performance trade-offs in IOs explicitly and empirically. This gap in the literature can be attributed to several difficulties associated with researching this issue. First, studies have analyzed member states' influence on IO staffing in institutional contexts where performance is difficult to observe, leading them to focus on staff appointment rather than retention. Second, studies have focused on areas where member states' preferences regarding specific staff positions are opaque, making it difficult to evaluate dynamics of patronage: for example, Kleine's (2013) study of the European Commission investigates member states' interest in staffing sectoral departments but not specific posts.

The case of UN peacekeeping allows circumventing these barriers. First, several indicators of peacekeeping missions' performance are observable. All peacekeeping leaders face the same primary task of reducing violence, and the decline in conflict is a standard criterion for assessing UN missions. There are also clear expectations regarding the safety of peacekeepers and civilians. Second, because UN peacekeeping needs the support of powerful countries for mission authorization, and because it relies on voluntary and mission-specific troop contributions, we can observe member states' influence regarding specific leadership positions. Such influence can be compared to the patronage dynamics in national bureaucracies: the way in which loyalists are appointed or kept in the job by an executive at the national level is similar to how citizens of powerful or pivotal states obtain or retain peacekeeping leadership posts in IOs. Third, the general procedures of appointments and contract

(non)renewal in UN peacekeeping are comparable to other IOs that operate under the unified personnel policies and practices of the International Civil Service Commission (Renninger 1977). In short, the UN peacekeeping bureaucracy combines transparency with regard to performance, power, and patronage that is rare elsewhere with dynamics of bureaucratic survival that are likely to exist in other IOs.

Performance and Accountability in UN Peacekeeping

UN peacekeeping operations typically have a dual leadership structure, divided along civilian and military lines. The senior civilian officer, the Special Representatives of the Secretary-General (SRSG), is the top diplomat responsible for overall mission leadership, liaising with conflict parties, and ensuring compliance with UN standards. The senior military leader, the Force Commander (FC), is responsible for the deployed forces, including the planning and execution of military operations. Both types of leaders are appointed and replaced by the UN Secretary-General, the head of the UN Secretariat and the organization's chief administrative official. In making these appointments, and in deciding whether to retain peacekeeping leaders, the Secretary-General has considerable discretion but, as set out in Paragraph 101 of the UN Charter, should consider appointees' ability to meet "the highest standards of efficiency, competence and integrity." In short, peacekeeping leaders should be hired and fired on merit.

However, member states may interfere in the appointment process in order to secure peacekeeping leadership posts for their nationals. Oksamytna, Bove, and Lundgren (2021) demonstrate that member states' institutional power, troop contributions, and proximity to the conflict-affected country increase the chances of securing such an appointment. States exert significant effort in placing their citizens in such positions, which may enable them to achieve certain goals: as UN officials know, "the amount of capital that individual Council members would be willing to invest in order to get their person into whatever high-level position is a signal."[2] Achieving those goals requires that nationals stay in those posts for some time, thus delivering a return on their state's investment of political or material capital.

Furthermore, peacekeeping leadership positions are a source of valuable diplomatic or military experience for individuals: such jobs are among the most prestigious in the UN bureaucracy and international diplomacy. For example, Sergio Vieira de Mello left his post of the UN High Commissioner for Human Rights, one of the highest-profile positions in the UN system, to serve as the SRSG in Iraq in 2003.[3] Accruing the benefits of exposure, training, and connections requires not only getting a peacekeeping leadership post but also keeping it for a respectable amount of time.

While both individual diplomats or commanders and their member states have an interest in durable tenure at the helm of a UN peacekeeping operation, the UN Secretariat needs to project an image of a responsible institution that holds its staff accountable for underperformance. The issues of performance and accountability in

peacekeeping are more complex than in many other areas of the UN's work, considering that UN troops can inflict harm by both using and refusing to use force. UN peacekeepers can thus be held responsible for both actions and omissions.

Starting with omissions, peacekeeping missions will be considered to underperform if they fail to achieve their main goals. All peacekeeping leaders face the same primary task of reducing violence, and the decline in armed conflict is a standard criterion for assessing peacekeeping performance (Di Salvatore and Ruggeri 2017; Bove, Ruggeri, and Ruffa 2020). Additionally, there are clear expectations regarding peacekeepers' contribution to civilian safety. A reduction of civilian victimization is an important consideration in assessing peacekeeping effectiveness (e.g., Hultman, Kathman, and Shannon 2013; Bove and Ruggeri 2016; Fjelde, Hultman, and Nilsson 2019; Bove, Ruggeri, and Ruffa 2020; Di Razza 2020). Protection of civilians, an integral element of multidimensional peacekeeping mandates since 1999 (Oksamytna 2021), can be seen as a legal obligation of peacekeepers—and, by extension, of their civilian and military leaders (Wills 2009; Di Razza 2020). Finally, peacekeeping leaders should ensure force protection and prevent attacks on peacekeepers. Avoiding such attacks is essential for not jeopardizing the achievement of core tasks, as they may cause some troop-contributing countries (TCCs) to withdraw their contingents or make peacekeepers reluctant to leave the base, decreasing the mission's ability to reduce violence and prevent civilian victimization. Both civilian and military leaders can be held responsible for force protection failures: whereas civilian heads of mission are "responsible for all aspects of mission management, including security aspects," military leadership also plays an important role in this regard through operational decisions (Willmot, Sheeran, and Sharland 2015, 16).

In terms of actions, peacekeepers' misconduct leads to a starkly negative evaluation of their performance in the eyes of the local population and global audiences. There is a growing literature on peacekeepers' accountability for sexual violence and abuse (e.g., Murphy 2006; Freedman 2018). Instances of gross violations committed personally by peacekeeping leaders, which are thankfully extremely rare, can result in criminal proceedings,[4] or administrative punishment (UN 2015). Of greater relevance is the expectation that leaders ensure that peacekeepers under their management or command do not engage in sexual exploitation or abuse (Zeid 2005).

Overall, peacekeeping leaders are expected to ensure that their missions work to minimize violence, civilian victimization, and security incidents while preventing sexual abuse by peacekeepers. If they do not meet these expectations, the UN Secretary-General will have to take corrective action. In ensuring accountability for poor mission outcomes, replacing an underperforming leader is one of the very few sanctions available to the Secretary-General. Leadership change is also a type of event for which data are publicly available and can be collected systematically.

Most often, replacements of underperforming leaders happen quietly when their contracts are not renewed or a suggestion is made that they leave before the end of the contract. Recently and unprecedentedly, there have been two cases of peacekeeping leaders' public dismissals. The SRSG in the Central African Republic,

Babacar Gaye of Senegal, was dismissed in 2015 over allegations of sexual abuse by peacekeepers in the mission under his leadership (Guardian 2015). The following year, after violence in South Sudan's capital, Juba, killed dozens of civilians and two peacekeepers, a UN investigation concluded that "a lack of leadership on the part of key senior Mission personnel had culminated in a chaotic and ineffective response" (UN 2016, §7). The Secretary-General discharged Kenyan FC Johnson Mogoa Kimani Ondieki. The Kenyan government's reaction illustrates member states' efforts to protect their nationals: after the FC's dismissal, Kenya withdrew its peace-keepers from South Sudan because "[t]he process leading to this unfortunate deci-sion not only lacked transparency but did not involve any formal consultation with the Government of Kenya" (Reuters 2016, §6). In Di Razza's (2020, 12) words, "the political backlash has been fierce."

After the incident, the SRSG also departed, but her replacement happened quietly. While the SRSG "also faced criticism for her handling of the Juba crisis," the Secretariat chose to "brand her departure as a voluntary and planned exit from the mission; a far cry from the public dismissal of the Force Commander" (Spink 2016, 24). All types of departures are recorded in our data. While peacekeeping leaders might choose to leave themselves, such events are likely rare and randomly distributed,[5] considering the benefits, both monetary and reputational, that accrue to the officeholders, as well as the stigma associated with a premature departure[6] and pride associated with the ability to endure in post.[7]

We acknowledge that peacekeeping performance does not depend solely on mission leadership but also reflects the quality of contributed troops, actions by the conflict parties, and other factors. Replacing leaders of poorly performing missions might not always be fair (Spink 2016), and peacekeeping officials have voiced "concerns about the attributability of success or failure to senior leaders" (Lottholz and von Billerbeck 2019, 29; see also Di Razza 2020). However, replacements of leaders of poorly performing missions would represent the Secretariat's efforts to promote, or at least appear as promoting accountability. If this were indeed the case, we would expect poor mission performance to be associated with shorter tenures of peacekeeping leaders, whom the Secretariat replaces earlier than leaders of solidly performing missions. This argument leads to the following hypothesis:

H1: Leaders in poorly performing missions have shorter tenures.

At the same time, in light of the literature on politics-performance trade-offs, we expect the Secretariat to consider factors other than performance in deciding how long a peacekeeping leader stays in post. For example, Secretary-General Boutros-Ghali discovered that "bureaucrats enjoying the protection of the Council's veto-wielding members raise a peculiar set of problems" (Salton 2017, 168), probably constraining his freedom to exercise formal powers of dismissal. A Secretary-General who displeases powerful member states risks a loss of funding (for example, the US did not pay its UN budget dues in full in the 1990s), institutional deadlock, or,

like in Boutros-Ghali's case, a denial of re-appointment. Replacing peacekeeping leaders from powerful countries before they have had a chance to spend some time in post can worsen the Secretariat's relations with those member states. In line with this dynamic, we expect that leaders hailing from powerful countries are more difficult to replace than leaders from less powerful states, including in situations when the mission underperforms. We thus seek to test the following second hypothesis:

**H2:** Leaders from powerful countries have longer tenures.

Next to the pressure from powerful countries, we expect the Secretariat to be susceptible to the influence of TCCs. The TCCs' ability to supply or withhold troops, which they provide voluntarily, gives them leverage, which has already been shown to matter for peacekeeping leadership appointments (Oksamytna, Bove, and Lundgren 2021). By similar logic, when contemplating replacement of a peace-keeping leader from a major TCC, the Secretariat has to consider the implications for that country's willingness to volunteer troops. The abovementioned example of Kenya's withdrawal from South Sudan following the dismissal of its FC is a case in point. Even in cases of quiet replacements, such decisions can have serious reper-cussions if the peacekeeping leader hails from a major TCC. India's withdrawal from Sierra Leone in 2000 was partly "prompted by Indian unhappiness with Secretary-General Kofi Annan's decision to replace General Jetley as the Force Commander" (Murthy 2007, 161). In line with this reasoning, we expect that influ-ence flowing from mission-specific patronage in the form of troop contributions can make leadership replacement more difficult, and also mitigate the negative effect of poor performance. We thus propose the third hypothesis:

H3: Leaders from large troop contributors have longer tenures.

Assessing whether mission performance, power, or troop contributions affect the length of tenure of civilian and military leaders of UN peacekeeping operations will reveal whether the UN Secretariat is able and willing to replace leaders that do not, for one reason or another, deliver results that are publicly expected, or whether the Secretariat prioritizes keeping powerful or important countries content by retaining their nationals in peacekeeping posts for a substantial amount of time.

Data and Methods

To evaluate our theoretical expectations, we construct a dataset on the tenure of 89 SRSGs and 149 FCs in thirty-eight UN peacekeeping missions between 1978 and 2017.[8] This dataset covers the majority of UN missions deploying during this time (see online appendix, Table A1) and records biographical details of leaders, the dates of tenure, and mission characteristics. Figures A1 and A2 in the online appendix provide details on the distribution of leadership posts across nationalities (see also Oksamytna, Bove, and Lundgren 2021).
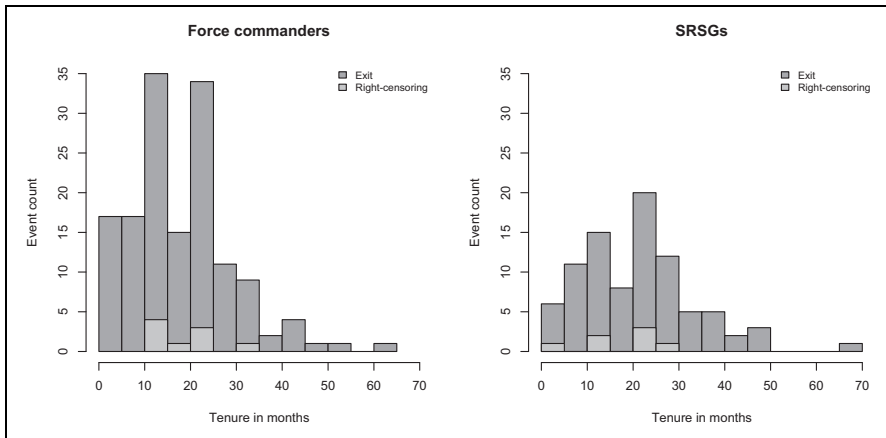
Figure 1. FCs and SRSGs tenures.

Given our interest in measuring the time to leader exit, we employ an event history framework (Cox and Oakes 1984; Freedman 2008). Event history analysis is conventionally used in International Relations to study the time it takes for an event to occur, such as the duration of a peace agreement or how fast the UN can generate sufficient troops for its peacekeeping missions (Lundgren, Oksamytna, and Coleman 2020). In our case, we study the duration of leaders' tenures. The unit of analysis is mission-leader-month. Leaders enter the risk set upon appointment and exit at the time of departure from post. We code the event history of leaders as 0 until they depart, 1 in the month of departure, and otherwise missing. Officials that remained in post at the end of 2017 are right-censored.

Figure 1 exhibits data on the length of tenure of FCs and SRSGs. We note that few FCs last longer than thirty-six months and extraordinarily few more than forty-eight months; the tenures of SRSGs are somewhat longer. The median FC tenure is twenty-one months, whereas that of an SRSG is twenty-five months. The distribution of FC tenures exhibits peaks at twelve and twenty-four months, likely indicating a pre-agreed rotation schedule, a pattern that is less clear for SRSGs. Overall, we observe that most leaders exit at other times, suggesting variability that may reflect replacement driven by considerations of performance, power, and contributions.

Figure 2 presents the associated Kaplan-Meier estimates. While the FCs and SRSGs have similar survival rates in the first year of tenure, SRSGs are subsequently at a lower risk. The twenty-four-month survival probability is 0.41 for FCs and 0.49 for SRSGs. At thirty-six months, the survival probabilities are estimated at 0.10 and 0.17, respectively. The difference is statistically significant at the $p < 0.05$ level. In other words, civilian leaders tend to stay in post longer than their military counterparts. While these statistics do not provide evidence in favor of any of our hypotheses, they suggest that the replacement of FCs and SRSGs are subject to partly
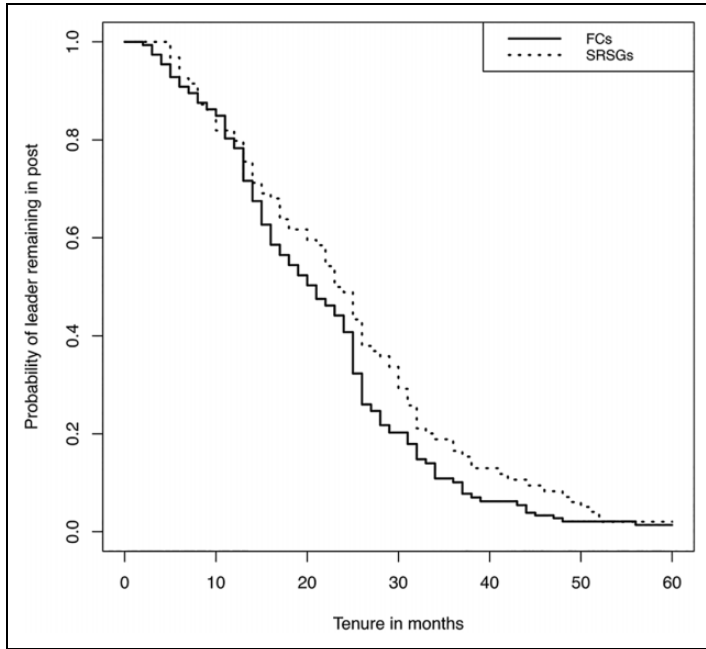
Figure 2. Kaplan-Meier estimates of leadership tenures.

diverging mechanisms, possibly reflecting differentiated responsibilities and varia-
tions in their countries of origin. (SRSGs are more likely to come from powerful
countries).

To evaluate our expectations regarding the determinants of FC and SRSG
tenures, we use Cox proportional hazards models, estimating the probability of
leader exit in a given month, provided that exit has not yet occurred. Since observa-
tions are clustered within missions, they are likely to exhibit correlated outcomes
due to unobserved factors that affect all leaders of the same mission. To model such
cluster-specific homogeneity, we cluster errors on missions and, in further exten-
sions, include shared frailty terms for missions, equivalent to random effects in a
multilevel framework (Therneau and Grambsch 2013). On the explanatory side, we
include covariates to represent our theoretical concepts of mission performance,
power, and contributions, as well as measures to control for confounding factors.

When it comes to measuring mission performance, we employ two separate
approaches. In the first, we create a *mission performance index*, weighing in three
key indicators of how well a mission meets the expectations about its ability to
prevent negative events: battle deaths, one-sided violence against civilians, and
attacks against mission. While these are imperfect proxies that might not capture

mission performance fully or accurately, the Secretariat cannot ignore such visible and salient measures when evaluating peacekeeping leaders' performance.[9] A broader measure of performance is motivated by an expectation that several factors simultaneously contribute to the perception of mission success (or a lack thereof). In the second, disaggregated approach, we examine whether and how *constituent indicators* of the index extend or shorten the duration of leaders' tenures. This approach also allows assessing the role of peacekeepers' discipline, as measured by the absence of sexual exploitation of the local population (Karim and Beardsley 2016). We do not include it in the main index because the data are less complete than for the other components; it also pertains to the behavior of peacekeepers *per se*, rather than outcomes that are not always within peacekeepers' control.

The construction of our mission performance index was inspired by several considerations (cf. Greco, Ishizaka, and Tasiou 2018). From a conceptual standpoint, we wanted an index that could help us identify leaders of underperforming missions, relative to missions that performed better with respect to the key goals of UN peacekeeping. From a methodological standpoint, we wanted a performance measure that allowed for comparison across leaders in different missions, took diverse local conditions into account, and was based on observable and readily available indicators.

The mission performance index incorporates three constituent indicators. The first and the second indicators relate to the mission's contribution to a positive security environment, operationalized as battle deaths and civilians killed in one-sided attacks, both from the Uppsala Conflict Data Program (UCDP) (Sundberg and Melander 2013).[10] The third indicator relates to performance in terms of the safety of personnel, operationalized as attacks against peacekeepers (Henke 2019). These are suitable performance indicators not only because they reflect the core components of peacekeeping mandates, but also because they tend to receive extensive media coverage, making bad performance visible to the public.

To enable comparison across leaders in different missions, the index is constructed with an emphasis on the relative deterioration in mission performance during a leader's tenure. This is achieved in four steps. First, for each leader, we create three time-series, representing the cumulative count of battle deaths, attacks against civilians, and attacks against the mission during the leader's tenure. Second, we standardize each leader's time-series of cumulative counts, so as to place them on a comparable scale. Third, for each leader, we sum the three standardized time-series, producing one leader-specific time-series that represents the accumulation of adverse events during the leader's tenure. Fourth, to ease interpretation, we inverse the scores (so that good performance equals high scores) and rescale so that 0 is the lowest value and 100 the highest value. Leaders in missions that have few or no adverse events will have scores at or close to 100, whereas leaders in missions that have many and substantial adverse events on these metrics will have lower scores. As with the other variables, the index is measured on a monthly basis (see Figure A5 in the online appendix).

The mean monthly mission performance index score is seventy-six for FCs and sixty-six for SRSGs. Good performers—leaders whose missions experience no dramatic worsening on the three variables—tend to have scores above ninety, whereas bad performers (here defined as the lowest quartile) have scores lower than fifty.

While we use a rather simple and intuitive measure, we recognize that performance indices can be constructed in several other ways. In our robustness checks, we present results for two alternative indices, which imply different assumptions about how adverse events affect the Secretariat's perceptions of mission performance.

More broadly, our indicators constitute a first step toward measuring mission performance, and a few caveats and clarifications are in order. First, both our index and the disaggregated measures focus on negative events, implying that "good performance" is only captured to the extent it involves the prevention of such events. In UN peacekeeping, this is a reasonable assumption, but other policy domains may require alternative approaches.

Second, it is a measure of mission performance and not of individual performance. We recognize that a mission's ability to stem violence or prevent other negative events does not depend only on leaders' actions. However, we expect that trends in their average survival in post, over many missions and longer periods of time, will be shaped by such factors on account of their overall responsibility for the mission.

Third, violence in the host country can be compounded by factors other than mission performance. A mission may lack the mandate or means to address violence; even a perfectly resourced operation experiences some adverse events. However, the selected indicators match what peacekeeping operations seek to accomplish through skillful "good offices" of the SRSG and an effective military strategy by the FC. Even if only a portion of the observed violence is related to mission performance, variation in that portion of violence, across missions and leaders, allows us to evaluate its correlation with the risk of replacement.

Fourth, there is a possibility of endogeneity. A leadership change can affect the level of violence in the host country if, for example, it temporarily reduces the mission's ability to operate effectively. However, our data indicate that SRSGs and FCs are rarely replaced at the same time and that leaders are often immediately followed by a successor. Another risk is that local actors intensify violence in anticipation of a leadership change. But given that most transitions are unforeseen and not publicly communicated before they are executed, this is likely rare. It is also possible that tenure, *per se,* affects the performance of the operation because leaders accumulate knowledge and learn how to make the mission successful. Yet, this is likely to have an impact only in the very long run. In light of these concerns, and although we control for a large number of potential confounding factors, our estimates do not necessarily demonstrate a causal mechanism and should thus not be interpreted in a causal fashion.

To evaluate H2, we assess the power of the leader's country based on two indicators. The permanent five members of the Security Council (*P5*) enjoy a disproportionate influence in UN peacekeeping (Allen and Yuen 2014). *Economic resources* provide another form of power that can be wielded to influence staffing decisions in IOs, for example, by supporting costly lobbying for appointments. We operationalize economic resources based on Gleditsch's (2002; updated through 2017) GDP data.

To evaluate H3, we measure mission-specific *troop contributions* based on monthly UN data. The variable records the number of troops supplied by the leader's country to the observed mission, reflecting its direct contribution. In our sensitivity tests, we evaluate the impact of *general troop contributions*, operationalized as the sum of troops provided by a country, across all UN missions and years, through to the year of observation.

In H2 and H3, we expect power and troop contribution to have an independent effect on leaders' tenure. Additionally, we expect them to mitigate or reduce the effect of bad performance on tenure by shielding nationals from accountability.

We include a vector of controls to adjust for possible confounding. Our performance measures emphasize longitudinal shifts in performance within an observed leader's tenure. Nevertheless, recognizing that missions experience different enabling and constraining circumstances, we account for missions' mandates, which Di Salvatore et al. (2020) show to have a considerable effect on peacekeeping outcomes. In particular, missions with a *Chapter VII* mandate are likely to have authorization for the use of force, thus making them riskier than other missions.

Since our sample includes terminated missions, we include a variable for *end of mission*, coded as 1 in the last month of the mission's existence. Since leaders' tenure ends at this date, regardless of performance, failing to control for this event would skew the results. Drawing on qualitative research, we also coded for leaders appointed in an *acting* (or interim) capacity; such leaders naturally have shorter tenures. In our extended tests, we also control for the impact of fixed-term contracts, by including variables coded 1 for the twelfth and twenty-fourth month of a leader's tenure.

While individual characteristics are difficult to observe, we control for one leader-level feature. *Previous experience* is coded as 1 for leaders who have previously headed or commanded a UN peacekeeping mission in our data and 0 otherwise. The expectation is that more experienced leaders are likely to survive longer in post.

Several variables are right-skewed. To reduce the risk that outlier observations are disproportionately influential, we log all continuous variables (after adding 1 to all values) in Table 1. All performance variables and troop contributions are measured at the monthly level; other variables have yearly measurements. Summary statistics (Table A2), correlation matrices (Table A3), and the distribution of key independent variables (Figure A3) are reported in the online appendix.

Table I. Cox Proportional Hazard Models of FC/SRSG Survival in Post.

	Dependent Variable							
	FC Exit				SRSG Exit			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Performance (index)	-0.014***	-0.012***	-0.010	-0.015***	-0.004	-0.004	-0.008	-0.004
	(0.005)	(0.005)	(0.033)	(0.005)	(0.004)	(0.004)	(0.027)	(0.003)
P5 country	-1.039***	2.554**	-1.028***	-1.019***	0.458	0.440	0.464	0.458
	(0.307)	(1.120)	(0.294)	(0.315)	(0.389)	(0.671)	(0.377)	(0.392)
GDP	0.211***	0.220***	0.227	0.211***	-0.053	-0.053	-0.071	-0.053
	(0.048)	(0.051)	(0.162)	(0.048)	(0.072)	(0.072)	(0.098)	(0.073)
Troop contributions	-0.038*	-0.048**	-0.038*	-0.072	0.016	0.015	0.014	0.014
	(0.022)	(0.024)	(0.023)	(0.111)	(0.048)	(0.046)	(0.047)	(0.075)
Previous experience	-0.131	-0.071	-0.127	-0.133	-0.315	-0.314	-0.316	-0.315
	(0.203)	(0.222)	(0.214)	(0.208)	(0.197)	(0.202)	(0.199)	(0.197)
Chapter VII	0.149	0.160	0.148	0.151	-0.061	-0.060	-0.046	-0.061
	(0.201)	(0.203)	(0.205)	(0.202)	(0.240)	(0.242)	(0.253)	(0.238)
Mission end	3.684***	3.659***	3.689***	3.669***	3.815***	3.816***	3.825***	3.815***
	(0.266)	(0.248)	(0.271)	(0.264)	(0.317)	(0.318)	(0.332)	(0.321)
Acting	3.669***	3.625***	3.668***	3.668***	1.673***	1.672***	1.670***	1.673***
	(0.564)	(0.557)	(0.563)	(0.567)	(0.374)	(0.372)	(0.372)	(0.375)
Performance × P5		-0.048***				0.0003		
		(0.013)				(0.006)		
Performance × GDP			-0.0003				0.0003	
			(0.003)				(0.002)	
Performance × Troop contributions				0.0005				0.00003
				(0.001)				(0.001)
Observations	2,496	2,496	2,496	2,496	1,705	1,705	1,705	1,705
Log Likelihood	-377.741	-376.314	-377.737	-377.685	-208.297	-208.296	-208.287	-208.297
LR Test	143.047***	145.901***	143.055***	143.159***	98.961***	98.961***	98.980***	98.961***

Note: Log hazards. Negative coefficients signify lower risk of exit. Robust errors clustered on missions. Two-tailed tests. Efron method for ties.
*p < 0.1.
**p < 0.05.
***p < 0.01

Results

We present our results in Tables 1 and 2. Models 1 to 4 in Table 1 relate to FCs and Models 5 to 8, which are analogously specified, to SRSGs. Models 1 and 5 are base models, including our composite measure of mission performance and variables representing power and contributions. In Models 2 to 4 and 6 to 8, we add interactions, seeking to gauge how the association between performance and survival is conditioned by power and contributions. In Table 2, we disaggregate the mission performance index and estimate coefficients for each constituent indicator, as well as for sexual exploitation by peacekeepers. The online appendix contains a wider set of results, as detailed below. We report log hazards; coefficients with a negative sign represent lower risk (and therefore longer tenures) and vice versa.

If H1 is correct, we would observe a negative association between *mission performance* and exit probabilities. Leaders in missions that perform well should be at a lower risk of exit; those in underperforming missions should be at a higher risk. As indicated by the negative coefficients on the mission performance variable in Model 1, our data support this intuition for FCs, which means that military leaders in missions that experience a larger number adverse events are at higher risk of replacement. The base SRSG model yields a negative point estimate, as expected, but the coefficient is not statistically significant. This suggests that the tenures of military and civilian leaders vary in their sensitivity to mission performance.

To illustrate the relationship between performance and FC tenures, Figure 3 plots the predicted hazards.[11] The slope indicates that good mission performance is associated with a lower hazard rate (and hence, a lower probability of exit). While dramatic shifts in performance can be associated with very substantive changes in the predicted hazard, most observations are focused in the middle of the range. In this range, all else equal, moving from the first quartile (sixty-three) to the third quartile of the performance index (ninety) is associated with a 15 percent reduction of the hazard rate for FCs. In substantive terms, an FC in a mission maintaining a first quartile performance level can look forward to an estimated twenty-three months on the job, compared with twenty-six months for an FC in a mission at the third quartile. (Figure A4 in online appendix illustrates this difference in the form of predicted Kaplan-Meier curves).

Consistent with H2, the P5 covariate is negatively associated with exits in Model 1, pointing to longer tenures for commanders from countries that are permanent members of the Security Council. This might suggest that in relatively rare cases when the P5 provide FCs (see Figure A2), they seek to protect their nationals to realize strategic objectives they have in the host country, including through the FC appointment, and shield the reputation of their military. At the mean level of mission performance, the predicted tenure is thirteen months longer for FCs who come from the P5, which is consistent with the expectation that the P5 lobby the Secretariat for longer tenures for their nationals (Figure 4). There is no evidence of a similar

systematic association for SRSGs, further reinforcing the impression that the tenures of these two leader types are governed by partly different mechanisms.

The positive coefficient on *GDP* in Model 1 indicates that FCs from more economically powerful countries have shorter tenures than FCs from other countries. As can be seen in Model 4, there is no significant association between GDP and SRSG tenures. These results are contrary to our expectation and suggest that the theorized mechanism—that economically powerful countries successfully lobby to retain their leaders in post—lacks support in the data.[12] Understanding the negative association between GDP and military leaders' tenures will require further research. It is possible, for example, that poorer countries, which might derive financial benefits from UN peacekeeping, or at least obtain other benefits of peacekeeping participation at a low cost (Coleman and Nyblade 2018), are more adamant about shielding their leaders in order to protect their reputation and ensure continuing invitations to participate in peacekeeping missions.

Examining H3, we find some evidence in favor of our proposition that mission-specific troop contributions extend leaders' tenures. The estimates vary in precision, but the negative coefficient in Model 1 indicates that larger contributions are associated with longer tenures for FCs (significant at the 90 percent level). Substantively, increasing contributions from 0 to 500 (6.21 when logged) troops reduces the FC hazard rate by about 20 percent. We observe no significant association for SRSGs. The stronger association for FCs is consistent with the informal understanding that FC appointments are closely linked to mission contributions (Jakobsen 2016), whereas those of SRSGs are less so.

Our theory suggested that performance, power, and contributions have independent effects on the length of tenure. It additionally suggests the possibility of conditional effects, i.e. that leaders from powerful countries or large TCCs are treated differently, despite serving in missions with the same level of performance. To gauge how the impact of performance is conditioned by power and contributions, we estimate interaction models (Models 2 to 4, 6 to 8 in Table 1). This allows us to examine if the effect of performance is mediated by political considerations. The negative coefficient on the interaction in Model 2 suggests that coming from a P5 country provides a further protective function for FCs, diminishing the impact of weak mission performance on the risk of exit.[13] By contrast, interacting performance with GDP or mission-specific troop contributions suggests that the coefficient for performance is not significantly altered by these factors. For SRSGs, we find no evidence that the coefficient for performance is conditioned by power or contributions.

Turning briefly to our control variables, we note that leaders appointed in an acting capacity have shorter tenures, as expected. All else equal, the predicted median tenure of an acting FC is a mere five months and twelve months for an acting SRSG. Leaders with previous experience have longer tenures, on average, but
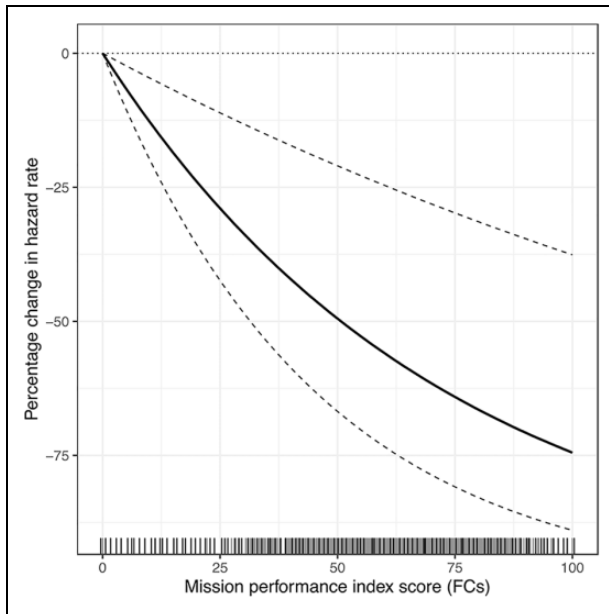
Figure 3. Percentage change in hazard rate for FCs as a function of mission performance. Predicted values with 95 percent confidence intervals. Calculations based on Model 1.

the estimates are too noisy to raise our confidence in a systematic association. Chapter VII mandates do not appear to be correlated with leaders' tenures.

Overall, our theoretical propositions receive mixed support in this first round of tests. With regard to FCs, the data are consistent with key parts of our theoretical argument. For SRSGs, the evidence is considerably weaker. None of the privileged variables is a systematic predictor of the duration of SRSG tenures. Thus, next to the results regarding FCs, a key finding is that the tenures of FCs and SRSGs are varyingly sensitive to observable mission conditions. The divergence between FCs and SRSGs underlines the unique responsibilities of the two roles: They are held accountable for different aspects of mission performance, a question to which we return below. It is also possible that SRSGs are evaluated on the basis of idiosyncratic and "positive" achievements, like promoting inter-ethnic understanding or resolving gridlocks between local actors, which are less likely to leave an imprint on a measure geared toward picking up adverse events.

To further deepen our analysis, we report, first, an extensive set of robustness tests, including alternative performance measures, and, second, an extension in which we disaggregate our performance index into its constituent parts.
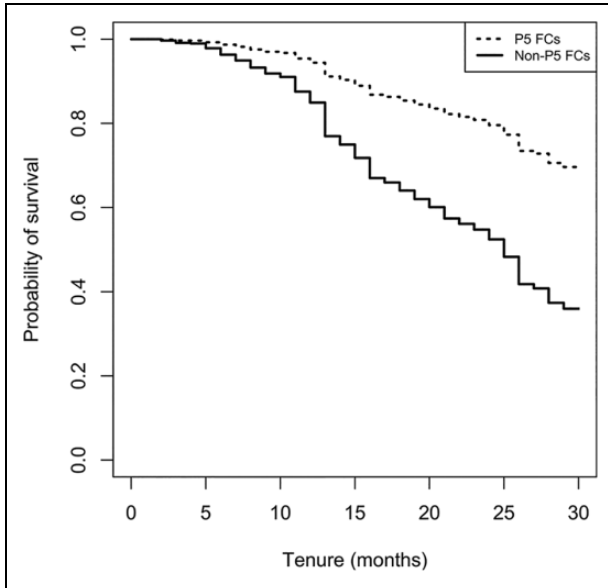
Figure 4. Predicted tenures of FCs, P5 and non-P5 countries. Prediction based on Model 1; all other covariates held at mean or reference level.

Robustness

To ensure that our results are not driven by particularities of model specification, sample selection, or operationalization, we undertake a number of robustness tests (Tables A4 through A15 in the online appendix).

First, we construct two alternative measures of mission performance and re-estimate our main models (Table A4). These measures differ from the base measure by emphasizing recent performance (defined as the last three months) and "good" and "bad" performance, respectively. The "good" variable is the count of recent months in which a mission has at least 50 percent fewer adverse events than its historical average. The "bad" variable is the count of months with at least 50 percent more adverse advents than the historical average. Given the way these measures are constructed, the number of observations shrinks, so we interpret the results with additional caution. Table A4 shows that FCs in missions that perform better than their historical average are more likely to stay in post, whereas FCs in missions that perform worse than this average are more likely to be replaced. Using these alternative measures reinforces our beliefs that the duration of FC tenures is sensitive to mission performance. As for SRSGs, recent "good" performance does not appear to affect tenures whereas "bad" performance lowers the risk of exit. Taken together, these results reinforce the conclusion that FCs and SRSGs differ not only with regard

to their average time in office but also with regard to the determinants of their tenures.

Second, we examine whether accountability rose to prominence after the end of the Cold War, as peacekeeping missions became more active and riskier. We re-estimate our models excluding observations prior to 1990. The results in Table A5 indicate that the results are indeed robust.

Third, we examine additional possible confounders (Table A6). Countries with extensive diplomatic networks can lobby in support of their nationals, while experienced diplomatic corps could translate into higher quality of appointees, leading to better overall performance. Per capita income could also confound the analyses if wealthier countries provide appointees with a potential for better performance due to investments in education and training. General troop contributions by TCCs, across all previous missions, could be another overlooked source of influence. Controlling for these additional variables, however, does not significantly affect our main results.

Fourth, we exclude the last six months of each mission from the sample (Table A7). When a mission is about to wrap up, the UN might be reluctant to replace top officials, regardless of mission performance, because of transition costs and the difficulty of finding qualified individuals for a job of very limited duration. All key results are robust to this alteration of the models.

Fifth, we test if our results hold up if we remove four observer missions from our sample (Table A8). These are smaller missions with limited mandates that may differ from the rest of the missions. Excluding them does not change the behavior of our key variables.

Sixth, our results could be driven by outlier observations. This concern is most clear with regard to the mission performance variable, which contains some outliers. Removing the observations with the ten lowest performance scores and the observations with perfect performance (no negative events) does not significantly change the results (Table A9). The main exception is that, with the best performers removed, we observe a positive relationship between performance and tenure length for SRSGs, suggesting wider support for our theory in this subset of the data. Removing observations that are statistically influential (deviance residual exceeding 2.5) yields similar results (Table A10).

Seventh, we replicate Table 1 using alternative modeling strategies. This involved a discrete time (logit) model and alternative methods for ties (Breslow rather than Efron) (Tables A11 to A12).

Eighth, we test for fixed-term contracts using discrete time logit models (Table A13). Including a time-varying variable coded as 1 in months twelve and twenty-four of a leader's tenure indicates that FCs, especially, are at higher risk in such months. This indicates the presence of formal rotations for FCs (as already illustrated by the descriptive figures).

Finally, all reported variables satisfy the proportional hazards assumption except GDP and previous experience in the main SRSG model (Model 4, Table 1).

Including time interactions for these variables, as the literature suggests (Box-Steffensmeier and Jones 2004), does not change the key results (Table A14). The results also indicate that SRSGs from countries with a higher GDP are likely to endure longer in post, but this pattern attenuates with time. To show that our interaction results are not sensitive to multicollinearity, we also present a table in which the performance variable has been standardized (Table A15).

Disaggregating Performance

While a broad index likely provides the most credible measure of how mission performance is perceived, disaggregating the index into its constituent parts, as we do in Table 2, allows us to identify specific drivers of early exits. We again estimate identically specified models for FCs and SRSGs and add two models that include a covariate for sexual exploitation, which was excluded from the main index for reasons of data availability.

The results suggest that the components of our performance index are varyingly associated with leader exits.[14] Leaders in missions that fail to stem violence, as measured by cumulative battle fatalities, have shorter tenures. The coefficient is positive in Model 9, indicating that increasing fatalities predict shorter tenures for FCs. For SRSGs, there is a positive association, but it is not statistically significant. Since violence reduction is the core goal of UN peacekeeping, these results strengthen our belief in a relation between mission performance and leader tenures. The stronger result for FCs is also plausible: battle deaths are linked to military performance, where FCs rather than SRSGs have more control. Figure 5 plots the predicted probabilities. Moving from a first quartile (0) to a third quartile fatality level (400 fatalities) is associated with an 18 percent higher risk of FC exit; moving to 1,000 fatalities increases the risk by 40 percent.

We note that for FCs, the other two performance indicators are estimated close to zero and with high variance, suggesting that their relationship to tenure varies considerably. This would imply, first, that the previously observed result for the performance index is largely driven by the strong impact of battle deaths and, second, that accountability is more selective than expected. Likewise for SRSGs, these results indicate that the weak association between the performance index and tenure is likely due to the underlying indicators pulling in different directions. While battle deaths and attacks against the mission are unrelated to tenure length, attacks against civilians are negatively associated with SRSG exits.

Finally and importantly, our results demonstrate that sexual exploitation by peacekeepers is not associated with shorter tenures of peacekeeping leaders. We recognize that the estimates are based on incomplete data, so we have limited statistical power, but the negative coefficient in Model 10 means that FCs in missions with an incidence of sexual violence have longer tenures than missions without, all else equal. The effect size is small and it is significant only at the 90 percent

level, but the results are clearly counter to theoretical as well as ethical expectations. Likewise, while the measure is positive for SRSGs, indicating that sexual exploitation by peacekeepers increases the risk of SRSG exit, it is not statistically significant. Despite the single example of dismissal (see above), these results lend credence to the accusations that the UN is not doing enough to stop sexual abuse by Blue Helmets (Zeid 2005; Freedman 2018).

Table 2. Survival of FCs and SRSGs (Disaggregated Performance Measures).

	Dependent Variable			
	FC Exit		SRSG Exit	
	(9)	(10)	(11)	(12)
Battle-related fatalities	0.402***		0.085	
	(0.116)		(0.060)	
Attacks against mission	−0.088		0.173	
	(0.108)		(0.161)	
Attacks against civilians	−0.141		−0.214*	
	(0.124)		(0.127)	
Sexual exploitation		−0.003*		0.001
		(0.001)		(0.0004)
P5 country	−0.947**	−1.835**	0.620*	0.244
	(0.376)	(0.873)	(0.322)	(0.710)
GDP	0.164***	0.239**	−0.073	−0.329***
	(0.052)	(0.113)	(0.064)	(0.115)
Mission troop contribution	−0.022	−0.110	−0.004	0.042
	(0.029)	(0.078)	(0.055)	(0.063)
Previous experience	−0.157	−0.632	−0.278	0.584*
	(0.200)	(0.391)	(0.193)	(0.349)
Chapter VII	0.247	0.908	−0.044	−0.557
	(0.173)	(0.593)	(0.221)	(0.401)
Mission end	3.732***	24.448***	3.757***	24.171***
	(0.217)	(0.514)	(0.322)	(0.404)
Acting	3.770***		1.666***	
	(0.532)		(0.327)	
Observations	2,496	918	1,705	749
Log Likelihood	−378.449	−82.452	−207.148	−56.022
LR Test	141.631***	31.133***	101.257***	24.824***

Note: Log hazards. Negative coefficients signify lower risk of exit. Robust errors clustered on missions. Two-tailed tests. Efron method for ties.
*$p < 0.1$.
**$p < 0.05$.
***$p < 0.01$.

Conclusion

Does performance of senior leaders in international organizations affect the length of their tenure? Are efforts to ensure performance accountability hampered by considerations of power or patronage? This is of crucial importance in UN peacekeeping, where missions can not only fail but also inflict harm on people they are sent to assist. Based on data from thirty-eight peacekeeping missions, 1978 to 2017, we investigated whether the UN Secretariat holds peacekeeping leaders accountable for mission performance and whether such decisions are influenced by political considerations.

Our findings suggest that the UN Secretariat balances performance and politics. On the one hand, the evidence demonstrates that military leaders' tenures are sensitive to overall mission performance, in particular to mounting battle deaths. This suggests that the UN Secretariat is making efforts to ensure performance accountability, especially with regard to the core peacekeepers' goal of violence reduction. On the other hand, these efforts appear to be frustrated by political concerns. We find that permanent Security Council membership and troop contributions lengthen Force Commanders' tenures, suggesting that adequately performing military leaders are unlikely to be prematurely replaced if their country is a major UN decision-maker or partner. We also find that the replacement risks that flow from mission
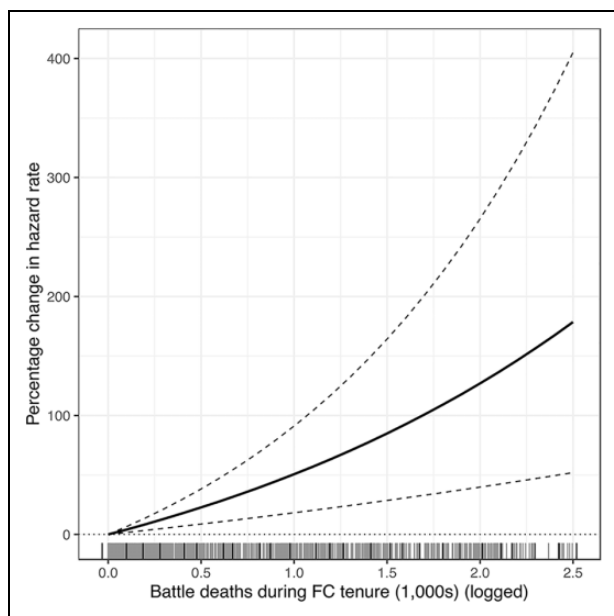
Figure 5. Percentage change in hazard rate for FCs as a function of battle deaths during tenure. Predicted values with 95 percent confidence intervals. Calculations based on model 9.

underperformance are reduced for Force Commanders from powerful countries. In addition, the differences between military and civilian leaders are more pronounced than we would expect. While civilian heads of mission are considered to have the overall responsibility for the operation, our findings indicate that mission performance does not affect the length of their tenure, in contrast to what we observe for Force Commanders.

These findings extend our understanding of IOs in three ways. First, we show that IOs can be subject to the type of politics-performance trade-offs that affect national civil services. IO bureaucracies are not impartial, technocratic bodies but arenas for political struggles. While in national bureaucracies, the fault lines are party-political, the divisions in IOs are between member states with divergent interests: IO secretariats need to anticipate how decisions affect their relationships with up to 193 member states, who wield different types of power and provide different kinds of support. In deciding on the tenures of peacekeeping leaders, the UN Secretariat weighs the benefits of replacing a poorly performing leader against the risks that their country might withdraw political or material support.

Second, by pointing to specific ways in which power and patronage operate in IOs, our findings contribute to the debate on accountability in global governance. The ability of powerful states to shield nationals from accountability can have deleterious consequences for IO performance and legitimacy. While the literature on unilateral influence argues that powerful states pursue their interests in IOs at the expense of the rest of the membership (Stone 2011), we extend this argument to cover pivotal countries on whose resources IOs depend. Major providers of voluntary resources can leverage their contributions to protect nationals from replacements. These findings also suggest that Grant and Keohane's (2005) distinction between hierarchical accountability, operating between levels of an IO bureaucracy, and supervisory accountability, operating between the international bureaucracy and its member state principals, is blurred. Member states, especially if they wield power or provide key resources, attempt interference that distorts hierarchical accountability.

Finally, our findings contribute to the debate on accountability in UN peacekeeping. In the wake of scandals triggered by sexual exploitation or failures to protect civilians, many wonder whether the Secretariat is willing and able to hold peacekeeping leaders accountable. While our results suggest cautious optimism, the Secretariat appears to be more concerned about the performance of military leaders than their civilian counterparts. This may, however, suggest that civilian leaders are evaluated based on achievements that are harder to quantify. Additionally, while the data on the relevant variable are very limited, it is concerning that sexual exploitation is not linked with the length of leaders' tenure.

Our study demonstrates that further research on performance and accountability in IOs is necessary. We suggest two avenues. First, while we detected politics-performance trade-offs in decisions on UN peacekeeping leaders' replacement, our knowledge of international bureaucracies can be improved

by investigating whether this dilemma is present in other IOs. Second, more precise and nuanced measures of IO leaders' performance can be developed. While our choice of variables was motivated by a belief that mission performance is a reasonable proxy for individual performance in the eyes of the UN Secretariat and the broader public, that UN missions' core objectives focus on security-related indicators, and that negative events have the most pronounced effect on perceptions of performance, we invite colleagues to think about other ways of measuring performance. These can improve our understanding of IO accountability, especially of civilian leaders who are expected to promote positive change in subtle and gradual ways by building relationships with national and international counterparts.

Declaration of Conflicting Interests

Funding

ORCID iD

Magnus Lundgren 🆔 https://orcid.org/0000-0002-9961-3645

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Also known as "patronage-expertise" or "loyalty-competence" trade-offs.
2. Interview with a UN peacekeeping official, January 2020. The interviews cited here were conducted in 2020 to 2021 for a project by one of the authors looking at a different aspect of UN peacekeeping, but some officials volunteered comments on the selection and tenure of UN officials and leaders.
3. Although the UN Assistance Mission for Iraq (UNAMI) is a political mission, the example demonstrates the value that officials attach to field leadership posts in difficult environments.
4. UN personnel enjoys various privileges and immunities that mean that most prosecutions take place in their home rather than host country (UN 2015).
5. In one tragic instance, a Brazilian FC committed suicide, instead of asking for reassignment or waiting for replacement, after "a week of intense public criticism of the Brazilian-led United Nations Mission in Haiti" (Thompson 2006, §4).
6. When the FC of the UN mission in the Democratic Republic of the Congo (MONUC) left his post after only three weeks, media coverage was starkly negative: "The hurried departure of the Spanish MONUC General Vicente Diaz de Villegas also created a bad

impression. He arrived in Congo at the end of September and didn't even last four weeks in the country before quitting his post" (Knaup 2008 §6).

7. A senior leader, who remained the head of a difficult mission for several years after several predecessors could not last a year, viewed the length of tenure as a achievement. Interview with a UN peacekeeping official, January 2021.

8. We extended the data presented in Bove, Ruggeri, and Zwetsloot (2016).

9. A similar approach has been used by Aaskoven and Nyrup (2019, 17) in their investigation of economic performance and bureaucratic promotions under the Nazi regime. They do not assume that senior bureaucrats "were promoted directly based on the economic data," only that such data were likely to serve as "a reasonable proxy for local economic performance, which the leadership of the Nazi Party was able to assess."

10. We refer the reader to this and the following study for a comprehensive discussion of the variables making up the index.

11. Simulations performed with Gandrud's (2015) procedures and software. All continuous covariates held at mean; dummies held at reference category.

12. The result is not due to multicollinearity. We ran diagnostic tests to detect potential multicollinearity in all our models. In Tables 1 and 2, variance inflation factors (VIFs) for all independent variables of theoretical interest are well within conservative limits (<5).

13. Due to the included interaction and the construction of the performance variable (only positive values), the coefficient on P5 shifts sign in Model 2. If the performance variable is standardized (Table A15) it remains negative.

14. It is not uncommon for individual indicators of composite indices to exhibit varying or no relationship to an external variable, even if a correlation between the index and the external variable exists. Cf. Greco, Ishizaka, and Tasiou (2018).

References

Aaskoven, Lasse, and Jacob Nyrup. 2019. "Performance and Promotions in an Autocracy: Evidence from Nazi Germany." Accessed 28 October 2020. https://dx.doi.org/10.2139/ssrn.3308807.

Allen, Susan Hannah, and Amy T. Yuen. 2014. "The politics of peacekeeping: UN Security Council oversight across peacekeeping missions." *International Studies Quarterly* 58 (3): 621-632.

Arias, Eric, and Alastair Smith. 2018. "Tenure, Promotion and Performance: The Career Path of US Ambassadors." *Review of International Organizations* 13 (1): 77-103.

Bauer, Michael W., and Jörn Ege. 2016. "Bureaucratic Autonomy of International Organizations' Secretariats." *Journal of European Public Policy* 23 (7): 1019-37.

Berlinski, Samuel, Torun Dewan, and Keith Dowding. 2007. "The Length of Ministerial Tenure in the United Kingdom, 1945–97." *British Journal of Political Science* 37 (2): 245-62.

Bove, Vincenzo, and Andrea Ruggeri. 2016. "Kinds of Blue. Diversity in UN Peacekeeping Missions and Civilian Protection." *British Journal of Political Science* 46 (3): 681-700.

Bove, Vincenzo, Andrea Ruggeri, and Chiara Ruffa. 2020. *Composing Peace: Mission Composition in UN Peacekeeping*. Oxford, UK: Oxford University Press.

Bove, Vincenzo, Andrea Ruggeri, and Remco Zwetsloot. 2016. "What Do We Know about UN Peacekeeping Leadership?" *International Peacekeeping* 24 (1): 17-23.

Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event history modeling: A guide for social scientists*. Cambridge, UK: Cambridge University Press.

Campbell, Susanna. 2018. *Global Governance and Local Peace: Accountability and Performance in International Peacebuilding*. Cambridge, MA: Cambridge University Press.

Chesterman, Simon. 2008. "Globalization Rules: Accountability, Power, and the Prospects for Global Administrative Law." *Global Governance* 14 (1): 39-52.

Coleman, Katharina, and Benjamin Nyblade. 2018. "Peacekeeping for Profit? The Scope and Limits of "Mercenary" UN Peacekeeping." *Journal of Peace Research* 55 (6): 726-41.

Cox, D. R., and David Oakes. 1984. *Analysis of Survival Data*. London, UK: Chapman and Hall.

Dahlström, Carl, and Mikael Holmgren. 2019. "The Political Dynamics of Bureaucratic Turnover." *British Journal of Political Science* 49 (3): 823-36.

Dijkstra, Hylke. 2016. *International Organizations and Military Affairs*. Abingdon, UK: Routledge.

Di Razza, Namie. 2020. *The Accountability System for the Protection of Civilians in UN Peacekeeping*. New York: International Peace Institute.

Di Salvatore, Jessica, Magnus Lundgren, Kseniya Oksamytna, and Hannah Smidt. 2020. "Introducing the Peacekeeping Mandates (PEMA) Dataset." *Social Science Research Network*. Accessed June 24, 2021. http://dx.doi.org/10.2139/ssrn.3703503.

Di Salvatore, Jessica, and Andrea Ruggeri. 2017. "Effectiveness of Peacekeeping Operations." *Oxford Research Encyclopedia of Politics*. Accessed October 28, 2020. https://doi.org/10.1093/acrefore/9780190228637.013.586.

Egorov, Georgy, and Konstantin Sonin. 2011. "Dictators and Their Viziers: Endogenizing the Loyalty–Competence Trade-off." *Journal of the European Economic Association* 9 (5): 903-30.

Fjelde, Hanne, Lisa Hultman, and Desirée Nilsson. 2019. "Protection Through Presence: UN Peacekeeping and the Costs of Targeting Civilians." *International Organization* 73 (1): 103-31.

Fleischer, Julia, and Nina Reiners. 2021. "Connecting International Relations and Public Administration: Toward A Joint Research Agenda for the Study of International Bureaucracy." *International Studies Review*. doi: 10.1093/isr/viaa097.

Freedman, David A. 2008. "Survival Analysis: A Primer." *The American Statistician* 62 (2): 110-19.

Freedman, Rosa. 2018. "UNaccountable: A New Approach to Peacekeepers and Sexual Abuse." *European Journal of International Law* 29 (3): 961-85.

Gallo, Nick, and David E. Lewis. 2012. "The Consequences of Presidential Patronage for Federal Agency Performance." *Journal of Public Administration Research and Theory* 22 (2): 219-43.

Gandrud, Christoper. 2015. "simPH: An R Package for Illustrating Estimates from Cox Proportional Hazard Models." *Journal of Statistical Software* 65 (3): 1-20.

Gleditsch, Kristian Skrede. 2002. "Expanded Trade and GDP Data." *Journal of Conflict Resolution* 46 (5): 712-24.

Grant, Ruth W., and Robert O. Keohane. 2005. "Accountability and Abuses of Power in World Politics.' *American Political Science Review* 99 (1): 29-43.

Greco, Salvatore, Alessio Ishizaka, and Menelaus Tasiou. 2018. "On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness." *Social Indicators Research* 141:61-94.

Guardian. 2015. "UN Peacekeeping Chief in CAR Sacked over Sex Abuse Claims." 12 August. Accessed July 20, 2018. https://www.theguardian.com/world/2015/aug/12/un-peacekeeping-chief-central-african-republic-resigns-babacar-gaye.

Hawkins, Darren G., David A. Lake, Daniel L. Nielson, and Michael J. Tierney, eds. 2006. *Delegation and agency in international organizations*. Cambridge, UK: Cambridge University Press.

Henke, Marina E. 2019. "UN Fatalities 1948–2015: A New Dataset." *Conflict Management and Peace Science* 36 (4): 425-42.

Hirschmann, Gisela. 2020. *Accountability in Global Governance: Pluralist Accountability in Global Governance*. Oxford, UK: Oxford University Press.

Hollibaugh, Gary E. Jr, Gabriel Horton, and David E. Lewis. 2014. "Presidents and Patronage." *American Journal of Political Science* 58 (4): 1024-42.

Hultman, Lisa, Jacob Kathman, and Megan Shannon. 2013. "United Nations Peacekeeping and Civilian Protection in Civil War." *American Journal of Political Science* 57 (4): 875-91.

Jakobsen, Peter Viggo. 2016. "Denmark and UN Peacekeeping: Glorious Past, Dim Future." *International Peacekeeping* 23 (5): 741-61.

Karim, Sabrina, and Kyle Beardsley. 2016. "Explaining Sexual Exploitation and Abuse in Peacekeeping Missions." *Journal of Peace Research* 53 (1): 100-15.

Kleine, Mareike. 2013. "Trading Control: National Fiefdoms in International Organizations." *International Theory* 5 (3): 321-46.

Knaup, Horand. 2008. "UN Peacekeepers' Credibility Is Damaged." 3 November. *Spiegel International*. Accessed October 27, 2020. https://www.spiegel.de/international/world/chaos-in-congo-un-peacekeepers-credibility-is-damaged-a-588091.html.

Kuyama, Sumihiro, and Michael Ross Fowler. 2009. "Introduction." In *Envisioning Reform: Enhancing UN Accountability in the Twenty-first Century*, edited by Sumihiro Kuyama and Michael Ross Fowler, 1-18. Tokyo: United Nations University Press.

Lottholz, Philipp, and Sarah von Billerbeck. 2019. *Senior Leadership Performance Management in International Organizations*. Reading, UK: University of Reading.

Lundgren, Magnus, Kseniya Oksamytna, and Katharina Coleman. 2020. "Only as Fast as Its Troop Contributors: Incentives, Capabilities, and Constraints in the UN's Peacekeeping Response." *Journal of Peace Research*. doi: 10.1177/0022343320940763.

Martin, Lisa. 1993. "International and Domestic Institutions in the EMU Process." *Economics and Politics* 5 (2): 125-44.

Murphy, Ray. 2006. "An Assessment of UN Efforts to Address Sexual Misconduct by Peace-keeping Personnel." *International Peacekeeping* 13 (4): 531-46.

Murthy, C. S. R. 2007. "Unintended Consequences of Peace Operations for Troop-contributing Countries from South Asia." In *Unintended Consequences of Peacekeeping*, edited by Chiyuki Aoi, Cedric de Coning, and Ramesh Thakur, 156-70. Tokyo: UN University Press.

Novosad, Paul, and Eric Werker. 2019. "Who Runs the International System? Nationality and Leadership in the United Nations Secretariat." *Review of International Organizations* 44 (1): 1-33.

Oksamytna, Kseniya. 2021. "Advocacy and Change in International Organizations: Protection, Communication, and Reconstruction in UN Peacekeeping." Manuscript in preparation.

Oksamytna, Kseniya, Vincenzo Bove, and Magnus Lundgren. 2021. "Leadership Selection in United Nations Peacekeeping." *International Studies Quarterly* 65 (1): 16-28.

Oksamytna, Kseniya, and Magnus Lundgren. 2021. "Decorating the 'Christmas Tree': The UN Security Council and the Secretariat's Recommendations on Peacekeeping Operations." *Global Governance* 27 (2): 226-250.

Parízek, Michal. 2017. "Control, Soft Information, and the Politics of International Organizations Staffing." *Review of International Organizations* 12 (4): 559-83.

Pollack, Mark. 2003. *The Engines of European Integration: Delegation, Agency, and Agenda Setting in the EU*. Oxford, UK: Oxford University Press.

Renninger, John P. 1977. "Staffing International Organizations: The Role of the International Civil Service Commission." *Public Administration Review* 37 (4): 391-96.

Reuters. 2016. "Kenya Says Withdrawing Its Troops from a U.N. Mission in South Sudan." 2 November. Accessed October 27, 2020. https://uk.reuters.com/article/uk-kenya-southsudan-idUKKBN12X2FA.

Salton, Herman T. 2017. *Dangerous Diplomacy: Bureaucracy, Power Politics, and the Role of the UN Secretariat in Rwanda*. Oxford, UK: Oxford University Press.

Spink, Lauren. 2016. "Accountability for Peacekeeping Failures Must Be Shared by the UN in New York." In *Global Peace Operations Review*, edited by Center on International Cooperation, 24-25. New York: Center on International Cooperation.

Stone, Randall W. 2011. *Controlling Institutions: International Organizations and the Global Economy*. Cambridge, UK: Cambridge University Press.

Sundberg, Ralph, and Erik Melander. 2013. "Introducing the UCDP Georeferenced Event Dataset." *Journal of Peace Research* 50 (4): 523-32.

Therneau, Terry M., and Patricia M. Grambsch. 2013. *Modeling Survival Data: Extending the Cox Model*. Ann Arbor, MI: Springer.

Thompson, Ginger. 2006. "Chief of U.N. Troops in Haiti Is Found Dead in Hotel Room." 8 January. *New York Times*. Accessed October 27, 2020. https://www.nytimes.com/2006/01/08/world/americas/chief-of-un-troops-in-haiti-is-found-dead-in-hotel-room.html.

United Nations. 2015. "Special Measures for Protection from Sexual Exploitation and Sexual Abuse: Report of the Secretary-general." *A/70/729*. Accessed June 24, 2021. https://undocs.org/A/70/729.

United Nations. 2016. "Executive Summary of the Independent Special Investigation into the Violence in Juba in 2016 and the Response by the United Nations Mission in South Sudan." *S/2016/924*. Accessed June 24, 2021. https://undocs.org/S/2016/924.

Verdirame, Guglielmo. 2011. *The UN and Human Rights: Who Guards the Guardians?* Cambridge, MA: Cambridge University Press.

Willmot, Haidi, Scott Sheeran, and Lisa Sharland. 2015. *Safety and Security Challenges in UN Peace Operations*. New York: International Peace Institute.

Wills, Siobhán. 2009. *Protecting Civilians: The Obligations of Peacekeepers*. Oxford, UK: Oxford University Press.

Woods, Ngaire, and Amrita Narlikar. 2001. "Governance and the Limits of Accountability: The WTO, the IMF, and the World Bank." *International Social Science Journal* 53 (170): 569-83.

Wouters, Jan, Nicolas Hachez, and Pierre Schmitt. 2011. "Managerial Accountability." In *International Organizations and the Idea of Autonomy*, edited by Richard Collins and Nigel D. White, 230-56. London, UK: Routledge.

Zeid, Ra'ad Zeid Al-Hussein. 2005. "A Comprehensive Strategy to Eliminate Future Sexual Exploitation and Abuse in United Nations Peacekeeping Operations." *A/59/710*. Accessed June 24, 2021. https://undocs.org/A/59/710.