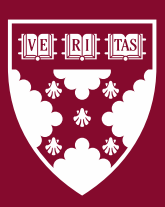
Working Paper 24-002

Managing Remote Work Quality: Evidence from Management Systems Standards Auditing

Ashley Palmarozzo
Michael W. Toffel



Harvard Business School

Managing Remote Work Quality: Evidence from Management Systems Standards Auditing

Ashley Palmarozzo
Harvard Business School

Michael W. Toffel
Harvard Business School

Working Paper 24-002

We gratefully acknowledge the research assistance provided by Melissa Ouellet.

Funding for this research was provided in part by Harvard Business School.

Managing Remote Work Quality: Evidence from Management Systems Standards Auditing

Ashley Palmarozzo
Harvard Business School
Boston MA 02163
apalmarozzo@hbs.edu

Michael W. Toffel†
Harvard Business School
Boston MA 02163
mtoffel@hbs.edu

July 12, 2023

Remote work has become more common, providing operational flexibility and productivity benefits, but questions remain about whether and how it affects quality. This study investigates the quality effects of remote work in a diagnostic service context in which remote work separates workers from the subject of their work: remotely auditing sites' compliance with management system standards. We analyze nearly 30,000 audits conducted in-person or remotely across thousands of sites around the world by auditors of one global company during 2019–2021, when remote auditing accelerated due to COVID-19 pandemic travel restrictions. We theorize that remote audits will be of lower quality (less comprehensive) than in-person audits because remote auditors face greater difficulties (a) obtaining information critical to detecting violations and (b) coordinating and exchanging information with fellow auditors. We find evidence of these theorized mechanisms: remote audits report fewer violations and quality problems are especially pronounced for (a) standards clauses in which auditors assess compliance via direct observation as opposed to document review and (b) audits conducted by multi-auditor teams. We also find the quality problems of remote audits partially mitigated when the auditors had previously conducted in-person audits of the site and when auditors' prior experience was more concentrated on the standard being audited. Understanding mechanisms by which remote work can erode quality and revealing attributes that exacerbate or attenuate such quality concerns can help companies and regulatory agencies better manage remote work.

Keywords: remote work, quality, audits, inspections, standards

† Corresponding author. We gratefully acknowledge the research assistance provided by Melissa Ouellet and the financial support from Harvard Business School's Division of Research and Faculty Development.

Managing Remote Work Quality:
Evidence from Management Systems Standards Auditing

1. Introduction

Remote work has become increasingly common across industries as varied as healthcare, education, information technology, and financial services (US Bureau of Labor Statistics 2022). Many expect remote work, accelerated by the COVID-19 pandemic, to endure and perhaps constitute 20% of workdays in the post-pandemic US economy (Barrero et al. 2021). Remote work has been shown to improve job satisfaction and productivity and reduce costs, but uncertainty remains about how it affects the quality of the work being done. The few studies of the quality implications of remote work compared situations—responding to phone calls or reviewing documents—in which workers changed their locations but not their work processes (Bloom et al. 2015, Choudhury et al. 2021).

But remote work often entails more fundamental changes in how work is conducted, with in-person interactions becoming technology-mediated via telephone, video, and online chats, as with bank tellers, customer service desks, healthcare appointments, and regulatory inspections. Moreover, when these transitions allow workers to work from home or "from anywhere," they insert physical separation not only between service providers and their "customers," but also between the service providers themselves. Both of these factors can reduce work quality.

We examine the quality implications of remote work in the context of auditing management system standards, a service context that—much like social audits of global supply chain factories, quality audits of franchisees to assess contract compliance, and regulatory inspections assessing occupational health and safety, environmental, food safety, and process quality practices—has traditionally involved onsite visits during which one or more auditors conduct face-to-face interviews with employees, conduct site tours to observe operations, and review physical documents. When conducting audits remotely, these interactions are mediated by technology such as telephones, video conferencing, and document sharing

applications. We study how the shift from in-person to remote work affects the quality of *management system standard certification assessments*, which are audits of the extent to which a site has implemented procedures required by one of six of the world's most popular voluntary management system standards governing quality, environment, and occupational health and safety.

Our research is timely, given the ongoing debate about the quality of remote workers conducting audits and regulatory inspections. Enthusiasts include both certification auditors and audited businesses, 80% of whom indicated in a 2021 survey that "remote procedures give the same confidence as onsite audits" and the majority of whom were satisfied with their remote audit experience and wanted to see "continued or increased use of remote activities in the future" (IAF, ILAC, and ISO 2021: 1–2). But this is not a consensus. A survey of environment, health, and safety (EHS) professionals indicated that while most had shifted at least some EHS audits from in-person to remote, "concern over the quality of results was by far the greatest impediment to starting a virtual auditing program"—in particular, concern that they might yield "less thorough/robust results" (NAEM 2021: 18). Another survey found internal auditors concerned that auditing their own company's operations remotely makes it more difficult to access information and produce quality audits (Internal Audit Foundation and AuditBoard 2021). Governments, too, are concerned, as exemplified by a US Department of Labor (2021: 2) report suggesting that remote inspections conducted by the Occupational Safety and Health Administration during the COVID-19 pandemic risked hazards for employees going "unidentified and unabated longer."

We theorize that two mechanisms will lead remote auditing to produce lower-quality (less comprehensive) audits by limiting an auditor's ability to gather sufficient evidence of violations and therefore documenting fewer violations than are truly occurring. First, auditors face greater difficulty accessing information at the audited site because they often rely on information that is physically embedded in the site to gather evidence of violations. While in-person auditing traditionally entails site tours to gather evidence of violations, remote auditing replaces these with video-mediated site tours, which do not convey certain contextual information such as smells, sounds, and fine-grained visual clues. Suspicious smells cannot be detected over Zoom, for example. Second, we theorize that remote auditing

will produce lower-quality audits because it becomes harder for auditors to access information gathered by their fellow audit team members. Audit team members frequently exchange information throughout the audit, coordinate individual activities, and assess audit progress in real time. We theorize that there will be less of this if auditors aren't actually onsite together, making it harder for them to piece together clues of noncompliance.

We also theorize that information access challenges with remote auditing will be attenuated—and thus the quality decrement of remote audits reduced—when auditors have greater prior in-person exposure to the audited site. Such auditors will be able to rely on experience-based heuristics that can prompt important questions even in the absence of onsite cues.

To test our theory, we obtained data from a large company that audits sites around the world to assess compliance with various management system standards. Specifically, we examine the 35,000 audits conducted in-person or remotely during 2019–2022 that pertained to six of the world's most popular management system standards: ISO 9001 Quality Management Systems standard, ISO 14001 Environmental Management Systems standard, ISO 27001 Information Technology Management Systems standard, OHSAS 18001 and ISO 45001 Occupational Health and Safety Management Systems standards, and ISO 13485 Medical Devices Quality Management Systems standards. Sixty-one percent of these audits were conducted fully in-person and 39% fully remotely. Most of these sites were in the United Kingdom, China, India, and the US; the rest were in 100 other countries.

Our empirical analysis indicates that remote audits yielded an average of 25% (or 0.40) fewer violations than in-person audits and we found evidence of each of our two theorized mechanisms. First, we investigated whether remote auditing quality suffers due to auditors facing greater challenges accessing information that leads to citations of violations of standards clauses, for which auditors primarily gather data via direct observations during site tours. Exploiting the fact that auditors rely on different auditing methods as the primary mechanism to identify violations of particular management system standard clauses, we examine whether the quality deficit of remote audits is especially pronounced for those standards clauses for which auditors tend to identify violations via direct observation—that is,

where remote auditors are at a particular disadvantage compared to in-person auditors—versus those clauses for which auditors tend to identify violations via document review—where remote and in-person audit access is more similar. Consistent with our theory, we find that the audit-quality decrement of remote audits is significantly larger for clauses for which auditors typically rely on direct observation to identify violations than for clauses for which auditors typically use document review to identify violations. Second, we find that remote auditing quality concerns are exacerbated when audits are conducted by teams of auditors compared to those conducted by a single auditor. This is consistent with the second theorized mechanism: remote auditing makes it harder for auditors to coordinate and share information so as to comprehensively identify violations.

We find two factors—both auditor attributes—that partially attenuate quality concerns of remote auditing. Specifically, we find that such concerns are less severe for audits conducted by auditors with (a) more prior in-person auditing exposure to the audited site or (b) more concentrated auditing experience with the management system standard being audited.

Our work contributes to the remote work literature by finding evidence that working remotely can impede quality. While prior research found no quality difference in remote work conducted by telephone customer service representatives or patent examiners (Bloom et al. 2015, Choudhury et al. 2021), remote workers in those contexts—responding to telephone inquiries or reviewing online documents—are no more separated from the callers or the patent documents than when they work in the office. Our context, however, separates auditors from in-person engagement with audited sites and with audit team members. Taken together, our results suggest that the quality implications of remote work depend on the extent to which working remotely changes work processes. Managers should therefore be especially careful to guard against quality loss when remotely conducting work, such as medical exams or safety inspections, that traditionally required gathering information in-person but where remote work separates the worker from the subject of their work.

Our research also contributes to the literature on monitoring quality. While there is significant scholarship on audit-, auditor-, and institution-level factors that affect audit quality for in-person audits,

our study is, to our knowledge, the first to examine whether varying the audit format affects quality. It also offers insights for quality monitoring beyond the auditing of management standards, such as assessing retail franchisees' compliance with franchiser terms and global supply chain factories' adherence to brands' supplier codes of conduct. Understanding the relationship between audit format and quality is especially timely given the growing interest in remote auditing in regulatory, supply chain, and standard certification contexts (IAF, ILAC, and ISO 2021, Mofid et al. 2021, US Department of Health and Human Services 2022). Given that remote auditing is useful in reducing travel time and costs and increasing flexibility, our results can guide audit providers on how to staff audits to minimize the associated quality concerns; for example, by focusing the use of remote audits on smaller sites that can be audited by one auditor. Moreover, as explained in the Discussion section, audit providers can use our results to develop a hybrid auditing approach, such as by performing document review remotely and observational activities onsite or by focusing remote audits on less risky sites.

2. Related Literature

Our work relates to the literature comparing work done remotely to that done in-person and to the literature on the effectiveness of monitoring the management practices of companies' business partners.

2.1. Remote Work

Allowing employees to work remotely (such as working from home and working from anywhere) has been found to lead to higher worker productivity and job satisfaction, lower attrition (Bloom et al. 2015, Choudhury et al. 2021),[1] and more efficient labor allocation and improved customer service (Sun et al. 2020). Another stream of remote work research has focused on banks, retailers, and hospitals offering online options to supplement their in-person service, which has been found to lead to increased customer retention, greater use by customers of *offline* service channels, and more overall customer interactions

[1] Tan and Netessine (2020) also find that introducing tabletop technology in restaurants to assist in ordering increased waiter productivity and sales per table.

(Buell et al. 2010, Campbell and Frei 2010, Xue et al. 2011, Bavafa et al. 2018, Bell et al. 2018, Rajan et al. 2019).[2]

Closer to our research are studies of how working remotely affects work quality; these have yielded mixed results. No difference in work quality was found between call center customer service representatives who worked from home versus the office (Bloom et al. 2015) or between patent examiners working from home versus "working from anywhere" (Choudhury et al. 2021).[3] But Bettinger et al. (2017) found worse educational outcomes when students took a course online compared to those who took it in-person, theorizing that the difference was due to the online students having less instructor oversight and thus being less motivated. Sun et al. (2020) found that hospital emergency room physicians remotely delivering healthcare services improved one dimension of service quality—reducing patient length-of-stay, both by reducing waiting time (treatment delays) and by more flexibly allocating the (remote) physicians to patients—with no erosion in care quality (readmission rates and in-hospital mortality rates). But another healthcare study found better patient health outcomes resulted when doctors supplemented in-person visits with virtual channels to communicate with patients (Bavafa et al. 2018).

Our study also examines the quality implications of remote work, but differs along two dimensions. First, prior studies theorize that differences in the quality of work performed remotely versus in-person result from differences in service delays (Sun et al. 2020) and oversight (Bettinger et al. 2017), whereas these dimensions do not meaningfully differ between remote versus in-person work in our auditing context. Second, prior studies that compare the quality of office work performed remotely versus in-person consider scenarios in which worker locations differed but the work did not entail in-person interaction with customers or colleagues either way. When Bloom et al. (2015) compared the quality of

[2] One study that examined the reverse transition—a retailer supplementing its online store by adding in-person showrooms—found that sales conversions (calls that resulted in customers placing orders) increased and product returns decreased, improvements that resulted from changes in customer behavior (Bell et al. 2018).
[3] Specifically, Bloom et al. (2015) found no difference in sales conversion rates or quality assessments of the recorded conversations and Choudhury et al. (2021) found no difference in rework or the number of citations patent examiners added during their patent reviews.

work between those working from the office to those working from home, the work entailed conducting sales calls via telephone. Similarly, Choudhury et al. (2021) compared remote patent examiners working from home to those working from anywhere, but in both cases the work entailed examining documents. We, however, compare work (auditing) conducted in-person at sites—with employees sharing information and providing site tours—to work conducted remotely—with auditors working offsite and engaging with onsite employees and with their own team members only via audio/video technology. Our setting therefore enables us to compare a particular form of remote work—where workers and customers are not co-located—to the in-person scenario in which they are co-located. This provides greater variation in coordination and information flow, which we theorize will affect work quality.

2.2. Monitoring Business Partners' Management Practices

Our work also relates to the literature on the effectiveness of firms monitoring the compliance of suppliers and other business partners with standards. Some buyers, for example, conduct monitoring to mitigate the risk of negative reputation spillovers; for example, from media reports that a fast food retailer franchisee's customers were made sick by contaminated food preparation practices or that an apparel brand's supplier used child labor. Our work is most closely related to those studies that identify factors that cause problems to go underreported (e.g., Gul et al. 2013, Short et al. 2016), which has so far largely been attributed to characteristics of individual auditors, auditing firms, audit arrangements, and institutional factors.

For example, lower-quality audits result when they are conducted (a) by auditors who are less-trained and less-experienced (Macher et al. 2011, Short et al. 2016) or have no experience auditing the site (Short et al. 2016, Ball et al. 2017), (b) by all-male audit teams (Short et al. 2016), or (c) later in an auditor's workday, when fatigue is more likely (Ibanez and Toffel 2020). Several audit firm characteristics have been found to erode monitoring quality; firms perform worse when they (a) have opportunities to cross-sell other services to those they are monitoring (Pierce and Toffel 2013), (b) face greater competition (Bennett et al. 2013), or (c) have performed the site's prior audit (Ibanez et al. 2023).

A key audit arrangement associated with worse audit quality is when the audit is paid for directly by the monitored entity rather than by another interested party (Duflo et al. 2013, Short et al. 2016).

These studies have all focused on audits performed in-person onsite. We examine whether and how audit quality varies between audits performed in-person and those performed remotely and how such differences are moderated by key attributes of individual auditors and their firms.

3. Theory

Management system standard certification audits assess the extent to which an organization is complying with the practices set forth in a management system standard. Such standards focus on areas such as quality management (ISO 9001), environmental management (ISO 14001), and occupational health and safety (ISO 45001/OHSAS 18001). Each standard has several clauses and sub-clauses (clauses, henceforth), which address different areas of best practices.[4] Companies may voluntarily adopt these standards and hire an auditor to certify their adherence for several reasons, including a desire to adopt best policies and procedures and a desire or a need to satisfy buyers or other stakeholders.

Auditors gather data to assess compliance with each clause by reviewing documents (such as policies, procedures, and training records), interviewing employees and managers, and directly observing processes and work activities (ISO 19011 Guidelines for Auditing Management Systems: Section 6.4.7). Instances of noncompliance are referred to as non-conformities or violations.

3.1. Quality Concerns of Remote Auditing

Service operations entail a coproduction process between service workers and customers, clients, patients, and so on (Vargo and Lusch 2004, Sampson and Froehle 2006). While remote work can reduce travel-related time and expenses and enable more efficient planning, remote work that physically separates workers from those coproducers tends to reduce the availability of critical information needed for such

[4] For example, ISO 9001's standard has clauses for best practices for leadership, operations, and performance evaluation.

coproduction, including access to diagnostic information ranging from patient symptoms to factory conditions. The result can be less-comprehensive and thus lower-quality work. In our context, this separation impairs auditors' ability to access information necessary to identify all the violations occurring at a site. Below, we argue that it is harder for remote auditors than for in-person auditors to (a) directly observe and detect some types of violations and (b) coordinate among audit team members, both of which lead us to predict that remote audits will be of lower quality than in-person audits.

3.1.1. Information access challenges. The extent to which conducting diagnostic work remotely impedes quality will vary with the data-gathering techniques used. Assessing compliance with some management standards clauses requires auditors to review documents—a similar process whether conducted remotely or in-person. Other types of information are harder to access remotely. Technology-mediated interactions tend to reduce participant engagement and important nonverbal communication compared to in-person interactions (Reid and Reid 2005, Bohannon et al. 2013).

Observing contextual attributes—which can include workers, equipment, procedures, and work settings—broadens the set of information that can be incorporated into assessments and decisions (Tyre and von Hippel 1997). In our context, auditors often rely on contextual knowledge gained at the audit site—by looking, listening, touching, and smelling—to better understand its operations and vulnerabilities and to assess compliance with the management system standard. Such on-the-spot observation can prompt ad hoc inquiries that can uncover violations. Site tours conducted remotely are limited by the audio-video technology—seldom as good as human senses—and to some extent by what the onsite employees choose to show. A recent survey revealed that EHS professionals were concerned that remote audit site tours risk resulting in assessments "limited to what those at the facility allow the auditor to see" and that some areas "may be overlooked…due to restriction of electronic equipment" in those areas (NAEM 2021: 19). These factors lead us to predict that a decline in audit quality associated with remote auditing will be especially pronounced for those standards clauses for which compliance is assessed via direct observation compared to those clauses assessed via document review.

3.1.2. Coordination challenges. Diagnostic service operations conducted remotely, such as remote auditing, entail physical separation not only between the remote workers and the subjects of their work, but also between coworkers conducting the work. Here, again, we theorize that separation reduces information flow between audit team members (virtual teams), reducing audit quality.

Studies have documented several challenges virtual teams face, many due to coordination problems that can arise because physical separation makes it harder for team members to gain a sense of shared knowledge (Cramton 2001). Uneven information distribution—or a lack of mutual knowledge—amongst team members makes it harder for them to identify and resolve problems. Virtual teams are more likely to report greater team conflict, possibly because they lack mutual knowledge—which they would otherwise have gained through in-person settings—to resolve issues early on (Hinds and Mortensen 2005). Virtual teams are also more likely to report less team trust (Peñarroja et al. 2013) and psychological safety (Gibson and Gibbs 2006), both of which can affect their ability to openly communicate and resolve problems with each other.

In our setting, audits may be conducted by one auditor or by a team of auditors. Auditor teams require frequent communication with one another to exchange information and coordinate activities. In particular, the rules governing auditors of management system standards require them to "confer periodically to exchange information, assess audit progress, and reassign work between the audit team members as needed" (ISO 19011: Section 6.4.4). Coordination is especially important because an auditor's next activity could depend on the outcome of another auditor's activity. We therefore expect that audit quality degradation associated with remote auditing will be exacerbated for audit teams.

3.2. Mitigating Remote Auditing Quality Concerns via Prior Site Exposure

With financial auditing, auditor familiarity with the site has long been suspected of reducing in-person scrutiny; some regulators even require or recommend rotation of lead auditors or audit firms after a

maximum number of years.[5] Research on in-person supplier auditing and regulatory inspections has found that auditors assigned to sites they have previously audited tend to conduct less-comprehensive audits (e.g., Short et al. 2016, Ball et al. 2017). One mechanism potentially driving this effect is "bounded awareness" (Chugh and Bazerman 2007), a cognitive bias that, in our context, suggests that auditors returning to the same site would overly focus on the problem areas they had identified in prior audits. Kumar and Chakrabarti (2012) argue that decision makers over-rely on information that supports their tacit knowledge gained from prior experiences, limiting their ability to recognize new problems.

Whereas these arguments suggest that in-person auditors returning to a site they've already audited tend to conduct lower-quality audits than they did previously, we are interested in a different comparison: How does the audit quality of remote auditors with experience conducting in-person audits at the site compare with that of remote auditors auditing a site for the first time? Here, an additional factor must be considered: memories of the site layout and work practices observed in-person along with tacit knowledge of the site's vulnerabilities can lead returning auditors to ask more questions that turn up more violations. Thus, having more in-person auditing experience at the site can help overcome the information access challenges associated with remote audits. We therefore expect that auditors with more prior in-person auditing experience at the audited site will conduct higher-quality remote audits of that site than auditors without that experience. In other words, the decrement in audit quality associated with remote auditing should be attenuated when auditors have more prior in-person exposure to the site.

4. Data and Measures

4.1. Empirical Context

We obtained data from a company that provides auditing, certification, training, and consulting to organizations around the world and that has requested anonymity as a condition of sharing data with us.

[5] For example, the Sarbanes-Oxley Act of 2002 requires US public companies to rotate lead audit partners every five years and the 2014 European Union (EU) Audit Regulation requires EU member states to create regulations mandating audit firm rotation every 10 years or sooner.

The company's auditing and certification practices are accredited by many leading accreditation bodies, which provides third-party assurance that its auditing meets international standards.

This company shared all certification audits it conducted from January 1, 2019 to October 7, 2021 pertaining to six of the most widely used management system standards, covering management of quality (ISO 9001), environment (ISO 14001), information security (ISO 27001), occupational health and safety (OHSAS 18001 and ISO 45001), and medical device quality (ISO 13485). In the course of obtaining and retaining certification, a site receives an *initial audit* to assess whether its management system meets the requirements of the given standard; if so, the audit firm issues a three-year certification.[6] Certified sites are subsequently subjected to *surveillance audits*, typically one a year. Auditors follow the same audit process in surveillance audits, but check a subset of areas within the standard.[7] In the third year, typically just before the certification term expires, a *re-certification audit* is conducted to provide another comprehensive assessment of the site's compliance; a successful audit results in a three-year renewal of certification. In our analysis, we focus only on surveillance audits, due their prevalence in the data and resulting statistical power; our dataset includes multiple surveillance audits for a given site being audited to a particular standard—in many cases, some onsite and others remote.[8]

Our raw data included 125,727 surveillance audits of 48,956 sites, addressing the six standards mentioned above, during the nearly three-year period of our sample. Because some sites adopted more than one standard, this initial dataset corresponded to 64,222 site-standard dyads, which correspond to all unique pairings of one site's audits of a particular management system standard (e.g., a site-standard dyad

---

[6] The initial audit is performed in two stages. In stage 1, auditors check whether the necessary procedures have been developed to meet the quality benchmark set forth in the standard. Stage 2 assesses whether those procedures were implemented and assesses the efficacy of their implementation. Failure to implement those procedures would result in a non-issuance of the certification until the site provides evidence that the failures have been rectified.

[7] All areas of the standard are audited across the two surveillance audits conducted over consecutive years during the three-year certification period.

[8] Because initial audits are typically conducted once per site-standard, we cannot compare in-person to remote initial audits of a given site-standard. Similarly, recertification audits are conducted roughly every three years and we have very few instances in our dataset of variation of in-person versus remote recertification audits of a given site-standard. We therefore exclude initial audits and recertification audits from our analysis and instead focus on surveillance audits.

includes all of a particular site's ISO 9001 audits). Because our specification (described below) includes site-standard–dyad fixed effects to enable comparisons of in-person audits to remote audits of the same site for a particular standard, our model requires at least two audits per site-standard dyad. This excludes 24,308 audits for which our data included only one surveillance audit for a given site-standard dyad. We then omit 5,197 audits of 1,080 sites (1,576 site-standards) with data anomalies that either resulted from data entry errors or did not reflect the typical course of audits according to our data provider.[9] We then omit sites coded as having two or more surveillance audits in a given year; because such audits tend to have a smaller scope than annual surveillance audits, including them in the sample could introduce bias into our analysis (e.g., if unobserved differences in such "mini-audits" are correlated with our focal variables and the number of violations). This resulted in excluding 28,408 surveillance audits,[10] including 4,427 hybrid audits performed partly in-person and partly remotely.[11] Finally, our use of Poisson regression omits audits when all of those corresponding to a site-standard reported zero violations, which resulted in additional 32,568 audits being dropped from our estimation sample.

Thus, the sample to which we apply our data analysis includes 35,247 surveillance audits of 16,986 site-standards at 14,615 sites. Of these, 54% assessed compliance to the ISO 9001 quality management standard, 18% to the ISO 14001 environmental management standard, 12% to the ISO 27001 information security management standard, 11% to the OHSAS 18001 or ISO 45001 occupational health and safety management standards, and 5% to the ISO 13485 medical device quality management

---

[9] Specifically, we omitted surveillance audits that (a) were missing a certificate number or auditor identifier, (b) prompted a change in certificate number, (c) were duplicates (same audit-format–site–standard–date), or (d) were associated with a site whose audits were conducted in an atypical sequence of initial, surveillance, and recertification audits or were missing audits from the typical sequence.

[10] We considered combining these multiple surveillance audits conducted in a calendar year for a single site-standard, thinking they might simply represent a surveillance audit split into parts. But our preliminary analysis led us to conclude that aggregated surveillance audits are not comparable to typical annual surveillance audits: the former required substantially more auditor-days than the latter, even after accounting for sites' employment level and industry, the factors that audit standards specify was determining the number of auditor-days to be allocated to each audit. We therefore dropped all site-standards that included multiple surveillance audits in a calendar year.

[11] 573 hybrid audits are same-date hybrid audits, with the same audit recorded twice in the data on the same day, once as in-person and once as remote. 3,854 hybrid audits are multi-date, with audits recorded on two days in a calendar year, one in-person and one remote.

standard. While these audits were conducted at sites in 109 countries, most took place in the United Kingdom (35% of sites), China (9%), India (8%), and the United States (7%).

Auditors assess a site's adherence to standards by reviewing its documents, interviewing its employees, and touring its facilities. Audits have traditionally been performed onsite, but a growing number have been conducted remotely, with the use of remote audits accelerating during the COVID-19 pandemic to reduce exposure and travel.[12] Remote audits are performed from auditors' homes or offices, using audio/video technology. Remote audits and in-person audits are intended to entail the same level of scrutiny and thus to result in equal audit quality.

We follow the approach of other empirical studies of auditor bias by controlling for factors that might be correlated with violations and, assuming all else equal, considering an establishment's audits that report fewer violations to be of lower quality than its audits that report more violations. This approach has been used by others in contexts including supplier codes of conduct (Short et al. 2016), vehicle inspections (Bennett et al. 2013, Pierce and Toffel 2013), and government inspections (Braithwaite and Makkai 1991, Gray and Shadbegian 2005, Duflo et al. 2013, Ibanez and Toffel 2020).

Auditors collect evidence to prove both compliance and noncompliance. When they do not find sufficient evidence of compliance with a clause, the site is provided with a second chance to furnish it, mitigating the risk that audits will report false violations. Specifically, management system auditing rules require auditors to review preliminary findings of noncompliance "with the auditee in order to obtain acknowledgement that the audit evidence is accurate and that the nonconformities are understood. Every attempt should be made to resolve any diverging opinions concerning the audit evidence or findings" (ISO 19011: Section 6.4.8). As a result, lower audit quality refers to auditors underreporting (not overreporting) violations.

[12] Starting in March 2020, an audit's format (remote or in-person) was determined by performing a risk assessment to assess whether an auditor can safely access the site in-person or whether it can be successfully audited remotely.

4.2. Dependent Variables

We measured the degree to which a site does not fully comply with all clauses of a management system standard as *violations*—the number of violations discovered and recorded in a surveillance audit. This approach has been used in contexts including supplier codes of conduct (Short et al. 2016), restaurant inspections (Ibanez and Toffel 2020), and government health and safety inspections (Braithwaite and Makkai 1991, Gray and Shadbegian 2005).[13] For example, violations identified during an audit of the ISO 9001 Quality Management System Standard could correspond to the standard's clauses addressing the site's operations, leadership, or personnel, such as if the site has not established clear communication channels with clients or customers or has not devoted resources to establish proper quality assessment processes.

Discussions with the company that provided our audit data indicated that auditors tend to rely on one of several methods: document review, direct observation, and interviews—to identify non-compliance with each clause within a standard. For example, the ISO 9001 Quality Management System Standard's clause 6.2 requires organizations to maintain documented information about their management system quality objectives and how to achieve them, which auditors verify by reviewing those documents. In contrast, ISO 9001 clause 8.5.1 requires that suitable infrastructure and environment be used for site operations, which auditors tend to verify by directly observing the production facility. Based on information obtained from our data provider for five of the six standards in our study (all except ISO 13485), we categorized each audit's *violations* into the number of violations associated with standard clauses for which violations were primarily detected by document review, by direct observation, or by employee interviews. Of the clauses in these five standards, 301 are primarily detected by document review (*document-review–based clauses*), 61 by direct observation (*direct-observation–based clauses*),

[13] Each violation is categorized as major or minor. A violation is classified as major when it impedes a management system's capability to achieve a site's objectives (e.g., financial success or product quality) and is otherwise classified as minor. In our sample, 97.5% of violations are minor. Below, we describe a robustness tests that indicate that our results are nearly identical when we exclude major violations.

and 44 by interviews (*interview-based clauses*), while 48 lacked a single primary detection mode (e.g., two were equally relied upon).[14]

In our analysis of how quality concerns associated with remote auditing vary by audit detection mode, we compare violations of direct-observation–based clauses to violations of document-review–based clauses, leaving aside interview-based violations, which we consider to be an intermediate category only somewhat affected by remote auditing. (We include interview-based clauses in a robustness test described below.)

4.3. Independent and Moderator Variables

We created *remote audit* as a binary variable coded 1 when an audit was conducted remotely and 0 when it was conducted in-person. To develop a measure of audit team size, we observed that 80% of audits were conducted by a single auditor and that, of the rest, the majority (76%) were conducted by a two-auditor team, 20% by three auditors, and 4% by teams of four to nine. To capture the primary variation in our data and avoid results potentially driven by outliers, we measured audit team size using a binary variable, *audit team,* coded 1 for audits conducted by more than one auditor and 0 when an audit was conducted by one auditor. (In unreported results, our results are robust to measuring team size as the actual number of auditors who conducted each audit.)

To measure an audit team's familiarity with the site from previous visits, we coded *average prior in-person site exposure (log)* as the sum of prior in-person audits any audit team member had conducted

[14] Examples of standards clauses primarily assessed via document review include requirements to establish a quality policy appropriate to the organization's purpose, context, and strategic direction and that includes a commitment to continual improvement of the quality management system (ISO 9001: 2015 clause "5.2.1 Establishing the quality policy") and requirements to maintain an up-to-date inventory of information and information processing facilities (ISO 27001: 2013 clause "A.8.1.1 Inventory of assets"). Standards clauses primarily assessed via direct observation include requirements for the site to have implemented production and service provision under conditions consistent with its documented control procedures (ISO 9001: 2015 clause "8.5.1 Control of production and service provision") and requirements to ensure that site entry is adequately controlled to permit access only to authorized personnel (ISO 27001: 2013 clause "11.1.2 Physical entry controls"). Standards clauses primarily assessed via interviews include requirements for the organization to monitor customers' perceptions of the extent to which their needs and expectations are being met (ISO 9001: 2015 clause "9.1.2 Customer satisfaction") and to monitor and evaluate the performance and effectiveness of the management system in a manner that ensures comparable and reproducible results (ISO 27001: 2013 clause "9.1 Monitoring, measurement, analysis and evaluation").

there before the focal audit divided by the audit team size, and then taking the log of the result to reduce skew.

4.4. Control Variables

We control for a number of audit-, auditor-, and site-level variables that may be correlated with our outcome measure. We measured a site's complexity as the number of *staff-days (log)* scheduled to perform an audit. Staff-days are determined using International Accreditation Forum (IAF) documentation and are based on site size (employment) and site risk (measured by industry). We use the natural log in our models to reduce skew. We created *COVID time period* as a binary variable coded 1 for audits conducted from March 2020 through December 2020 and 0 otherwise.[15]

We created *prior remote site exposure* as a binary variable coded 1 when at least one audit team member had previously conducted a remote audit of the site and 0 when none had done so.[16] We created *focal standard advanced training* as a binary variable coded 1 if any audit team member other than the audit team leader held a "lead assessor" qualification for the focal audit's standard, and 0 otherwise.[17] We coded *maximum auditing experience (log)* as the maximum number of audits any auditor on the team had conducted since the start of our sample period up until the start of the focal audit, which we log (after adding 1 to accommodate zeros) to reduce skew. We created *female on audit team* as a binary variable coded 1 if at least one audit team member is female, and 0 for all male audit teams.[18] We created *percent outsourced* as the percentage of audit team members who were contract employees rather than employees of the certifying organization.

---

[15] Because our specifications also include year fixed effects (described below), this effectively creates a separate dummy to divide 2020 into two elements.
[16] We chose this measure instead of measuring *average prior remote site visits (log)* because the binary variable captures 87% of variation in *average prior remote site visits (log)*.
[17] We recoded this variable to 0 for the 3,474 audits for which this information was unavailable and created a dummy variable coded 1 to flag those observations and 0 otherwise.
[18] This variable is missing for 11,324 audits and is then set to 0 and modeled with a binary variable equal to 1 when missing.

We created *multi-standard audit* as a binary variable coded 1 if an audit for a different standard was conducted at the same time as the focal audit, and 0 otherwise. We also created audit-year fixed effects to control for secular factors and audit-sequence fixed effects to reflect how many audits a site has received from the audit company before the focal audit (1st audit, 2nd audit, etc.).[19]

Summary statistics and correlations are reported in Tables 1 and 2.

5. Empirical Models and Results

5.1. Empirical Specification

We estimate four models to test our theory. Model 1 predicts *violations* as a function of *remote audit* and other controls (described below) to test the main effect of remote auditing on audit quality. Model 2 also includes an interaction term between *remote audit* and *direct-observation–based clauses* to assess whether quality concerns with regard to remote audits are more severe for management standard clauses whose assessment relies on auditors' direct observation as opposed to document review. Model 3 instead includes an interaction term between *remote audit* and *audit team* to assess whether quality concerns with regard to remote audits are more severe for audits conducted by teams compared to those conducted by a single auditor. Model 4 instead includes an interaction term between *remote audit* and *average prior in-person site exposure (log)* to assess whether quality concerns with regard to remote audits are attenuated when the remote audit team includes members who have conducted in-person audits of the site before.

The unit of observation for Models 1, 3, and 4 is an establishment's audit conducted on a given date to assess the implementation of a particular standard. These models include site-standard–dyad fixed effects (e.g., Site A–ISO 9001) to control for time-invariant factors (observed and unobserved) associated with each combination of site and standard that could affect compliance, such as industry and location.

[19] Before creating audit-sequence fixed effects, we winsorized the sequence count at the 95th percentile (20th audit) to minimize the influence of outliers.

Model 2 differs in a few ways from the other models. Here, we divide each audit into two units of observation that are identical except that in one, *direct-observation–based clause* is coded 1 and the dependent variable *violations* includes only violations of standards clauses whose detection mode is direct observation, while in the other, *direct-observation–based clause* is coded 0 and the dependent variable *violations* includes only violations of standards clauses whose detection mode is document review. We include the interaction term *remote audit × direct-observation–based clause* to assess whether violations detected via direct observation are more vulnerable to quality concerns associated with remote auditing than violations detected by document review. This model includes site–standard–detection-mode fixed effects (e.g., Site A–ISO 9001–direct-observation–based) to hold constant time-invariant characteristics specific to a site, standard, and audit detection mode that could affect compliance. The main effect of *direct-observation–based clause* is therefore absorbed by these fixed effects.[20]

All four models include all of the following control variables. We include *staff-days (log)* because more time at the site provides more opportunities to uncover violations. To account for the possibility that an auditor's advanced qualifications for the audit's standard could affect audit qualify, we include *focal standard advanced training*. To ensure that a team's prior experience remotely auditing a site does not confound our testing whether a team's prior in-person site exposure moderates the relationship between remote audits and audit quality, we include *prior remote site exposure*. Since cumulative experience influences audit outcomes (Macher et al. 2011, Short et al. 2016, Ball et al. 2017), we include *maximum auditing experience (log)*. We include *female on audit team* as research has found that audit teams with at least one woman find more violations (Short et al. 2016). We include *percent outsourced* because research suggests that audit firms differ in the number of violations they record (Ibanez et al. 2023). We

[20] Our modeling specification here is akin to separately estimating the same model on two samples—one in which violations are recorded primarily through direct observation (where *direct observation* equals 1) and one in which violations are recorded primarily through document review (where *direct observation* equals 0)—and then performing a Chow test for coefficient equality (Chow 1960). Our choice to estimate one model that combines both samples allows for easier interpretation of results. Doing so assumes no differential correlation between other control variables and violations that stem from observation vs. document review detection modes.

include *multi-standard audit* to account for the possibility that sites' violation counts could be affected by their being audited against several standards simultaneously.

All models also include audit year fixed effects to control for secular factors that might affect auditor stringency and/or violation counts. While year fixed effects can control for differences between years, the COVID-19 pandemic that began in March 2020 might be another temporal factor that affected audit results (e.g., because some sites operated with fewer workers), so we include *COVID time period.* We include audit-sequence fixed effects because sites might become more compliant the more they are audited, having had more opportunities to learn about and resolve violations.

5.2. Identification Strategy

5.2.1. Remote audit assignment. If our results are to provide an unbiased estimate of the effect of remote audits on violation counts, unobserved variables should not be correlated with both an audit's outcome and the decision to perform the audit remotely. Through conversations with the data provider, we learned that remote audits were first adopted in 2019 to cut costs and gained much wider use with the COVID-19 pandemic. Of the remote audits in our 35,247 estimation sample, 0.2% occurred from January 2019 to March 2020 (that is, before the pandemic became widespread), 53.1% during the remainder of 2020, and 44.7% in 2021.

To determine if a particular audit would be remote or in-person, the auditing company used a two-step process. First, assuming country laws allowed auditors to access a site for an in-person audit, the auditing company used their standardized "Pre-visit COVID-19 checklist" to determine whether their auditors could conduct an in-person audit without significant risk of COVID-19 exposure.[21] Second, the company assessed whether the site had the IT infrastructure to safely and rigorously facilitate a remote audit. Any site that could not be visited in-person safely and did not have sufficient infrastructure for a

[21] Questions in this checklist include those regarding the number of positive COVID-19 cases among site employees in the past weeks, whether the site has implemented social distancing guidelines in all areas, and whether there is a dedicated area where auditors can maintain physical distance from others.

remote audit would not receive an audit until it was again safe to do so in-person. Moreover, audited sites would not receive a remote audit if they did not agree to it.

5.2.2. Auditor assignment. There are three reasons why we believe the assignment of auditors to audits is plausibly exogenous to our focal variables and violation counts. First, we learned through conversations with the audit company that auditors were assigned to audits (of any format) based on their availability and that for sites in certain more complex industries, such as aerospace and medical devices, auditors also needed industry-specific auditing qualifications. This provides no evidence that the auditor assignment process might bias our results; for example, there is no evidence that auditors thought to be "more stringent" were disproportionately assigned to in-person audits.

Second, including site-standard fixed effects in our model specifications holds constant the minimum auditing qualifications needed to audit a particular site for a particular standard, which mitigates the possible concern that arises in a cross-site analysis that sites in more complex industries both require in-person audits (due to their complexity) and require audit teams with more qualifications (who are more likely to find more violations due to their additional qualifications).

Third, we learned that while the audit company first decided whom to assign to a particular audit and then whether the audit would be in-person or remote, the company did so using the decision process described above that does not take into account which auditors had been assigned. This further mitigates concerns that individual auditor attributes might influence their assignment to audits or how the audit was conducted.

5.3. Results

Given that our dependent variable (*violations*) is a count, we estimate all models using a Poisson pseudo-maximum likelihood fixed-effects regression.[22] We report standard errors clustered at the audited site. We interpret effect sizes in terms of incidence rate ratios (IRR) and average marginal effects (AME).

[22] Poisson panel estimators are consistent even when data are not distributed Poisson provided the conditional mean is correctly specified, and impose weaker distributional assumptions than negative binomial regression (Cameron and Trivedi 2010).

Table 3 reports our regression results. Before addressing the results concerning our theorized relationships, the statistically significant effects of several control variables on audit quality warrant discussion. In Model 1, the positive coefficient on *staff-days (log)* indicates that audits for which more staff-days are allotted tend to yield more violations, possibly because the audited sites are larger (in terms of employment) and/or their management system is more complex, which presents more opportunities for auditors to uncover violations. The positive coefficient on *female on audit team* indicates that the presence of a female auditor on an audit team is associated with higher audit quality, a finding consistent with Short et al. (2016). Similarly, the negative coefficient on *percent outsourced* means that audit teams with a larger proportion of auditors who are not direct employees of the audit provider report fewer violations, a result consistent with Ibanez et al. (2023).

Model 1 tests the overall effect of remote work on audit quality. The negative and statistically significant coefficient on *remote audit* ($\beta = -0.292$, $p<0.01$, IRR=0.75, AME=0.40) indicates that remote audits yield an average of 25% fewer violations than in-person audits, or an average of 0.40 fewer violations. Average predicted violations are 1.18 for remote audits and 1.58 for in-person audits.

The results of Model 2 indicate that audit quality is worse for violations of clauses for which auditors rely on direct observation to detect violations, compared to document review (the omitted category).[23] The statistically significant negative coefficient on the main effect of *remote audit* ($\beta = -0.227$ $p<0.01$, IRR = 0.800) reveals that for assessments of standards clauses assessed primarily via document review, remote audits yield significantly fewer violations than in-person audits. One possible explanation is that compliance areas primarily audited through document review use secondary audit methods (interviews or direct observation) to confirm the presence of violations and the increased difficulty in obtaining information in remote audits through these secondary methods makes it more difficult to gather enough evidence to record such violations. The statistically significant negative

[23] Recall that the unit of analysis for Model 2—unlike Models 1, 3, and 4—is an audit detection mode: each audit is represented by two observations, one in which *violations* includes only clauses that tend to be detected by direct observations and a second in which *violations* includes only clauses that tend to be detected by document review.

coefficient on the interaction term *remote audit × direct-observation–based clauses* (β = -0.425, $p<0.01$, IRR = 0.654) indicates that the quality of remote audits is further eroded for standards clauses for which auditor detection relies on direct observations of the site, consistent with our theory. Figure 1 depicts average marginal effects that indicate that, for violations primarily detected via document review, remote audits report an average of 1.05 such violations compared to 1.31 for in-person audits, a reduction of 0.26 violations or a 20% decline. Standards clauses detected via auditors' direct site observations suffer a greater average "remote penalty" of 0.38 violations—a 48% decline—as remote audits report an average of 0.42 such violations compared to 0.80 for in-person audits.

Model 3 tests whether the decrement in remote audits' violations (compared to in-person audits) is larger for auditor teams than for auditors working alone. The statistically significant negative coefficient on *remote audit × audit team* (β = -0.298, $p<0.01$, IRR = 0.74) indicates that it is; the quality decrement of remote audits conducted by audit teams is above and beyond the quality decrement that occurs among audits conducted by a single auditor (*remote audit* β = -0.243, $p<0.01$, IRR = 0.784). Figure 2 reports average marginal effects: for audits conducted by a single auditor, remote audits result in an average of 1.20 violations compared to 1.53 for in-person audits, a decline of 0.33 violations or 22%. For team audits, remote audits report an average of 1.02 violations whereas in-person audits report 1.75, a decline of 0.73 violations, or 42%. These findings are consistent with our second proposed mechanism: remote audit teams experience coordination problems. In in-person settings, these meetings tend to occur more frequently and at more natural points throughout the audit due to the auditors' colocation. However, in remote audits, the lack of auditor colocation potentially reduces the frequency and natural timing of these meetings, which undermines work quality. Note that with in-person auditing, more violations are reported by auditor teams than by single auditors (an average of 1.75 versus 1.53), whereas with remote auditing, fewer violations are reported by auditor teams than by single auditors (an average of 1.02 versus 1.20).

Model 4 reports results of testing whether the quality decrement of remote auditing is attenuated for auditors who had conducted more in-person audits at the site. We begin by interpreting the negative coefficient on the main effect of *average prior in-person site exposure* (β=-0.195, $p<0.01$, IRR = 0.823) as indicating lower-quality in-person audits when staffed by audit teams with more in-person audit experience at the site, a result that is consistent with findings that auditors who return to a site are less stringent than auditors new to a site (Short et al. 2016, Ball et al. 2017). Turning to the interaction effect, the statistically significant positive coefficient on *remote audit × average prior in-person site exposure (*β=0.087, $p<0.01$, IRR = 1.091) provides evidence that the audit-quality decrement of remote auditing (and of auditors returning to a site) is attenuated when the audit team has more in-person auditing exposure to the site.[24] The IRRs for these effects indicate that each one-unit increase in *average prior in-person site exposure (log)* is associated with a nearly 18% reduction in in-person audit violations (an "auditor familiarity penalty"), but this effect is attenuated by 9% for remote audits. The average marginal effects, displayed in Figure 3, indicate that, for example, tripling the (unlogged) average number of prior in-person audits from approximately two to six reduces the average violation gap between remote and in-person audits by 0.16 violations.[25] Figure 3 indicates that the moderating effect of more prior in-person exposure on audit quality is driven by a steeper decline in in-person audit quality compared to a modest (nearly flat) decline in remote audit quality. These results indicate that prior in-person site exposure only partially attenuates quality concerns for remote auditing and that greater in-person exposure does not

[24] Because the maximum value of unlogged *average prior in-person site exposure* is 48, one concern is that outlier values drive our results for this variable (even after taking its natural log). To explore this concern, we estimated a model that top-codes it at the 95th percentile (an average of 9 prior in-person site visits) before taking its natural log. Using this top-coded variable yields results consistent with our primary results, which indicates that our results are not driven by outlier values of *average prior in-person site exposure*.

[25] We arrive at the estimation of 0.16 violations as follows. An increase from 1 to 2 log points of *average prior in-person site exposure (log)*—corresponding to an increase of roughly 2 to 6 prior in-site audit visits—corresponds to a decline in *predicted average violations* of 0.29 for in-person audits (from roughly 1.62 to 1.33) and 0.13 for remote audits (from roughly 1.20 to 1.07). The difference in differences between 0.29 and 0.13 is 0.16. These values come from Figure 3 and we calculate the x-axis values of roughly 1 to 2 log points corresponding to values of roughly 2 to 6 unlogged prior in-person site visits because our logged measure is calculated after adding 1 to the raw values and ln(2 visits+1) ≈1 log point and ln(6 visits +1) ≈ 2 log points.

increase remote audit quality but instead attenuates the "familiarity penalty" that motivates calls for auditor rotation.

5.4. Extension

Our main analysis revealed that auditors' prior in-person exposure with the audited site attenuates quality concerns associated with remote auditing. We now explore a second element that might also affect remote auditing quality: the extent to which the auditors' experience was concentrated on the management standard of the focal audit. A long literature has documented performance benefits from workers focusing on specific tasks (e.g., Taylor 1911, Fayol 1967); we explore whether this applies to remote auditing. To measure the extent to which the audit team included an auditor highly focused on the management system being audited, we calculated the proportion of each audit team member's prior auditing experience that focused on that standard and assigned the maximum value to the audit team's *maximum management standard focus* for the audit. To estimate the influence of auditor focus on the quality of remote audits, we modify Model 1's specification by adding *maximum management standard focus* and its interaction with *remote audit.* We report the results as Model 5, finding a positive and statistically significant coefficient on that interaction term *remote audit* × *maximum management standard focus*, indicating that *maximum management standard focus* attenuates the quality detriment of remote auditing. Accumulating auditing experience with a given standard appears to partially mitigate remote auditors' information access challenges. The small, nonsignificant coefficient on the main effect of *maximum management standard focus* yields no evidence that such focus improves in-person auditing.

5.5. Robustness Tests

Our results are robust to several alternative measures and samples. We find results nearly identical to those of our primary models when we use average instead of maximum values of auditing experience. Because minor violations account for 97.5% of all violations in our sample, we explored whether the anomalous presence of major violations might be biasing our results by re-estimating all models using a

dependent variable that includes only minor violations; our primary results are robust to this alternative (Table A1 in the Online Appendix).

We assessed whether auditor discretion might influence our findings. Auditors have discretion about whether to assess some standard clauses during a site's first or second surveillance audit in the three-year certification cycle. We refer to these as "discretionary clauses," in contrast to the "mandatory clauses" that must be audited in every surveillance audit. Our results could be biased if remote auditors chose to audit fewer or less-violation-prone discretionary clauses, which would mean that our observation of fewer violations in remote audits could thus be attributable to that choice rather than to our posited mechanisms. We therefore re-estimate our models but exclude from our sample those audits whose sequence poses the greatest risk of such behavior. Specifically, we identify and exclude from our sample the 742 site-standard dyads (corresponding to 1,674 audits, or 5% of our sample) whose first audit is remote and second audit is in-person because this sequence could enable auditors to choose fewer discretionary clauses in the remote audit and then audit a larger set of discretionary clauses in the in-person audit.[26] In unreported results, we find that our primary results are robust to the exclusion of the audit sequence at greatest risk of this possible source of endogeneity. We also re-estimate our models with the dependent variable being the number of violations not associated with discretionary clauses. By testing our theory on violations that stem from mandatory clauses only, we avoid the concern that auditor discretion might be correlated with remote versus in-person audits, which could bias our results. We first omit from the sample all audits of standard ISO 13485 because we could not obtain a list of mandatory clauses (and find our main results robust to this omission).[27] We then modify our dependent variable to

[26] Note that auditors do not have this option if the first audit is in-person and the second is remote, because auditors decide how to allocate discretionary clauses to surveillance audits at the time of the first surveillance audit. When the first surveillance audit is in-person, auditors will likely not anticipate the second to be remote and so would likely allocate an equal number of discretionary clauses to both surveillance audits.

[27] We excluded ISO 13485 audits here because we did not obtain from our data provider a list of which of its clauses are mandatory. As a robustness test, we estimated our primary models on a sample that excludes those audits, which yielded results very similar to our main results estimated on the full sample.

include only violations of mandatory clauses. These models, reported in Table A2 in the Online Appendix, indicate that our primary results are robust to the omission of discretionary clauses.

We examined whether a key decision in our sample construction might have affected our results. The largest omission of audits (unrelated to model specification decisions) occurred when we dropped the 3,552 site-standards (corresponding to 15,946 surveillance audits) for which two or more surveillance audits occurred in the same year, because this differs from the typical practice of annual surveillance audits. For this robustness test, we included those audits in our sample after summing the violations for any site-standard's pair of surveillance audits occurring in the same year and referring to them as 7,973 "bundled audits." Adding these to our primary sample of 35,247 audits yielded an expanded sample of 43,220. Before estimating our models on this increased sample, we added two variables to account for two possible effects of these additional audits on quality. First, we include the binary variable *bundled audit*, which equals 1 for a bundled audit and 0 otherwise. Second, because the two portions of the bundled audit could be performed using both remote and in-person audit formats, we include *hybrid audit* which equals 1 if the bundled audit was partially remote and partially in-person. The results, reported in Table A3 in the Online Appendix, confirm that our main results are robust to including these audits.

Recalling that Model 2 tests the information access mechanism by comparing the effect of remote audits on standards clauses whose violations are primarily detected by direct observation to that of clauses whose violations are primarily detected by document review, we also include, as a robustness test, a third record per audit reflecting clauses whose violations are primarily detected from interviews. To do so, we not only expand the sample but also include the interaction between *remote audit* and *interview-based clauses*. Our results, reported in Table A4 in the Online Appendix, yield a statistically significant negative coefficient on *remote audit × direct-observation–based clauses* ($\beta = -0.423$, $p<0.01$, IRR = 0.655) that reinforces our primary finding (in Table 3, Model 2) that remote audit quality concerns are worse for clauses for which discovery of violations relies on direct observation, even when the model includes interview-based audits. The nonsignificant (and slightly positive) coefficient on *remote audit × interview-*

based clauses yields no evidence that remote auditing impedes the reporting of violations of clauses that rely primarily on interviews.

6. Discussion and Conclusion

We find that audit quality is negatively impacted when audits are conducted remotely and our analysis yields evidence that is consistent with our two theorized mechanisms: difficulty in information access and within-team coordination. Remote auditors appear to have problems gathering information in ways that proxy in-person auditors' direct observations during site tours, problems which we attribute to both their senses and their agency being hampered due to technology mediation. We also find that remote audits conducted by multiple auditors suffer more quality degradation than those conducted by individual auditors, suggesting that the collaboration necessary within audit teams to process the evidence they gather is especially difficult when conducting audits remotely. Finally, we theorize and find evidence that an audit team with more experience conducting in-person audits of the site attenuates the quality decrement of a remote audit. Our empirical extension indicates that quality concerns for remote audits are somewhat mitigated when the audit team includes at least one auditor whose experience concentrated on the standard being assessed.

Our findings contribute to the remote work literature by showing that work quality can suffer when working remotely changes how the work is conducted; in our case, by separating the workers from the evidence they are assessing and from the coworkers with whom they need to collaborate. These contextual differences likely explain why we find that working remotely degrades quality while prior studies did not find such quality differences in remote work settings—answering customer calls and reviewing patent documents—that had neither of these attributes (Bloom et al. 2015, Choudhury et al. 2021). Given that many other work settings besides auditing—from telemedicine to drone-based crop monitoring to regulatory inspections—require direct observation to gather critical information, our results suggest that managers in those domains should take special care to control the quality of remote work.

The literature on the quality of business monitoring has focused on how the quality of in-person assessments such as quality auditing and regulatory inspections are affected by individual and organizational attributes (e.g., Macher et al. 2011, Bennett et al. 2013, Duflo et al. 2013, Pierce and Toffel 2013, Short et al. 2016) and by scheduling (e.g., Ibanez and Toffel 2020). We contribute to this work by going beyond the in-person context to examine how audit quality is affected by remote work. Our findings are especially relevant to this literature stream given the increase in and expected persistence of remote audits in a variety of contexts, both within our context of management system certification assessments (IAF, ILAC, and ISO 2021) and well beyond, including various forms of regulatory inspections (Mofid et al. 2021, US Department of Health and Human Services 2022). Our results also support those concerned about the comprehensiveness of remote inspections (e.g., US Department of Labor 2021) and suggest caution to remote auditing's enthusiasts (e.g., IAF, ILAC, and ISO 2021). We seek to lend nuance to the debate about remote work by identifying mechanisms that moderate quality concerns with respect to remote auditing; these mechanisms suggest actions companies and regulators can take, such as avoiding larger remote auditing (or inspection) teams and developing better technologies and procedures for remote site tours. Moreover, our results can help companies and regulators design hybrid or blended auditing or inspection approaches to capture some of the benefits of remote auditing—such as increased flexibility, improved productivity, and reduced travel time and cost—while limiting its negative quality impacts. For example, hybrid schemes might pair remote document review with in-person direct observations or target fully in-person (that is, more comprehensive) audits to high-risk sites, fully remote audits to lower-risk sites, and hybrid audits to intermediate-risk sites.

While our results are based on the remote work context of auditing adherence to management system standards, we suspect that they generalize to other contexts in which working remotely reduces or eliminates in-person information exchange. Examples include teaching, conducting medical exams, auditing working conditions at global supply chain factories, assessing restaurant franchisees' adherence to franchisors' procedural requirements, and government inspections that assess compliance with occupational safety, environmental, and restaurant hygiene regulations.

Our study has limitations that highlight future research opportunities. While our sample period includes (but is not limited to) the COVID-19 pandemic and our data provider indicated that remote auditing was only used when in-person audits were infeasible, it is possible that the pandemic affected site operations in ways that could alter compliance with management system standard requirements. For example, it is possible that sites audited remotely during the COVID-19 pandemic were operating with lower-than-usual volumes, which could naturally reduce the number of violations. To address this concern, we include a *COVID-19 time period* dummy variable (equal to 1 during March–December 2020) to hold constant the pandemic's effect (if any) on audit outcomes and we have no reason to believe that operational levels varied by audit format, which would have biased our results. It is also possible that sites audited remotely during the height of the pandemic were operating with unusually strong adherence to management system standards requirements, which might explain our finding that remote audits yielded fewer violations. However, we have no reason to believe that and, in fact, suspect the opposite; namely, that sites for which COVID made in-person audits infeasible were likely to be especially focused on production practices and perhaps less focused on the procedural requirements of management standards, which would have made them more prone to violations. Our results could therefore be viewed as conservative estimates of quality degradation associated with remote audits. Future work could compare remote to in-person audit quality at a time when the operational effects of COVID-19 are less pronounced.

It is also possible that remote auditing yielded less comprehensive audits due to auditors' increased mental exhaustion from doing a much larger portion of their work online. While our study cannot account for such potential mental costs of remote work, future research could.

References

Ball G, Siemsen E, Shah, R (2017) Do plant inspections predict future quality? The role of investigator experience. *Manufacturing & Service Operations Management* 19(4): 534–550.

Barrero JM, Bloom N, Davis SJ (2021) Why working from home will stick. NBER Working Paper No. 28731, April 2021.

Bavafa H, Hitt LM, Terwiesch C (2018) The impact of e-visits on visit frequencies and patient health: Evidence from primary care. *Management Science* 64(12): 5461–5480.

Bell DR, Gallino S, Moreno A (2018) Offline showrooms in omnichannel retail: Demand and operational benefits. *Management Science* 64(4): 1629–1651.
Bennett VM, Pierce L, Snyder JA, Toffel MW (2013) Customer-driven misconduct: How competition corrupts business practices. *Management Science* 59(8): 1725–1742.
Bettinger EP, Fox, L, Loeb S, Taylor ES (2017) Virtual classrooms: How online college courses affect student success. *American Economic Review* 107(9): 2855–2875.
Bloom N, Liang J, Roberts J, Ying ZJ (2015) Does working from home work? Evidence from a Chinese experiment. *Quarterly Journal of Economics* 130(1): 165–218.
Bohannon LS, Herber AM, Pelz JB, Rantanen EM (2013) Eye contact and video-mediated communication: A review. *Displays* 34(2): 177–185.
Braithwaite J, Makkai T (1991) Testing an expected utility model of corporate deterrence. *Law & Society Review* 25(1): 7–40.
Buell RW, Campbell D, Frei FX (2010) Are self-service customers satisfied or stuck? *Production and Operations Management* 19(6): 679–697.
Cameron AC, Trivedi PK (2010) *Microeconometrics Using Stata*, revised edition (Stata Press, College Station, TX).
Campbell D, Frei F (2010) Cost structure, customer profitability, and retention implications of self-service distribution channels: Evidence from customer behavior in an online banking channel. *Management Science* 56(1): 4–24.
Choudhury PR, Foroughi C, Larson B (2021) Work-from-anywhere: The productivity effects of geographic flexibility. *Strategic Management Journal* 42(4): 655–683.
Chow GC (1960) Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 28(3): 591–605.
Chugh D, Bazerman MH (2007) Bounded awareness: What you fail to see can hurt you. *Mind & Society* 6(1): 1–18.
Cramton CD (2001) The mutual knowledge problem and its consequences for dispersed collaboration. *Organization Science* 12(3): 346–371.
Duflo E, Greenstone M, Pande R, Ryan N (2013) Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India. *Quarterly Journal of Economics* 128(4): 1499–1545.
Fayol H (1967) *General and Industrial Management* (Pitman, London).
Gibson CB, Gibbs JL (2006) Unpacking the concept of virtuality: The effects of geographic dispersion, electronic dependence, dynamic structure, and national diversity on team innovation. *Administrative Science Quarterly* 51(3): 451–495.
Gray WB, Shadbegian R (2005) When and why do plants comply? Paper mills in the 1980s. *Law & Policy* 27(2): 238–261.
Gul FA, Wu D, Yang Z (2013). Do individual auditors affect audit quality? Evidence from archival data. *Accounting Review* 88(6): 1993–2023.
Hinds PJ, Mortensen M (2005) Understanding conflict in geographically distributed teams: The moderating effects of shared identity, shared context, and spontaneous communication. *Organization Science* 16(3): 290–307.
IAF, ILAC, and ISO (2021) Use of remote techniques supported by IAF/ILAC/ISO survey. Retrieved March 9, 2023, https://ilac.org/latest_ilac_news/use-of-remote-techniques-supported-by-iaf-ilac-iso-survey/.
Ibanez MR, Palmarozzo A, Short JL, Toffel MW (2023) Second- vs. third-party audit quality: Evidence from global supply chain monitoring. Working paper, Harvard Business School, Boston.
Ibanez MR, Toffel MW (2020) How scheduling can bias quality assessment: Evidence from food-safety inspections. *Management Science* 66(6): 2396–2416.
Internal Audit Foundation and AuditBoard (2021) The remote auditor: Challenges, opportunities, and new ways of working (The Internal Audit Foundation and AuditBoard, Lake Mary, FL).

https://www.theiia.org/en/content/research/foundation/2021/the-remote-auditor-challenges-opportunities-and-new-ways-of-working/.
Kumar AJ, Chakrabarti A (2012). Bounded awareness and tacit knowledge: Revisiting Challenger disaster. *Journal of Knowledge Management* 16(6): 934–949.
Macher JT, Mayo JW, Nickerson JA (2011) Regulator heterogeneity and endogenous efforts to close the information asymmetry gap. *Journal of Law & Economics* 54(1): 25–54.
Mofid S, Bolislis WR, Brading C, Hamilton N, Hardit C, Nagaoka M, Parain J, Zanta M, Kühler TC (2021) The utility of remote inspections during the COVID-19 health emergency and in the postpandemic setting. *Clinical Therapeutics* 43(12): 2046–2063.
NAEM (2021) The emergence of virtual EHS audits (The National Association for Environmental, Health & Safety and Sustainability Management (NAEM), Washington, DC).
Peñarroja V, Orengo V, Zornoza A, Hernández A (2013) The effects of virtuality level on task-related collaborative behaviors: The mediating role of team trust. *Computers in Human Behavior* 29(3): 967–974.
Pierce L, Toffel MW (2013) The role of organizational scope and governance in strengthening private monitoring. *Organization Science* 24(5): 1558–1584.
Rajan B, Tezcan T, Seidmann A (2019) Service systems with heterogeneous customers: Investigating the effect of telemedicine on chronic care. *Management Science* 65(3): 1236–1267.
Reid DJ, Reid FJ (2005) Online focus groups: An in-depth comparison of computer-mediated and conventional focus group discussions. *International Journal of Market Research* 47(2): 131–162.
Sampson SE, Froehle CM (2006) Foundations and implications of a proposed unified services theory. *Production and Operations Management* 15(2): 329–343.
Short JL, Toffel MW, Hugill AR (2016) Monitoring global supply chains. *Strategic Management Journal* 37(9): 1878–1897.
Sun S, Lu SF, Rui H (2020) Does telemedicine reduce emergency room congestion? Evidence from New York State. *Information Systems Research* 31(3): 972–986.
Tan TF, Netessine S (2020) At your service on the table: Impact of tabletop technology on restaurant performance. *Management Science* 66(10): 4496–4515.
Taylor FW (1911) *The Principles of Scientific Management* (Harper, New York).
Tyre MJ, von Hippel E (1997) The situated nature of adaptive learning in organizations. *Organization Science* 8(1): 71–83.
US Bureau of Labor Statistics (2022) US business response to the COVID-19 pandemic—2021. Report, US Department of Labor, Washington, DC. https://www.bls.gov/news.release/pdf/covid2.pdf.
US Department of Health and Human Services (2022) Conducting remote regulatory assessments. Report, Food and Drug Administration, US Department of Health and Human Services, Washington, DC.
US Department of Labor (2021) COVID-19: Increased worksite complaints and reduced OSHA inspections leave US workers' safety at increased risk. Report, Office of Inspector General—Office of Audit, US Department of Labor, Washington, DC. https://www.oig.dol.gov/public/reports/oa/2021/19-21-003-10-105.pdf.
Vargo SL, Lusch RF (2004) Evolving to a new dominant logic for marketing. *Journal of Marketing* 68(1): 1–17.
Xue M, Hitt LM, Chen P (2011) Determinants and outcomes of internet banking adoption. *Management Science* 57(2): 291–307.

Table 1. Summary Statistics

	N	Mean	SD	Min	Max
Violations	35,247	1.42	1.65	0	58
Violations of document-review–based clauses	35,247	1.02	1.38	0	56
Violations of direct-observation–based clauses	35,247	0.15	0.46	0	9
Violations of interview-based clauses	35,247	0.10	0.34	0	4
Remote audit	35,247	0.39	0.49	0	1
Audit team	35,247	0.20	0.40	0	1
Average prior in-person site exposure	35,247	2.89	3.22	0	48
Average prior in-person site exposure (log)	35,247	1.10	0.71	0	3.89
Focal standard advanced training [a]	31,335	0.23	0.42	0	1
Staff-days	35,247	1.56	1.07	0.5	30
Staff-days (log)	35,247	0.29	0.53	-0.69	3.40
COVID time period	35,247	0.31	0.46	0	1
Prior remote site exposure	35,247	0.15	0.35	0	1
Maximum auditing experience	35,247	344.17	211.02	0	1048
Maximum auditing experience (log) [b]	35,247	5.53	1.00	0	6.96
Female on audit team [a]	22,131	0.24	0.43	0	1
Percent outsourced	35,247	0.27	0.42	0	1
Multi-standard audit	35,247	0.10	0.30	0	1
Maximum management standard focus	35,247	0.56	0.29	0	1
Year	35,247	2019.91	0.79	2019	2021
Audit sequence [c]	35,247	6.52	3.63	1	16

[a] Variables for which missing values were replaced with 0 in the versions used in the regression models.

[b] We avoid missing logged values by adding 1 to the raw value before taking the natural log.

[c] Winsorized at the 95th percentile.

Table 2. Correlations

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	Violations	1.00											
(2)	Remote audit	-0.12	1.00										
(3)	Audit team	0.02	-0.12	1.00									
(4)	Average prior in-person site exposure (log)	-0.01	-0.34	0.06	1.00								
(5)	Focal standard advanced training [a]	0.04	-0.08	0.65	0.02	1.00							
(6)	Staff-days (log)	0.17	-0.05	0.48	0.02	0.42	1.00						
(7)	COVID time period	-0.06	0.38	0.00	-0.11	0.03	0.01	1.00					
(8)	Prior remote site exposure	-0.05	0.28	-0.05	0.13	-0.07	-0.01	-0.07	1.00				
(9)	Maximum auditing experience (log)	-0.10	0.03	0.19	0.32	0.05	-0.05	0.01	0.15	1.00			
(10)	Female on audit team [a]	0.03	0.03	0.08	-0.03	0.07	0.03	0.02	0.01	0.04	1.00		
(11)	Percent outsourced	-0.02	-0.05	0.08	0.05	0.00	0.09	-0.02	-0.04	-0.15	-0.22	1.00	
(12)	Multi-standard audit	-0.04	-0.06	0.22	0.11	0.19	-0.07	-0.01	-0.07	0.11	0.00	0.01	1.00
(13)	Maximum management standard focus	0.03	0.05	-0.05	-0.20	0.01	0.08	0.02	-0.08	-0.24	0.05	0.03	-0.22

N = 35,247 audits.

[a] Variables in which missing values were replaced with 0.

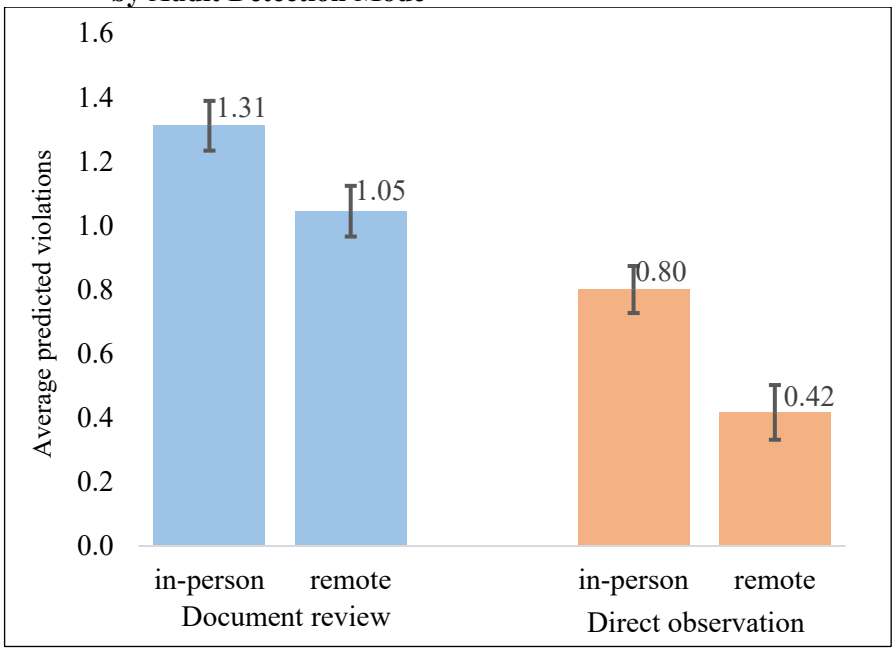
Table 3. Regression Results

	Dependent variable: Violations				
	(1)	(2)	(3)	(4)	(5)
Remote audit	-0.292**	-0.227**	-0.243**	-0.392**	-0.377**
	(0.026)	(0.029)	(0.027)	(0.040)	(0.042)
Remote audit × Direct-observation–based clauses		-0.425**			
		(0.045)			
Remote audit × Audit team			-0.298**		
			(0.046)		
Remote audit × Average prior in-person site exposure (log)				0.087**	
				(0.026)	
Remote audit × Maximum management standard focus					0.145**
					(0.055)
Audit team	0.036	0.072	0.134**	0.031	0.030
	(0.041)	(0.047)	(0.045)	(0.041)	(0.041)
Average prior in-person site exposure (log)	-0.135**	-0.127**	-0.137**	-0.195**	-0.136**
	(0.021)	(0.023)	(0.021)	(0.028)	(0.021)
Focal standard advanced training [a]	0.041	0.024	0.042	0.041	0.035
	(0.034)	(0.040)	(0.034)	(0.034)	(0.034)
Staff-days (log)	0.422**	0.432**	0.424**	0.423**	0.422**
	(0.051)	(0.060)	(0.052)	(0.051)	(0.051)
COVID time period	-0.065+	-0.241**	-0.072+	-0.068+	-0.066+
	(0.039)	(0.042)	(0.039)	(0.039)	(0.039)
Prior remote site exposure	-0.038	-0.044	-0.031	-0.051+	-0.030
	(0.029)	(0.032)	(0.029)	(0.029)	(0.029)
Maximum auditing experience (log)	-0.033**	-0.044**	-0.033**	-0.029*	-0.030*
	(0.012)	(0.014)	(0.012)	(0.012)	(0.013)
Female on audit team [a]	0.082*	0.075*	0.087*	0.081*	0.077*
	(0.035)	(0.038)	(0.035)	(0.035)	(0.035)
Percent outsourced	-0.109**	-0.119**	-0.107**	-0.112**	-0.113**
	(0.039)	(0.044)	(0.040)	(0.040)	(0.040)
Multi-standard audit	-0.022	-0.023	-0.010	-0.026	-0.021
	(0.041)	(0.046)	(0.042)	(0.041)	(0.041)
Maximum management standard focus					0.056
					(0.059)
Site-standard fixed effects	Included		Included	Included	Included
Site–standard–detection-mode fixed effects		Included			
Year fixed effects	Included	Included	Included	Included	Included
Audit-sequence fixed effects	Included	Included	Included	Included	Included
Observations (audits)	35,247	37,631	35,247	35,247	35,247
Number of sites	14,615	13,296	14,615	14,615	14,615
Number of site-standards	16,986	15,388	16,986	16,986	16,986

Poisson regression coefficients with standard errors are clustered by audited site; **p<0.01; *p<0.05; +p<0.10. The unit of analysis for Models 1, 3, and 4 is the site-standard-date and, for Model 2, is the site–standard–detection-mode–date. Model 2's observation count is less than twice as large as the other models because it excludes ISO 13485 audits (for which we were unable to obtain data to map its clauses to primary audit detection mode) and because the inclusion of the three-way fixed effect (site–standard–detection-mode) results in this model dropping observations that report zero violations for all instances of a site's various audits of a given standard using a particular detection mode (e.g., if Site A's ISO 9001 audits always yielded zero violations of observation-based clauses). In Model 2, *observation-based clauses* is interpreted as the effect compared to the omitted category of *document-review–based clauses*; the main effect of *observation* is absorbed by the inclusion of site–standard–detection-mode fixed effects.

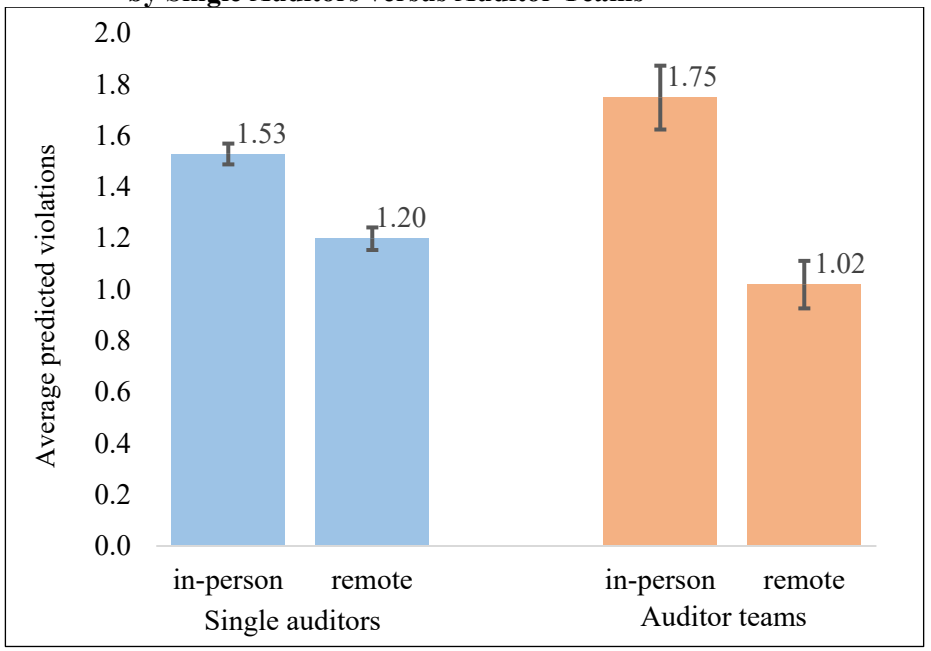
[a] Variables in which missing values were replaced with 0. All regression models include dummy variables to flag observations for which missing values of these variables were recoded to 0.

Figure 1. Average Predicted Violations for Remote vs. In-person Audits, by Audit Detection Mode



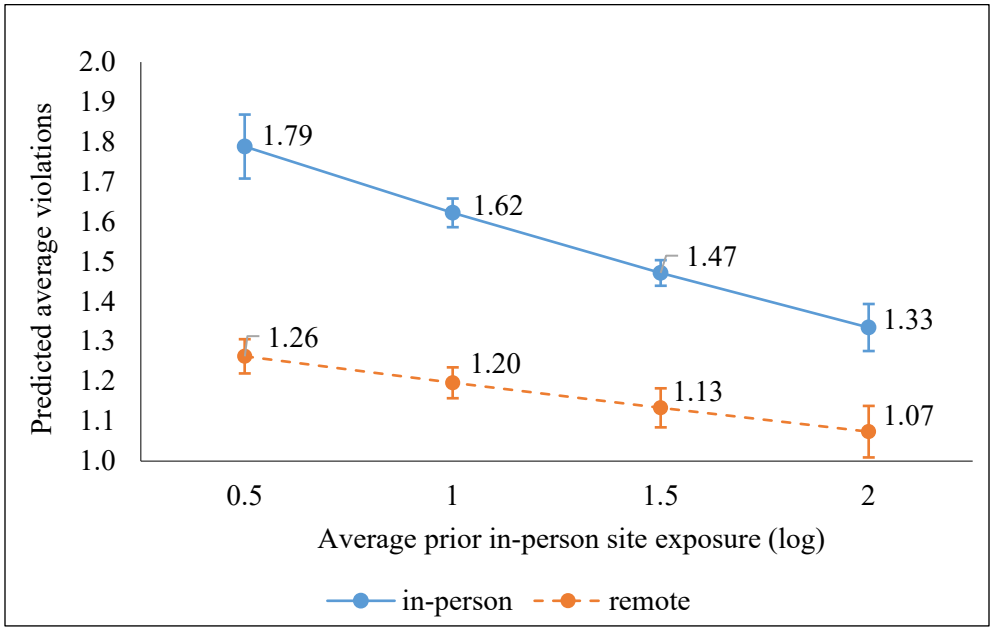
This figure relies on results from Table 3, Column 2, and depicts predictive average effects and their 95% confidence intervals. The decrement of remote audits for violations identified primarily through the observation of site processes is significantly larger than the decrement of remote audits for violations identified primarily through document review.

Figure 2. Average Predicted Violations for Remote vs. In-person Audits, by Single Auditors versus Auditor Teams



This figure relies on results from Table 3, Column 3, and depicts predictive average effects and their 95% confidence intervals. The gap in predicted average violations between remote and in-person audits is significantly larger for auditor teams than for single auditors.

Figure 3. Average Predicted Effects of Remote vs. In-person Audits, by Various Levels of Prior In-person Site Exposure



Predictive average effects and their 95% confidence intervals based on results reported in Column 4 of Table 3. The audit quality gap is attenuated when auditors have more prior in-person site exposure. The x-axis is a log scale because average prior in-person site exposure (prior audit) is logged in the model.

Online Appendix

. Table A1. Robustness Test: Predicting Minor Violations

	Dependent variable: Minor violations			
	(1)	(2)	(3)	(4)
Remote audit	-0.294**	-0.227**	-0.246**	-0.396**
	(0.026)	(0.029)	(0.027)	(0.039)
Remote audit × Direct-observation–based clauses		-0.425**		
		(0.045)		
Remote audit × Audit team			-0.292**	
			(0.046)	
Remote audit × Average prior in-person site exposure (log)				0.089**
				(0.025)
Audit team	-0.034	0.072	0.131**	0.029
	(0.041)	(0.047)	(0.045)	(0.057)
Average prior in-person site exposure (log)	-0.128**	-0.127**	-0.130**	-0.190**
	(0.020)	(0.023)	(0.020)	(0.027)
Focal standard advanced training [a]	0.040	0.024	0.042	0.041
	(0.034)	(0.040)	(0.035)	(0.034)
Staff-days (log)	0.434**	0.432**	0.437**	0.435**
	(0.051)	(0.060)	(0.052)	(0.051)
COVID time period	-0.066+	-0.241**	-0.073+	-0.069+
	(0.039)	(0.042)	(0.039)	(0.039)
Prior remote site exposure	-0.028	-0.044	-0.021	-0.042
	(0.029)	(0.032)	(0.029)	(0.029)
Maximum auditing experience (log)	-0.033**	-0.044**	-0.033**	-0.029*
	(0.012)	(0.014)	(0.012)	(0.012)
Female on audit team [a]	0.075*	0.075*	0.080*	0.074*
	(0.034)	(0.038)	(0.034)	(0.034)
Percent outsourced	-0.115**	-0.119**	-0.113**	-0.118**
	(0.040)	(0.044)	(0.040)	(0.040)
Multi-standard audit	-0.017	-0.023	-0.005	-0.022
	(0.041)	(0.046)	(0.041)	(0.041)
Site-standard fixed effects	Included		Included	Included
Site–standard–detection-mode fixed effects		Included		
Year fixed effects	Included	Included	Included	Included
Audit-sequence fixed effects	Included	Included	Included	Included
Observations (audits)	35,172	37,478	35,172	35,172
Number of sites	14,586	13,263	14,586	14,586
Number of site-standards	16,949	15,345	16,949	16,949

Poisson regression coefficients with standard errors are clustered by site; **p<0.01; *p<0.05; +p<0.10. The unit of analysis for Models 1, 3, and 4 is the site-standard-date and, for Model 2, is the site–standard–detection-mode–date for Column 2. Model 2's observation count is less than twice as large as the other models because it excludes ISO 13485 audits (for which we were unable to obtain data to map its clauses to primary audit detection mode) and because the inclusion of the three-way fixed effect (site–standard–detection-mode) results in this model dropping observations that report zero violations for all instances of a site's various audits of a given standard using a particular detection mode (e.g., if Site A's ISO 9001 audits always yielded zero violations of observation-based clauses). In Model 2, *observation-based clauses* is interpreted as the effect compared to the omitted category of *document-review–based clauses*; the main effect of *observation* is absorbed by the inclusion of site–standard–detection-mode fixed effects.

[a] Variables in which missing values were replaced with 0. All regression models include dummy variables to flag observations for which missing values of these variables were recoded to 0.

Table A2. Robustness Test: Predicting Violations of Mandatory Clauses

	Dependent variable: Violations of mandatory clauses			
	(1)	(2)	(3)	(4)
Remote audit	-0.346**	-0.254**	-0.283**	-0.441**
	(0.035)	(0.037)	(0.036)	(0.053)
Remote audit × Direct-observation–based clauses		-0.506**		
		(0.052)		
Remote audit × Audit team			-0.388**	
			(0.063)	
Remote audit × Average prior in-person site exposure (log)				0.082*
				(0.035)
Audit team	-0.010	-0.007	0.113+	-0.013
	(0.057)	(0.058)	(0.061)	(0.057)
Average prior in-person site exposure (log)	-0.104**	-0.099**	-0.106**	-0.160**
	(0.027)	(0.028)	(0.027)	(0.037)
Focal standard advanced training [a]	0.049	0.038	0.052	0.047
	(0.050)	(0.051)	(0.051)	(0.050)
Staff-days (log)	0.581**	0.584**	0.581**	0.580**
	(0.068)	(0.071)	(0.068)	(0.068)
COVID time period	-0.204**	-0.231**	-0.212**	-0.207**
	(0.051)	(0.052)	(0.051)	(0.051)
Prior remote site exposure	-0.005	-0.020	0.002	-0.016
	(0.039)	(0.040)	(0.039)	(0.039)
Maximum auditing experience (log)	-0.043**	-0.049**	-0.042**	-0.039*
	(0.016)	(0.017)	(0.016)	(0.016)
Female on audit team [a]	0.086+	0.099*	0.085+	0.086+
	(0.045)	(0.045)	(0.045)	(0.045)
Percent outsourced	-0.086+	-0.091+	-0.085	-0.090+
	(0.052)	(0.053)	(0.052)	(0.052)
Multi-standard audit	-0.015	-0.019	-0.001	-0.017
	(0.059)	(0.061)	(0.059)	(0.059)
Site-standard fixed effects	Included		Included	Included
Site–standard–detection-mode fixed effects		Included		
Year fixed effects	Included	Included	Included	Included
Audit-sequence fixed effects	Included	Included	Included	Included
Observations (audits)	24,581	27,622	24,581	24,581
Number of sites	10,615	10,412	10,615	10,615
Number of site-standards	11,834	11,609	11,834	11,834

Poisson regression coefficients with standard errors are clustered by site; **p<0.01; *p<0.05; +p<0.10. See notes for Table A1.

[a] Variables in which missing values were replaced with 0. All regression models include dummy variables to flag observations for which missing values of these variables were recoded to 0.

Table A3. Robustness Test: Expanded Sample to Include Bundled Audits

	Dependent variable: Violations			
	(1)	(2)	(3)	(4)
Remote audit	-0.328**	-0.282**	-0.294**	-0.375**
	(0.023)	(0.026)	(0.023)	(0.034)
Remote audit × Direct-observation–based clauses		-0.344**		
		(0.037)		
Remote audit × Audit team			-0.233**	
			(0.043)	
Remote audit × Average prior in-person site exposure (log)				0.039+
				(0.020)
Audit team	0.043	0.078+	0.125**	0.041
	(0.036)	(0.041)	(0.039)	(0.036)
Average prior in-person site exposure (log)	-0.140**	-0.131**	-0.140**	-0.167**
	(0.017)	(0.019)	(0.017)	(0.023)
Focal standard advanced training [a]	0.045	0.041	0.047+	0.045
	(0.028)	(0.033)	(0.028)	(0.028)
Staff-days (log)	0.432**	0.386**	0.431**	0.433**
	(0.046)	(0.053)	(0.047)	(0.046)
COVID time period	0.016	-0.087**	0.009	0.016
	(0.029)	(0.032)	(0.029)	(0.029)
Prior remote site exposure	-0.053*	-0.065**	-0.050*	-0.060**
	(0.023)	(0.025)	(0.023)	(0.023)
Maximum auditing experience (log)	-0.030**	-0.040**	-0.031**	-0.028*
	(0.011)	(0.013)	(0.011)	(0.011)
Female on audit team [a]	0.094**	0.097**	0.097**	0.093**
	(0.028)	(0.031)	(0.028)	(0.028)
Percent outsourced	-0.103**	-0.098**	-0.099**	-0.106**
	(0.033)	(0.037)	(0.033)	(0.033)
Multi-standard audit	0.002	0.007	0.007	0.000
	(0.035)	(0.037)	(0.035)	(0.035)
Hybrid audit	0.193**	0.199**	0.186**	0.182**
	(0.033)	(0.036)	(0.033)	(0.033)
Bundled audit	0.173**	0.283**	0.174**	0.174**
	(0.048)	(0.054)	(0.048)	(0.048)
Site-standard fixed effects	Included		Included	Included
Site–standard–detection-mode fixed effects		Included		
Year fixed effects	Included	Included	Included	Included
Audit-sequence fixed effects	Included	Included	Included	Included
Observations (audits)	43,220	47,484	43,220	43,220
Number of sites	16,988	15,634	16,988	16,988
Number of site-standards	20,538	18,673	20,538	20,538

Poisson regression coefficients with standard errors are clustered by site; **p<0.01; *p<0.05; +p<0.10. See notes for Table A1.

[a] Variables in which missing values were replaced with 0. All regression models include dummy variables to flag observations for which missing values of these variables were recoded to 0.

Table A4. Robustness Test of Model 2: Include All Three Audit Detection Modes

	Dependent variable: Violations
	(1)
Remote audit	-0.235**
	(0.029)
Remote audit × Direct-observation–based clauses	-0.423**
	(0.045)
Remote audit × Interview-based clauses	0.009
	(0.047)
Audit team	0.072
	(0.047)
Average prior in-person site exposure (log)	-0.124**
	(0.022)
Focal standard advanced training [a]	0.022
	(0.039)
Staff-days (log)	0.438**
	(0.059)
COVID time period	-0.230**
	(0.041)
Prior remote site exposure	-0.045
	(0.031)
Maximum auditing experience (log)	-0.041**
	(0.013)
Female on audit team [a]	0.102**
	(0.038)
Percent outsourced	-0.109*
	(0.043)
Multi-standard audit	-0.025
	(0.045)
Site–standard–detection-mode fixed effects	Included
Year fixed effects	Included
Audit-sequence fixed effects	Included
Observations (audits)	43,699
Number of sites	13,641
Number of site-standards	15,891

Poisson regression coefficients with standard errors are clustered by site; **p<0.01; *p<0.05; +p<0.10. The unit of analysis is the site–standard–detection-mode. The omitted category for detection mode is *document review.* The main effect of *observation* is absorbed by the inclusion of site–standard–detection-mode fixed effects.

[a] Variables in which missing values were replaced with 0. The model includes dummy variables to flag observations for which missing values of these variables were recoded to 0.