What Bishops Knew: Sexual Abuse and Institutional Response in the American Catholic Church*

Diego Battiston† Jordi Blanes i Vidal‡ David Cai§

August, 2026

Abstract

The Catholic Church is one of the world's largest, most enduring institutions. Its response to clergy sexual abuse is a leading case of how a self-governing organisation manages internal misconduct. We build a new micro-level panel linking documented allegations against U.S. Catholic clergy to the administrative actions taken against them, drawn from the Church's own annual clergy directory. Our first finding is that these actions (e.g., administrative leave, retirement, removal from active clergy) are associated with large reductions in subsequently documented victims. Second, we find that these actions were consistently underused, with the majority of accused priests remaining in their assignments or being moved to other dioceses. Our main finding is that institutional response varied enormously by bishop, dwarfing variation by diocese or across calendar time, including around the 2002 reform charter. We conclude that individual leadership, not national policy or geographical differences, drove how forcefully the Church acted. Our findings shed light on how self-governing institutions trade the informational advantage of self-policing against the incentive to conceal.

JEL classification: D23, K42, M51, Z12.

Keywords: Organisational Misconduct, Institutional Self-Governance, Sexual Abuse, Catholic Church, Difference-in-Differences.

*We thank Ayusmita Mukherjee, Ruta Rupeikyte, and Aleksandr Vedishchev for excellent research assistance. All errors remain our own.

†School of Economics, University of Edinburgh. Email: diego.battiston@ed.ac.uk.

‡Department of Management, London School of Economics and Centre for Economic Policy Research. Email: j.blanes-i-vidal@lse.ac.uk.

§Department of Management, London School of Economics. Email: J.Cai18@lse.ac.uk.

1 Introduction

The Catholic Church is one of the world's largest and most enduring institutions, counting roughly 1.4 billion members worldwide, about one in six people on the planet.[1] It is also one of the oldest continuously operating organisations in the world, predating every existing nation-state. Its internal governance is a hierarchy of dioceses answering to bishops, who are appointed by Rome but operate with substantial autonomy from it. This arrangement has its own literature in economics [Iannaccone, 1998]. An institution this size has consequences that reach far beyond its own membership.

Like other powerful institutions, the Church largely governs itself. Police departments, militaries, hospitals, universities, and churches maintain their own systems for detecting misconduct and imposing punishment, largely insulated from external review. This arrangement embodies a fundamental trade-off between information and incentives. The institution sits closest to the information, and can act faster than any external process. But the institution also bears the costs of enforcement (reputational damage, financial liability, the disruption of internal networks), while the harm from misconduct falls largely on outsiders. A self-governing institution can therefore act quickly, but may prefer not to act at all. How institutions react internally to detecting misconduct is a difficult empirical question because the available data is typically suppressed.

The clergy abuse crisis in the U.S. Catholic Church is a rare case where this veil has been forcibly lifted. The first prominent case was the 1984–85 criminal prosecution of Fr. Gilbert Gauthe in the Diocese of Lafayette, Louisiana, which established that a priest had serially abused children and that diocesan officials knew of his conduct and reassigned him regardless [Berry, 1985, 1992]. Especially after the Boston Globe's 2002 Spotlight investigation, a flood of disclosure followed: grand jury investigations, attorney-general reports, litigation, diocesan self-disclosures, and independent archives together documented thousands of credibly accused clergy across nearly every U.S. diocese [Rezendes, 2002, John Jay College of Criminal Justice, 2004, 2011]. This record makes the American case one of the best-documented, longest-running accounts of institutional misconduct and response available to social scientists.

The crisis is not confined to the United States: comparable inquiries have documented similar patterns of abuse and cover-up elsewhere.[2] But the U.S. Church is where the crisis first became public, where documentation is most extensive, and where the reform response discussed below was most codified, putting it at the forefront of both the scandal and the institutional reckoning with it. The Church's own National Review Board concluded that the graver institutional failure was not the abuse itself but the response to it.[3] Crucially, the record reveals not only who offended, but what the organisation knew

[1] Its statistical yearbook additionally reports over 5,000 hospitals, tens of thousands of dispensaries and care homes, and more than 150,000 primary and secondary schools operating under its auspices across every inhabited continent [Central Office of Church Statistics, Secretariat of State, 2025, Agenzia Fides, 2024].

[2] Notably the Royal Commission into Institutional Responses to Child Sexual Abuse in Australia [Royal Commission into Institutional Responses to Child Sexual Abuse, 2017], and government-commissioned inquiries in Ireland [Commission to Inquire into Child Abuse, 2009, Commission of Investigation into the Catholic Archdiocese of Dublin, 2009] and Germany [Dreßing et al., 2018].

[3] Bishops' "reliance on secrecy and an undue emphasis on the avoidance of scandal... ultimately bred far greater scandal

and when: allegations reported privately to officials, surfaced in lawsuits, or established in criminal proceedings can be dated and matched to the administrative actions (or inaction) that followed. Yet this record has been analysed almost entirely case-by-case, through journalism and litigation, rather than as a comparable, quantitative account of how a self-governing institution used the information it possessed.

A recurring theme in that record is that dioceses possessed, but did not consistently use, tools that could have limited an accused priest's access to potential victims. Even when a diocese did act, it sometimes reassigned the priest to a new parish or diocese rather than removing him from ministry. This narrative, captured by the phrase "passing the trash," has shaped public perception of the crisis and the policy response to it, above all the U.S. bishops' 2002 Dallas Charter.[4] Yet its basic empirical claims have not, to our knowledge, been tested with systematic data.

This paper answers these questions using a newly constructed micro-level panel built around roughly 8,000 cases identified by BishopAccountability.org. For each accused cleric, its archive assembles a narrative case history drawn from diocesan disclosures, grand jury and attorney-general reports, court records, and media coverage. Under an explicit, versioned coding protocol, we convert these narratives into (i) a count of documented victims per priest-year and (ii) an event panel of *church-knowledge* (CK) episodes: the dated points at which a diocese, religious order, or civil authority is documented to have learned of an allegation. Church knowledge can arise through a criminal proceeding, a civil lawsuit, a private report to Church officials, or the priest's inclusion on a diocesan list of credibly accused clergy. We combine this with an independent source for the administrative side of the panel: each priest's assignment status in a given year (active in ministry, on leave, retired, or absent from the directory). This is taken directly from the Official Catholic Directory, the Church's own annually published roster of clergy assignments. The resulting panel therefore pairs a directly observed administrative record with independently documented outcomes: the victims and church-knowledge episodes drawn from BishopAccountability.org.

We use these data to answer five questions. First, did the administrative tools available to a bishop actually reduce further documented abuse when they were used? Second, were these tools used, and used promptly, once a diocese had knowledge of an allegation, and, conditional on acting, did the diocese respond by removing the priest from ministry or by moving him elsewhere? Third, how much of the variation in institutional response was attributable to *who* the bishop was and *which* diocese he led, rather than a common response shifting with the broader reform environment over time? Fourth, was there a discontinuity in response around the 2002 Dallas Charter, or did any change instead unfold gradually and unevenly across dioceses? A fifth, more speculative exercise asks how many additional documented victims a counterfactual policy of uniformly prompt response might have avoided.

and reputational injury" than open acknowledgment of the abuse would have caused on its own [National Review Board for the Protection of Children and Young People, 2004].

[4]Formally the *Charter for the Protection of Children and Young People*; it established a nominal "zero tolerance" norm for substantiated abuse [United States Conference of Catholic Bishops, 2002]. We describe its provisions and enforcement in more detail in Section 2.

First, the tools worked: priest-years on leave, retired, or absent from the record show substantially and significantly fewer documented victims than priest-years in active ministry, with no evidence of pre-trends. Second, the tools were underused: even the most consequential response, a priest disappearing from the record entirely, followed a first church-knowledge episode only around 20 percentage points of the time, and the diocese-initiated statuses of retirement and leave rose by even less. Church knowledge was also associated with a smaller decline in the probability that a priest held a senior title such as monsignor, more likely reflecting being passed over for promotion than a title being stripped. Dioceses also transferred some priests to a different diocese, consistent with "passing the trash," but this explains only a modest share of the larger decline in same-diocese presence; most of that decline instead reflects the priest disappearing from the record altogether. Third, bishop- and diocese-level responsiveness varies enormously, an order of magnitude more than the heterogeneity across calendar years, which is statistically indistinguishable from zero, pointing to individual leadership rather than a common secular trend. Fourth, we find no evidence that the Dallas Charter produced a discontinuous shift in response: year-to-year variation is an order of magnitude smaller than cross-bishop variation, and shrinks to essentially zero once sampling noise is accounted for. Fifth, had every bishop responded at the rate of the 90th-percentile bishop, we calculate that on the order of 330 additional documented victims could have been avoided.

We use stacked difference-in-differences designs anchored, respectively, on assignment-status transitions and on church-knowledge dates, following the recent literature on staggered-adoption event studies [Callaway and Sant'Anna, 2021, Sun and Abraham, 2021, Goodman-Bacon, 2021, Cengiz et al., 2019]. For the bishop- and diocese-heterogeneity analysis, we recover unit-specific treatment effects by fully interacting the stacked-DiD specification with bishop, diocese, and year fixed effects, then apply empirical-Bayes shrinkage to separate signal from noise.[5] The same specification's year effects also test for a discontinuity coinciding with the 2002 Dallas Charter (Section 5.3).

This paper makes three contributions. First, we contribute to the literature on why organisations fail to manage internal misconduct [Garicano and Rayo, 2016], including how misconduct comes to light [Dyck et al., 2010] and how organisations handle misconduct by their own members [Adams-Prassl et al., 2024]. We provide, to our knowledge, the first systematic, quantitative account of an institution's own documented response to a sustained wave of internal allegations, built from individual administrative actions rather than policy announcements or case studies.

Second, most of the economics-of-religion literature studies the effect of religion on outcomes such as education, income, health, and crime [Iannaccone, 1998]. We instead show that how a religious institution is internally managed is itself a first-order object of study, and extend the outcomes this literature considers to a very different margin: documented child sexual abuse.

Third, we contribute to the literature on middle managers, which has shown that manager identity matters for productivity in public-sector and for-profit settings [Bertrand and Schoar, 2003, Lazear

[5] We additionally validate the resulting dispersion estimates against a model-free split-sample reliability check that correlates independently estimated effects across random halves of the underlying data (Section 5.3).

et al., 2015, Fenizia, 2022]. We show that individual bishops, the Church's own middle managers between Rome and parish priests, matter just as much for how forcefully misconduct was addressed. This holds even in an organisation whose managers answer, at least nominally, to a higher calling than career incentives or a bottom line.

Our findings speak to an active policy debate: in 2025, Washington State attempted to require clergy to report child abuse disclosed even during confession. This prompted a Vatican-backed threat of excommunication and a federal civil-rights investigation into the law as "anti-Catholic"; courts ultimately blocked enforcement of the confession-specific provision, and Washington was not an isolated case.[6] Our results suggest that mandating disclosure to the criminal justice system is likely necessary: even the most severe internal response followed a documented allegation only around 20 percent of the time, and whether it did so at all depended heavily on which bishop happened to be in charge.

The remainder of the paper proceeds as follows. Section 2 describes the institutional environment and the tools available to a bishop. Section 3 describes the data and its construction. Section 4 sets out the empirical strategy. Section 5 presents results, and Section 6 concludes. All tables and figures are collected at the end of the paper.

2 Institutional Background

The Catholic Church is organised hierarchically: at its head is the Pope in Rome, whose Curia includes the Congregation for the Doctrine of the Faith, the body with ultimate authority over clerical discipline. Beneath Rome, the United States is divided into dioceses and archdioceses, each headed by a bishop or archbishop appointed by the Pope, who holds canonical authority over the clergy and parishes in his territory. Most priests are diocesan priests: they are incardinated to a specific diocese and assigned by its bishop to parishes, schools, or other posts. A parallel structure exists for priests, brothers, and sisters in religious orders (Jesuits, Franciscans, Benedictines, and the like), who answer primarily to their own order's superior, not the local bishop, even when serving in his territory. This distinction matters: it determines who has the authority to act on a given allegation. Our panel pools both diocesan and religious clergy, but the relevant decision-maker, a bishop or a religious superior, differs across them.

Upon receiving an allegation, a bishop (or religious superior) has several administrative tools short of a full canonical trial. He can restrict a priest's faculties and access to ministry, an informal step that can be taken unilaterally and revised as an investigation proceeds. He can remove the priest from the diocese's public parish or clergy directory, which curtails his ability to be assigned or recognised as active clergy, though it does not, by itself, revoke his clerical status. He can place the priest on administrative leave pending investigation, a formally documented but in-principle reversible step, or accept or require his retirement, which is typically treated as final. He can also strip an honorific senior title, such as monsignor, a largely symbolic sanction. The most severe outcome, laicization (formal

[6] New Hampshire, North Carolina, Oklahoma, Texas, and West Virginia already require clergy to report abuse learned in confession without exception. Bills that would have removed the clergy-confession exemption were introduced but did not pass in California (2019), Delaware, Hawaii, Utah, and Vermont (2023), and Montana (2025).

removal from the clerical state), requires action by the Vatican's Congregation for the Doctrine of the Faith and is not separately observed in our directory-based status variable. Historically, dioceses have also been accused of quietly reassigning a priest to a new parish or diocese rather than removing him from ministry, a "passing the trash" pattern we test directly in Section 5.2.

The public record of this crisis did not begin with a single moment. Internal diocesan correspondence that later surfaced in litigation shows that Bishop Gerard Frey of Lafayette was privately informed of Fr. Gilbert Gauthe's abuse roughly a decade before Gauthe's 1984 indictment, evidence that private church knowledge of individual cases can substantially predate any public reckoning [Berry, 1992]. Gauthe's prosecution first broke the story into the Catholic press, but it was the 1992 case of Fr. James Porter of the Diocese of Fall River, Massachusetts, eventually linked to dozens of victims, that first brought clergy abuse to national television coverage, culminating in a 1993 guilty plea. It took another decade, and the Boston Globe's 2002 Spotlight investigation, for these individual scandals to become a sustained, nationwide reckoning [Rezendes, 2002].

The 2002 Dallas Charter (formally, the *Charter for the Protection of Children and Young People*) was adopted by the U.S. Conference of Catholic Bishops after the Spotlight revelations. It established a nominal "zero tolerance" norm (permanent removal from ministry following one substantiated act of abuse of a minor), along with diocesan review boards, victim-assistance coordinators, and mandated reporting requirements. Enforcement, however, remained heavily discretionary at the level of the individual bishop and diocese: the Charter set expectations but did not create a centralized enforcement mechanism, diocesan review boards' composition and independence varied, and a bishop retained wide latitude over the speed and thoroughness of an investigation and over which of the tools described above, if any, to apply and when. U.S. bishops are appointed by Rome and periodically rotate across dioceses or retire over the sample period, so "who is bishop of diocese d at the time of a given allegation" varies both across space and, for a fixed diocese, over time. This is the identifying variation we exploit in Section 5.3 to separate individual-bishop from diocese-level and calendar-time responsiveness.

Our data draw on two distinct sources. BishopAccountability.org, an independent nonprofit archive founded after the Spotlight investigation, compiles diocesan self-disclosed lists of credibly accused clergy, attorney-general and grand-jury reports, court and settlement records, and media coverage into a narrative case file for each accused cleric. These case files are the source for the abuse-incident and church-knowledge data used throughout this paper, but the administrative actions we study are not drawn from them at all. A priest's assignment status in a given year (active in ministry, on leave, retired, or absent from the roster) is instead determined directly from the Official Catholic Directory, the Church's own annually published, diocese-by-diocese listing of every priest's assignment. The Directory is compiled and published by the Church itself as a routine administrative record, independent of any particular allegation, investigation, or disclosure process. A priest's disappearance from, or reappearance in, this listing is therefore a direct read of the institution's own personnel record, not an inference drawn from litigation or media coverage.

3 Data

The primary dataset is a priest-year panel spanning 1949 to 2018 (of which 57 years are covered by a surviving Directory volume), covering approximately 8,000 cases identified by BishopAccountability.org. Each priest's panel runs from his first directory year through whichever of the following three years comes earliest: the year he died, the year 2018 (the last year covered by the sample), or exactly seven years after the year he was last observed in the directory.[7] Years for which the underlying Official Catholic Directory volume does not exist are dropped for all priests, so that the panel's implicit "at risk of being observed" window is common across priests rather than driven by idiosyncratic archival gaps.[8] For each priest-year we determine two things: the first is an assignment status, taken directly from the Official Catholic Directory, active in a U.S. diocesan post, on leave, retired, or absent from the directory. We further split absence into a temporary interior gap versus the tail period after the priest was last seen, since these plausibly reflect very different underlying situations.[9] The second is a count of documented victims, taken separately from the BishopAccountability.org case narratives.

Converting the case narratives into a priest-year victim count requires resolving two kinds of ambiguity in how sources describe abuse. The first concerns timing: a source may report abuse in a single year (e.g., one victim in 1950), across a specific multi-year range (e.g., one victim "in 1950–1955"), or only over an approximate period (e.g., "in the early 1950s," which we code by convention to a specific five-year span, 1950–1954). The first case is unambiguous; in the second, the source does not usually indicate whether the abuse was continuous throughout the stated range or concentrated near its endpoints, and in the third, even the range's boundaries are uncertain. This ambiguity compounds when a range contains multiple victims: a report of six victims abused "in 1950–1955" could mean one new victim each year, all six victims in every year of the range, or any pattern in between, and the source rarely specifies which. Reports are also frequently vague about the count itself, which we resolve to a conservative lower bound (e.g., "several" is coded as 2, "more than 6" as 6, and "dozens" as 24).

We use three victim-count measures to handle this ambiguity, summarized in Table 1. **Measure 1 (Maximal)** sums, for each priest-year, the full documented victim count of every abuse incident whose stated range includes that year; in the example above, all six victims would be assigned to each of the six years 1950–1955.[10] It is computed over the full priest-year panel described above. **Measure 2 (High Confidence)** restricts this count to victims from abuse incidents whose timing is precisely

[7] The seven-year window is an arbitrary but conventional choice; our results are not sensitive to the exact threshold.

[8] Thirteen annual volumes are missing entirely: 1950, 1953, 1959, 1996, 1997, 2000, 2001, 2002, 2003, 2006, 2007, 2011, and 2012. Because no priest is observed in these years, they are treated as not having existed, rather than as years of absence, and are removed from every priest's panel.

[9] The panel is constructed by matching each priest's entries across annual Directory volumes by name and surname. This matching is imperfect, since surnames are sometimes misread across scanned volumes; we use an algorithm designed to minimize both missed matches and false-positive matches across different priests, but it inevitably leaves occasional, idiosyncratic gaps in a priest's observed presence that reflect matching failures rather than a true absence from ministry. We therefore distinguish an interior gap, in which a priest is observed both before and after a missing year (for example, present in 1960 and 1962 but not 1961), from a tail gap, in which a priest is not observed in a given year and never observed again thereafter; the two are much more likely to reflect, respectively, a matching artifact and a genuine departure.

[10] Counts are winsorized at 15 to limit the influence of a small number of priests with extreme, and sometimes only loosely bounded, victim counts.

dated, an exact year or an explicit stated period, rather than only approximately known, such as a decade or a range inferred from an assignment or tenure window. Because a priest is only included in this measure's sample if he has at least one such precisely dated victim, Measure 2 is computed over a smaller, restricted subsample of priests, and correspondingly of priest-years (87,064 priest-years), rather than over the same full priest-year panel used for Measure 1 (126,164 priest-years). The difference in sample size therefore reflects a difference in *which priests* are included, not a difference in how the underlying panel is constructed. **Measure 3 (Fractional)** instead divides each incident's documented victim count evenly across its candidate years, and, separately, across candidate postings when a priest served in two overlapping assignments without source material pinning the incident to one rather than the other. In the example above, each of the six years 1950–1955 would instead be assigned one victim ($6/6 = 1$), constructed to have positive mass in exactly the priest-years where Measure 1 is positive, hence the identical sample size but much lower per-year mean. All three measures are far from normally distributed, with means well under one victim per priest-year and a long right tail, consistent with the underlying data reflecting a relatively small number of priests with extensive documented abuse histories against a much larger group with one or a few documented victims.

From the same case narratives we construct an event panel of church-knowledge (CK) episodes: dated instances at which a diocese, religious order, or civil authority is documented to have learned of an allegation against a priest. Each CK episode is classified by *knowledge type* into one of five categories recorded in the underlying coding protocol: a criminal proceeding (`police`), a civil lawsuit or settlement (`civil`), a documented diocesan or order action such as removal from ministry (`action`), a private report made directly to Church officials (`private`), or inclusion on a diocesan or attorney-general public list of credibly accused clergy (`list`). We order CK episodes within priest so that we can distinguish a priest's *first* CK from subsequent ones, and for each CK episode we observe the priest's diocese and assignment status immediately before and after, which lets us construct the diocese-transfer measure used in Section 5.2.

Because both the abuse incidents and the church-knowledge episodes are hand-coded from free-text narrative sources under an explicit, versioned coding protocol, we emphasize two features of the construction that bear on interpretation. First, the coding rules are conservative by design: victim counts default to the lowest defensible figure consistent with the source language (e.g., a bare plural is coded as "2+" rather than an arbitrary larger number), and church-knowledge dates are only coded when a source directly documents an awareness event, rather than being inferred from, say, the abuse date alone. This conservatism should if anything work against finding the effects we report. Second, our source material is itself the product of disclosure (diocesan self-disclosure, litigation discovery, or investigative reporting). Both abuse and church-knowledge are therefore necessarily *documented* abuse and *documented* knowledge; we return to the implications of this in Section 6. The assignment-status side of the panel is not subject to this concern: it is read directly off the Official Catholic Directory's own published listings rather than hand-coded from disclosure-based narratives, so it does not share the same potential dependence on litigation or media disclosure.

4 Empirical Strategy

We use two related stacked difference-in-differences designs.

For the claim that tools reduce victims (Section 5.1), the event is a transition in a priest's assignment status: onto leave, into retirement, or out of the public directory. We estimate the association between assignment status and documented victims in a priest-year panel,

$$y_{it} = \sum_{s} \beta_s \, \mathbf{1}[\text{status}_{it} = s] + \gamma \left(\log(\text{years since ordination}_{it} + 1) \times \text{recorded_ordination}_i\right) + \alpha_i + \delta_t + \varepsilon_{it},$$

with priest and year fixed effects (α_i, δ_t) and standard errors clustered by priest, estimated by Poisson pseudo-maximum-likelihood, OLS on levels, OLS on $\log(y+1)$, and a linear probability model for $\mathbf{1}[y_{it} \geq 1]$ (Table 2). Identification of β_s comes from within-priest variation in status over time. These coefficients should therefore be read as describing how a given priest's documented-victim rate changes around his own status transitions, net of a flexible age-at-ordination control, rather than as a comparison across priests who ever versus never experience a given status. We complement this with an event-study version that traces out victims relative to the year of the status transition (Figure 1). This lets us check directly for pre-trends that would undercut a causal reading of the levels regression.

For the claims that tools are underused and that response varies by bishop and diocese (Sections 5.2–5.3), the event is a priest's church-knowledge date. We build two such designs. A *first-CK* anchor is a priest's earliest documented, exactly-dated church-knowledge episode, excluding cases where an earlier but only approximately dated episode may in fact have come first, and excluding episodes whose only basis is the priest's inclusion on a public list. A *subsequent-CK* anchor is a documented, exactly-dated CK episode that is not a priest's first, preceded by no other CK of his own in the previous seven years and by no prior spell on leave.

For each such anchor we build a stack of priest-years. The treated observation is the anchored priest; controls are matched to him on the anchor year, on his year-by-year presence or absence in the public directory over the years leading up to the anchor, and on entry cohort. For first-CK anchors, controls are priests who had not yet had a documented church-knowledge episode as of the anchor year; because nearly every priest in the underlying case archive eventually has one, these controls typically differ from the treated priest mainly in timing, since their own first CK arrives more than seven years later. For subsequent-CK anchors, controls are priests with at least one earlier CK of their own who remain free of any new CK within seven years on either side of the anchor, and who, like the anchored priest, have never previously been placed on leave. The same treated-versus-control assignment is used for every outcome we study (no access, directory removal, retirement, leave, senior title); we do not construct a separate control group for each outcome that excludes priests experiencing some other, competing status change. We estimate the average treatment effect on the treated (ATT) via a stacked regression with stack-by-relative-time and priest-within-stack fixed effects (our "spec-10b"). This stacked design avoids the negative-weighting problem that can afflict two-way fixed-effects estimators when treatment

is staggered and effects are heterogeneous across cohorts. Each stack's control group is drawn only from not-yet-treated priests observed in the same calendar window as that stack's own treated priest.

To recover bishop-, diocese-, and year-specific treatment effects we augment this specification with a full set of interactions: $\text{bishop}_n \times \text{treated} \times \text{post}$, $\text{diocese}_n \times \text{treated} \times \text{post}$, and $\text{anchor_year}_n \times \text{treated} \times \text{post}$. These are estimated jointly in a single regression with stack-by-relative-time fixed effects absorbed, so that each unit's coefficient nets out the other two dimensions of heterogeneity, rather than conflating, say, a bishop effect with the diocese he happens to lead. Unit-specific coefficients are estimated with heterogeneous precision, since bishops and dioceses vary widely in how many stacks they contribute, so we apply empirical-Bayes shrinkage: for each unit class we compute the equal-weighted method-of-moments shrinkage factor $\hat{\tau}^2 = \max(0,\ \widehat{\text{Var}}(\hat{\beta}) - \overline{\widehat{se}^2})$ and shrink each unit's raw estimate toward the grand mean in proportion to its own sampling variance, $\hat{\theta}_u = \bar{\beta} + \frac{\hat{\tau}^2}{\hat{\tau}^2 + \widehat{se}_u^2}(\hat{\beta}_u - \bar{\beta})$. We report both the raw and shrunk standard deviation of unit effects (Table 5). We impose a minimum of five contributing stacks per bishop and, symmetrically, per diocese, and experiment with alternative thresholds as a robustness check, to confirm that our conclusions are not an artifact of including units estimated off very few stacks.

Empirical-Bayes shrinkage is model-based: it relies on the assumption that sampling variances are correctly estimated and that the underlying unit effects are drawn from a common distribution. We therefore validate the resulting dispersion estimates with a model-free split-sample reliability check. We randomly split the stacks contributing to each bishop's (diocese's) estimate into two halves, re-estimate the fully-interacted regression separately on each half, and correlate the two independent estimates of the same underlying bishop (diocese) effect. A high split-half correlation is only possible if a meaningful share of the cross-sectional variation in raw estimates reflects a genuine, persistent unit effect rather than noise. Noise-driven variation would be independent across the two halves, and so uncorrelated. Under the assumption that estimation noise is independent across the two halves, the covariance of the two halves' estimates recovers the variance of the underlying fixed effect itself, giving an "implied" standard deviation, $\widehat{SD}(\theta) = \sqrt{\max(0,\ \widehat{\text{Cov}}(\hat{\theta}^{(1)}, \hat{\theta}^{(2)}))}$, with confidence intervals obtained by a pairs bootstrap (Table 6, Figure 10).

Finally, Section 5.4 combines the priest-year victim-reduction estimates of Table 2 with the estimated cross-bishop dispersion in responsiveness from Table 5 to construct a counterfactual: how many additional documented victims would plausibly have been avoided had every bishop responded to first church knowledge at the rate of a high-responsiveness benchmark bishop. The construction is described in full in Section 5.4 and in Table 7's notes.

5 Results

5.1 US bishops had internal tools to reduce the number of victims

Table 2 shows that priest-years spent on leave, retired, or absent from the record are associated with significantly fewer documented victims than priest-years in active ministry. Relative to a sample mean

of 0.51 victims per priest-year, being on leave is associated with 0.41 fewer victims (OLS), a 14.3 log-point reduction, and an 11.9 percentage-point lower probability of any documented victim. Retirement is associated with a smaller but still significant reduction (0.21 victims, 6.4 log points, 5.3pp). The period after a priest was last seen in the directory (effectively, having left active ministry altogether) is associated with the largest reduction of all (0.52 victims, 18.3 log points, 15.6pp). A temporary interior gap in the directory record, by contrast, shows only a marginally significant association, consistent with such gaps often reflecting incomplete record-keeping rather than a genuine removal from ministry. It is reassuring that this placebo-like category behaves differently from the three status changes that plausibly reflect an actual administrative decision.

The magnitudes line up with a simple opportunity-access mechanism: leave and post-departure absence, which most directly curtail a priest's contact with potential victims, show the largest reductions, while retirement (which in practice sometimes still permits some residual contact within a diocese) shows a smaller one. Figure 1 shows the corresponding event studies: victims decline sharply and durably in the years following each type of status transition, with no evidence of pre-trends. This supports a causal reading of these associations, rather than reverse causality in which priests are moved onto leave or into retirement only after their documented abuse had already stopped for unrelated reasons.

5.2 Tools are underused

Given that these tools work, how often are they actually used? Figure 2 and Table 3 show that even the most consequential response is used only a minority of the time. A first CK raises the probability that a priest becomes absent from the public record by 19.93 percentage points (off a base rate of 40.99%), and of directory removal by 13.54pp (off 33.77%). These are not cheap, cosmetic outcomes: a priest who disappears from the record has typically left active ministry entirely, through resignation or laicization, the same status that Section 5.1 shows carries the largest reduction in documented victims of any we study. The critical fact, however, is the level of this response rather than its shape: even for this most severe outcome the increase is only around 20 percentage points, not 100. In the large majority of cases, church knowledge is not followed by the priest disappearing from the record at all. The formally diocese-initiated statuses of retirement and leave rise by even less (1.12pp and 5.26pp, respectively). Church knowledge is also associated with a smaller but statistically significant decline in the probability that a priest holds a senior honorific title such as monsignor (-0.98pp for a first CK). Our stacked difference-in-differences design identifies a treatment-control gap rather than a within-priest loss event, and such titles are typically conferred later in a career rather than revoked; this more plausibly reflects accused priests being passed over for promotion than titles being formally stripped. Subsequent CK episodes show a similar pattern, with somewhat larger base rates, consistent with priests who have already had one CK being further along in this process.

This pattern suggests that when the record does change, it is disproportionately because the priest himself has left ministry or been laicized once his case became public, rather than because a bishop formally placed him on leave or accepted his retirement. (Appearing on a diocesan list, for instance, is

itself often the CK event under our knowledge-type coding.) The formally diocese-initiated statuses of retirement and leave remain comparatively rare responses to church knowledge, and, as noted above, no response at all remains the modal outcome across every status we observe.

Does the response at least strengthen with repeated allegations against the same priest? Figure 4 traces the no-access ATT separately by CK ordinal (1st through 4th-or-later). If anything, the pattern is flat to mildly declining across ordinals, offering little support for a story in which dioceses ratchet up their response as evidence against a priest accumulates; a priest facing his fourth documented church-knowledge episode is not visibly more likely to lose access than one facing his first.

Finally, we ask whether dioceses respond to church knowledge by relocating a priest rather than removing him. Table 4 shows that they do, at least in part. The ATT on being reassigned to a *different* diocese around church knowledge is positive, and for subsequent CKs clearly significant (0.4–1.1pp, off a base rate of only 2–4%); this is essentially unchanged whether unknown post-CK locations are dropped or conservatively coded as a diocese change. This is consistent with the "passing the trash" pattern: bishops did, on average, respond to some allegations by transferring the priest elsewhere rather than removing him. At the same time, the ATT on remaining in the *same* diocese's active roster is far larger in magnitude, negative, and highly significant (roughly −19 to −20pp on base rates of 38–55%), for both first and subsequent CKs. Transfer to a different diocese therefore accounts for only a modest share of the overall decline in same-diocese presence; most of it instead reflects the priest disappearing from the record altogether (via loss of access, directory removal, leave, or retirement), as documented above. We stress that this evidence speaks specifically to reassignment *across* dioceses, and it does not rule out a priest being moved between parishes or posts *within* the same diocese, which our diocese-level transfer measure cannot detect.

5.3 Bishop heterogeneity

If tools work but are underused on average, does the shortfall vary systematically by who is in charge? Figures 5–9 plot empirical-Bayes-shrunk bishop- and diocese-level effects, ranked, together with shrinkage confidence intervals, for the baseline specification, a robustness check restricted to the largest connected set of bishops and dioceses, a specification augmented with priest characteristics, and a placebo cohort pre/post trend check. Across specifications the ranking of bishop effects is stable and shows wide dispersion: some bishops' dioceses respond to church knowledge far more decisively than others, even after shrinking each estimate toward the grand mean in proportion to its own sampling noise. This ranking is not materially disturbed by restricting to the most densely connected subsample or by adding priest-level controls, which argues against the dispersion being an artifact of compositional differences in which priests and dioceses happen to be well identified.

Table 5 quantifies this dispersion and compares it against calendar-year heterogeneity: the empirical-Bayes-shrunk standard deviation of bishop effects is 23.84 percentage points (raw: 37.79pp, 201 bishops), and of diocese effects 43.54pp (raw: 47.02pp, 54 dioceses). Both are far larger, and far more robust to shrinkage, than the year effects, whose raw dispersion of 12.93pp shrinks essentially to zero (0.00pp across

64 years) once sampling noise is accounted for. In other words, once we condition on the stacked-DiD design's own controls for calendar time, there is no detectable residual *secular* trend in how forcefully the Church responded to church knowledge. Nearly all of the meaningful variation is cross-sectional, tied to individual bishops and dioceses rather than to when in the sample period the episode occurred. This is a striking result given the prominence of the 2002 Dallas Charter as a nominal turning point: whatever effect the Charter and the broader reform environment had on norms, it did not translate into a detectable common shift in actual response rates once individual bishop and diocese identity are held fixed. This is consistent with the Charter's enforcement having remained, in practice, a matter of local discretion rather than centralized mandate.

Empirical-Bayes shrinkage is itself a model-based construct, so Table 6 and Figure 10 report a model-free check: correlating bishop and diocese effects estimated independently on two random halves of the underlying stacks. Bishop effects correlate strongly and precisely across halves ($\hat{\rho} = 0.535$, 95% CI $[0.345, 0.704]$, $N = 64$ bishops observed in both halves), implying an underlying bishop-effect standard deviation of 22.74pp (CI $[16.01, 28.48]$) that closely matches the empirical-Bayes estimate of 23.84pp, cross-validating it with an entirely different, assumption-light method. The corresponding diocese-level check is directionally consistent but far less precise ($\hat{\rho} = 0.161$, 95% CI $[-0.324, 0.775]$, $N = 19$ dioceses meeting the five-stacks-per-half minimum in both halves), reflecting the much smaller number of dioceses with enough stacks to support a split-sample comparison. We do not read the resulting implied diocese SD as informative given how wide its confidence interval is, and instead treat the empirical-Bayes diocese estimate in Table 5 as the more reliable of the two for that unit.

5.4 Victims avoided

Sections 5.1–5.3 together imply a simple counterfactual exercise: some bishops responded to church knowledge far more decisively than others, and more decisive response is associated with fewer subsequent documented victims. How many additional victims might have been avoided had the less responsive bishops matched the more responsive ones? We construct this counterfactual as follows. Using the empirical-Bayes-shrunk bishop-level no-access ATTs from the first-CK, largest-connected-set specification underlying Figure 6 (bmin = 5), we set a benchmark responsiveness level δ_P equal to the P-th percentile of the bishop-level distribution and compute each bishop's shortfall $\Delta_b = \max(0,\ \delta_P - \hat{\theta}_b)$. We apply this shortfall as a constant extra probability of no access over a fixed seven-year post-first-CK window for every focal priest under bishop b, giving $n_b \times 7 \times (\Delta_b/100)$ extra no-access priest-years in aggregate. We then convert extra no-access priest-years into avoided victims using the blended per-year victim effect from Table 2's OLS-count column, weighting the leave, retired, and absent (tail) coefficients by the empirical mix of these three underlying statuses actually observed among post-CK no-access years for treated priests, since a shift into "no access" status can correspond to any of the three.

Table 7 reports the results for three benchmark percentiles, with 95% confidence intervals from a 1,000-draw bootstrap over bishops. Under our primary benchmark, the 90th-percentile bishop, the

average bishop's shortfall is 14.08 percentage points, implying 704 extra no-access priest-years and 333.6 avoided documented victims (95% CI $[225.7, 527.6]$) in aggregate across the sample. The 75th- and 95th-percentile benchmarks give a similar order of magnitude: 247.4 avoided victims (CI $[155.8, 327.7]$) and 389.5 (CI $[268.6, 782.3]$), respectively. The estimates widen, as expected, as the benchmark is pushed further into the tail of an increasingly thinly estimated part of the bishop-effect distribution. We emphasize that this exercise is a mechanical extrapolation of the estimated cross-bishop dispersion documented in Section 5.3, not a structural policy simulation: it assumes that a less responsive bishop could have matched a more responsive one's behaviour, without any offsetting change in, for instance, the rate at which allegations were reported to begin with. It also inherits the sampling uncertainty of the underlying bishop-level effects, reflected in the fairly wide confidence intervals; read cautiously, it nonetheless puts a concrete, if approximate, number on the human cost implied by the bishop-level heterogeneity documented above.

6 Conclusion

Using a newly constructed micro panel of documented clergy sexual abuse and diocesan administrative response, we show that the tools available to Catholic bishops (restricting access, removing a priest from the public directory, placing him on leave, or accepting his retirement) are associated with large and significant reductions in subsequently documented victims. Yet even the most consequential of these tools, a priest disappearing from the public record entirely, followed a documented allegation only around 20 percentage points of the time, far short of a systematic response, and there is no evidence that dioceses ramped up their response as allegations against a given priest accumulated. We also find some support for the popular narrative that dioceses responded to allegations by transferring priests to new postings, though this explains only a modest share of the much larger decline in the probability that priests remained listed within the same diocese; most of that decline instead reflects priests disappearing from the record altogether. The substantial cross-bishop and cross-diocese dispersion in responsiveness is reliable enough to survive an independent split-sample check, and far larger than any detectable calendar-time trend, pointing to individual institutional leadership, rather than a shared secular shift in norms, as the dominant driver of how forcefully the Church acted once it knew. A back-of-the-envelope counterfactual suggests the human stakes of that dispersion are not small: on the order of a few hundred additional documented victims, in our sample alone, if the least responsive bishops had matched a high-responsiveness benchmark.

These findings come with important caveats. Our victim and church-knowledge data reflect only *documented* abuse and *documented* knowledge, both of which are themselves outcomes of institutional and legal disclosure processes. These processes may correlate with the very bishop- and diocese-level factors we study. If more responsive bishops are also more likely to have their dioceses' records surface in litigation or attorney-general investigations, our bishop-level dispersion estimates could partly reflect disclosure intensity rather than underlying behaviour, though it is not obvious this would bias the sign

of our main results. This concern applies less to the administrative-action side of our panel: a priest's assignment status is read directly off the Official Catholic Directory's own published listings, rather than reconstructed from litigation or media disclosure. It is therefore not itself a product of the same disclosure process we worry may be correlated with bishop responsiveness. The coding of free-text case narratives into structured victim-count and church-knowledge variables, while governed by an explicit and versioned protocol, necessarily involves judgment in ambiguous cases; we have tried throughout to favour conservative coding rules that if anything work against our findings. Our identification of church-knowledge dates, while conservative, cannot rule out earlier informal or undocumented awareness, which would tend to attenuate our estimated treatment effects toward zero rather than inflate them. And the victims-avoided counterfactual of Section 5.4 is a mechanical extrapolation, not a structural policy simulation, and should be read as illustrative of scale rather than as a precise policy-relevant estimate.

We view the paper's central contribution as demonstrating that these questions are answerable with comparable, quantitative evidence, rather than case-by-case narrative alone, and it also serves as a template for extending this approach to other institutional settings in which an organisation's own personnel records can be linked to documented misconduct. Natural extensions include exploiting bishop transitions within a diocese to further sharpen the separation of bishop- from diocese-specific responsiveness, examining whether observable bishop characteristics (age at appointment, prior diocese, seminary cohort) predict the dispersion we document, and extending the victims-avoided exercise to alternative, non-mechanical counterfactuals.

References

Abi Adams-Prassl, Kristiina Huttunen, Emily Nix, and Ning Zhang. Violence against women at work. *The Quarterly Journal of Economics*, 139(2):937–991, 2024. doi: 10.1093/qje/qjad045.

Agenzia Fides. Vatican – catholic church statistics 2024. Agenzia Fides, October 2024. URL `https://www.fides.org/en/news/75542-VATICAN_CATHOLIC_CHURCH_STATISTICS_2024`. Released for World Mission Sunday, 20 October 2024; data updated to 31 December 2022.

Jason Berry. Pedophile priest: Study in inept church response. *National Catholic Reporter*, June 1985.

Jason Berry. *Lead Us Not into Temptation: Catholic Priests and the Sexual Abuse of Children.* Doubleday, New York, 1992.

Marianne Bertrand and Antoinette Schoar. Managing with style: The effect of managers on firm policies. *Quarterly Journal of Economics*, 118(4):1169–1208, 2003.

Brantly Callaway and Pedro H. C. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021. doi: 10.1016/j.jeconom.2020.12.001.

Doruk Cengiz, Arindrajit Dube, Attila Lindner, and Ben Zipperer. The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454, 2019. doi: 10.1093/qje/qjz014.

Central Office of Church Statistics, Secretariat of State. Annuarium statisticum ecclesiae 2023 (statistical yearbook of the church 2023). Technical report, Libreria Editrice Vaticana, Vatican City, 2025. Data on Catholic population as of 30 June 2023; other data as of 31 December 2023.

Commission of Investigation into the Catholic Archdiocese of Dublin. Report of the commission of investigation into the catholic archdiocese of dublin (the murphy report). Technical report, Department of Justice and Equality, Dublin, 2009. Submitted July 2009; published 26 November 2009; chaired by Judge Yvonne Murphy.

Commission to Inquire into Child Abuse. Report of the commission to inquire into child abuse (the ryan report). Technical report, Stationery Office, Dublin, 2009. 5 volumes; chaired by Justice Sean Ryan.

Harald Dreßing, Hans Joachim Salize, Dieter Dölling, Dieter Hermann, Andreas Kruse, Eric Schmitt, and Britta Bannenberg. Sexueller missbrauch an minderjährigen durch katholische priester, diakone und männliche ordensangehörige im bereich der deutschen bischofskonferenz (mhg-studie). Technical report, Deutsche Bischofskonferenz, Bonn, September 2018.

Alexander Dyck, Adair Morse, and Luigi Zingales. Who blows the whistle on corporate fraud? *Journal of Finance*, 65(6):2213–2253, 2010. doi: 10.1111/j.1540-6261.2010.01614.x.

Alessandra Fenizia. Managers and productivity in the public sector. *Econometrica*, 90(3):1063–1084, 2022. doi: 10.3982/ECTA19244.

Luis Garicano and Luis Rayo. Why organizations fail: Models and cases. *Journal of Economic Literature*, 54(1):137–192, 2016. doi: 10.1257/jel.54.1.137.

Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, 2021. doi: 10.1016/j.jeconom.2021.03.014.

Laurence R. Iannaccone. Introduction to the economics of religion. *Journal of Economic Literature*, 36 (3):1465–1495, 1998.

John Jay College of Criminal Justice. The nature and scope of sexual abuse of minors by catholic priests and deacons in the united states, 1950–2002. Technical report, United States Conference of Catholic Bishops, Washington, DC, 2004.

John Jay College of Criminal Justice. The causes and context of sexual abuse of minors by catholic priests in the united states, 1950–2010. Technical report, United States Conference of Catholic Bishops, Washington, DC, May 2011.

Edward P. Lazear, Kathryn L. Shaw, and Christopher Stanton. The value of bosses. *Journal of Labor Economics*, 33(4):823–861, 2015.

National Review Board for the Protection of Children and Young People. A report on the crisis in the catholic church in the united states. Technical report, United States Conference of Catholic Bishops, Washington, DC, February 2004.

Michael Rezendes. Church allowed abuse by priest for years. *The Boston Globe*, January 2002. Globe Spotlight Team investigation.

Royal Commission into Institutional Responses to Child Sexual Abuse. Final report. Technical report, Commonwealth of Australia, Canberra, 2017. 17 volumes.

Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199, 2021. doi: 10.1016/j.jeconom.2020.09.006.

United States Conference of Catholic Bishops. Charter for the protection of children and young people. United States Conference of Catholic Bishops, June 2002.

7 Tables and Figures

Table 1: VICTIM COUNTS PER PRIEST-YEAR, BY MEASURE

	N	Mean	SD	Min	Max
Measure 1 (Maximal)	126,164	0.512	1.900	0	15
Measure 2 (High Confidence)	87,064	0.189	0.719	0	10
Measure 3 (Fractional)	126,164	0.086	0.355	0.00	15.00

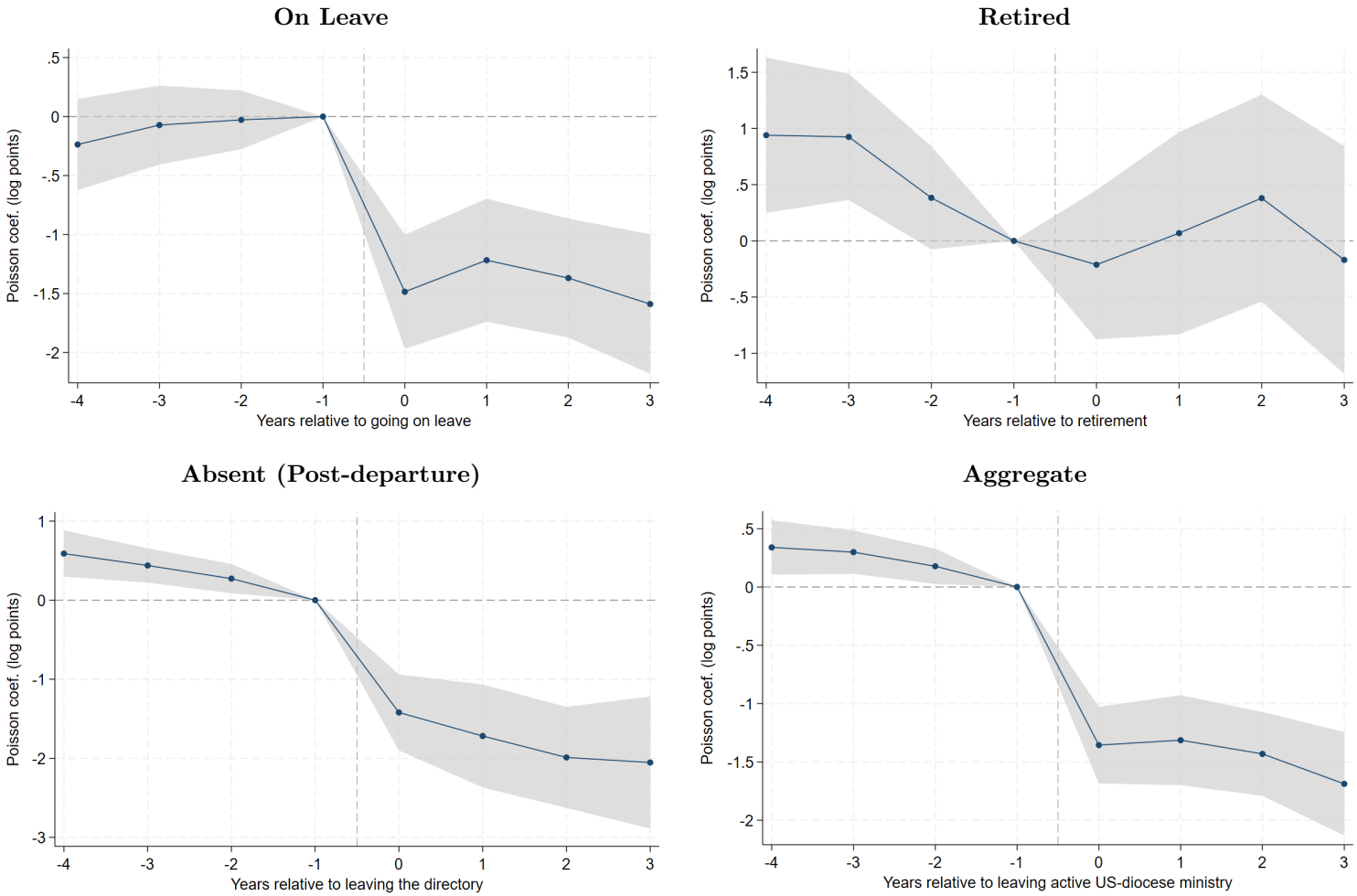
Notes: The underlying sample is the priest-year panel used throughout the paper: a panel dataset of all priests appearing in the case archive compiled by BishopAccountability.org (approximately 8,000 cases). For each priest, the panel begins in the year the priest first appears in the Official Catholic Directory and ends in whichever of the following three years comes earliest: the year the priest died, the year 2018 (the last year covered by the sample), or exactly seven years after the year the priest was last observed in the directory. Measure 1 (Maximal) and Measure 3 (Fractional) are both computed over this full priest-year panel (126,164 priest-years) and therefore share the same sample by construction. Measure 2 (High Confidence) is computed over a smaller, restricted sample: only the priest-years belonging to priests who have at least one documented victim meeting a stricter, higher-confidence identification standard are included (87,064 priest-years). The lower sample size for Measure 2 therefore reflects a difference in which priests are included, not a difference in how the priest-year panel itself is constructed; the panel construction rule described above is identical across all three measures.

Table 2: ASSIGNMENT STATUS AND DOCUMENTED VICTIMS

	(1) Poisson Victims	(2) OLS Victims	(3) OLS $\log(\text{Victims}+1)$	(4) LPM $\mathbf{1}[\text{Victim} \geq 1]$
On leave	-0.4975***	-0.4100***	-0.1431***	-0.1191***
	(0.1240)	(0.0614)	(0.0145)	(0.0101)
Retired	-0.9534***	-0.2084***	-0.0636***	-0.0534***
	(0.1974)	(0.0537)	(0.0123)	(0.0082)
Absent (interior gap)	-0.0845	-0.0551	-0.0248**	-0.0278***
	(0.0688)	(0.0325)	(0.0091)	(0.0079)
Absent (after last seen)	-1.1660***	-0.5214***	-0.1832***	-0.1564***
	(0.1706)	(0.0448)	(0.0106)	(0.0074)
Log years since ordination × recorded	0.5962***	0.1176***	0.0280**	0.0102
	(0.0638)	(0.0329)	(0.0091)	(0.0074)
Mean of dep var	0.5124	0.5124	0.2018	0.1905

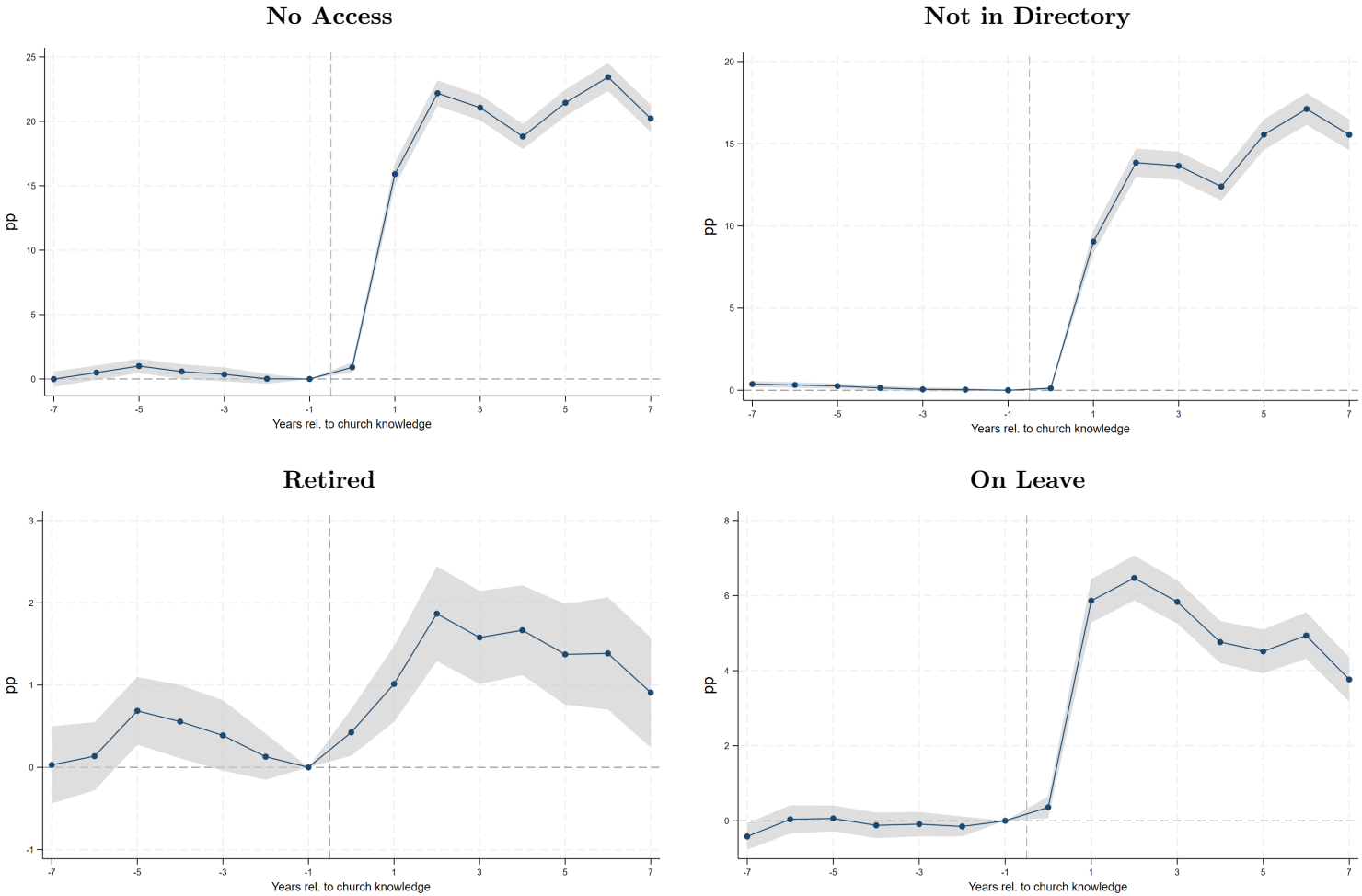
Notes: Dependent variable is the number of children who could have been abused by the priest in a given year (sum across overlapping incidents), using all documented victims, winsorized at 15. Column 3 uses $\log(\text{Victims}+1)$; column 4 uses an indicator for at least one victim. Sample: BishopAccountability-named priests; panel runs from each priest's first directory year through $\min(\textit{last_seen}+7,\ \textit{died_year},\ 2018)$. Years for which the Directory volume does not exist are excluded for all priests. The omitted reference category is U.S.-diocese assignment. The age control is the interaction of an indicator for a recorded ordination year with $\log(\text{years since ordination}+1)$. All columns include priest and year fixed effects, with standard errors in parentheses clustered at the priest level. N priest-years: 126,163 (Poisson column: 126,163). N priests: 4,108.

Figure 1: STACKED-DID EVENT STUDY OF ASSIGNMENT-STATUS TRANSITIONS ON VICTIMS



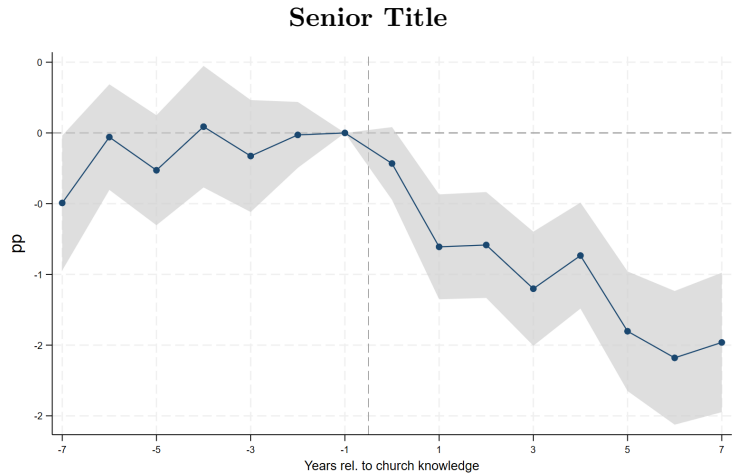
Notes: Each panel is a stacked event study of Measure 1 (Maximal) documented victims around a transition in assignment status. "On leave" and "retired" are read from the priest's listed status in the Official Catholic Directory (leave includes explicit leave, unassigned, awaiting-assignment, and similar language; retired is any status text indicating retirement); "absent" means the priest does not appear in the Directory at all that year. The three named panels correspond to three separate transitions – into on-leave status, into retirement, and into post-departure absence (leaving the directory after last being seen) – and the Aggregate panel pools all three. The event window runs from four years before the transition (relative year −4) through three years after (relative year +3), counted in consecutive *observed* panel years rather than consecutive calendar years, so that years for which no Directory volume survives do not count against the window. A treated priest spends the four years before the transition in an active U.S.-diocese post and the transition year plus the following three years in the target status, has at least one documented church-knowledge episode dated in the four years before the transition, and was never on leave or retired before the transition. A control priest remains in an active U.S.-diocese post throughout the same eight-year window, was never on leave or retired over that window, and likewise has at least one documented church-knowledge episode in the corresponding pre-window years. Each treated priest is matched to up to ten controls sharing his transition year, his cumulative documented-victim count as of the year before the transition (grouped into a small number of bins), and his exact count of pre-window church-knowledge episodes; controls may be reused across different treated priests. Coefficients are estimated by a stacked Poisson (pseudo-maximum-likelihood) regression with priest-within-stack and stack-by-relative-year fixed effects, weighting each control by the inverse of the number of controls matched to its treated priest, with the

Figure 2: STACKED-DID ATT AROUND FIRST CHURCH KNOWLEDGE



Notes: Each panel is a stacked event-study estimate of the effect of a first church-knowledge episode on the named outcome, plotted by year relative to the anchor (relative year 0 is the anchor year itself); the underlying stack runs from seven years before the anchor through seven years after. *First church knowledge* is a priest's earliest documented, exactly-dated church-knowledge episode, excluding cases where an earlier but only approximately dated episode may in fact have come first, and excluding episodes whose only basis is the priest's inclusion on a public list (Section 4). Treated priests are first-CK anchors; controls are priests who had not yet had a documented church-knowledge episode as of the anchor year, matched to the anchored priest on anchor year, on his year-by-year presence or absence in the public directory over the pre-anchor years, and on entry cohort. "No Access" pools the on-leave, retired, and absent-from-directory outcomes; "Not in Directory" means the priest does not appear in that year's Official Catholic Directory at all; "Retired" and "On Leave" are read from the priest's listed status (leave includes explicit leave, unassigned, awaiting-assignment, and similar language). Coefficients are estimated with priest-within-stack and stack-by-relative-year fixed effects, with standard errors clustered at the priest-within-stack level.

Figure 3: STACKED-DID ATT AROUND FIRST CHURCH KNOWLEDGE: SENIOR TITLE



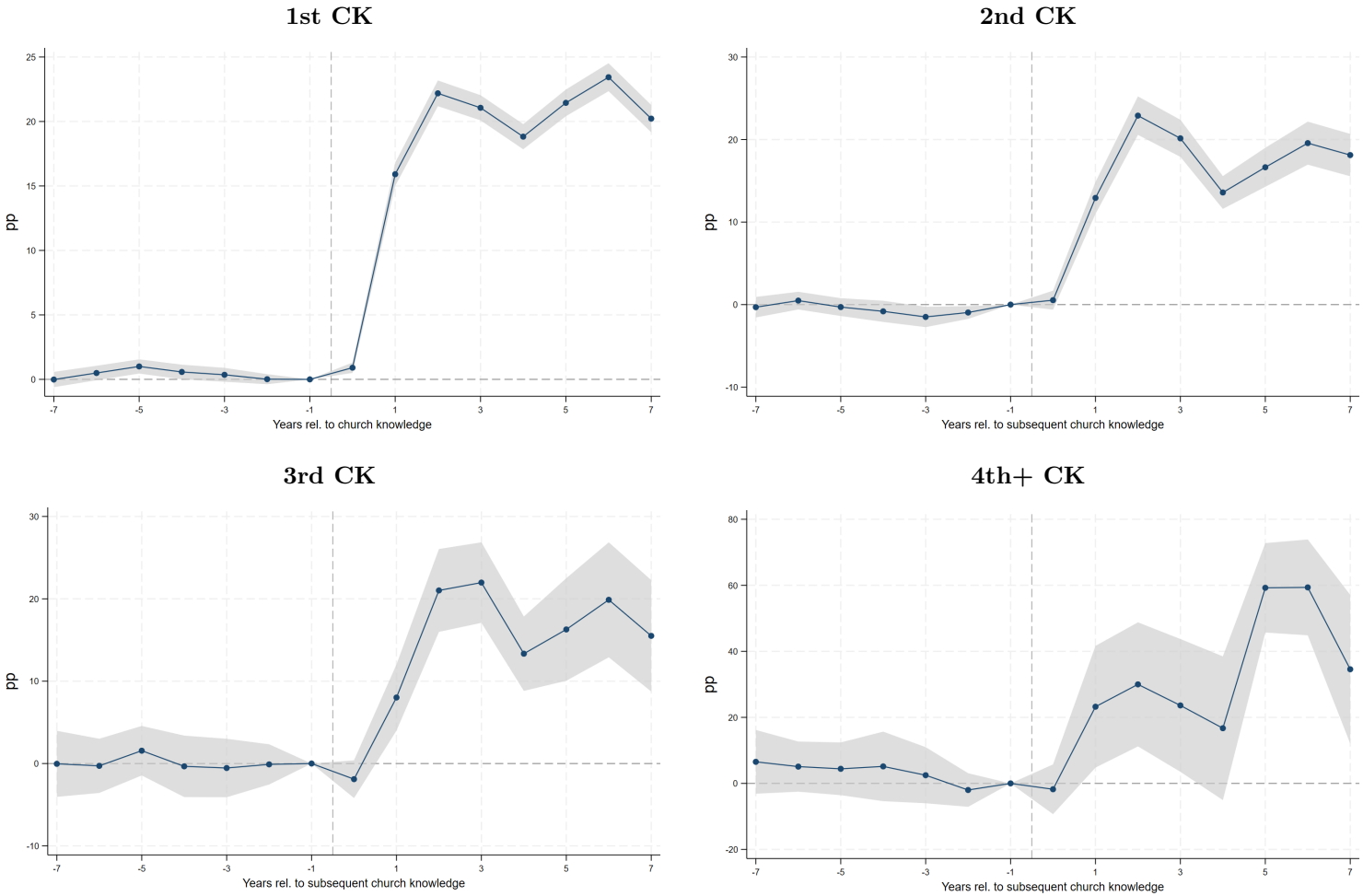
Notes: Same first-CK stacked event-study design as Figure 2, with the outcome instead being whether the priest holds a senior honorific title such as monsignor. See the notes to Figure 2 for the definitions of the treated and control groups and the estimation window.

Table 3: POOLED ATT: FIRST CHURCH KNOWLEDGE VS. SUBSEQUENT CHURCH KNOWLEDGE

	1st CK	Subsequent CKs
No access	19.93***	17.93***
	(0.35)	(0.69)
Mean of dep var	40.99	59.85
Not in directory	13.54***	9.48***
	(0.31)	(0.64)
Mean of dep var	33.77	49.79
Retired	1.12***	3.28***
	(0.20)	(0.54)
Mean of dep var	4.72	8.94
On leave	5.26***	5.16***
	(0.20)	(0.34)
Mean of dep var	2.50	1.12
Senior title	−0.98***	−0.40
	(0.14)	(0.28)
Mean of dep var	3.73	3.43

Notes: Each cell reports the pooled average treatment effect on the treated (ATT, in percentage points) from the same stacked difference-in-differences design described in Section 4, collapsing the event study to a single treated×post coefficient rather than tracing it out by relative year; the row below each estimate is the outcome's mean in that column's estimation sample. The *1st CK* column uses the first-CK design (treated priests are first-CK anchors; controls have not yet had a documented church-knowledge episode as of the anchor year). The *Subsequent CKs* column pools all subsequent-CK anchors regardless of ordinal (treated priests have at least one earlier CK of their own; controls likewise have at least one earlier CK but remain free of any new CK within seven years on either side of the anchor). Because the two columns use different control designs, the ATTs are not directly comparable to each other in levels. Standard errors, in parentheses, are clustered at the priest-within-stack level. The 1st CK column's estimation sample comprises 3,162 focal (treated) stacks; the Subsequent CKs column pools 1,925 focal stacks across all qualifying-subsequent-CK ordinals.

Figure 4: NO-ACCESS ATT BY CHURCH-KNOWLEDGE ORDINAL



Notes: Each panel restricts the subsequent-CK design of Table 3 to anchors of a given ordinal and plots the resulting event study for the no-access outcome (defined in the notes to Figure 2). The ordinal counts a priest's *qualifying* subsequent-CK anchors – those that already satisfy the subsequent-CK eligibility rule of Section 4 (no other CK of the priest's own in the previous seven years, and no prior spell on leave) – in ascending order; it need not equal the literal count of every documented church-knowledge episode a priest has, since intervening episodes that fail the eligibility rule are not counted. "1st CK" uses the first-CK design of Figure 2; "2nd," "3rd," and "4th+ CK" use the subsequent-CK design restricted to a priest's first, second, and third-or-later qualifying subsequent anchor, respectively, using the same matching procedure throughout – only the ordinal filter changes across panels.

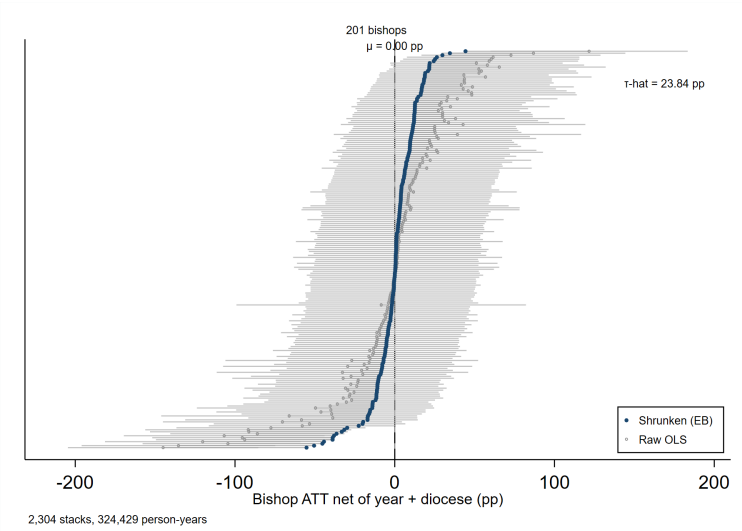
Table 4: DIOCESE TRANSFER AROUND CHURCH KNOWLEDGE, BY SCOPE AND UNKNOWN-LOCATION CONVENTION

	First CK		All subsequent	
	Unk. dropped	Unk. = diff.	Unk. dropped	Unk. = diff.
Diff.	0.42	0.43*	1.08**	1.07**
	(0.22)	(0.22)	(0.35)	(0.35)
Mean of dep var	4.05	4.11	2.24	2.45
Same	−20.36***	−20.35***	−19.02***	−19.00***
	(0.38)	(0.38)	(0.71)	(0.71)
Mean of dep var	54.94	54.91	37.78	37.70

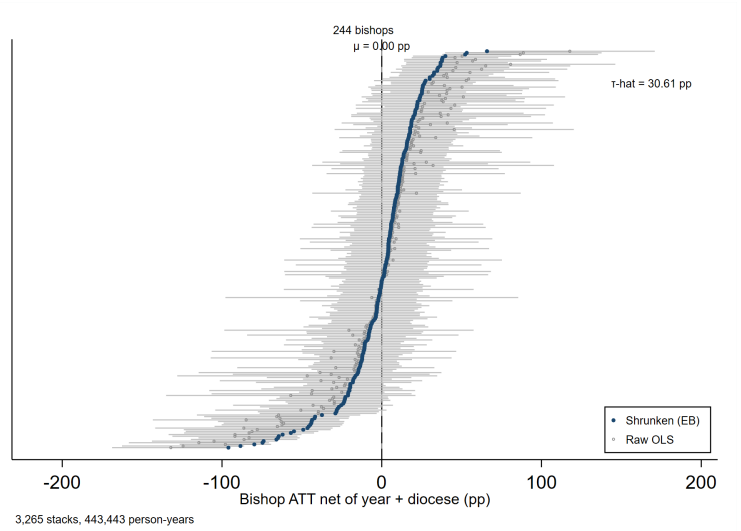
Notes: Each cell reports a pooled stacked-DiD ATT (in percentage points, from the same design as Table 3) for whether a priest's diocese in a given year differs from (*Diff.*) or matches (*Same*) a reference diocese established for him at or shortly before the anchor: his diocese in the anchor year itself if it is a valid U.S.-diocese assignment, or otherwise his most recent valid diocese in the seven years before the anchor. Years in which the priest has no ministry access (on leave, retired, or absent from the directory) are coded as 0 for both *Diff.* and *Same* under both conventions below, so they lower both shares rather than being excluded from the sample; the comparison is genuinely ambiguous, and handled differently across conventions, only for priest-years in which the priest remains listed with some diocese but that diocese cannot be resolved against the reference. *Unk. dropped* excludes exactly those unresolvable active priest-years from the estimation sample; *Unk. = diff.* instead retains them, coded as differing from the reference diocese. *First CK* and *All subsequent* use the same two matched designs, and treatment/control definitions, as the corresponding columns of Table 3. Standard errors, in parentheses, are clustered at the priest-within-stack level.

Figure 5: BISHOP-LEVEL EMPIRICAL-BAYES-SHRUNK EFFECTS, BASELINE

Panel A: First CK Only



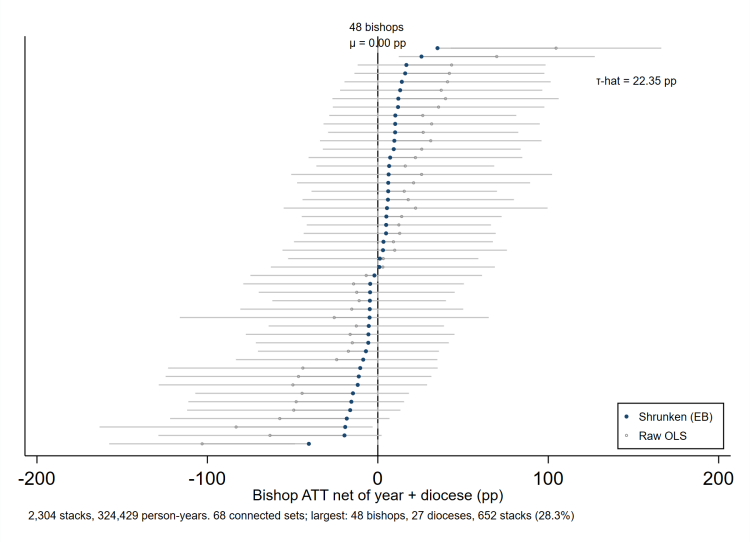
Panel B: First + Subsequent CK



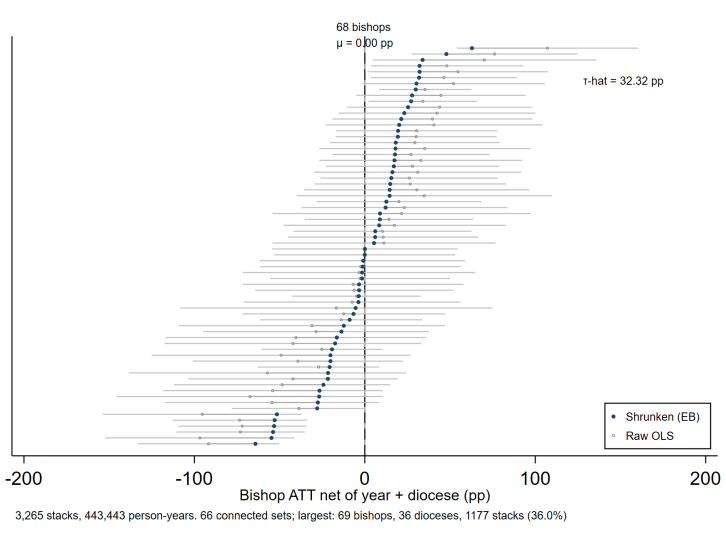
Notes: Panel A restricts to the first-CK design (Section 4); Panel B pools the first-CK and subsequent-CK designs. Each bar is an empirical-Bayes-shrunk estimate of one bishop's contribution to the no-access ATT (Figure 2), net of diocese and calendar-year effects, from a single stacked regression that estimates bishop, diocese, and calendar-year effects jointly and shrinks each raw estimate toward the sample average in proportion to its own sampling variance (Section 4); horizontal bars are shrinkage-based 95% confidence intervals. Only bishops contributing at least five stacks are shown. This is the baseline specification: no sample restriction beyond the five-stack minimum, and no controls beyond the treated×post interactions described in Section 4. The split-sample reliability check reported in Table 6 was computed for Panel A of this figure only; it was not separately computed for Panel B.

Figure 6: BISHOP-LEVEL EMPIRICAL-BAYES-SHRUNK EFFECTS, LARGEST CONNECTED SET ONLY

Panel A: First CK Only



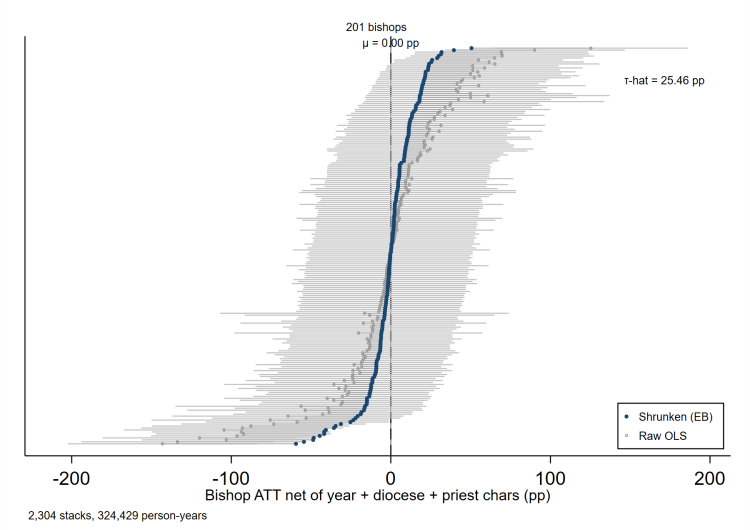
Panel B: First + Subsequent CK



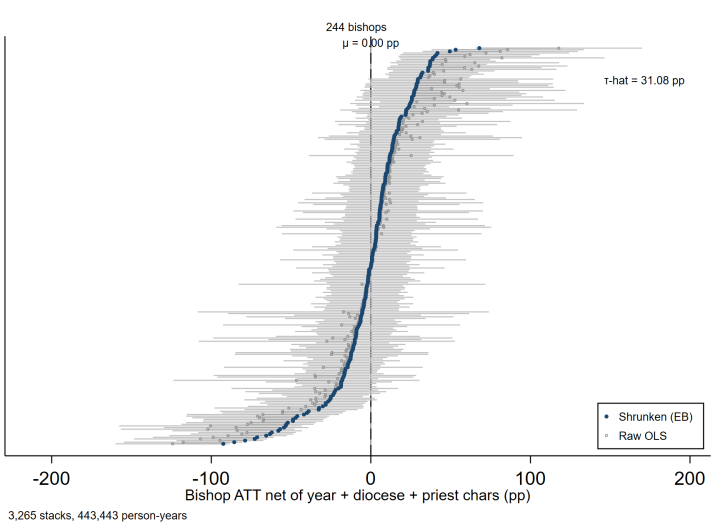
Notes: Same estimation as Figure 5, with bishops (and, where relevant, dioceses) further restricted to the largest mutually connected component of the bishop-diocese network implied by the five-stack-minimum sample: two bishops are connected if they are ever observed leading the same diocese, directly or through a chain of shared dioceses, and only the single largest such connected group is plotted. This restriction matters because a bishop's effect is only separately identified from the diocese he leads when the data link him to other bishops through diocesan turnover. Panel A of this figure (bmin = 5) is the specification used for the victims-avoided counterfactual of Section 5.4. No split-sample reliability check was separately computed for this specification; Table 6 reports reliability for the baseline specification of Figure 5 only.

Figure 7: BISHOP-LEVEL EMPIRICAL-BAYES-SHRUNK EFFECTS, BASELINE + PRIEST CHARACTERISTICS

Panel A: First CK Only



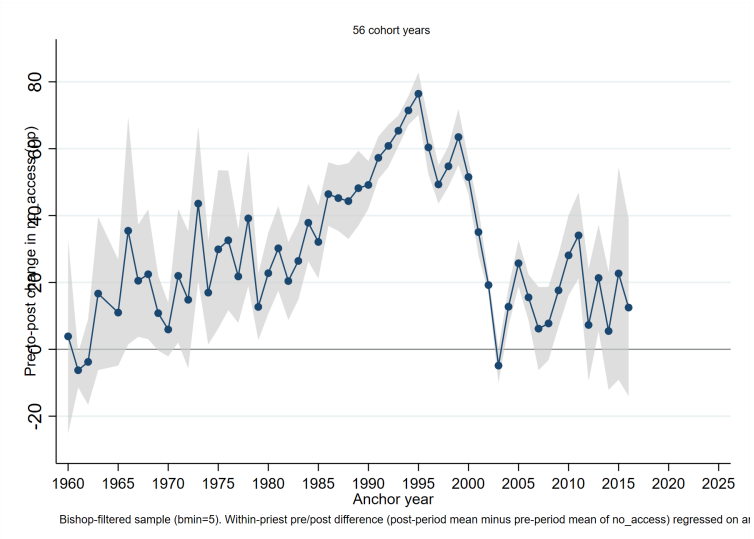
Panel B: First + Subsequent CK



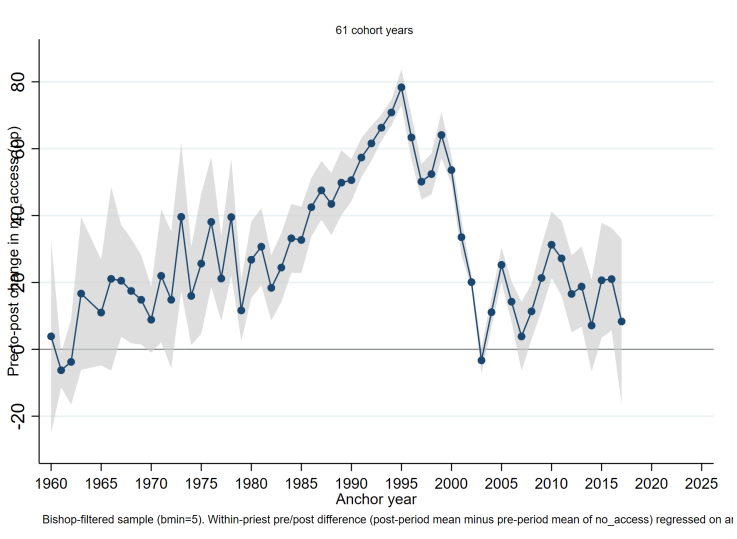
Notes: Same estimation as Figure 5 (baseline sample, no connected-set restriction), augmented with four additional priest-level controls, each interacted with treated×post: the channel through which the diocese first learned of the allegation (a private report to Church officials versus a public channel such as a lawsuit, criminal proceeding, or public list); whether the priest's cumulative documented-victim count as of the anchor was above the sample median; whether the priest already held a senior honorific title at the time of the anchor; and whether the priest was a diocesan priest rather than a member of a religious order. No split-sample reliability check was separately computed for this specification.

Figure 8: COHORT PRE/POST TREND, BISHOP-FILTERED SAMPLE

Panel A: First CK Only



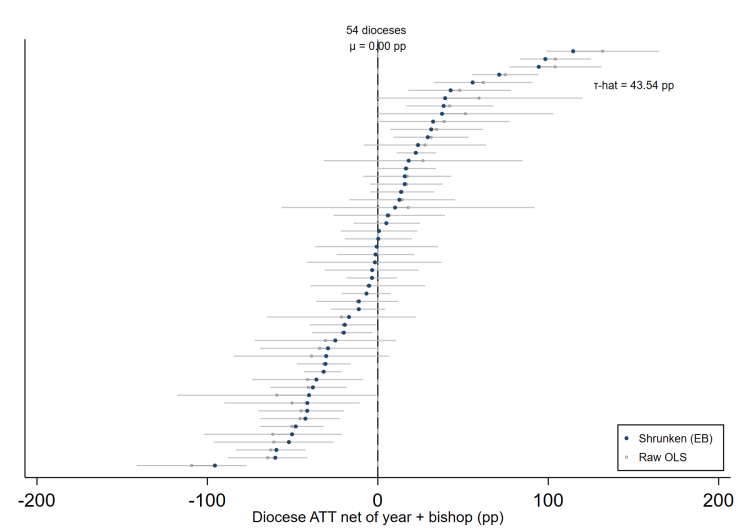
Panel B: First + Subsequent CK



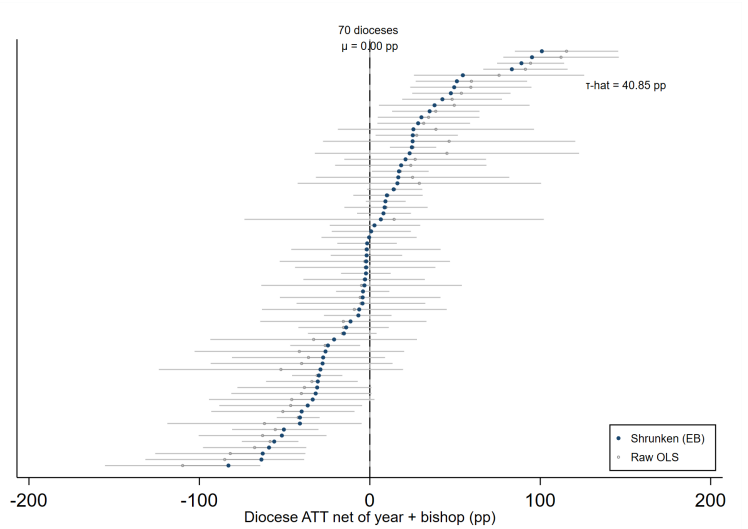
Notes: Unlike Figures 5–7 and 9, this figure does not estimate bishop- or diocese-specific effects and has no control group. For the same bishop-filtered sample of treated (anchored) priests used in the baseline specification (at least five stacks under their bishop), each priest-stack's own change in the no-access rate from before to after the anchor is computed, and these individual pre/post changes are regressed on a full set of anchor-year dummies (no constant). The resulting coefficients test whether the treated group's own average response shifted systematically with the calendar year of the anchor – a placebo check for a secular or Dallas-Charter-related trend that a bishop- and diocese-blind design could still pick up. Only anchor years contributing at least five stacks are shown. No split-sample reliability check applies to this design.

Figure 9: DIOCESE-LEVEL EMPIRICAL-BAYES-SHRUNK EFFECTS

Panel A: First CK Only



Panel B: First + Subsequent CK



Notes: The diocese-level analog of the baseline specification of Figure 5: the same jointly estimated regression, reporting diocese rather than bishop coefficients, with no connected-set restriction and no priest-characteristic controls. Only dioceses contributing at least five stacks are shown. The split-sample reliability check reported in Table 6 was computed for Panel A of this figure only; it was not separately computed for Panel B.

Table 5: DISPERSION OF BISHOP, DIOCESE, AND YEAR FIXED EFFECTS (FIRST-CK DESIGN)

	N units	Raw OLS SD (pp)	EB-shrunk SD (pp)
Bishop	201	37.79	23.84
Diocese	54	47.02	43.54
Year	64	12.93	0.00

Notes: Each row reports the dispersion, across units of the named type, of the empirical-Bayes-shrunk bishop, diocese, and year effects from the baseline specification of Figure 5, Panel A (first-CK design only), as described in Section 4. *Raw OLS SD* is the standard deviation of the unshrunk coefficients; *EB-shrunk SD* is the standard deviation after empirical-Bayes shrinkage, which pulls each unit's estimate toward the sample average in proportion to its own sampling variance and so removes part of the dispersion attributable to sampling noise alone. *N units* is the number of bishops, dioceses, or years contributing at least five stacks to the underlying regression, the minimum required for a unit to be included.

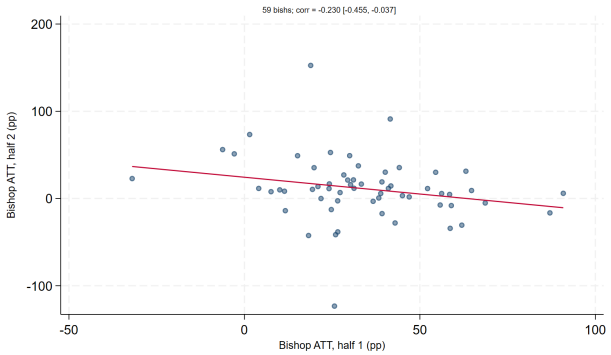
Table 6: SPLIT-SAMPLE RELIABILITY OF BISHOP AND DIOCESE EFFECTS (RANDOM HALF OF STACKS), WITH IMPLIED SD OF THE UNDERLYING FIXED EFFECT

Unit	N (both halves)	corr	95% CI	Implied SD (pp)	95% CI
Bishop	64	0.535	[0.345, 0.704]	22.74	[16.01, 28.48]
Diocese	19	0.161	[-0.324, 0.775]	12.93	[0.00, 26.13]

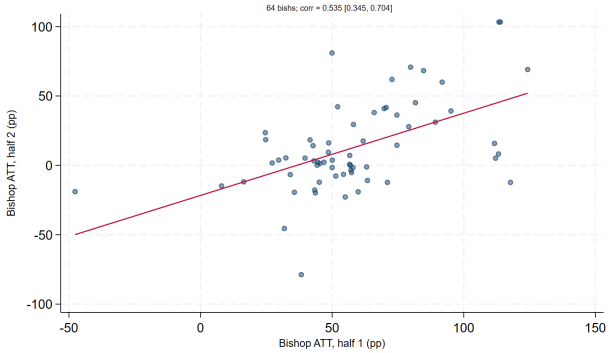
Notes: For each unit type, the stacks underlying the first-CK design of Figure 5, Panel A, are randomly split in half (by stack, so every priest-year within a stack goes to the same half), and the baseline bishop or diocese effect is re-estimated independently on each half, re-applying the five-stacks-per-unit minimum within each half. *N (both halves)* is the number of units meeting that minimum in both halves. *corr* is the correlation across halves of the two independent (unshrunk) estimates of the same underlying unit effect; a high correlation indicates that cross-sectional variation in the raw estimates reflects a genuine, persistent unit effect rather than sampling noise, since noise would be uncorrelated across independently drawn halves. *Implied SD* recovers the standard deviation of the underlying effect from the covariance of the two halves' estimates (Section 4); it does not rely on the shrinkage model behind the empirical-Bayes estimates of Table 5 and so serves as a model-free cross-check on that table's shrunk standard deviations. 95% confidence intervals, in brackets, are from a 1,000-draw pairs bootstrap. This random-half-of-stacks split is one of two split-sample checks we run; Figure 10 also reports a second, deterministic split by even/odd anchor year, which behaves differently (see that figure's notes) and is not the basis for the numbers reported here.

Figure 10: SPLIT-SAMPLE RELIABILITY OF BISHOP EFFECTS

Panel A: Even/Odd Anchor Year



Panel B: Random Half of Stacks



Notes: Both panels correlate independent estimates of the same bishop effect across two halves of the underlying stacks, from the baseline first-CK specification of Figure 5, Panel A. Panel B (*Random Half of Stacks*) uses the random split described in the notes to Table 6, and is the basis for that table's reported correlation and implied standard deviation. Panel A (*Even/Odd Anchor Year*) instead splits stacks deterministically by whether the anchor year is even or odd; we include it only as a cautionary comparison, since a deterministic split by calendar year can pick up genuine common time variation as if it were sampling noise, and it is not used to construct any reliability estimate reported elsewhere in the paper.

Table 7: COUNTERFACTUAL VICTIMS AVOIDED UNDER A BENCHMARK-BISHOP RESPONSE RATE

Benchmark bishop	Bishop ATT gap (pp)	Extra no-access years	Victims avoided
75th percentile	9.52	522	247.4 [155.8, 327.7]
90th percentile (primary)	14.08	704	333.6 [225.7, 527.6]
95th percentile	16.76	821	389.5 [268.6, 782.3]

Notes: Counterfactual: every bishop responds to first church knowledge at the rate of the P-th percentile bishop (empirical-Bayes-shrunk no-access ATT, net of diocese and anchor-year controls, largest connected set, bmin=5). Gap applied as a constant extra probability of no access over a fixed 7-year post-anchor window per focal priest, converted to avoided victims via the blended catholic6 Table 4 (v1, OLS, col. 2) per-year effect, weighted by the empirical leave/retired/absent mix among observed post-anchor no-access years. 95% CI in brackets from a 1,000-draw nonparametric bootstrap over bishops.