Road to Recovery: What Is Associated with Species Improvement, Recovery and Extinction under the Endangered Species Act?*

Eyal Frank
University of Chicago, NBER and CEPR

October 20, 2026

Abstract

I assemble the federal record of reported status, reclassification, recovery and extinction for 1,758 domestic entities ever listed under the Endangered Species Act with the Fish and Wildlife Service as the lead federal agency. I code the status the agency reported in 10 recovery reports to Congress, from 1990 through FY2009–10, and the operative rule behind every removal and reclassification to the end of 2025, and I link them to threats named at listing, critical habitat designations, and spending—each observed before the outcome. One key finding is that the reported status is persistent: most entities keep their label from one report to the next, and improving is the least persistent label. Threats, critical habitat status, and spending provide low predictive power about the next status. Removals, either for recovery or extinction, within 20 years of listing are rarely observed.

JEL codes: Q57, Q58, K32.

Keywords: Endangered Species Act, species recovery, extinction, critical habitat, conservation spending.

*Corresponding author email: eyalfrank@uchicago.edu. All remaining errors are my own.

1 Introduction

The Endangered Species Act of 1973 (ESA) is the main federal law for protecting imperiled plants and animals in the United States. Listing a species triggers prohibitions on harming it, a duty for federal agencies to consult before they act in ways that could affect it, and, in most cases, the designation of critical habitat (CH) and a recovery plan. The statute's stated goal is to bring each species to the point where these protections are no longer needed.

How often recovery happens is contested. Benefits from listing arrive slowly, if at all, as wildlife populations recover. Debates over reforming the ESA raise arguments about how often and how fast listed species recover, depending on whether they seek to weaken or strengthen the act, yet the evidence for this debate is mixed. Of the more than 1,000 species Male and Bean (2005) follow through the FWS reports for 1988–2002, 52 percent were repeatedly improving or at least not declining, while 35 percent were still declining 13 or more years after protection. The systematic evidence on which features of a species and of its protection go with improvement, recovery or extinction rests on FWS reports that end by the mid-2000s (Taylor et al., 2005; Male and Bean, 2005; Gibbs and Currie, 2012), and it rarely treats recovery and extinction together.

Measuring what the record shows is hard for four reasons. First, the status FWS reports is an agency judgment, not a population count. Report editions changed their definitions and instructions, some labels repeat from one edition to the next, and an "unknown" status may say more about monitoring effort than about the status of the species. Second, protection follows distress—FWS designates CH, writes plans and spends money on the species it judges most at risk, so a raw association between spending and decline describes the agency's response, not the effect of the policy. Third, the terminal outcomes are rare and compete with each other. A species removed as extinct can no longer be removed as recovered, and a comparison that censors extinctions overstates recovery. Fourth, the roster itself is a moving target. A list of the species protected today omits those already removed, and the current federal databases drop entities that were later split, merged, or removed from their publicly available databases.

In this paper, I document what the federal record shows about the correlates of reported status, reclassification, recovery and extinction under the ESA. I build a historical roster of every domestic entity (species, subspecies or distinct population segment) ever listed with FWS as sole or joint lead, 1,758 entities in all, including those since removed. For each entity I code the status FWS reported in every Recovery Report to Congress that carries a species-level status—10 reporting periods from 1990 through FY2009–10. The FWS changed the definitions behind its status labels several times, so I group the reports into periods with the same definitions, which I call reporting regimes, and compare species' statuses only within a regime. I also follow first removals and reclassifications between endangered and threatened, coded from the operative final rule, until the end of 2025. I link these outcomes to the threats named at listing, CH status, spending within consistent reporting regimes, consultations of federal agencies with the FWS, and taxonomic group—each dated before the outcome it describes. I make no causal claim and offer no forecasting tool, and "status" throughout means the status FWS reports, not a measured population growth rate.

I begin with the raw table of transitions in reported status between consecutive comparable reports, with unknown kept as its own state. I then ask whether three groups of predictors (threats named at listing, critical habitat and spending) help predict the next reported status beyond a benchmark of prior status, years since listing and taxon. I add each group to the benchmark on its own and measure how much it improves the predictions. Next, I turn my attention from the population proxies as reported by the FWS to whether the statutory protections are still in place for the species. I measure the share of listed entities removed from the list for recovery, and the share removed for extinction, within a given number of years from listing. Because an entity can leave the list only once, these two outcomes, and removals for other reasons, are treated as competing risks. Entities still listed at the end of 2025 count for the years they were observed. Similarly, I study whether species shift between the two listing statuses of Threatened and Endangered. In practice, the awarded statutory protections remain nearly unchanged but an uplisting or downlisting event reflects an assessment made by the FWS about the status of the species. As a check that these associations do not depend on the form of the model, I refit them with a random forest classifier and ask which predictors it relies on when predicting status for entities it was not trained on.

The reported status of a species is persistent over time. Within each reporting regime most entities reported as stable or declining keep that label in the next report, and the share that stays stable runs from 0.627 to 0.776 (Figure 1). Improving is the least persistent label, or close to it, in every regime. Once the prior status is known, threats, CH and spending add little to the held-out fit of next status. In fact, of the 41 comparisons I study, only 6 remain statistically distinguishable from zero after correcting for the number of tests (Holm's procedure).[1] The comparisons that I can reject a null of no association for are all in the wildlife population statuses, and in each one the predictor shifts probability between a declining and an unknown reported status, not toward improvement. Compared with entities that had no CH in force, entities whose CH had come into force within the previous five years are more likely to be reported as declining in the next report, and less likely to be reported as unknown. So are entities with more cumulative spending. The shift in the probability of declining is 0.216 for recent CH and 0.254 for spending, but neither association keeps its sign in every reporting regime. Whether entities are grouped by taxon, by listing era or by the threats named at listing, few in any group were removed for recovery or for extinction within 20 years of listing. Among the taxonomic groups, the cumulative incidence of recovery removal at 20 years is highest among vertebrates, at 0.020, and no contrast in removal or reclassification incidence passes Holm's procedure. The comparison of removal after the CH landmark is uninformative because all 14 exits in its window fall among entities whose CH state is missing.

The checks in Appendix C and in the Robustness Supplement leave these facts in place and locate the status associations. The associations of CH and spending with declining and unknown hold on the entities with FWS as sole lead and on the entities on the current list, but they come

[1]The Holm's procedure is a slightly less conservative version of the Bonferroni correction for multiple hypothesis testing.

from the later listings: none of them holds among the entities listed before 1990. The opposite sign of recent spending appears only when cumulative spending is in the same model. Whether the next status is known at all shows the same pattern as the difference between declining and unknown. The removal facts do not change on the sole-lead or current entities, or when the extinctions last sighted before listing are excluded, with and without the bound for removals that state no sighting. A future draft of this manuscript will add three additional analyses for what might predict the population statuses: threat comparisons on the listings that name their primary threat explicitly, the comparison of the threat flags with rule-coded threats, and the role of land use (public versus private).

This paper builds on a set of studies that relate the status FWS reports to features of the species and its protection (Hatch et al., 2002; Miller et al., 2002; Taylor et al., 2005; Male and Bean, 2005; Kerkvliet and Langpap, 2007; Gibbs and Currie, 2012; Langpap and Kerkvliet, 2012). Their windows end between 1995 and 2006. Departing from this work, I contribute in three ways. First, I use every recovery report with species-level status, and I follow the legal outcomes from 1973 to the end of 2025, more than a decade past the end of the status record. Second, I study status change and the terminal outcomes in one framework, treating recovered and extinct removals as competing outcomes and tracking reclassifications as dated legal events. To the best of my knowledge, none of the studies in this literature does this. Third, I add a species-level history of consultations between the FWS and other federal agencies, which I found in no earlier study of status or removal. However, because the consultation records start in 2008, they enter only as exploratory groups in the removal analyses, and I do not include them when estimating the status model. Even though threat coding from listing documents is not new (Wilcove et al., 1998; Leu et al., 2019); what I add is the link from the threats named at listing to later status and removal.

Future versions of this paper will add five additional components. First, the status series ends with the FY2009–10 report. A later version extends it past 2010 by coding the population estimates and trend statements in FWS 5-year reviews into the same four categories, validated against the recovery reports in the years both exist; this adds more than a decade of status observations. Second, the FWS posts recovery plans for species. I will extract the projected time to recovery and the projected cost from each recovery plan and test out of sample how well species characteristics predict those costs. I will also use recovery-plan existence and its age as predictors. Third, the analysis will measure the public share of each entity's range and CH, first from the current layers and then, once earlier and removed polygons are rebuilt from the coordinates published in Federal Register rules, as a dated predictor of status and removal. Fourth, the threat comparisons here use the threat flags of Leu et al. (2019) matched to my roster. In the next version, the primary threats will be coded from each final listing rule. Fifth, the reclassification histories of entities will use every Federal Register issue published after 1980 and before the 1996 CFR edition.

The remainder of the paper is organized as follows: Section 2 describes listing, recovery planning, reclassification and removal. Section 3 reviews the related literature. Section 4 describes the data and Section 5 the empirical approach. Section 6 reports the facts. Section 7 concludes.

2 Background

This section describes how a species enters the list, what listing sets in motion, how FWS reports on listed species and how a species leaves the list or changes status on it. Each step leaves a dated legal or administrative record, and those records are the data I use.

2.1 Listing

The ESA protects two classes of species. A species is endangered if it is in danger of extinction, and threatened if it is likely to become endangered within the foreseeable future, in each case across all or part of its range.[2] The statute's "species" includes any subspecies of fish, wildlife or plants and, for vertebrates, any distinct population segment (DPS), so the unit that is listed is often smaller than a biological species.[3] I call each listed unit an entity. FWS leads for most terrestrial and freshwater entities and NMFS for most marine and anadromous ones, and some entities are shared between the two agencies. I keep every entity for which FWS is sole or joint lead.

FWS lists an entity by final rule in the Federal Register, usually after a proposed rule and public comment and often after a petition. The decision must rest solely on the best available scientific and commercial data and on five statutory factors. These are the destruction, modification or curtailment of the entity's habitat or range, overuse for commercial, recreational, scientific or educational purposes, disease or predation, the inadequacy of existing regulation, and other natural or manmade factors.[4] In its discussion of these factors, the final rule sets out the agency's account of why the entity was in trouble at listing. I take the threats at listing from that discussion. The rules have grown more detailed over time: Leu et al. (2019) find that the number of threats named per listing rose from 1.5 in 1975 to 3.7 in 2017, and Wilcove et al. (1998) note that notices for species listed before 1980 named fewer threats. I therefore read a record of threats at listing also as a record of how thoroughly each era wrote its rules, and I report threat coverage by listing era.

Two timing details matter for the clock I use. First, the 1973 Act carried over the lists made under the Endangered Species Preservation Act of 1966 and the Endangered Species Conservation Act of 1969, so some entities were protected before the 1973 Act took effect on 28 December 1973. Second, FWS can list an entity on an emergency basis for up to 240 days, without the usual notice and comment.[5] An emergency rule is temporary and is not a final listing, so I start an entity's clock only when a final rule lists it (Section 4).

2.2 What Listing Sets in Motion

I use three obligations that listing triggers as predictors. The first is critical habitat. FWS is to designate, at listing and to the extent it judges prudent and determinable, the areas essential to the

[2] ESA §3(6) and §3(20), 16 U.S.C. §1532.

[3] ESA §3(16). FWS and the National Marine Fisheries Service (NMFS) set out how they recognize a DPS in a joint policy of 1996 (61 FR 4722).

[4] ESA §4(a)(1) and §4(b)(1)(A), 16 U.S.C. §1533.

[5] ESA §4(b)(7).

entity's conservation, and it may exclude areas after weighing economic and other impacts.[6] CH is designated, revised and occasionally removed by its final rules. The second is consultation. Every federal agency must consult FWS to ensure that the actions it funds, permits or carries out are not likely to jeopardize a listed entity or destroy or adversely modify its habitat, in particular, its CH.[7] The third is spending. Since the 1988 amendments, FWS has reported to Congress each fiscal year the money that federal agencies and states spent on each listed entity.[8] Listing also brings the prohibition on taking endangered animals, which FWS may extend to threatened animals by rule under section 4(d). I do not use the prohibition as a predictor, because it applies at listing and varies little in a way I can date.[9]

2.3 Recovery Plans, Reports and Reviews

The statute's goal is recovery: to bring an entity to the point at which its protections are no longer necessary.[10] FWS is to write a recovery plan for each listed entity unless it finds that a plan would not promote conservation. Since 1988, each plan must describe site-specific management actions, give objective and measurable criteria for removal from the list, and estimate the time and cost of the measures needed to meet them.[11] Plans are revised, and one plan can cover several entities. I will use those projected times and costs in the recovery-plan component that follows in a later version of this paper.

The 1988 amendments also asked FWS to report to Congress every two years on its recovery plans and on the status of the species they cover.[12] The Recovery Reports to Congress that answered this request printed, for each listed entity they covered, a one-word assessment of its status since the previous report: improving, stable, declining or unknown. Labels outside these four, where an edition uses them, I keep as their own codes. The editions from 1990 through FY2009–10 carry this species-level status, except for a 1996 summary with no species table. The FY2011–12 report prints a species table without a status column, and later editions drop the table. I read the assessments as FWS judgments, not counts, and the definitions and instructions behind them changed across editions. Male and Bean (2005) report that FWS did not assess the status of 40.3 percent of listed species in each report they studied. I treat each edition's labels as an administrative record and fix, before any tabulation, which editions share definitions (Section 5).

FWS must also review every listed entity at least once every five years and decide whether it should be removed, reclassified or left as it is.[13] The 5-year reviews often state population estimates and trends, and FWS began writing them in volume in the mid-2000s. In a later version, I will

[6] ESA §3(5), §4(a)(3) and §4(b)(2). The 1978 amendments first required CH to be specified at listing; the 1982 amendments gave the requirement its current "maximum extent prudent and determinable" form.

[7] ESA §7(a)(2), 16 U.S.C. §1536. Formal consultation ends with a biological opinion, which may find jeopardy or adverse modification and must then suggest reasonable and prudent alternatives (ESA §7(b)(3)(A)).

[8] ESA §18, 16 U.S.C. §1544.

[9] ESA §9 and §4(d).

[10] ESA §3(3), which defines "conservation" as the use of all methods and procedures needed to reach that point.

[11] ESA §4(f)(1).

[12] ESA §4(f)(3).

[13] ESA §4(c)(2).

code these statements into the four status categories to extend the status series past 2010.

2.4 Reclassification and Removal

An entity leaves the list or changes status on it only by final rule, judged on the same five factors as listing. A downlisting moves an entity from endangered to threatened and an uplisting from threatened to endangered. A removal, which FWS calls a delisting, ends protection. I date each of these events by the effective date of its final rule. The regulations name the grounds for removal. The entity is extinct, it has recovered, or new information shows that it does not meet the definition of an endangered or threatened species, or of a species, for example after a taxonomic revision.[14] Courts can also remove listings. Because the wording of the grounds changed twice, I classify each removal from the operative final rule as recovered, extinct or other, and treat the other removals as exits rather than successes.

Two features of extinction removals shape how I use them—and why I keep the last confirmed sighting as its own date. First, a removal for extinction dates the agency's recognition of the loss, not the loss itself, and it can follow the last confirmed sighting by decades. Second, FWS often removes several extinct entities in a single rule, so the timing of extinction removals partly reflects when the agency processed them. Extinction is not a marginal reason for removal: Abbitt and Scott (2001) note that 7 of the 27 species delisted by the late 1990s had gone extinct. The legal outcomes I study are therefore few, dated by the agency and competing with each other, and the empirical approach is built around that (Section 5).

3 Related Literature

This paper relates to four strands of work on the ESA: studies of the status FWS reports, studies of removal from the list, studies of the threats behind listings, and studies of spending and recovery plans. In each case I state what the closest papers measure and where this paper departs from them.

3.1 Reported Status

I begin with seven studies that relate the status FWS reports to features of the species and of its protection. Hatch et al. (2002) and Miller et al. (2002) use the 1996 report in a single cross-section. Taylor et al. (2005) use reports for 1990–2004 and compare an early period with a late one. Male and Bean (2005) use reports for 1988–2002 with expenditures for 1989–2002. Kerkvliet and Langpap (2007) use reports for 1990–2002, and Langpap and Kerkvliet (2012) reports for 1990–2004, both for vertebrates (author draft, Sec. II; Table 1). Gibbs and Currie (2012) use reports for

[14] 50 CFR 424.11(e), as revised in 2024 (89 FR 24300). Before 2019 the grounds were extinction, recovery and error in the original data for classification (50 CFR 424.11(d), 2018 edition). The 2019 revision (84 FR 45020) dropped recovery as a separate ground, and the 2024 revision restored it. Delisting and reclassification rest on the same five factors as listing (ESA §4(c)(2); 50 CFR 424.11(c), (e)).

1988–2006 and find that recovery plans, CH and funding together explain less than 13 percent of the variation in status (abstract). The windows therefore end between 1995 and 2006. Ferraro et al. (2007) are often read with these studies, but their outcome is different: the change in NatureServe endangerment scores over 1993–2004, for listed and unlisted vertebrates (author draft).

The findings in this literature on CH disagree. Taylor et al. (2005) find CH associated with a higher chance of improving in their late period and a lower chance of declining in the early one (Table 1). Male and Bean (2005) find a marginal association ($P = 0.063$, Table 1), and Kerkvliet and Langpap (2007), who measure CH as days since designation, find no statistically significant association (author draft, Table III). The studies differ in window, sample, the coding of CH and the controls, making it difficult to draw strong conclusions or compare the findings to one another.

The studies also differ in how they treat a status of unknown. Taylor et al. (2005) drop species with no score in a period, and Kerkvliet and Langpap (2007) and Langpap and Kerkvliet (2012) omit uncertain status because it cannot be ordered (author draft, Sec. II; fn. 7). Hatch et al. (2002) and Miller et al. (2002), in contrast, keep unknown as a category, and Male and Bean (2005) report that FWS did not assess 40.3 percent of listed species in each report. Keeping unknown is therefore not new. What I add is to carry unknown as a state in the transition table, so that a reader can see where species with unknown status came from and where they go.

Relative to these studies, I use every report edition with species-level status, a window that ends with FY2009–10, and I compare editions only within reporting regimes whose definitions I fix from the reports' front matter before any tabulation. I make the transition table, rather than a summary score over many reports, the first fact, because a score that sums labels across reports cannot separate a species that moved from one that kept a repeated label.

3.2 Removal from the List

None of the studies above, as I read them, treats recovered and extinct removals as competing outcomes. Taylor et al. (2005) fold extinct and delisted species into the sample without a separate outcome. Kerkvliet and Langpap (2007) and Langpap and Kerkvliet (2012) place extinct at the bottom of an ordered scale and pool recovered with improving, because few species had recovered (author draft, Sec. II). Male and Bean (2005) and Miller et al. (2002) exclude extinct and delisted species. Two studies compare recovered species with others rather than follow status over time. Abbitt and Scott (2001) compare recovered or recovering units with declining ones and drop extinct removals by design and Haines et al. (2021) compare recovered species with species still listed. Greenwald et al. (2019) study extinctions under the ESA.

Departing from this work, I follow every entity from listing to its first removal or to the end of 2025 and estimate the cumulative incidence of recovery and of extinction as competing risks. Removals for other reasons, such as error or taxonomic revision, compete with both and are not counted as successes. I also track downlistings and uplistings between endangered and threatened as dated legal events, which none of the studies I inspected models as an outcome.

3.3 Threats at Listing

Wilcove et al. (1998) code the threats to 1,880 imperiled species from listing notices and several other sources, as of 1996, and count all known threats without separating major from minor ones. Leu et al. (2019) code six threat categories as present or absent from final listing rules for species listed 1975–2017. They have no primary threat, exclude the statutory factor on inadequate regulation and potential threats, and report raw agreement between coders of 88 percent, without a measure that corrects for chance agreement (such as Cohen's kappa). They find that the number of threats per listing rose from 1.5 in 1975 to 3.7 in 2017. Neither study relates listing-time threats to later status or removal.

I use the Leu et al. (2019) flags, matched to my roster, as the threat block of this draft. The comparisons are therefore read by the presence of each broad threat, not by a primary threat. The link from threats named at listing to later status and removal is what this paper adds. The rule-coded primary threats, with an explicit or inferred flag and double-coded reliability, follow in the next version.

3.4 Spending, Land, Consultations and Recovery Plans

Several of the status studies find spending associated with status. Male and Bean (2005) find FWS and NOAA funding associated with status ($F = 11.27$, Table 1). Kerkvliet and Langpap (2007) find that spending lowers the probability of declining and does little for improvement (author draft, Sec. IV), and Miller et al. (2002) find that the share of the plan's requested budget actually spent is higher for improving species. Because FWS spends more on species in distress, none of these associations isolates an effect of spending, and I read them, like my own estimates, as descriptions of where money went. I did not find a study that sets its spending comparisons within reporting regimes defined from the reports' own instructions. Most of their windows end before FY2012, when the expenditure reports first let agencies report costs they cannot assign to one species in a multi-species category. I compare spending only within a regime.

Moore et al. (2022) model realized annual species-level spending from the expenditure reports for 1993–2016 as a function of species characteristics and project it to listings induced by climate change (Table 3). They do not use recovery-plan costs or timelines, and their fit statistics are in sample. Miller et al. (2002) use the budget in each of 243 recovery plans as the denominator of a spending share and do not predict plan costs. Restani and Marzluff (2001) find that priority rank explains less than 5 percent of the variation in spending across 85 bird species. Gerber (2016) compares spending with the recovery budgets in plans. In the recovery-plan component of the next version, I test out of sample how well species characteristics predict projected plan costs, which departs from all of these.

Land ownership is not new as a predictor of status. Hatch et al. (2002) place each species in one of four bins of federal jurisdiction over its primary habitat, taken from plan text in 1998. They find that species found only on federal land are more likely to be improving, although the term adds nothing once plan variables enter (Table 3). Ferraro et al. (2007) tried federal and state land shares

by state and dropped them (author draft), and Clancy et al. (2020) measure the public share of species' ranges. What I add is a public share of each entity's range and CH, built for every listed entity and used with status and removal together. The land shares are deferred from this draft and follow in the next version.

Finally, Malcom and Li (2015) describe the FWS consultation records for 2008–2015. I found no study that builds a species-level history of section 7 consultations and relates it to later status or removal. Because the records start in FY2008 and the status record ends with FY2009–10, consultations enter this draft only as exploratory groups in the removal analyses.

4 Data

I build every series from public federal documents: the Federal Register, the Code of Federal Regulations, FWS reports to Congress, FWS's Environmental Conservation Online System (ECOS), and agency records released under the Freedom of Information Act. None of the sources is restricted. Appendix A describes how each series was built; here I describe what each measures, its coverage and its known weaknesses.

4.1 The Roster of Listed Entities

The unit of analysis is the listed entity: a species, subspecies or DPS as it was listed. The roster contains every domestic entity ever listed with FWS as sole or joint lead, including those since removed, 1,758 entities in all, of which 1,739 have FWS as sole lead. I start from the ECOS exports of listed and formerly listed entities. ECOS, however, drops entities that were later replaced or removed, and it files their old listings under the entities that replaced them; the 1978 listing of the green sea turtle and the 2003 gray wolf DPSs are examples. I therefore reconstruct the roster from the annual editions of the lists in 50 CFR 17.11 and 17.12 from 1996 onward, matched against the ECOS records, and restore each missing entity I identify from its own listing and removal rules (12 entities). The match leaves 264 domestic CFR entries unresolved, and future versions of the paper will review those one by one. For 1973–1995, the only check is the names printed in the 1990, FY1992 and 1994 recovery reports and the rules they cite. An entity that was listed and then replaced before 1996 would be missed if none of those reports names it. The roster may therefore be incomplete for entities replaced before 1996.

Three rules fix what counts as an entity. First, an entity created by a split, a merger or a new DPS is its own roster entry, linked to its predecessors, and it enters the analysis only at its own listing. Second, experimental populations designated under section 10(j) are not separate entities. Each is recorded as an action linked to its listed parent, so a taxon is not counted twice. Third, I keep every entity even when some of its information cannot be checked against the rule that set its status. Such values are taken from ECOS and flagged, and every headline result is repeated on the entities whose information was all checked. Dropping the unchecked entities would remove mostly the entities that left the list, which is the type bias that the historical roster is designed to avoid.

Clock and cutoff.—Each entity's clock starts at the effective date of the first final rule that listed it, read from the rule's own statement of dates. Entities first protected under the 1966 or 1969 Acts start on 28 December 1973, with the earlier date kept. An emergency listing is temporary and does not start the clock; an entity listed only by emergency rule has no start date and enters no analysis (1 entity). The cutoff is 31 December 2025. Removals and reclassifications effective on or before it count, entities still listed then are censored, and the 3 entities first listed after it stay on the roster but enter no analysis of the legal outcomes. I also establish the first year, October 1996, from which every change between consecutive CFR editions matches a recorded rule and every recorded rule matches a change. From that year the event record is complete. Entities listed earlier enter the risk set for removal in that year, with time still measured from their own listing.

4.2 Reported Status

The status outcome comes from the Recovery Reports to Congress. I obtained 11 reports covering 12 reporting periods, and 10 of those periods carry a species-level status column: 1990, FY1992, 1994, FY1997–98, FY1999–2000 and the biennial reports from FY2001–02 through FY2009–10. The 1996 report is a summary with no species table, and the FY2011–12 report prints a species table without a status column. The status series therefore ends with FY2009–10. Each record is the label printed for an entity in an edition, 10,336 records in all, and each edition's period ends on the date the table states.

I parse the later editions from their text and check them with a second reader and a random sample of rows. The labels map to improving, stable, declining and unknown, with unknown always its own category. A few editions print other labels, such as extinct or presumed extinct, in captivity, or presumed extirpated in the United States, and these keep their own codes. An entity missing from an edition is recorded as absent, not unknown. Printed names are matched to roster entities through one shared name table and one deterministic matcher, and unmatched or ambiguous names are reported by edition rather than dropped.

The labels have three weaknesses, and the analysis is built around each. First, definitions and instructions changed across editions. I sorted the editions into reporting regimes from their front matter alone, before any tabulation, and I pool transitions only within a regime (6 regimes). Second, the tables print a one-word label with no supporting text and never mark a label as carried forward from the previous edition. I cannot separate a repeated assessment from a repeated label, so I record how many consecutive editions have repeated each label and describe it as label repetition. Third, coverage changes by edition. The 1990 table covers only species with an approved recovery plan, and the only copy of the 1994 report I could find, the official FWS file and its single archived capture, lacks printed pages 50 and 51. The 74 entities whose 1994 rows fall on those pages are coded missing for that edition and left out of the edition pairs that include it. Because the table is ordered by taxon, the gap is concentrated in crustaceans, insects, arachnids and plants rather than spread at random.

4.3 Removals and Reclassifications

The legal outcomes come from the operative final rules, retrieved from the official Federal Register issues on GovInfo. For each removal, I record the effective date and the reason (recovered, extinct or other), coded from the rule's own text with the supporting passage, never from the reason field in ECOS. Of the 136 first removals by the cutoff, 67 are for recovery, 31 for extinction and 38 for other reasons. A removal whose rule could not be retrieved or coded keeps an unknown reason (2 removals) and is never imputed, censored or moved to "other". For extinction removals I also record the last confirmed sighting year the rule reports.

Reclassifications come from the final downlisting and uplisting rules that ECOS links to each entity, with effective dates read from the rules. Each carries a scope: a change for the whole entity, or for part of its range only. The primary analysis counts whole-entity changes (44 first downlistings and 5 first uplistings), and partial changes enter a sensitivity analysis. An entity's reclassification history counts as certified when the annual CFR editions confirm its status from listing, or, for entities listed before October 1996, when the rules found by a fixed search agree with the CFR at that year.

4.4 Threats at Listing

In this draft, the threat block uses the flags of Leu et al. (2019), who code six threat categories as present or absent from the final listing rules of species listed from 1975 to 2017. I match their records to my roster one to one on name, listing year and geography (1,438 entities), and I drop their column on current listing status, which is an outcome. I collapse the six categories into three broad groups: habitat loss and modification (habitat modification and pollution), biological threats (interactions with other species), and other threats (overuse and environmental and demographic stochasticity).

The flags have four limits. They name no primary threat. They exclude the statutory factor on inadequate regulation and any threat a rule describes as only potential. They do not cover entities listed before 1975 or after 2017, and entities they do not cover, or that I cannot match one to one, are excluded from the threat comparisons and counted by era and taxon rather than treated as having no threat. Finally, later rules name more threats (Leu et al., 2019), so I report threat coverage by listing era. Primary threats coded from each listing rule, with an explicit or inferred flag and a double-coded reliability sample, will replace the flags in the next version.

4.5 Critical Habitat

The CH predictor is whether CH was in force for an entity on a given date and for how long. The current ECOS CH layer cannot date it: it omits CH that was removed or revised and the CH of removed entities, and its date field holds no usable dates. I therefore read the effective date of every CH designation, revision and removal from the final rules in the Federal Register. CH is in force at a date if the latest action effective by then was a designation or a partial revision of CH in

force, and not in force after a removal of all of it. By the cutoff, 521 entities have at least one CH final rule. An entity whose search finds no CH action and no contrary evidence is coded as never having CH. An entity with evidence of CH, such as a CH entry in the CFR or a current polygon, but no dated rule is coded missing, never "none". CH that was designated and fully removed before 1996, or within 2019–2024, would leave no trace in either check; a sensitivity analysis treats the never-CH entities as missing.

4.6 Spending

Spending comes from the annual reports of federal and state expenditures on listed species that the ESA has required since 1988. The measure is the printed species-specific total, federal plus state, in Table 1 of each report. Land acquisition, reported separately, is never added to it in the main specification, and a robustness check adds it. I convert amounts to 2025 dollars with the BEA GDP implicit price deflator averaged over each fiscal year.

The species tables are usable from FY1998 to FY2022. The FY1989–95 reports give only summary totals and lists of the top species, which I never allocate to individual species. The FY1996 and FY1997 reports survive only as scans whose tables do not reconcile to their printed totals, so those years are missing for every entity. Two entity-level limits remain. Salmon and steelhead population units are combined in FY1999 and FY2000, and before FY2003 a species with several listed populations is reported as one row. I never split these rows among entities.

The reports also changed what they ask agencies to report. I read each report's front matter for changes in what counts as species-specific spending, and I find one break. From FY2012 agencies may report costs they cannot assign to a single species in a multi-species category—before, they had to assign such costs to species or leave them out. Other changes in the reports, such as the FY2001–07 category of expenses not identified with any species, sit outside the species rows. The generic caution that every report since FY1998 carries about comparing years dates no change. The two regimes are FY1998–2011 and FY2012–2022. The three-year moving average and the cumulative total are computed within one regime, and a total that would cross FY2012 is undefined rather than summed.

The timing rules of the plan fix which spending the status models can use. A status transition uses spending dated by the end of the earlier edition's period. The transition table and the models built on it use only pairs of editions within one status reporting regime, and pairs that straddle a regime break appear in a separate, labelled table. The confirmatory status models therefore use spending from 1998 through 2004, the year of the latest reference date in a within-regime pair (30 September 2004), with the moving average first defined at the FY1999–2000 edition. No reference date falls in the FY2012–2022 spending regime, so spending from those years enters no confirmatory model; it appears only in the descriptives and the exploratory analyses.

4.7 Consultations, Taxon and Land

Section 7 consultations come from the 2021 FOIA release of FWS's formal and informal consultations, supplemented by the Defenders of Wildlife records for activities the release lacks (169,806 consultations). The release covers consultations concluded from FY2008, so earlier years are missing, not zero. A consultation names species, not listed entities, so I match it to the species and attach it to that species' entities (0.974 of species names match). About half of the informal records give no conclusion type. Because the status record ends with FY2009–10, the only edition pair with a reference date in FY2008 or later runs from FY2007–08 to FY2009–10, and that pair crosses a status reporting regime. The status model with consultations is therefore not estimable, and I report it as such. Consultation counts for that pair appear only descriptively, in the separate table of cross-regime pairs. Consultations still enter the removal analyses as exploratory groups at landmarks in FY2008 or later.

Taxonomic group is a fixed attribute of each entity, from FWS's species taxonomy records, grouped into six classes for the status models (mammals, birds, reptiles and amphibians, fishes, invertebrates and plants) and three for incidence (vertebrates, invertebrates and plants).

Land ownership is deferred from this draft. The next version of the manuscript will include the public land shares of each entity's range and CH, from the current ECOS polygons and the Protected Areas Database of the United States (PAD-US). Recovery plans are not a predictor in this draft. The only dated source on plans covers entities listed today, so using it would date plans only for survivors.

5 Empirical Approach

The paper documents facts about the federal record; it does not estimate the effect of any protection. An experiment that could identify such effects would assign CH, spending or the timing of listing at random across species of similar condition. Nothing in the ESA's history approximates that assignment: FWS designates CH, writes plans and spends where it judges the need greatest, so every protection is chosen in part on the expected outcome. I therefore make no causal claims. What makes the facts below credible is timing and measurement. Every predictor is dated before the outcome it describes, every outcome is measured the same way within a stated regime.

5.1 Transitions in Reported Status

The first fact is the raw transition table. For entity i with a status record in edition r and in the next edition r', I estimate

$$P_{kl}(r, r') = \Pr\left(S_{ir'} = l \mid S_{ir} = k\right), \qquad k, l \in \{\mathrm{I}, \mathrm{S}, \mathrm{D}, \mathrm{U}\}, \tag{1}$$

where S_{ir} is the status reported for entity i in edition r (improving, stable, declining or unknown). The table pools edition pairs only within a reporting regime, a set of consecutive editions whose

definitions, categories and instructions do not differ in substance. Pairs that straddle a regime break are shown in a separate table and are never pooled. Unknown is a state like the others—the table shows where entities with unknown status come from and where they go. A label outside the four categories enters as its own row and column. I report counts and row shares, with 95 percent intervals from 2,000 bootstrap draws that resample entities. Next to the transitions, I report the share of records in each status by edition, which counts published records rather than movements and so can repeat a label from one edition to the next.

5.2 Associations of Next Status with Lagged Predictors

I then ask which features of an entity and its protection go with its next reported status, beyond what its prior status, age on the list and taxon already say. I estimate multinomial logits of the form

$$\Pr\left(S_{ir'} = l \mid \cdot\right) = \frac{\exp\left(\alpha_{l,S_{ir}} + \gamma_{l,A_i(\tau_r)} + \delta_{l,g(i)} + \eta_{l,rr'} + X_i(\tau_r)'\beta_l\right)}{\sum_m \exp\left(\alpha_{m,S_{ir}} + \gamma_{m,A_i(\tau_r)} + \delta_{m,g(i)} + \eta_{m,rr'} + X_i(\tau_r)'\beta_m\right)}, \tag{2}$$

with stable as the base outcome. Here $\alpha_{l,S_{ir}}$ is a set of indicators for the prior status, and $\gamma_{l,A_i(\tau_r)}$ are bands of years since listing at τ_r, the end of the period edition r covers (0–4, 5–9, 10–19 and 20 or more years). The terms $\delta_{l,g(i)}$ are indicators for six taxonomic groups, and $\eta_{l,rr'}$ are edition-pair indicators, so that every comparison is made within a pair of editions whose labels and coverage are the same. Together these terms form the benchmark model. $X_i(\tau_r)$ is one block of predictors, each dated no later than τ_r. The threat block is the presence of any threat named at listing in each of three broad groups (habitat loss and modification, biological threats, and other threats). The CH block is whether CH was in force at τ_r and for how long. The spending block is the three-year moving average and the cumulative total of species-specific spending, both within one reporting regime and in constant dollars. Because these models use only within-regime edition pairs, the spending block draws on spending from 1998 through 2004 alone. This follows from the timing rules rather than from a choice (Section 4).

Each block is added to the benchmark on its own, on the pairs where the block is defined, and the benchmark is refitted on exactly those pairs; a model with all three blocks is reported alongside, outside the headline contrasts. I report average differences in predicted probability rather than coefficients: from the reference level to each other level for banded and grouped variables, and from the 25th to the 75th percentile for spending. I also ask whether each block improves prediction. Using five-fold cross-validation, with each entity kept within one fold, I compare the benchmark's out-of-sample log loss and Brier score (the mean squared Each model is estimated separately within each reporting regime, and I call an association stable only if the within-regime estimates share its sign.

The coefficients β_l describe how the agency's assessments co-move with the predictors, several of which are the agency's own earlier actions. A positive association between spending and a later improving status is consistent with spending helping, with FWS spending more on species it expects to recover, and with improvement being easier to detect where money buys monitoring.

5.3 Recovery, Extinction and Reclassification as Competing Risks

For the legal outcomes the unit is the entity's first listing spell. Let T_i be the time from listing to the first effective removal and J_i its reason: recovered, extinct or other. I estimate the cumulative incidence

$$F_j(t \mid g) = \Pr\left(T_i \leq t,\ J_i = j \mid G_i = g\right), \qquad j \in \{\text{rec}, \text{ext}\}, \tag{3}$$

the probability that an entity in group g is removed for reason j within t years of listing, with the Aalen–Johansen estimator. Removals for other reasons compete with both outcomes, and entities still listed on 31 December 2025 are censored. Entities that left the list with no operative removal notice I could find are excluded from these analyses and from the landmark analysis below, and counted. A Kaplan–Meier curve that censors the competing causes would overstate recovery, and I show it only as a labelled contrast. Groups are fixed at listing so that no group is defined by survival—taxon (vertebrates, invertebrates and plants), listing era (1967–1982, 1983–1995 and 1996–2025), and, for each broad threat group, whether the listing named such a threat. I compare groups at 10, 20 and 30 years after listing, with 20 years the headline, and only where every compared group still has at least 20 entities at risk. If the event roster is complete only from a year later than 1973, entities listed earlier enter the risk set in that year, with time still measured from listing.

The estimator is that of Aalen and Johansen (1978), who give a product-limit estimator of the transition probabilities of a Markov chain from censored data. Removal from the list is the simplest such chain: an entity starts listed and moves once, to recovered, extinct or other. In that case the estimator has a simple form. Let $t_1 < t_2 < \ldots$ be the times at which some entity is removed, n_k the number of entities still listed and observed just before t_k, and d_{jk} the number removed for reason j at t_k. Then

$$\hat{F}_j(t \mid g) = \sum_{t_k \leq t} \hat{S}(t_{k-1} \mid g)\, \frac{d_{jk}}{n_k}, \tag{4}$$

where $\hat{S}$ is the product-limit estimate of the probability of still being listed, with removals of every kind counted as events. Each removal for reason j adds its share of the risk set, weighted by the probability of having stayed listed until then. An entity still listed at the cutoff leaves the risk set without an event, and an entity that enters late joins it when it enters; Geskus (2011) gives the estimator under such late entry and censoring together. Andersen et al. (1993) and Kalbfleisch and Prentice (2002) treat the estimator and its variance, Putter et al. (2007) set out its use with competing risks, and I compute it with the Stata routine of Coviello and Boggess (2004). The model of Fine and Gray (1999) relates the same cumulative incidence to covariates, and Appendix Section C.4 uses it as a check. Cumulative incidence with competing exits is not confined to medicine. Heisey and Patterson (2006) review its methods for cause-specific mortality in wildlife studies, and O'Neill (2019) estimates the cumulative incidence of exits from unemployment to several destinations.

The reason not to use one minus a Kaplan–Meier curve is what censoring assumes. The es-

timator of Kaplan and Meier (1958) treats a censored entity as one that could still be removed for recovery later, at the rate of the entities still at risk. An entity removed for extinction cannot be. If extinctions and other removals are treated as censored, one minus the curve estimates the probability of recovery in a world in which no entity goes extinct or leaves the list for any other reason. As Gooley et al. (1999) show, it overstates the cumulative incidence of recovery once any competing removal has occurred. That world is never observed, and Tsiatis (1975) shows that the probability of recovery in it is not identified from competing-risks data without assumptions about how the times to recovery and to extinction are related, which the data cannot test. The Aalen–Johansen estimator avoids both problems. It drops an entity from the risk set when it leaves for any reason and weights each recovery by the probability of still being listed, so the incidences of recovery, extinction and other removal add up to the probability of removal of any kind.

Reclassifications use the same machinery. For entities whose first spell opens endangered, I estimate the cumulative incidence of the first whole-entity downlisting to threatened; for those opening threatened, of the first uplisting to endangered. A removal of any kind competes with the reclassification, which does not itself end the spell. These analyses require each entity's reclassification history to be certified from its listing, and entities without that certification are excluded and counted. Beside it I report an exploratory version, labelled as such in every table and figure, on 1,615 entities. It adds the 882 entities listed earlier whose recorded rules reach their status in the 1996 CFR edition, the earliest in the source, and whose search through 1980 found no unrecorded change of status. The check rests mainly on the first condition, which catches any missed net change of status before 1996 and does not depend on the search. It has two limits. It cannot exclude a pair of changes that offset each other anywhere from listing to 1996, so an entity's first reclassification may be dated too late. And the search is only a supplement: on known rules from the years in which it finds text, through 1980, it found 5 of 6. The two versions are never pooled, and the exploratory sample is never described as certified. Both follow from the failed search and my decision to add the second.

One grouping needs a landmark. Whether an entity ever had CH depends on how long it survived on the list, because a long-listed entity has more time to receive it, so grouping on "ever had CH" would select on the outcome. I instead condition on being still listed 10 years after listing and group entities by whether CH was in force at that date:

$$F_j^{\ell}(t \mid g) = \Pr\left(T_i \le \ell + t,\ J_i = j \mid T_i > \ell,\ G_i^{\ell} = g\right), \qquad \ell = 10, \tag{5}$$

where G_i^{ℓ} uses only information dated by ℓ years after listing. Entities removed before the landmark are reported in a table of earlier exits, and a landmark of 5 years serves as a check. An entity whose CH history could not be completed has no CH state at the landmark and is left out of both groups, and I count the exits after the landmark that fall among these entities. Where those exits make up all or most of the exits, the comparison of the observed groups says nothing about the missing group, and I read it as uninformative about CH, not as a null difference.

The terminal outcomes are few. Extinction comparisons can detect only large differences,

roughly a doubling of incidence or more, and I describe a null difference in extinction as imprecise rather than as evidence of no difference.

5.4 A Flexible Check

Equation (2) imposes a functional form. As a check that the status associations do not depend on it, I fit a random forest to the same pairs and predictors, with 500 trees, no tuning and probability output. I evaluate it only on held-out whole entities, or on later editions in a time split, never on random rows. I report held-out log loss against the multinomial models, permutation importance on held-out data for single predictors and for each block, and accumulated local effects for spending and years since listing, drawn within the observed range. The forest is a check on the associations, not a forecasting tool, and I do not report impurity-based importance.

5.5 Inference and the List of Contrasts

Status tables and figures cluster on the entity: standard errors in the multinomial models are clustered by entity, and the intervals for transition shares come from the entity bootstrap. Incidence tables and figures use 95 percent percentile intervals from 2,000 bootstrap draws that resample entities, which I prefer to asymptotic variances because extinction groups have few events. FWS often removes or reclassifies several entities in one rule, so a second bootstrap resamples those batch rules as clusters, and where it gives an interval at least 25 percent wider I report the wider one.

The main facts I report (those I call "headline facts") are the diagonal and off-diagonal shares of the transition table within each regime, plus four families of contrasts. The first holds the block contrasts of the status models, the second recovery and extinction incidence at 20 years by taxon, era and threat, and the third downlisting and uplisting incidence at 20 years by the same groups. The fourth is recovery and extinction incidence after the landmark, by CH in force. Within each family I report unadjusted 95 percent intervals and Holm-adjusted p-values, and I call a contrast statistically distinguishable from zero only if it passes Holm's procedure at 5 percent. Everything else is labelled exploratory.

The procedure of Holm (1979) controls the family-wise error rate, the probability of rejecting at least one true null hypothesis in a family. With m contrasts in a family, it orders their p-values from smallest to largest, $p_{(1)} \leq \cdots \leq p_{(m)}$, and compares $p_{(k)}$ with $\alpha/(m-k+1)$. It rejects in that order and stops at the first contrast that fails, so the smallest p-value faces the Bonferroni threshold α/m and each later one a looser threshold. A contrast's Holm-adjusted p-value is the smallest α at which the procedure would reject it. The Bonferroni correction of Dunn (1961) tests every contrast at α/m. Holm's procedure controls the same error rate, under any dependence among the tests, and rejects every contrast that Bonferroni rejects and sometimes more. Benjamini and Hochberg (1995) control a different quantity, the false discovery rate, which is the expected share of false rejections among all rejections. Their procedure rejects more in large families, at the cost of some false findings. Romano and Wolf (2005) control the family-wise error rate with the joint distribution of the test statistics, estimated by resampling, which gains power when the tests are correlated.

List et al. (2019) discuss these corrections for economics, and Clarke et al. (2020) implement the Romano–Wolf correction in Stata.

Holm's procedure fits the facts here for three reasons. First, the families are short lists fixed in the analysis plan, and every contrast that passes is reported as a headline fact, so one false headline costs more than one missed one. That is the error the family-wise rate bounds. Second, the contrasts in a family share entities and are correlated in ways that differ across families, and the procedure needs no assumption about that dependence. Third, it needs only the p-value of each contrast, so it applies in the same way to the status models and to the incidence comparisons.

5.6 Checks

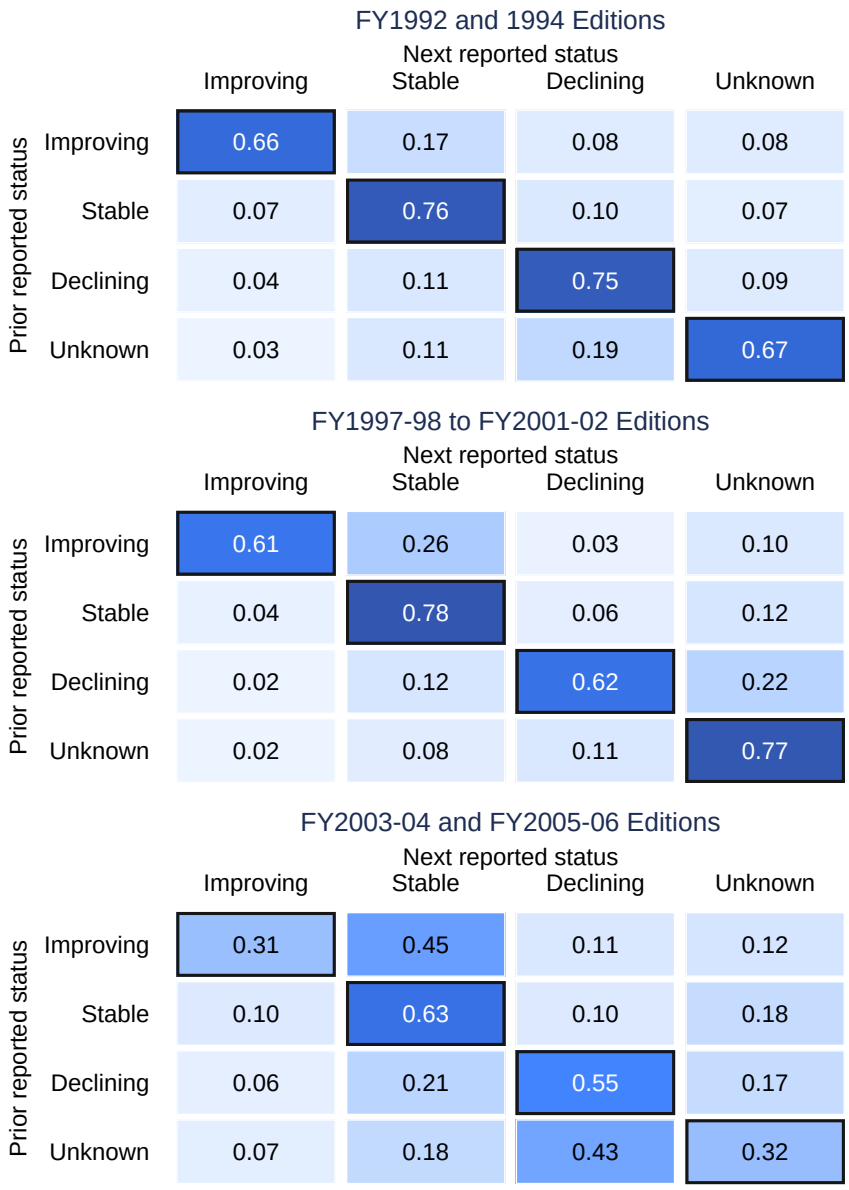
Each headline fact is accompanied by robustness checks, reported in Appendix C and in the Robustness Supplement. First, a timing audit confirms that no predictor window ends after its reference date and no status record follows a removal. Second, I reconcile the historical roster with the current list and report every fact on the full roster and on the entities listed today, which shows how much survivor selection would change it. Third, I show every status fact within reporting regimes, and spending associations within each spending regime that has pairs. Fourth, because the reports print a one-word label and never mark a label as carried forward, I separate records by how many consecutive editions have repeated the label, and I describe this as label repetition, not as evidence that FWS carried an assessment forward. Fifth, I keep unknown explicit and model being assessed (any known status) separately from the status reported. Sixth, for extinction I compare the rule clock with the last confirmed sighting, remove entities last seen before they were listed, and bound the removals whose rule gives no sighting date. Finally, every headline fact is repeated for the entities with FWS as sole lead.

6 Results

I report the facts in the following order: transitions in reported status first, then the associations of next status with lagged predictors, then the legal outcomes. Figure 1 shows the first fact, the transitions in reported status, and Figures 2 and 4 draw the headline contrasts of the status models and of removal with their unadjusted intervals. Appendix Tables E2 and E3 collect every headline contrast with its unadjusted interval and its Holm-adjusted p-value.[15] Of the 41 tested contrasts, 6 pass Holm's procedure at 5 percent, and all of them are in the status models, where they are differences in predicted probability from 0.157 to 0.254 in absolute value. No contrast of recovery, extinction or reclassification incidence passes.

[15]The paper's Appendix numbers its sections, tables and figures with the letters A to C. Each companion document has a letter of its own: D for the Disclosures, which record every contact the project had with the outcome data before and after the analysis plan was frozen; E for the Exhibit Appendix, which holds the tables and figures the paper does not show; R for the Robustness Supplement; and T for the Technical Appendix.

Figure 1: Next Reported Status Given the Prior Status, within Each Reporting Regime



Notes: Each cell is the share of pairs with the prior status in its row that have the next status in its column, within one reporting regime; each row sums to one, and a darker blue is a larger share. The outlined diagonal is persistence. Status is the status the U.S. Fish and Wildlife Service reports, not a population trend. No test is reported. Pair counts and 95% intervals: Appendix Table E1.

6.1 Transitions in Reported Status

The first fact is persistence. Figure 1 shows, within each reporting regime, the share of entities that keep the status they had in one edition when the next edition reports them, for the pairs of consecutive comparable editions in which both labels are improving, stable, declining or unknown. The full table (Appendix Table E13) holds all 4,094 pairs. Only the regimes with more than one edition contain such pairs: the FY1992 and 1994 reports, the editions from FY1997–98 to FY2001–02, and the FY2003–04 and FY2005–06 reports. Figure 1 names each by its editions, and the Appendix tables call them groups 2, 3 and 4. In the two earlier regimes the diagonal dominates; in the last, it dominates for stable and declining. Most entities reported as stable or declining carry the same label into the next edition. The share that stays stable runs from 0.627 in the last regime to 0.776 in the middle one, and the share that stays declining from 0.547 to 0.748. The intervals on these shares are narrow except in the smallest rows.

Movements between labels are uneven. Entities reported as declining rarely move to improving in the next edition; when they leave declining, they move mostly to stable or to unknown. Improving is the least persistent label, or close to it, in every regime, and in the FY2003–04 to FY2005–06 pair more improving entities are reported as stable next than stay improving (a share of 0.312 stay). Unknown is persistent in the two earlier regimes, but in the last one most entities with an unknown status receive a known one in the next edition, mostly declining. Because the reports print a one-word label and never mark it as carried forward, part of this persistence may be a repeated label rather than a repeated assessment, and the table describes published records, not populations. Labels outside the four categories are shown as their own rows in Appendix Section E.2, as unmapped codes, and enter no model.

The shares of records by status in each edition (Appendix Section E.2) count published records rather than movements, so they also include entities that appear in one edition only. They show the same feature from another angle: improving holds the smallest share of records in every edition, and stable, declining and unknown divide most of the rest.

6.2 Associations of Next Status with Lagged Predictors

The status models ask whether threats, CH or spending go with the next reported status once the prior status, years since listing, taxon and edition pair are held fixed. Figure 2 shows the average differences in predicted probability for the 21 headline contrasts, each from the model that adds its block to the benchmark on the pairs where the block is defined, and Appendix Table E4 reports them with their Holm-adjusted p-values. The 6 that pass Holm's procedure come in pairs: each moves probability between declining and unknown, with improving left nearly unchanged.

First, entities whose CH came into force less than five years before the reference date are more likely to be reported as declining next, and less likely to be reported as unknown, than entities with no CH in force. The two differences are close to mirror images: 0.216 in the probability of declining (95 percent interval 0.161 to 0.272) and −0.240 in the probability of unknown (−0.274 to −0.206). Entities whose CH has been in force for five years or more do not differ from those with none in

any outcome that passes Holm. Second, entities at the 75th percentile of cumulative spending are more likely than those at the 25th to be reported as declining next and less likely to be reported as unknown, by the largest margin of any headline contrast (0.254 in probability). The difference in unknown is −0.247 (−0.333 to −0.160). The three-year moving average of spending, in the same model, has differences of the opposite signs: −0.164 in declining (−0.217 to −0.112) and 0.157 in unknown (0.086 to 0.228). The spending block draws on spending from FY1998 through FY2004 alone, as the timing rules imply (Section 4). Third, no threat group passes Holm. The flags for biological and habitat-loss threats show differences close to zero for every outcome, and the flag for other threats shows small differences that do not survive the adjustment.

None of the contrasts that pass Holm is stable. I consider an association stable only if the estimates within each reporting regime share its sign (Appendix Table E25). Of the 21 headline rows, 2 meet that standard and 19 do not, and the two that meet it are differences in unknown that are not distinguishable from zero after adjustment. The pooled differences therefore summarize regimes that disagree in sign, and the spending block can be estimated in only two regimes.

The blocks add little to the benchmark out of sample. The change in held-out multinomial log loss relative to the benchmark is 0.002 for threats, -0.020 for CH and -0.014 for spending, and -0.041 for the model with all three blocks, against a benchmark loss of 0.947 (Appendix Table E17; negative values are better fits).

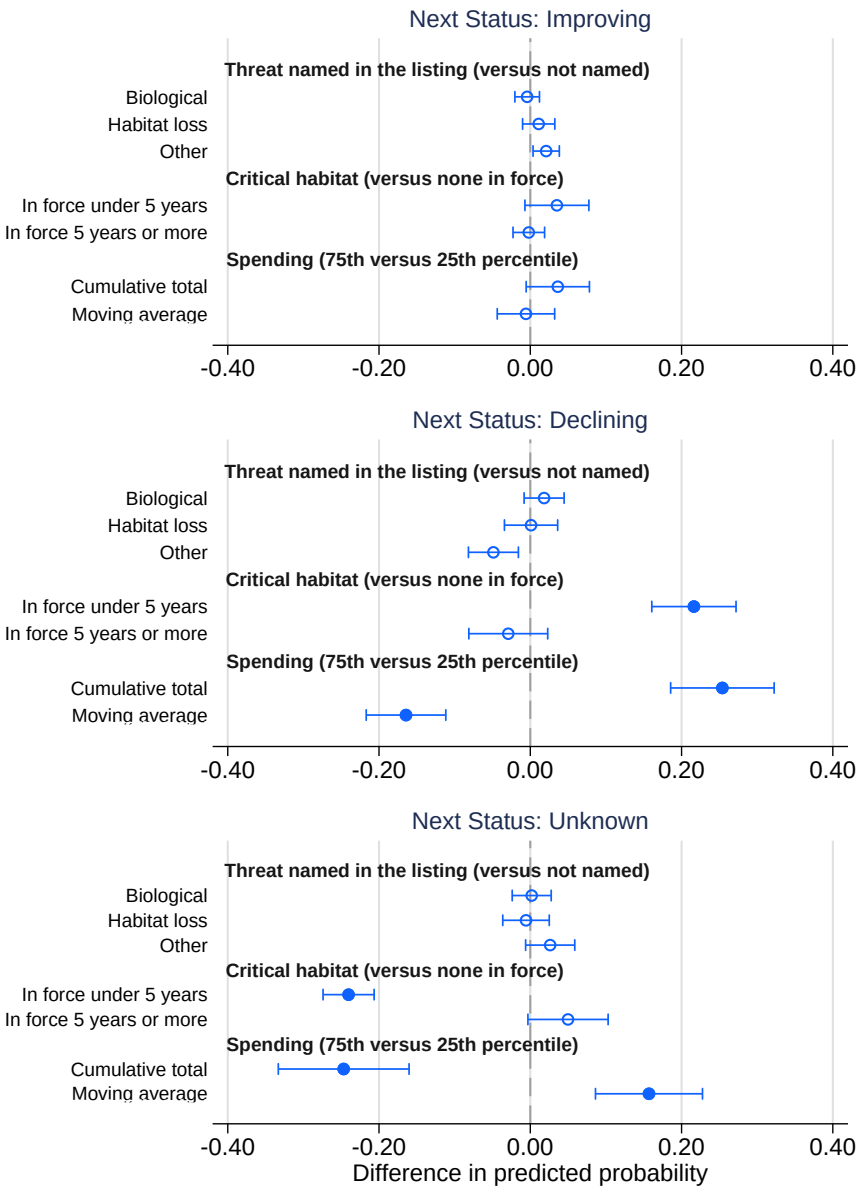
I read the CH and spending patterns as facts about assessment, not about recovery. The differences sit almost entirely in unknown and declining: newly designated CH and more cumulative spending go with fewer entities whose status is unknown and more whose status is declining. That pattern is consistent with attention producing assessments, with FWS designating CH and spending where it already sees decline, and with assessed entities more often being found in decline. The design cannot separate these readings, and the instability across regimes argues against reading any of them as a stable feature of the record. The earlier studies disagree on CH (Taylor et al., 2005; Male and Bean, 2005; Kerkvliet and Langpap, 2007; Gibbs and Currie, 2012), and these results do not settle that disagreement.

The status model with consultation counts is not estimable, as Section 5 anticipated, and no such model was fitted.

6.3 A Flexible Check on the Status Associations

A random forest fitted to the same pairs and predictors fits the held-out data better than the multinomial models (Table 1). On the 1,615 pairs from 959 entities where every block is defined, its held-out log loss is 0.905, against 1.040 for the benchmark and 0.999 for the model with all blocks. The time split could not be computed, because the planned rule leaves no later edition to test on. Its permutation importance tells the same story as the held-out fit of the multinomial blocks (Table 2). Permuting the benchmark variables raises the held-out loss by 0.481, most of it from the prior status (0.397). Permuting CH or spending raises it by 0.068 and 0.065, and permuting the threat flags by 0.017. The accumulated local effects of spending and of years since

Figure 2: Associations of the Next Reported Status with Threats, Critical Habitat and Spending



Notes: Points are differences in the predicted probability that the next reported status is improving, declining or unknown (0.01 is one percentage point), with unadjusted 95% intervals; the dashed line marks no difference. A filled marker passes Holm's procedure at 5% in the family of 21 contrasts; a hollow marker does not, even where its unadjusted interval excludes zero. Pair counts and p-values: Appendix Table E4.

listing are drawn within the observed range (Figure 3). The status an entity had in the last report is by far the best guide to the next one, and the forest finds no threat pattern that the linear models missed.

Table 1: Held-Out Fit of the Random Forest against the Regression Benchmarks

	Entity folds	Difference from benchmark
Random forest	0.905	-0.135
Benchmark regression model	1.040	0.000
Regression model with all three blocks	0.999	-0.041

Sample: pairs on which every block is defined. Entries are held-out log loss (lower is better), from 5-fold cross-validation by entity (Appendix Section B.3).

Table 2: Importance of Predictor Blocks and Single Predictors in the Random Forest

	Entity folds	SD over permutations
Block: Benchmark	0.481	0.011
Block: Threats	0.017	0.003
Block: Critical habitat	0.068	0.006
Block: Spending	0.065	0.005
Predictor: Prior status	0.397	0.014
Predictor: Years since listing, band	0.036	0.002
Predictor: Taxon	0.041	0.002
Predictor: Report edition pair	0.074	0.006
Predictor: Habitat-loss threat	0.005	0.001
Predictor: Biological threat	0.009	0.002
Predictor: Other threat	0.003	0.001
Predictor: Critical habitat	0.068	0.005
Predictor: Spending, moving average	0.023	0.002
Predictor: Spending, cumulative	0.028	0.003

Sample as the fit table. Entries are the mean increase in held-out log loss when the predictor or block is permuted, over 20 permutations, with the standard deviation over permutations; no interval. Near zero means the fit does not depend on it.

6.4 Recovery and Extinction as Competing Risks

Removal for recovery or extinction is rare within 20 years of listing, in every group (Figure 4; Appendix Tables E5–E9). Of the 136 first removals by the cutoff, 67 are for recovery and 31 for extinction, and the analysis sample holds 1,721 entities. Entities listed before October 1996 enter the risk set only then, so for the earliest era the 20-year horizon counts few years of observation. At 20 years the cumulative incidence of recovery removal is a few percent at most in every group, and

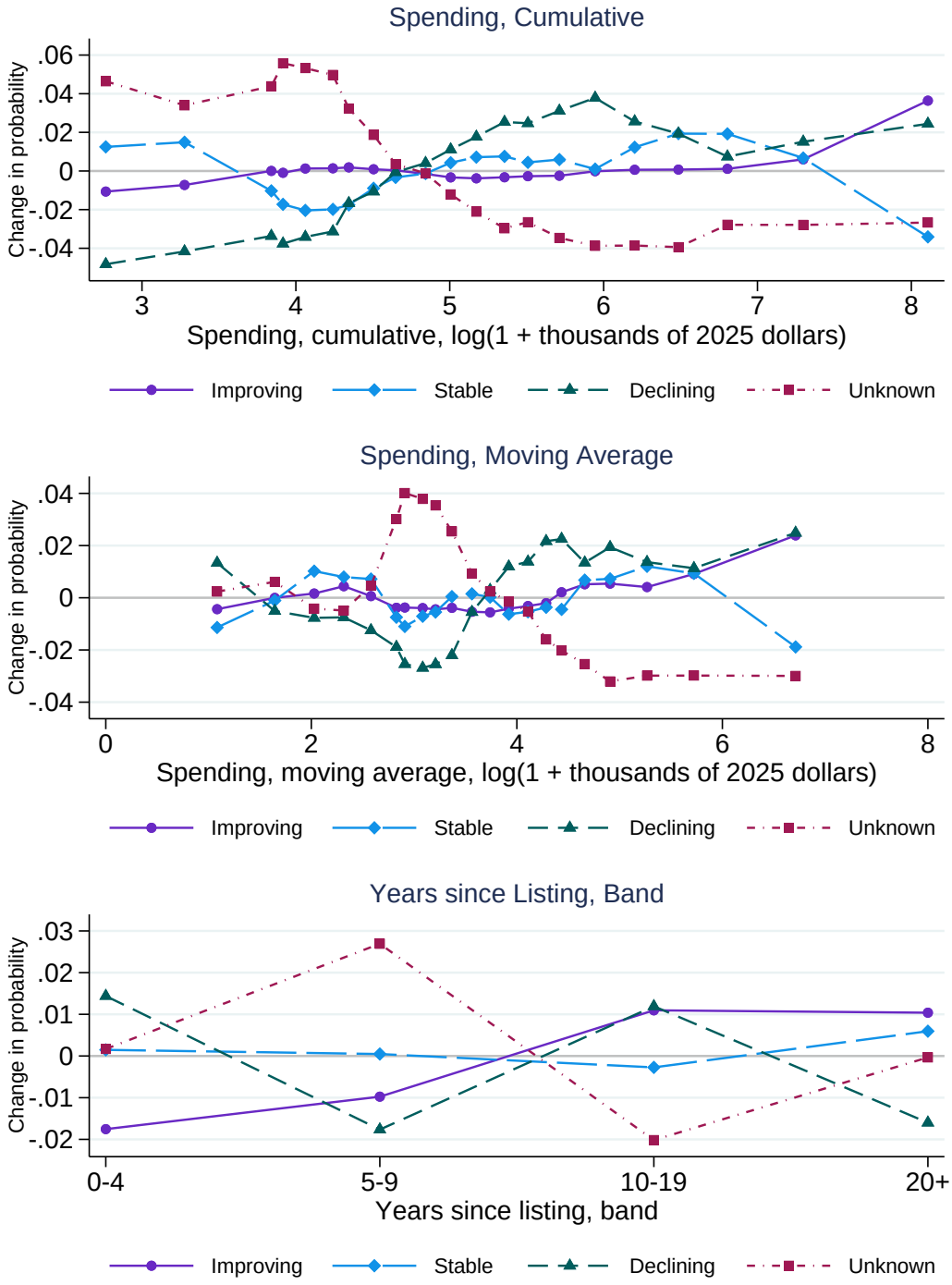


Figure 3: Association of Spending and Listing Age with the Predicted Next Reported Status. Panels: cumulative spending, the three-year average of spending and listing age, each drawn within the observed range. Curves: the centred change in the random forest's predicted probability of each next status (Appendix Section B.3).

the incidence of extinction removal is lower still. Vertebrates have the highest recovery incidence of the three taxonomic groups, 0.020 against 0.007 for plants and 0.000 for invertebrates. Entities listed from 1996 on have a higher recovery incidence than those listed earlier, 0.018 against 0.000 for 1967 to 1982 listings. The highest extinction incidence in these groups is 0.004, among vertebrates. Neither difference passes Holm's procedure: of the 14 contrasts, 0 pass. The extinction contrasts can detect only large differences, so the absence of a difference in extinction is imprecise, not evidence that the groups are alike.

The threat flags do not separate the groups either. At 20 years the differences in recovery and extinction incidence between entities whose listing named a threat group and those whose listing did not are small for all three flags, and none passes Holm's procedure. The threat comparisons leave out 285 entities that the flags do not cover, so they describe listings from 1975 to 2017 only.

6.5 Removal after the CH Landmark

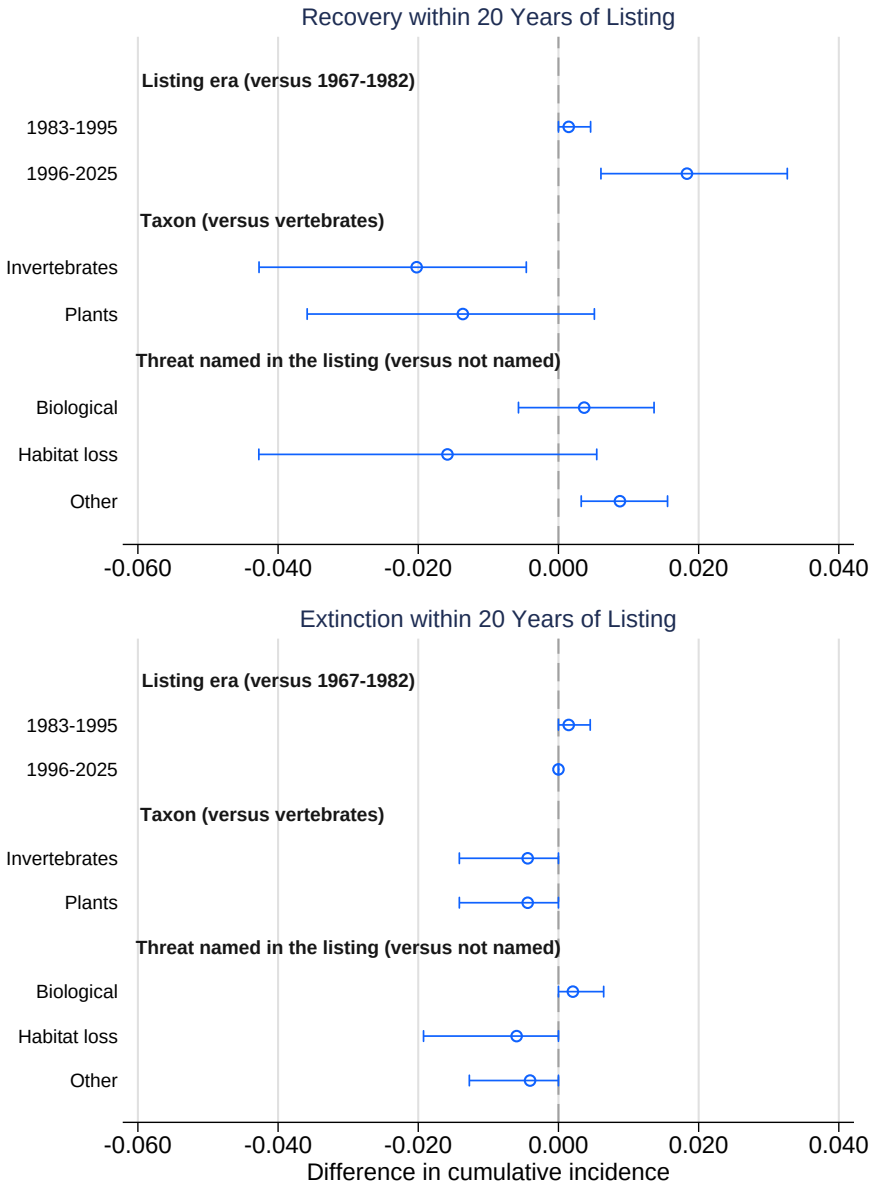
The landmark comparison is uninformative about CH (Appendix Table E10). Among the 1,559 entities still listed 10 years after listing, none of those with an observed CH state is removed in the next 10 years; 652 without CH in force at the landmark and 360 with CH in force remain at risk at that horizon. The 160 entities whose CH state at the landmark is missing hold 14 exits in that window (8 recovered, 1 extinct and 5 other), out of 14 exits in all, and the observed groups hold none. Because all exits fall among the entities whose CH state is missing, comparing the observed groups says nothing about whether CH goes with removal, and I do not read it as a null result. The 10 entities removed before the landmark are listed in Appendix Section E.6.

6.6 Reclassification

The planned reclassification analysis is uninformative, as Section 5 anticipated (Appendix Tables E51 and E52). Its cohorts, drawn from the 733 entities with a certified history, hold 560 entities for downlisting and 171 for uplisting, and they record 2 first downlistings and 1 first uplisting. Every contrast by listing era and by taxon is withheld by the at-risk rule, and none of the others passes Holm's procedure.

The exploratory analysis on the 1,615 entities that pass the weaker check, which was not planned in advance, has more events, 34 first downlistings and 5 first uplistings (Appendix Tables E56 and E57). It is never pooled with the planned analysis. Downlisting within 20 years of listing is rare in every group, and it is most common among vertebrates (0.026) and among entities listed from 1967 to 1982 (0.032), both from the exploratory analysis. Uplisting has fewer events still, and its intervals are wide. No contrast passes Holm's procedure (0 of 14), and because offsetting changes cannot be excluded, a first reclassification in this sample may be dated too late.

Figure 4: Main Results on Removal: Differences in Cumulative Incidence by Group



Notes: Points are differences in the cumulative incidence of each outcome between a group and its reference group, with unadjusted 95% intervals; the dashed line marks no difference. Intervals are bootstrap percentile intervals, so they need not be centred on the estimate, and they are skewed where incidence is near zero. A filled marker passes Holm's procedure at 5% within its family, and a hollow marker does not, even where its unadjusted interval excludes zero. Extinction differences rest on few events and are imprecise. A point at zero without an interval marks a comparison with no removals in either group. The critical habitat comparison is not plotted: no species with an observed critical habitat state is removed in the next 10 years, so the comparison cannot tell the groups apart. Downlisting, uplisting and the p-values are in Appendix Table E3.

6.7 Checks

The checks for each headline fact are in Appendix C and in the Robustness Supplement, and three of them change how the facts read. First, the associations of CH and spending with declining and unknown hold on the entities with FWS as sole lead, on the entities on the current list and on the pairs one edition apart, but not among the entities listed before 1990 or from 1967 to 1982, so they come from the later listings (Robustness Supplement Section R.3). Second, without the cumulative total the moving average of spending shows no difference in declining or unknown, so its opposite signs appear only alongside cumulative spending (Robustness Supplement Section R.3). Third, a model of whether the next status is known at all repeats the pattern between declining and unknown, which fits the reading of these associations as facts about assessment (Appendix Section C.3). The removal facts do not change on the sole-lead or current entities, when the extinctions last sighted before listing are excluded, with and without the bound for removals that state no sighting, in hazard models or within listing eras (Appendix Section C.4 and Robustness Supplement Section R.4). Three checks are owed and listed in Robustness Supplement Table R9, among them the threat comparison on the listings that name their primary threat explicitly.

7 Conclusion

Judging whether the ESA recovers species is hard, because the federal record of status, protection and removal was never built to answer that question. In this paper I assembled that record for 1,758 entities and documented what it shows, without any claim about the effect of protection.

The reported status of a listed entity changes little from one report to the next. Within every reporting regime most entities keep their label, improving is the least persistent label, and few entities reported as declining are reported as improving next. Once the prior status is known, threats, CH and spending add little to the fit of next status out of sample. The associations that survive adjustment for multiple comparisons all move probability between declining and unknown. They go with newly designated CH and with spending (cumulative and recent spending take opposite signs), and none keeps its sign across reporting regimes. These associations hold on the entities with FWS as sole lead and on those listed today, and they come from the later listings: none of them holds among the entities listed before 1990. Removal for recovery or extinction within 20 years of listing is rare in every group I compare, and no contrast in removal or reclassification incidence is statistically distinguishable from zero. The comparison of removal after the CH landmark is uninformative because the exits fall among the entities whose CH history could not be completed. The planned reclassification analysis is also uninformative because the search for earlier reclassification rules failed its recall check.

I caution against reading these facts as evidence that protection does or does not work. Every predictor is chosen in part by FWS on the expected outcome, the outcome is an assessment by the same agency, and the pattern in declining and unknown fits attention producing assessments as well as it fits a change in the species. A null difference in extinction is imprecise, and the removal

comparisons rest on few events. The facts describe the record as the agency kept it.

References

Aalen, Odd O. and Søren Johansen, "An empirical transition matrix for non-homogeneous Markov chains based on censored observations," *Scandinavian Journal of Statistics*, 1978, *5* (3), 141–150. No DOI assigned; JSTOR stable URL.

Abbitt, Robbyn J. F. and J. Michael Scott, "Examining differences between recovered and declining endangered species," *Conservation Biology*, 2001, *15* (5), 1274–1284.

Andersen, Per Kragh, Ørnulf Borgan, Richard D. Gill, and Niels Keiding, *Statistical Models Based on Counting Processes*, Springer, 1993.

Benjamini, Yoav and Yosef Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1995, *57* (1), 289–300.

Clancy, Niall G., John P. Draper, J. Marshall Wolf, Umarfarooq A. Abdulwahab, Maya C. Pendleton, Soren Brothers, Janice Brahney, Jennifer Weathered, Edd Hammill, and Trisha B. Atwood, "Protecting endangered species in the USA requires both public and private land conservation," *Scientific Reports*, 2020, *10* (1), 11925.

Clarke, Damian, Joseph P. Romano, and Michael Wolf, "The Romano–Wolf multiple-hypothesis correction in Stata," *The Stata Journal*, 2020, *20* (4), 812–843.

Coviello, Vincenzo and May Boggess, "Cumulative incidence estimation in the presence of competing risks," *The Stata Journal*, 2004, *4* (2), 103–112.

Dunn, Olive Jean, "Multiple comparisons among means," *Journal of the American Statistical Association*, 1961, *56* (293), 52–64.

Ferraro, Paul J., Craig McIntosh, and Monica Ospina, "The effectiveness of the US endangered species act: An econometric analysis using matching methods," *Journal of Environmental Economics and Management*, 2007, *54* (3), 245–261.

Fine, Jason P. and Robert J. Gray, "A proportional hazards model for the subdistribution of a competing risk," *Journal of the American Statistical Association*, 1999, *94* (446), 496–509.

Gerber, Leah R., "Conservation triage or injurious neglect in endangered species recovery," *Proceedings of the National Academy of Sciences*, 2016, *113* (13), 3563–3566.

Geskus, Ronald B., "Cause-specific cumulative incidence estimation and the Fine and Gray model under both left truncation and right censoring," *Biometrics*, 2011, *67* (1), 39–49.

Gibbs, Katherine E. and David J. Currie, "Protecting Endangered Species: Do the Main Legislative Tools Work?," *PLOS ONE*, 2012, *7* (5), e35730.

Gooley, Ted A., Wendy Leisenring, John Crowley, and Barry E. Storer, "Estimation of failure probabilities in the presence of competing risks: new representations of old estimators," *Statistics in Medicine*, 1999, *18* (6), 695–706.

Greenwald, Noah, Kieran F. Suckling, Brett Hartl, and Loyal A. Mehrhoff, "Extinction and the U.S. Endangered Species Act," *PeerJ*, 2019, *7*, e6803.

Haines, Aaron M., Matthias Leu, Delaney M. Costante, Tyler C. Treakle, Carli Parenti, Jennifer R. B. Miller, and Jacob W. Malcom, "Benchmark for the ESA: Having a Backbone Is Good for Recovery," *Frontiers in Conservation Science*, 2021, *2*, 630490.

Hatch, Leila, María Uriarte, Daniel Fink, Laura Aldrich-Wolfe, Richard G. Allen, Colleen Webb, Kelly Zamudio, and Alison Power, "Jurisdiction over endangered species' habitat: The impacts of people and property on recovery planning," *Ecological Applications*, 2002, *12* (3), 690–700.

Heisey, Dennis M. and Brent R. Patterson, "A review of methods to estimate cause-specific mortality in presence of competing risks," *Journal of Wildlife Management*, 2006, *70* (6), 1544–1555.

Holm, Sture, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, 1979, *6* (2), 65–70. No DOI assigned; JSTOR stable URL.

Kalbfleisch, John D. and Ross L. Prentice, *The Statistical Analysis of Failure Time Data*, 2nd ed., Wiley, 2002.

Kaplan, Edward L. and Paul Meier, "Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association*, 1958, *53* (282), 457–481.

Kerkvliet, Joe and Christian Langpap, "Learning from endangered and threatened species recovery programs: A case study using U.S. Endangered Species Act recovery scores," *Ecological Economics*, 2007, *63* (2-3), 499–510.

Langpap, Christian and Joe Kerkvliet, "Endangered species conservation on private land: Assessing the effectiveness of habitat conservation plans," *Journal of Environmental Economics and Management*, 2012, *64* (1), 1–15.

Leu, Matthias, Aaron M. Haines, Courtney E. Check, Delaney M. Costante, Jessica C. Evans, Margaret A. Hollingsworth, Isabel T. Ritrovato, Ann Marie Rydberg, Alexander M. Sandercock, Kayli L. Thomas, and Tyler C. Treakle, "Temporal analysis of threats causing species endangerment in the United States," *Conservation Science and Practice*, 2019, *1* (8), e78.

List, John A., Azeem M. Shaikh, and Yang Xu, "Multiple hypothesis testing in experimental economics," *Experimental Economics*, 2019, *22* (4), 773–793.

Malcom, Jacob W. and Ya-Wei Li, "Data contradict common perceptions about a controversial provision of the US Endangered Species Act," *Proceedings of the National Academy of Sciences*, 2015, *112* (52), 15844–15849.

Male, Timothy D. and Michael J. Bean, "Measuring progress in US endangered species conservation," *Ecology Letters*, 2005, *8* (9), 986–992.

Miller, Julie K., J. Michael Scott, Craig R. Miller, and Lisette P. Waits, "The Endangered Species Act: Dollars and Sense?," *BioScience*, 2002, *52* (2), 163.

Moore, Frances C., Arianna Stokes, Marc N. Conte, and Xiaoli Dong, "Noah's Ark in a Warming World: Climate Change, Biodiversity Loss, and Public Adaptation Costs in the United States," *Journal of the Association of Environmental and Resource Economists*, 2022, *9* (5), 981–1015.

O'Neill, Donal, "A new competing risks decomposition: application to the effect of cutting unemployment benefit on unemployment durations," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 2019, *68* (3), 793–807.

Putter, Hein, Marta Fiocco, and Ronald B. Geskus, "Tutorial in biostatistics: competing risks and multi-state models," *Statistics in Medicine*, 2007, *26* (11), 2389–2430.

Restani, Marco and John M. Marzluff, "Avian Conservation under the Endangered Species Act: Expenditures versus Recovery Priorities," *Conservation Biology*, 2001, *15* (5), 1292–1299.

Romano, Joseph P. and Michael Wolf, "Stepwise multiple testing as formalized data snooping," *Econometrica*, 2005, *73* (4), 1237–1282.

Taylor, Martin F. J., Kieran F. Suckling, and Jeffrey J. Rachlinski, "The Effectiveness of the Endangered Species Act: A Quantitative Analysis," *BioScience*, 2005, *55* (4), 360–367.

Tsiatis, Anastasios, "A nonidentifiability aspect of the problem of competing risks," *Proceedings of the National Academy of Sciences*, 1975, *72* (1), 20–22.

Wilcove, David S., David Rothstein, Jason Dubow, Ali Phillips, and Elizabeth Losos, "Quantifying Threats to Imperiled Species in the United States," *BioScience*, 1998, *48* (8), 607–615.

Appendix

A How the Data Were Built

In the main text I describe what each series measures and its weaknesses (Section 4). Here I describe how each was built and checked, so a reader can judge the construction without rerunning it. Four rules apply throughout. Every series is built by code from public documents saved unchanged, with a note recording each document's source and retrieval date. A missing value is never a zero. Nothing is filled from a document dated after the fact it records. And the seats that coded predictors were masked from the outcome data, with every lapse in that masking recorded (Disclosures Section D.2).

Appendix Table A1 lists the sources. All are public. Coverage is the span of documents the project holds, not the span of entities they cover, which I give in each subsection.

A.1 Roster, Clock and Spells

I begin from the ECOS exports of every entity with a listing record and keep domestic entities with FWS as sole or joint lead. I screen out experimental populations under section 10(j) as separate entities on three signals—a description that names an experimental population, an ECOS 10(j) publication, and the absence of a final listing rule of its own. I keep each as an action linked to its parent. I restore entities that ECOS no longer carries from the annual CFR lists. For each pair of consecutive editions from 1996 onward I compare the entries, match every entity that appears, disappears or changes status to a ledger entity, and look up the rule behind each unmatched change. Restored entities get their own identifiers and their own listing and removal rules. Before 1996 the same check uses the entity names printed in the 1990, FY1992 and 1994 recovery reports.

An entity's clock starts on the effective date stated in the DATES line of its first final listing rule. Where the line gives a delay ("effective 30 days after publication"), the date is publication plus the delay. Where no rule text is available, I use the ECOS listing date and flag it, and every headline fact is repeated on entities whose dates come from rule text. The first year from which the event record is complete, October 1996, comes from the same CFR comparison. It is the first year after which every net change between editions matches a recorded rule and every recorded rule matches a change.

The construction leaves three residuals, which I report rather than resolve by assumption. First, 264 domestic entries in the CFR editions from 1996 onward could not be matched to a ledger entity by the rules above and were not reviewed one by one. Second, an entity listed and replaced or vacated before 1996 that none of the three early recovery reports names would be missing. Third, some fields could not be verified from an operative rule: the listing citation, the effective date or the scope of a listing. These entities stay on the roster with the field carried from ECOS and a flag, and every headline fact is repeated on the entities with no such flag. Entities listed only by emergency rule have no clock and enter no analysis (1 entities), and the 3 entities first listed after the cutoff stay on the roster without entering the analyses of the legal outcomes.

Table A1: Sources of the Data

Series	Source	Documents held	Feeds
Roster, clock, spells	ECOS exports; 50 CFR 17.11 and 17.12 (GovInfo); final listing rules (Federal Register)	CFR editions 1996–2025; rules from 1967 on	every analysis
Reported status	FWS Recovery Reports to Congress	eleven reports, 1990 to FY2011–12; status through FY2009–10	transitions and status models
Removals, reclassifications	Final removal and reclassification rules (Federal Register, GovInfo)	rules effective through 31 December 2025	removal and reclassification incidence
Threats at listing	Leu et al. (2019) published data	listings 1975–2017	threat block
Critical habitat	CH final rules (Federal Register, GovInfo); ECOS designation events	rules through 31 December 2025	CH block
Spending	FWS expenditure reports to Congress; BEA GDP deflator	FY1998–FY2022 usable	spending block
Consultations	2021 FOIA release (FWS); Defenders of Wildlife records	conclusions from FY2008	exploratory landmark groups (status model not estimable)
Taxon	FWS species taxonomy records	current	benchmark and groups
Land	current ECOS range and CH polygons; PAD-US	deferred to the next version	none in this draft

Notes: The table lists every source used in this draft. ECOS is the FWS Environmental Conservation Online System; GovInfo is the official online edition of the Federal Register and the Code of Federal Regulations; PAD-US is the Protected Areas Database of the United States. The FY1996 and FY1997 expenditure reports are held but unusable (Appendix Section A.6).

A.2 Recovery-Report Status

I transcribed the three scanned editions (1990, FY1992 and 1994) from the page images, row by row, with a separate count of rows on every page. The scans' text layers (1990 and FY1992) and a fresh optical character reading at the scans' native resolution (1994) served as a second reader. Rows are matched one to one by name within a page, never by status, and every disagreement between the two readers went to a third reader who saw the page but not the first two readings. So did a seeded sample of rows the second reader could not match; the other unmatched rows have not all had a blind reading. The later editions have a text layer. I parse them and check them with a second, independent text extractor, page-by-page row counts and a random sample of 30 rows per edition, and I compare each edition's row count with the total the report prints where it prints one. Hand transcriptions are stored with hashes so that any later change is visible.

Each edition's reporting period ends on the date its table states. Where a report covers two periods in one table, the earlier column's end date is inferred and flagged. The regime table records, for each pair of consecutive editions, whether the definitions, categories or instructions in the front matter differ in substance, with the page that shows it. It was built from the front matter alone and fixed before any tabulation. Printed names are matched to entities with the shared name table, which holds every scientific and common name and synonym from ECOS and the listing rules with their validity dates, and one deterministic matcher. No fuzzy match is accepted without review. The labels map to the four categories by a fixed table, with unknown always separate. A printed label outside the four (extinct or presumed extinct, in captivity, presumed extirpated in the United States, and two codes used only in 1994) keeps its own code. Because no edition marks a label as carried forward, I record for each record how many consecutive editions have printed the same label, and I call it label repetition. The entities whose 1994 rows fall on the missing pages 50 and 51 are coded missing by source for that edition, never imputed, and listed by taxon (74 entities).

Two coverage limits follow from the documents. The 1990 table covers only the species with an approved recovery plan, so an entity's first record can come from a later edition even if it was listed before 1990. The FY2011–12 report prints its species table without a status column—the series ends with FY2009–10.

A.3 Removals and Reclassifications

I code each removal from its operative final rule, retrieved from the official Federal Register issue on GovInfo by citation. The coder records the reason (recovered, extinct or other, never taken from the reason field in ECOS), the effective date and the passage that supports both, and, for an extinction, the last confirmed sighting the rule reports. Rules give the sighting as a year, so I keep it at that precision. A removal whose rule cannot be retrieved or coded keeps an unknown reason and is never imputed. Only a rule that removes the whole entity ends its spell. An independent reviewer, on the same vendor's model, recoded a blind sample of removals, including every extinction and every code the first coder marked as less certain, without seeing the first codes (70 of 70 agreed). The 2 entities that left the ECOS list without any operative notice I could find have no removal

date or reason. They are excluded from the removal analyses and listed (Disclosures Section D.1).

Reclassifications are the final downlisting and uplisting rules that ECOS links to each entity, with effective dates read from the rules. I code their scope, whole entity or part of the range, from the rule's text—where the text does not settle it, the change is presumed to cover the whole entity and flagged. An entity's reclassification history from listing is certified when its status in every CFR edition from October 1996 on agrees with the rules found. For entities listed earlier, the rules found from listing to October 1996 must also agree with the CFR at that year, under a search protocol fixed before any reclassification estimate. The protocol searches ECOS and GovInfo by scientific name, common name and CFR entry. The CFR serves only as a check and never supplies a missing change. A reclassification made by a rule that ECOS does not label as one would be missed, and the CFR comparison is the check against that. Two residuals remain: a change reversed within one CFR year, and offsetting changes before October 1996 that no source records.

Before any reclassification estimate, both steps of the search protocol were checked on known roster rules published before 1995, in this order. The page screen first kept 16 of 17 such rules, and what it missed was lost to a misread running head on a scanned page. With the reading fixed, it kept 17 of 17. The GovInfo search returned only 5 of the 16 rules it was run for. It finds Federal Register text through 1980 and none from then to 1993, and it misses the full text that GovInfo holds from 1994. Most of its misses fall after 1980, which no fix to the screen can repair. Of the known rules from the years in which it finds text, it returned 5 of 6. No entity listed before October 1996 is therefore certified (944 entities), and certification covers the 733 entities listed later. A separate flag marks the 882 earlier-listed entities that pass a weaker check: the status at listing and every reclassification rule ECOS records lead to the status in the 1996 CFR edition, the earliest in the source, and the search through 1980 found no rule changing the status that ECOS omits. With the certified entities they form the 1,615 entities of the exploratory reclassification analysis. The flag rests mainly on the CFR comparison, which catches any missed net change before 1996 without the search. It cannot exclude a pair of changes that offset each other anywhere from listing to 1996, and the search through 1980 is a supplement with the recall given above. The flag never stands in for certification.

A.4 Threats at Listing

The threat flags are the published data of Leu et al. (2019), downloaded from the repository their paper names, pinned to one commit and checked against its hash. I use them internally with citation and do not redistribute the raw file. Their column on current listing status is an outcome, so it is dropped on reading and never loaded. A Leu record is matched to a roster entity only if the name, the listing year and the geography all agree and the match is one to one. Ambiguous and unmatched records get missing flags, never zeros. On the accepted roster, 1,438 Leu records match an entity. Every one of the 71 records with no name match is outside the roster's scope: either a species with NMFS as lead, or a coral with no roster name in its genus, checked by hand. That leaves 0 in-scope records unmatched. The name key ignores bracketed synonyms, and one spelling

variant is linked by hand; the matcher is not loosened otherwise. The six Leu categories collapse into the three broad groups of the plan: habitat modification and pollution; interactions with other species; and overuse and environmental and demographic stochasticity.

The flags cover entities listed from 1975 to 2017, so earlier and later listings have no threat record and are excluded from the threat comparisons, counted by era and taxon. They also leave out the statutory factor on inadequate regulation and any threat a rule calls only potential.

The coding of primary threats from the listing rules, which replaces these flags in the next version, follows a codebook frozen before production coding. Coders see only the listing rule. The masked scripts that prepare and read the coding packets work from an explicit list of permitted files that excludes every outcome file, and they stop if a path falls outside it. The model coders' own sessions are not confined that way: each is audited from its transcript after the run, before its codes are accepted, and an operating-system read boundary is planned before the second coder runs. Four proposals from the calibration round that would change how the primary threat is chosen were deferred to a later version of the codebook and are not used.

A.5 Critical Habitat

I find each CH final rule from the designation events in ECOS and the rules themselves, for every roster entity regardless of its current status. Its effective date comes from the rule's DATES line on GovInfo. I code each rule as a designation, a partial revision or a removal of all of the entity's CH. CH is in force on a date if the latest action effective by then left it in force. An entity with no CH action and no contrary evidence is coded as never having CH. Contrary evidence is a CH entry in any CFR edition, a current CH polygon or an ECOS designation event, and an entity with such evidence but no dated rule is coded missing.

Actions that take effect on the same day apply in a fixed order: designation, then revision, then removal. The band of years since designation counts from the first CH final rule, including after a removal and a later designation.

I do not use the current ECOS CH layer to date CH. Its date field holds no usable dates in any feature, and it omits CH that was removed or revised and the CH of entities no longer listed, so dating CH from it would tie the predictor to survival on the list. The layer will serve only the land shares of the next version. One residual remains. CH designated and fully removed before 1996, or designated and removed between 2019 and 2024, would leave no trace in the CFR or in the current polygons. An entity whose search missed such a rule would then be coded as never having CH. A sensitivity analysis treats every entity coded as never having CH as missing instead.

A.6 Spending

I convert the expenditure reports to text and parse their species tables row by row. Each table-year is kept when its rows sum to the table's printed total within rounding, or when the gap is a documented error in the source. Within each row I check that the FWS and other federal amounts sum to the federal amount and that federal and state amounts sum to the species total, and the

few rows that fail are flagged rather than corrected. A printed zero is a reported zero; a species absent from a year's table is missing. The FY1996 and FY1997 reports survive only as scans whose tables do not reconcile with their printed totals, and a second optical reading was worse, so those years are missing for every entity. The FY1989–95 reports print only totals and lists of the top species, and I never allocate them to species.

Printed names are matched to entities with the same name table and matcher as the status reports. Experimental populations under section 10(j) are not roster entities, so the amounts printed for them belong to their listed parent and are summed into it (577 entity-year amounts). A version that leaves them out is an exploratory sensitivity. Rows printed for one entity under variant names, or for several of its populations, are summed too, because each table's printed total is the sum of its rows. A row that fits no single entity stays ambiguous and is never summed into any. Rows that combine several listed entities stay combined and attach to no single entity: the salmon and steelhead units in FY1999 and FY2000, and, before FY2003, species with several listed populations. Land acquisition is reported in a separate table (Table 3 of each report) and kept apart from the species-specific amount. Its rows follow the same rules: amounts for a 10(j) population and rows of one entity under variant names are summed into the entity, and a row that fits no single entity stays ambiguous and is never summed. From FY2012 the reports add a multi-species row, which I never allocate to the species it covers.

The regime break at FY2012 rests on the front matter of each report, read for changes in what counts as species-specific spending, with the page cited for every candidate break. The other candidates change only the format, the unit of listing or a category outside the species rows. Amounts are deflated with the BEA GDP implicit price deflator, averaged over the fiscal year, with 2025 as the base. The three-year moving average needs at least two observed years in one regime, and the cumulative total needs every year from the later of the regime's start and the entity's listing, so neither is computed across FY2012 or over a missing year. Under the plan's timing rules a status transition uses spending dated by the end of the earlier edition's period, and only within-regime edition pairs enter the confirmatory models, so those models use spending from 1998 through 2004 only (Section 4).

A.7 Consultations

The FOIA workbooks of formal and informal consultations are the primary source, and the Defenders of Wildlife records add only activities the workbooks lack. The unit is a consultation and a species it names. A conclusion is assigned to a species only when the record carries a single conclusion type, because the lists of conclusions and species in a record are not aligned. I match species names to ECOS species by exact normalized common or scientific name, and a matched species' consultations attach to each of its listed entities. Placeholders such as "no species present" are labelled as such, and a name shared by several ECOS species is left unassigned rather than guessed. Dates outside the source's range, or a conclusion before its start, are set to missing and flagged, never repaired. About half of the informal records give no conclusion type, so conclusions

describe formal consultations far better than informal ones.

A.8 Land

The land shares are deferred from this draft because the PAD-US ownership layer could not be obtained in time. In the next version, each current ECOS range and CH polygon will be intersected with PAD-US land in public ownership, and the share will be the public area over the polygon's area. Each share will carry the vintage of the polygon and of the ownership layer, and a flag for entities whose range or CH was ever revised, for the check on never-revised entities.

A.9 Coverage Limits in One Place

The limits above recur across the paper, so I collect them here with the choice made for each. None is filled by assumption.

The roster is checked against the CFR from 1996 onward, with 264 domestic CFR entries unresolved, and before 1996 against the names in three recovery reports. The 2 entities that left the list with no operative notice are excluded from the removal analyses. The status series runs from 1990 to FY2009–10; the 1990 table covers only species with an approved plan, and the 1994 edition lacks two pages (74 entities coded missing). The threat flags cover listings from 1975 to 2017 and name no primary threat. CH dates rest on the rules found, with a residual for CH designated and removed before 1996 or between 2019 and 2024. Reclassification histories are certified only for entities listed from October 1996 on, because the Federal Register search failed its recall check, mainly because it finds no text after 1980. The planned analysis on them is uninformative, and a labelled exploratory analysis adds the earlier-listed entities that pass a weaker check. Spending is usable from FY1998 to FY2022, in two regimes, with combined rows never split. Consultations start in FY2008, and the land shares are deferred. Recovery plans enter no analysis in this draft.

B Estimation Details

In the main text I describe the estimators and the inference (Section 5). Here I collect the details that the notes to the tables and figures used to carry and that the main text does not give, so that each note can stay short.

B.1 Intervals and the Holm Families

An interval is flagged unreliable when more than 5 percent of its bootstrap draws fail to estimate, and an unreliable contrast is left out of the Holm adjustment. A contrast withheld by the at-risk rule still counts toward the size of its family, with a p-value of one and no reported p-value. Its groups are shown and flagged as thin. In the bootstrap that resamples batch rules, the clusters are the rule that causes each entity's first exit, and censored entities are singletons. A table marks each row whose reported interval is the wider batch-rule interval. Plan-family clustering is not yet applied in this draft, a limitation on the intervals of the removal, landmark and reclassification tables and figures.

B.2 The Status Models

Each entry in the status-model tables is the average, over the estimation sample, of the difference in the predicted probability of a next status, from the multinomial logit pooled over reporting regimes with edition-pair indicators. Its interval is a delta-method normal 95 percent interval from standard errors clustered by entity, unadjusted for multiplicity. The reference levels are no critical habitat in force and the threat flag not set; for spending, the 75th percentile of log(1+spending) in the estimation sample is compared with the 25th. Prior status, years since listing and taxon group are the benchmark variables and appear in every model, and the threat block uses the planned threat form. Coefficients are not reported. In the held-out fit, a difference is the model's loss minus the benchmark's on the same pairs, so a positive difference is a worse fit.

B.3 The Random Forest Check

The forest is fitted on the pairs on which every block is defined, with land and consultations excluded, and the multinomial benchmarks are refitted on the same folds. Its fit on entity folds is the mean over the folds of the cross-validation by whole entity. A cell of the time split is blank where the planned rule leaves no test pairs. In the permutation importance, every column of a block is permuted jointly. Each entry is the mean over permutations of the increase in held-out multinomial log loss, and the standard deviation is taken over permutations. Versions of the forest that add land or consultations are exploratory. The importance and the accumulated local effects describe the fitted forest. They are associations, not measures of any change in status.

B.4 Clocks for the Removal Analyses

The removal analyses measure time from the entity's own listing. A sensitivity check repeats them under two other clocks. The inherited clock measures time from the inherited origin, the later of the earliest predecessor's listing in the crosswalk and 28 December 1973, with entry at the entity's own listing. The inherited-affected sample is restricted to split, merged and distinct population segment entities that have a predecessor, on the inherited clock. Under every clock, entities listed before the roster is complete enter the risk set when it becomes complete, with time still counted from listing.

C Robustness

This appendix reports the approved set of planned checks. Every other planned check is in the Robustness Supplement, a separate document, with its table and its reading; none is dropped. Appendix Table C1 lists every run of the estimation routines on a restricted sample or with an option switched, with the rows each used, including the runs the Robustness Supplement reports. Estimated in that table means that the run completed, not that every model in it fitted. Like the headline facts, every check describes associations in the record. The runs do not adjust their own comparisons for multiplicity, so I compare each with the headline result rather than read it as a test.

Table C1: Robustness Runs and the Number of Rows Each Used

	Pairs or rows	Estimated
S11: 5 cohort 1990plus	2,085	1
S11: 5 cohort pre1990	2,009	1
S3: 5 without repeated labels (first appearances)	601	1
S3b: C1 tables by label age, descriptive only, no model (plan S3): first	601	0
S3b: C1 tables by label age, descriptive only, no model (plan S3): repeated_twice_plus	744	0
S3b: NOT ESTIMABLE (0 pairs): C1 tables by label age, descriptive only, no model (plan S3): repeated_once	0	0
S4: 5 gap band 1	3,002	1
S4: NOT ESTIMABLE (0 pairs): 5 gap band 3	0	0
S4: NOT ESTIMABLE B3/B4: 0 spending-bearing pairs; 5 gap band 2	0	0
S17 item7: C4 threat groups, era1	.	1
S17 item7: C4 threat groups, era2	.	1
S17 item7: C4 threat groups, era3	.	1
S17 item7: NOT DEFINED (DL-090): explicit-primary subset, 0 rows	.	0

EXPLORATORY: C5 consultation landmark 10 (any consultation since FY2008 at T0+10y)	.	1
EXPLORATORY: C5 consultation landmark 5 (any consultation since FY2008 at T0+5y)	.	1
S10: 4 current-roster subset	3,787	1
S10: 4 sole-lead subset	4,072	1
S6: B0 to B4 for known status	4,094	1
S17: 7 listing era 1	991	1
S17: 7 listing era 2	2,356	1
S17: 7 listing era 3	747	1
S17: NOT DEFINED (DL-090): explicit-primary subset, 0 rows	.	0
S7: 6 identical to primary (window FY1998)	.	0
S7: 6 primary window (for the FY1996 reading)	4,094	1
S7: 6 spending regime R1	3,481	1
S7: 6 with land acquisition (combined measure replaces the primary one, DL-050)	4,094	1
S7: NOT ESTIMABLE: v3 spending_regime_id has only R1 (0 R2 pairs); a covariate fact, not an outcome	0	0
S8: 6 without the cumulative total	4,094	1
R2: 2 C5 with landmark 5	.	1
R12: 12 C4R down, counting partial-range reclassifications	.	1
R12: 12 C4R up, counting partial-range reclassifications	.	1
S23: C4R down, whole-entity reclassifications	.	1
S23: C4R up, whole-entity reclassifications	.	1
S10: 4 current-roster subset, C4 C5 C4R	.	1
S9: 8 excluding last sighting before listing	.	1
S9: 8 rule clock, all entities (primary sample)	.	1
S9: 8 bound: extinctions without a stated sighting treated as sighted before listing (removed)	.	1

S10: 4 sole-lead subset, C4 C5 C4R	.	1
S4: NOT ESTIMABLE B3/B4: 0	0	0
spending-bearing pairs; 5 gap band 2 (rerun under the not-estimable rule)		
S1b: future exposure, diagnostic only	4,094	1

One row per run: the routine, the restriction or option changed, rows used, and whether the routine ran (1 = the routine ran; some models within a run may not have been estimated, Appendix Section C).

C.1 The Record and the Timing Rules

The timing audits find no violation. No exit precedes its entry, no entry precedes the listing date, and no status record falls after the first effective removal or before the listing date in the first-spell file (Appendix Table C2). Removals that took place before the roster window opens fall outside every risk set and are counted on their own row. The D4.2 routine counts the same removals as unobserved, together with any removal dated on the entry date, so the C4 and C5 populations are unchanged (Appendix Table C3). The roster reconciles: entries equal exits, plus entities still listed at the cutoff, plus entities with a missing event code, with nothing left over (Appendix Tables C4 and C5). Every printed record is either in the first-spell file or carries an audit flag, and the few printed before the listing date or after removal are flagged rather than used. The reclassification dates pass the same audit (Appendix Table C6). A handful of reclassifications retained after the cutoff are flagged rather than counted as violations. No analysis file carries the last status before removal as a variable: it is a retrospective description, never a predictor.

Table C2: Timing Audit Repeated on the Analysis Files

group	item	n	share	note
S1	exits: analysis exit before entry (negative duration)	0	.	
S1	exits: removed before the roster window opens (outside every risk set; reported,	25	.	entry is the roster start; counted apart from negative durations
S1	exits: entry before T0	0	.	
S1	status: record after the first effective removal	0	.	
S1	status: first-spell record before T0	0	.	
S1	status: next period end not after the period end	0	.	

Every count must be 0; a violation stops the run (Class C).

Table C3: Removals before the Roster Window: Strict Count against the Unobserved Count of the Risk-Set Audit

group	item	n	share	note
S1 and D4.2	D4.4: removals before the roster window (strict, t_end < entry)	25	.	
S1 and D4.2	D4.4: removals on the entry date (t_end == entry)	1	.	
S1 and D4.2	D4.2 cif_sample_audit, removals: unobserved (stop <= entry, dropped before every	26	.	D4.2 uses <=, D4.4 counts < and ties apart; C4 and C5 populations are unchanged

D4.2 counts rows with stop at or before entry; the difference from the strict count is the ties. No id-level listing.

Table C4: Roster Reconciliation and Printed-Record Audit

group	item	n	share	note
roster	entities in the roster frame	1,758	.	
roster	no T0 (emergency-only or never-listed review): in no risk set	6	.	
entries equal exits plus still listed	entities with a first effective removal	136	.	
entries equal exits plus still listed	entities censored at the cutoff (still listed)	1,611	.	
entries equal exits plus still listed	entities with missing event code	11	.	
entries equal exits plus still listed	entries minus (exits plus still listed plus missing)	0	.	
roster against current list	current_roster disagrees with censored at the cutoff	11	.	entities not censored but on the current list, or censored but off it
removal against its rule	removals without an operative rule id	0	.	

group	item	n	share	note
removal against its rule	removals reported with the operative notice not located	2	.	
removal against its rule	removals with the reason not yet coded	2	.	
dates	entry before T0	0	.	
dates	analysis exit before entry	25	.	
first year the event roster is complete	presence: earliest date complete from (days since 1960)	13,423	.	01oct1996
first year the event roster is complete	status changes: earliest date complete from (days since 1960)	13,423	.	01oct1996
printed records	record rows in analysis_status_records	10,443	.	
printed records	printed records dated before the entity's T0	13	.	
printed records	printed records dated after the first effective removal	6	.	
printed records	record rows from a printed row covering several entities	18	.	
printed records	missing-by-source placeholder rows	74	.	
manifest keys	printed_records_before_t0 (manifest)	13	.	
manifest keys	printed_records_after_removal (manifest)	6	.	
entity-edition records	rows in analysis_status (first-spell file)	10,410	.	
entity-edition records	printed records in neither the first-spell file nor an audit flag	0	.	should be 0

Counts only. C2 stays entity-level on analysis_status. The two manifest keys equal the flagged record counts.

Table C5: Printed Rows and Entity Records by Edition

edition_seq	printed_rows	records	multi_ entity_row	source_ missing	period_ before_t0	period_after_ removal
1	328	329	2	0	0	0
2	680	681	2	0	1	0
3	781	858	6	74	0	0
4	1,107	1,108	2	0	11	0
5	1,182	1,183	2	0	0	3
6	1,227	1,228	2	0	1	0
7	1,234	1,234	0	0	0	0
8	1,251	1,251	0	0	0	0
9	1,258	1,258	0	0	0	1
10	1,312	1,313	2	0	0	2

Table C6: Reclassification Timing Audit

group	item	n	share	note
S24	entities	1,758	.	
S24	first down_whole date before T0	0	.	
S24	first down_whole date after the first removal	0	.	
S24	first up_whole date before T0	0	.	
S24	first up_whole date after the first removal	0	.	
S24	first down_partial date before T0	0	.	
S24	first down_partial date after the first removal	0	.	
S24	first up_partial date before T0	0	.	
S24	first up_partial date after the first removal	0	.	
S24	whole downlisting from an opening status of threatened with no earlier uplisting	0	.	
S24	whole uplisting from an opening status of endangered with no earlier downlisting	0	.	
S24	reclassification date on an entity with no T0, emergency-only or post-cutoff lis	0	.	
same-day with the first removal (removal wins, D<T strict)	downlisting	0	.	
same-day with the first removal (removal wins, D<T strict)	uplisting	0	.	
S24	retained reclassification after the cutoff (flagged, not a violation)	3	.	

Violations must be 0 and stop the run if not. Counts by direction of same-day ties; no outcome shares.

C.2 Reported Status and the Pairs

Selection into pairs is similar across groups (Appendix Table C7). The share of first-spell records with an eligible next record differs little by prior status, taxon or listing era, and most records without one have their next record in another reporting regime. The pairs therefore lose the regime boundaries, not a particular kind of entity.

The reports never mark a label as carried forward, so whether a label was reassessed cannot be identified from them (Appendix Tables C8 and C9). A label can only be counted as a first appearance or as a repeat of the label printed before.

Table C7: Selection into Pairs

group	item	n	share	note
records	first-spell records	10,410	.	
records	pair eligible	4,094	0.393	
not eligible	no consecutive next record (absent from the next edition or last edition)	1,386	.	
not eligible	next record in another reporting regime	4,880	.	
not eligible	same regime, next record present, not eligible (status or D3.8 rule)	50	.	
prior status: share pair eligible	5	0	0	of 1 rows
prior status: share pair eligible	CA	23	0.397	of 58 rows
prior status: share pair eligible	D	1,227	0.406	of 3,022 rows
prior status: share pair eligible	EX	81	0.448	of 181 rows
prior status: share pair eligible	I	308	0.362	of 852 rows
prior status: share pair eligible	NA	0	0	of 1 rows
prior status: share pair eligible	S	1,180	0.364	of 3,245 rows

group	item	n	share	note
prior status: share pair eligible	U	1,274	0.430	of 2,964 rows
prior status: share pair eligible	XU	1	0.250	of 4 rows
prior status: share pair eligible	missing	0	0	of 82 rows
taxon: share pair eligible	birds	343	0.400	of 858 rows
taxon: share pair eligible	fishes	401	0.408	of 983 rows
taxon: share pair eligible	invertebrates	531	0.359	of 1,479 rows
taxon: share pair eligible	mammals	217	0.388	of 560 rows
taxon: share pair eligible	plants	2,408	0.400	of 6,024 rows
taxon: share pair eligible	reptiles_amphibians	194	0.383	of 506 rows
listing era: share pair eligible	1967-1982	991	0.390	of 2,538 rows
listing era: share pair eligible	1983-1995	2,356	0.408	of 5,779 rows
listing era: share pair eligible	1996-2025	747	0.357	of 2,093 rows
listing era: share last record before removal	1967-1982	55	0.022	of 2,538 rows
listing era: share last record before removal	1983-1995	43	0.007	of 5,779 rows

group	item	n	share	note
listing era: share last record before removal	1996-2025	19	0.009	of 2,093 rows

Counts and shares of first-spell records. Prior status, taxon and listing era of pair-eligible against other records.

Table C8: Label Repetition by Edition Sequence

group	item	n	share	note
verbatim-repeat table	records with a carried_forward value	0	.	NOT IDENTIFIABLE from the reports: a label with no text; carried_forward is null; no proxy
label age (label repetition, not carry-forward)	records	10,410	.	
label age, all records	first	4,395	1	of 4,395 rows
label age, all records	repeated_once	2,455	1	of 2,455 rows
label age, all records	repeated_twice_plus	3,253	1	of 3,253 rows
label age, all records	missing	307	1	of 307 rows
label age, share pair eligible	first	1,589	0.362	of 4,395 rows
label age, share pair eligible	repeated_once	1,225	0.499	of 2,455 rows
label age, share pair eligible	repeated_twice_plus	1,122	0.345	of 3,253 rows
label age, share pair eligible	missing	158	0.515	of 307 rows

Label age counts repeats of the printed label across the edition sequence; it is not evidence of carry-forward.

Table C9: Records by Label Age and Edition

edition_seq	recordsfirst	recordsmissing	recordsrepeated_once	recordsrepeated_twice_plus
1	326	1	.	.
2	403	1	274	.
3	342	76	227	213
4	544	76	193	283
5	303	56	451	369
6	469	37	149	567
7	397	29	330	478
8	688	13	172	377
9	430	10	417	399
10	493	8	242	567

C.3 The Status Models

None of the contrasts that pass Holm's procedure in the headline status models is stable. The plan calls an association stable only if the estimates within each reporting regime share its sign, and Section 6 reports the within-regime estimates (Appendix Table E25).

A model of whether the next status is known at all reproduces the pattern of the status models: newly designated CH and cumulative spending go with a known status, and the moving average of spending with an unknown one. This is the same pattern that the headline differences show between declining and unknown.

C.4 Recovery and Extinction

Fine-Gray and cause-specific Cox models on the same groups (Appendix Table C10) agree in direction. Their unadjusted intervals put the hazard of recovery lower among invertebrates than among vertebrates and lower where the listing named a habitat-loss threat, the same directions as the headline differences in recovery incidence, which do not pass Holm's procedure. They put the hazard of extinction higher among invertebrates than among vertebrates, while the headline difference in extinction incidence at 20 years is slightly negative; the hazard ratio covers the whole follow-up and rests on few events. Some extinction ratios are printed as zero, with an interval of zero.

Excluding the extinctions last sighted before listing, the contrasts look as they do in the headline: no contrast passes Holm's procedure in any family where the run reports it. In this run, as in the other S9 runs and the sole-lead run (Robustness Supplement Section R.4), the same contrasts have small unadjusted p-values, recovery among entities listed from 1996 on and among entities whose listing named an other threat, as do the differences by taxon in removals for other reasons. The C5 comparison at a five-year landmark shows no contrast with a small p-value.

Table C10: Fine-Gray and Cause-Specific Cox Hazards for the Main Comparison Groups

group	item	n	share	note
stcrreg, cause 1, landmark 0	taxon3 level 2.taxon3	61	0.083	ratio vs lowest level; 95% CI 0.01 to 0.61; 1,721 entities
stcrreg, cause 1, landmark 0	taxon3 level 3.taxon3	61	0.720	ratio vs lowest level; 95% CI 0.38 to 1.36; 1,721 entities
stcrreg, cause 1, landmark 0	listing_era3 level 2.listing_era3	61	0.978	ratio vs lowest level; 95% CI 0.49 to 1.95; 1,721 entities

group	item	n	share	note
stcrreg, cause 1, landmark 0	listing_era3 level 3.listing_era3	61	2.104	ratio vs lowest level; 95% CI 0.86 to 5.16; 1,721 entities
stcrreg, cause 1, landmark 0	threat_any_habitat_loss level 1.threat_any_ habitat_loss	47	0.526	ratio vs lowest level; 95% CI 0.28 to 0.98; 1,436 entities
stcrreg, cause 1, landmark 0	threat_any_biological level 1.threat_any_ biological	47	1.208	ratio vs lowest level; 95% CI 0.66 to 2.20; 1,436 entities
stcrreg, cause 1, landmark 0	threat_any_other level 1.threat_any_other	47	0.806	ratio vs lowest level; 95% CI 0.46 to 1.41; 1,436 entities
cox, cause 1, landmark 0	taxon3 level 2.taxon3	61	0.083	ratio vs lowest level; 95% CI 0.01 to 0.61; 1,721 entities
cox, cause 1, landmark 0	taxon3 level 3.taxon3	61	0.703	ratio vs lowest level; 95% CI 0.37 to 1.33; 1,721 entities
cox, cause 1, landmark 0	listing_era3 level 2.listing_era3	61	1.056	ratio vs lowest level; 95% CI 0.50 to 2.22; 1,721 entities
cox, cause 1, landmark 0	listing_era3 level 3.listing_era3	61	2.488	ratio vs lowest level; 95% CI 0.91 to 6.77; 1,721 entities
cox, cause 1, landmark 0	threat_any_habitat_loss level 1.threat_any_ habitat_loss	47	0.529	ratio vs lowest level; 95% CI 0.29 to 0.98; 1,436 entities
cox, cause 1, landmark 0	threat_any_biological level 1.threat_any_ biological	47	1.201	ratio vs lowest level; 95% CI 0.66 to 2.19; 1,436 entities
cox, cause 1, landmark 0	threat_any_other level 1.threat_any_other	47	0.810	ratio vs lowest level; 95% CI 0.46 to 1.42; 1,436 entities
stcrreg, cause 2, landmark 0	taxon3 level 2.taxon3	24	3.522	ratio vs lowest level; 95% CI 1.11 to 11.14; 1,721 entities
stcrreg, cause 2, landmark 0	taxon3 level 3.taxon3	24	0.000	ratio vs lowest level; 95% CI 0.00 to 0.00; 1,721 entities
stcrreg, cause 2, landmark 0	listing_era3 level 2.listing_era3	24	2.109	ratio vs lowest level; 95% CI 0.85 to 5.23; 1,721 entities

group	item	n	share	note
stcrreg, cause 2, landmark 0	listing_era3 level 3.listing_era3	24	0.000	ratio vs lowest level; 95% CI 0.00 to 0.00; 1,721 entities
stcrreg, cause 2, landmark 0	threat_any_habitat_loss level 1.threat_any_ habitat_loss	11	1.013	ratio vs lowest level; 95% CI 0.29 to 3.49; 1,436 entities
stcrreg, cause 2, landmark 0	threat_any_biological level 1.threat_any_ biological	11	0.421	ratio vs lowest level; 95% CI 0.17 to 1.04; 1,436 entities
stcrreg, cause 2, landmark 0	threat_any_other level 1.threat_any_other	11	1.103	ratio vs lowest level; 95% CI 0.51 to 2.40; 1,436 entities
cox, cause 2, landmark 0	taxon3 level 2.taxon3	24	3.461	ratio vs lowest level; 95% CI 1.04 to 11.49; 1,721 entities
cox, cause 2, landmark 0	taxon3 level 3.taxon3	24	0.000	ratio vs lowest level; 95% CI 0.00 to 0.00; 1,721 entities
cox, cause 2, landmark 0	listing_era3 level 2.listing_era3	24	6.873	ratio vs lowest level; 95% CI 0.28 to 166.08; 1,721 entities
cox, cause 2, landmark 0	listing_era3 level 3.listing_era3	24	0.000	ratio vs lowest level; 95% CI 0.00 to 0.00; 1,721 entities
cox, cause 2, landmark 0	threat_any_habitat_loss level 1.threat_any_ habitat_loss	11	0.974	ratio vs lowest level; 95% CI 0.28 to 3.34; 1,436 entities
cox, cause 2, landmark 0	threat_any_biological level 1.threat_any_ biological	11	0.433	ratio vs lowest level; 95% CI 0.18 to 1.06; 1,436 entities
cox, cause 2, landmark 0	threat_any_other level 1.threat_any_other	11	1.166	ratio vs lowest level; 95% CI 0.51 to 2.66; 1,436 entities
stcrreg, cause 1, landmark 10	ch_state_t10 level 2.ch_ state_t10	25	0.190	ratio vs lowest level; 95% CI 0.03 to 1.37; 1,399 entities
cox, cause 1, landmark 10	ch_state_t10 level 2.ch_ state_t10	25	0.189	ratio vs lowest level; 95% CI 0.03 to 1.37; 1,399 entities
stcrreg, cause 2, landmark 10	ch_state_t10 level 2.ch_ state_t10	20	0.000	ratio vs lowest level; 95% CI 0.00 to 0.00; 1,399 entities
cox, cause 2, landmark 10	ch_state_t10 level 2.ch_ state_t10	20	0.000	ratio vs lowest level; 95% CI 0.00 to 0.00; 1,399 entities

group	item	n	share	note

Descriptive contrasts between groups fixed at listing (C4) or at the landmark (C5); no causal reading. Cause 1 recovery, cause 2 extinction; share holds the hazard ratio. Not estimated with fewer than 3 events.