Supplementary Information for:

Gender Inequality in Academic Promotion: A Trajectory-Based Analysis

Rongkang Pei, Zeyu Lyu*, Guolong Wang, Zhichao Wang

Corresponding author: lyu.zeyu.e8@tohoku.ac.jp

This PDF file includes: Supplementary Text; Supplementary Figures S1–S8; Supplementary Tables S1–S14; Supplementary References.

Contents

1 Supplementary Text

1.1 Supplementary Note 1: Overview and Supplementary Information for Main Results

1.1.1 Scope and organization of the Supplementary Information

This Supplementary Information provides additional documentation supporting the main-text results, including: (i) sample construction and observability criteria for the two promotion stages; (ii) descriptive statistics for gender composition and promotion durations by stage and discipline; (iii) operational definitions and coding rules for key variables (promotion stages, cohort bins, discipline aggregation, and covariates); (iv) detailed modeling specifications and inference procedures (including the construction of cohort-specific marginal gender effects and cell-balanced weighting); and (v) robustness checks and sensitivity analyses (e.g., alternative discipline mappings and restrictions for limited-exposure cohorts). Unless otherwise noted, all analyses focus on PhD conferral cohorts within 1970–2024, grouped into five-year bins in cohort-based figures and models, consistent with the main text.

Supplementary Table S1. Sample construction and coverage of key variables.

Inclusion / filtering step	N retained	% of previous	Key criteria / variables
Full PhD-holder cohort with verifiable PhD year	78,208	–	PhD year verified
Restrict PhD conferral to 1970–2024	74,617	95.4	Cohort window used in main analyses
Valid research field label available (1970–2024 subset)	74,344	99.6	Field label parsed and mappable
Exclude residual discipline category "Other"	73,995	99.5	Remove sparse residual group
Stage 1 observed transitions (completeness)	20,916	28.3	Observed AP start year
Stage 2 observed transitions (completeness)	10,818	51.7	Observed FP start year
Stage 1 regression sample (complete covariates)	20,762	99.3	Complete covariates for WLS/OLS
Stage 2 regression sample (complete covariates)	10,731	99.2	Complete covariates for WLS/OLS

Notes. The sample construction follows a sequential filtering process starting from the full API-retrieved cohort. Stage 1 transitions refer to the path from PhD completion to Associate Professor; Stage 2 transitions refer to the path from Associate Professor to Full Professor. Transitions are defined by record completeness (i.e., an observed start year for the next rank), yielding the observed-transition samples reported in this table. The non-negative-duration restriction (duration ≥ 0) is applied subsequently when constructing the final analytical samples used for descriptive statistics and regression analyses. Covariates for the regression samples include gender, institutional sector, and academic discipline.

1.1.2 Analytic sample and key variables

The analytic sample is constructed from researcher profiles retrieved via the *researchmap* Web API. The full PhD-holder cohort comprises $N = 78{,}208$ researchers with verifiable PhD completion years. Gender is observed for the full cohort (no missing values in the analytic extract), and women represent 17.30% ($N = 13{,}531$) of the PhD-holder cohort.

Stage-specific samples are defined by record completeness of each transition endpoint in the career-history records, rather than by assuming unobserved endpoints are non-events. Specifically, Stage 1 transitions require an observed PhD year and a subsequent observed Associate Professor start year, whereas

Stage 2 transitions require an observed Associate Professor start year followed by an observed Full Professor start year. Under these completeness criteria, the stage-specific samples with known gender comprise $N = 20{,}916$ (3,765 women; 17,151 men) for Stage 1 and $N = 10{,}818$ (1,711 women; 9,107 men) for Stage 2.

Promotion durations are constructed as year differences between the recorded start years of successive ranks. After excluding cases with inconsistent temporal ordering (i.e., negative promotion durations), the final analytic samples for duration-based analyses comprise $N = 19{,}596$ (3,383 women; 16,213 men) for Stage 1 and $N = 10{,}643$ (1,674 women; 8,969 men) for Stage 2.

We report the full filtering pipeline, inclusion steps, and coverage of key variables in Supplementary Table S1, including the counts retained after: (i) restricting PhD conferral years to 1970–2024; (ii) retaining records with valid research-field information; (iii) constructing stage-specific observability samples for Stage 1 and Stage 2; and (iv) excluding logically inconsistent or non-interpretable timeline values.

Supplementary Table S2. Descriptive statistics of promotion durations by stage and gender.

Stage	Gender	*n*	Mean (SD)	Median [P25, P75]
Stage 1	Female	3,383	7.42 (5.51)	6 [3, 10]
Stage 1	Male	16,213	8.49 (5.38)	8 [5, 12]
Stage 2	Female	1,674	8.08 (4.30)	7 [5, 10]
Stage 2	Male	8,969	8.48 (4.59)	8 [5, 11]

Notes. Promotion durations (in years) are computed as year differences between the recorded start years of successive career stages. This table reports the final analytic samples after applying record completeness criteria (i.e., observed transition endpoints) and excluding cases with inconsistent temporal ordering (negative durations). SD = standard deviation; P25/P75 = 25th and 75th percentiles.

1.1.3 External benchmark against official statistics

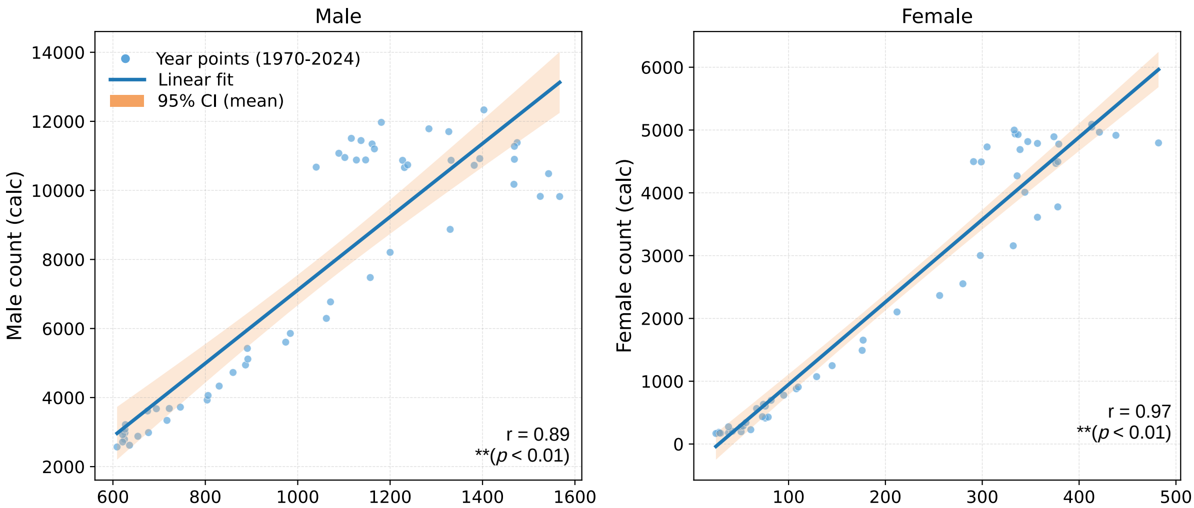
As an external validity check on cohort coverage, we benchmark the annual gender-specific PhD counts derived from the processed *researchmap* cohort against official statistics from e-Stat (MEXT series on doctoral graduates). We align the two sources by calendar year (1970–2024) and compare year-by-year counts separately for men and women. Supplementary Figure S1 shows strong concordance between the series (men: $r = 0.89$, $p < 0.01$; women: $r = 0.97$, $p < 0.01$). This agreement indicates that, at the aggregate time-series level relevant for cohort-based analyses, the processed *researchmap* cohort closely tracks the official gender distribution in doctoral production in Japan.

1.1.4 Descriptive statistics: promotion durations by stage and gender

Promotion durations are summarized separately for the two observed transitions: Stage 1 and Stage 2. Because raw career-history records can contain rare but extreme values (e.g., unusually long gaps, delayed reporting, or partially documented trajectories), we prioritize robust summaries of central tendency and dispersion. Accordingly, Supplementary Table S2 reports both mean (SD) and median with the interquartile range (P25–P75), which together capture typical promotion pacing while remaining interpretable under skewed duration distributions.

Consistent with the main-text descriptive comparisons, the Stage 1 observed-transition sample shows a larger gender contrast in promotion timing than Stage 2. In Stage 1, women have a shorter observed duration than men (median: 6 vs. 8 years; mean: 7.42 vs. 8.49 years). In Stage 2, the corresponding differences are smaller (median: 7 vs. 8 years; mean: 8.08 vs. 8.48 years). These descriptive summaries

Supplementary Figure S1. External benchmark consistency check against official statistics (1970–2024).



Notes: Scatterplots compare annual gender-specific PhD counts from the processed *researchmap* cohort (y-axis) with official e-Stat/MEXT counts (x-axis), shown separately for men (left) and women (right). Solid lines denote linear fits; shaded bands indicate 95% confidence intervals. Reported values are Pearson correlations with two-sided p-values.

provide the baseline distributional context for the cohort- and discipline-stratified analyses reported in the main text.

1.1.5 Gender composition by discipline and stage-entry blocks

To document how representation changes across the observable promotion pipeline, Supplementary Table S3 reports gender composition by aggregated discipline group under three blocks: (i) the overall cohort with valid field labels in the full API pull; (ii) the Stage 1 entry block, defined by observability of an Associate Professor start year following the PhD year; and (iii) the Stage 2 entry block, defined by observability of a Full Professor start year following an Associate Professor start year. This block structure mirrors the funnel-style representation in the main text and makes explicit that "stage entry" is defined by recorded observability rather than by assuming unobserved endpoints are non-events.

Across all blocks, women remain underrepresented in STEM relative to Humanities and Interdisciplinary fields. In the overall cohort (field valid), women account for 11.9% in STEM (6,311/53,100), 31.2% in Humanities (6,034/19,362), and 20.1% in Interdisciplinary fields (1,025/5,104). The same ordering holds at Stage 1 entry and Stage 2 entry, where female representation generally declines further at the later-stage entry point. For clarity and because cell sizes are sparse, the residual "Other" category is not shown in the table body and is excluded from discipline-stratified regression analyses; the corresponding counts are reported in the table notes.

1.1.6 Discipline aggregation (Humanities, STEM, Interdisciplinary)

To facilitate cohort-by-field comparisons while avoiding sparse cells, we aggregate *researchmap* field labels into three discipline groups used throughout the main analyses: Humanities, STEM, and Interdisciplinary (1). The aggregation is anchored to the platform's native field taxonomy. For each profile, we

Supplementary Table S3. Supplementary Table S3. Gender composition by discipline group and stage-entry block.

Discipline group	Female	Male	Total	Female %
Overall cohort (field valid)				
STEM	6,311	46,789	53,100	11.9
Humanities	6,034	13,328	19,362	31.2
Interdisciplinary	1,025	4,079	5,104	20.1
Stage 1 entry				
STEM	1,340	10,985	12,325	10.9
Humanities	2,053	4,666	6,719	30.6
Interdisciplinary	333	1,385	1,718	19.4
Stage 2 entry				
STEM	567	5,683	6,250	9.1
Humanities	959	2,505	3,464	27.7
Interdisciplinary	163	854	1,017	16.0

Notes. "Stage entry" blocks are defined by the observability of each transition endpoint in the standardized career-history records. The residual discipline category "Other" is excluded from the table for clarity and from discipline-stratified regression analyses due to sparse cell counts. For reference, the excluded "Other" counts are: Overall cohort: 111 women and 242 men (total 353; 31.4% female); Stage 1 entry: 28 women and 49 men (total 77; 36.4% female); Stage 2 entry: 15 women and 26 men (total 41; 36.6% female).

parse the reported field string by splitting on the delimiter "|" and then apply a deterministic mapping to classify each segment as Humanities-oriented or STEM-oriented using a predefined lookup table.

Discipline assignment follows a transparent rule hierarchy. If the field string contains only Humanities-related segments, the profile is assigned to Humanities; if it contains only STEM-related segments, it is assigned to STEM. If at least one Humanities-related segment and at least one STEM-related segment co-occur within the same field string, the profile is classified as Interdisciplinary. Field strings that cannot be mapped deterministically are labeled "Other" and are excluded from discipline-stratified regression analyses due to sparse and heterogeneous content. Supplementary Table S4 summarizes these rules.

Because the complete mapping involves Japanese field strings and is lengthy (covering all raw field labels), we provide the full label-level mapping and rule notes as machine-readable Supplementary Data (CSV), rather than as a multi-page LaTeX table. This design preserves reproducibility while keeping the Supplementary Text focused on the operational rules used in the main analyses.

Supplementary Table S4. Field-label aggregation rules for discipline groups.

Assigned group	Rule (applied to researchmap field strings)
Humanities	Field string contains only Humanities-related segments (after splitting by the delimiter "|"), and no STEM-related segment.
STEM	Field string contains only STEM-related segments, and no Humanities-related segment.
Interdisciplinary	Field string contains at least one Humanities-related segment and at least one STEM-related segment.
Other	Field string cannot be mapped under the above rules, or corresponds to residual/uncategorized labels.

Notes. The leading segment before the delimiter "|" is used as the unit of mapping. The complete raw-label mapping with rule annotations (covering all raw field labels) is provided as Supplementary Data.

1.2 Supplementary Note 2: Data dictionary and variable operationalization

1.2.1 Data source and raw-field dictionary

This note provides a structured data dictionary for the *researchmap* fields used in the main analyses and documents the operational definitions of promotion stages, cohort bins, discipline aggregation, and path-structure covariates. All constructions follow the same multi-stage data-processing pipeline described in the main-text Methods.

Supplementary Table S5. Data dictionary for key *researchmap* fields used in this study.

Data component	Raw field(s)	Description and usage in this study
Degree records	Degree title; awarding institution code; award year (if reported)	Inputs to doctoral degree identification and PhD completion-year assignment. When multiple doctoral degrees exist for one researcher, the earliest completion year is used.
Educational histories	Institution codes; start/end years	Used to infer PhD completion year for researchers lacking an explicit award year in the degree record by matching doctoral degree–granting institution codes to the corresponding education-history institution codes.
Career-history (CV) records	Raw appointment texts; year ranges; institution strings	Inputs to Associate Professor and Full Professor identification. Raw texts are processed by LLM-based structured extraction, hard alignment constraints, semantic similarity aggregation, and systematic manual verification (see Methods).
Research field labels	Field label strings (delimiter "|")	Inputs to discipline aggregation. The leading segment preceding "|" is parsed as the mapping unit; multiple segments are used to classify interdisciplinary cases.
Name strings	Kanji name; romaji name	Inputs to gender inference. Kanji name is prioritized when available; romaji name is used when kanji is unavailable.
Institution attributes	Institution-type crosswalk (National/ Public/ Private/ Joint/ Missing)	Used to normalize university sector and construct institutional mobility variables (origin vs. destination sector) in extended controlled specifications.

Notes. Raw-field names are shown conceptually to emphasize reproducible constructs rather than API-specific key strings. Intermediate variables and processing outputs are documented in the Supplementary Data and code.

1.2.2 Operationalization of promotion stages and promotion timelines

Consistent with the main text, we define two promotion stages based on the *observability* of transition endpoints in standardized career-history records. Stage 1 (PhD to Associate Professor) requires an observed PhD completion year and a subsequent observed Associate Professor (AP) start year; Stage 2 (Associate Professor to Full Professor) requires an observed AP start year followed by an observed Full Professor (FP) start year. Promotion timelines are measured in years as differences between the recorded start years of adjacent stages:

$$t_{\mathrm{doc}\to\mathrm{ap}} = \text{AP\_start\_year} - \text{PhD\_year}, \quad t_{\mathrm{ap}\to\mathrm{fp}} = \text{FP\_start\_year} - \text{AP\_start\_year}. \tag{S1}$$

Negative values (inconsistent ordering) are excluded, consistent with the filtering pipeline reported in Supplementary Table S1.

For Associate Professor and Full Professor, the year in which a given academic title was first recorded in the standardized career-history records is treated as the promotion year to that title. Because intermediate years may be missing in CV-style records, the corresponding timeline is assumed to span from the first to the last recorded year (inclusive), even when intermediate years are not explicitly documented, following the operational assumptions in the main-text Methods.

Supplementary Table S6. Core analytic variables and stage-specific observability criteria.

Variable	Type / Unit	Definition (aligned with Methods)
PhD Completion Year	Year (integer)	Verified PhD award year assigned via the two-step rule (direct adoption if reported; otherwise inferred via institutional record matching).
AP Appointment Year	Year (integer)	First recorded year of a validated Associate Professor appointment in standardized career-history records.
FP Appointment Year	Year (integer)	First recorded year of a validated Full Professor appointment in standardized career-history records.
Stage 1 Duration ($t_{doc \to ap}$)	Years (continuous)	Time from PhD completion to first Associate Professor appointment; included only if $\Delta t \geq 0$.
Stage 2 Duration ($t_{ap \to fp}$)	Years (continuous)	Time from Associate Professor to Full Professor appointment; included only if $\Delta t \geq 0$.
PhD Cohort	Categorical (5-year bins)	PhD completion year grouped into five-year intervals (1970–1974, . . . , 2020–2024), consistent with the main text models.
Gender (Female)	Binary (0/1)	Indicator variable for gender (Female = 1; Male = 0), inferred from personal names using the methodology described in Note 1.2.7.

Notes. Stage-specific samples are defined by *observability* (presence of both start and end dates) rather than treating missing data as non-events, consistent with the survival analysis framework in the main text.

1.2.3 Doctoral degree identification and assignment of PhD completion year

Doctoral degree identification constitutes a parallel branch of the career-stage pipeline in the main-text Methods. Degree titles recorded in *researchmap* were encoded as vector representations and matched against prototypical semantic representations of doctoral degrees to identify candidate PhD records. All algorithmically classified doctoral degrees were then subjected to systematic manual verification. Through this procedure, we confirmed 151,446 valid doctoral degree records corresponding to 147,051 researchers.

PhD completion years were assigned using two complementary approaches, consistent with the main text. First, when a degree awarded year is explicitly reported in a doctoral degree record, that year is directly adopted; in cases where multiple doctoral degrees are recorded for the same individual, the earliest completion year is used. Second, for researchers lacking an explicit degree awarded year, doctoral degree–granting institution codes (provided by the *researchmap* Web API) are matched to corresponding institution codes in educational history records, and the latest recorded end year under the matched institution is assigned as the doctoral completion year. Using these procedures, we identified PhD completion years for 78,208 researchers, of whom 74,589 obtained their doctoral degrees between 1970 and 2024 and constitute the final analytical sample used in this study.

1.2.4 Career-history processing and identification of AP/FP appointments

The identification and validation of Associate Professor and Full Professor appointments follow the unified multi-stage data-processing pipeline described in the main-text Methods. Raw CV texts are first standardized via punctuation normalization and whitespace removal. We then employed GPT-5 Nano with a fixed prompt template to extract academic titles from career descriptions in a structured format. The model is instructed to return only titles explicitly stated in the text (e.g., Associate Professor, Professor) while ignoring preceding modifiers related to departments, disciplines, or institutions. All model outputs are returned in a strictly defined JSON structure to ensure determinism and reproducibility.

To minimize noise introduced by automated extraction, a career record is retained for analysis only if the extracted title can be strictly aligned with the original appointment text or corresponding metadata. This hard alignment constraint is implemented as a strict filter: any output failing alignment is discarded and does not enter manual correction or probabilistic weighting procedures. Titles that pass verification are then encoded as vector representations and aggregated based on semantic similarity to consolidate cross-linguistic and synonymous expressions. All aggregated records undergo full manual review to confirm that they correspond to formal Associate Professor or Full Professor titles.

To prevent honorary, temporary, or otherwise non-equivalent titles from being misclassified as formal career stages, we trace preliminarily identified AP/FP records back to their original CV texts. Records containing markers such as honorary, acting, temporary, or assistant are excluded. After cleaning and standardization, the final dataset comprises 51,157 researchers associated with 51,169 Associate Professor records and 44,836 researchers associated with 44,857 Full Professor records.

1.2.5 Discipline aggregation: Humanities, STEM, and Interdisciplinary

Researchers' disciplinary affiliations are derived from the research field labels reported in their *researchmap* profiles (1). Each field label is first parsed by retaining its leading segment, defined as the substring preceding the delimiter "|", which captures the primary disciplinary designation used by the platform. These parsed segments are then mapped to two broad orientations—Humanities versus STEM—using a predefined lookup table reflecting their conventional disciplinary orientation. Profiles containing multiple field segments are classified as Interdisciplinary, while the leading segment is retained as a reference main discipline. Labels that cannot be mapped deterministically are assigned to a residual "Other" category, and missing field information is treated as unknown. The residual "Other" category is documented in Supplementary Data and Supplementary Table S3, but is excluded from discipline-stratified regression analyses due to sparse cell counts.

Supplementary Table S7. Rules for mapping *researchmap* field strings to aggregated discipline groups.

Assigned group	Rule (applied to field label strings)
Humanities	Field string contains only Humanities-related segments (after splitting by "|"), and no STEM-related segment.
STEM	Field string contains only STEM-related segments, and no Humanities-related segment.
Interdisciplinary	Field string contains at least one Humanities-related segment and at least one STEM-related segment.
Other	Field string cannot be mapped under the above rules (residual/uncategorized labels).

Notes. The leading segment before "|" is used as the unit of mapping. The full raw-label mapping is provided in Supplementary Data.

1.2.6 Path-structure covariates

In extended controlled specifications (reported as robustness checks), we incorporate path-structure variables consistent with the main text, including institutional mobility, regional mobility, pre-tenure trajectories, and indicators for same-year promotion. We summarize the covariate blocks below.

Supplementary Table S8. Summary of path-structure covariates and construction windows.

Covariate block	Values	Construction and time window (reference)
Institutional mobility	Same type; Down→Up; Up→Down; Unknown	Normalize university sector (National/Public/Joint/Private; ordered as National>Public/Joint>Private) for origin and destination institutions, then compare sectors within each stage window (Stage 1; Stage 2). Reference: Same type (if present).
Regional mobility	Same region; Local→Metro; Metro→Local; Other; Unknown	Map zones to macro regions (metro areas vs. local Japan vs. overseas), then classify moves within each stage window (Stage 1; Stage 2). Reference: Same region (if present).
Pre-tenure trajectories (pre-AP)	None; Postdoc; Assistant; Lecturer; Multiple roles	Derived from pre-AP role spells in the standardized career histories; window is before first AP appointment. Reference: None (if present).
Same-year promotion	0/1	Stage 1: $1[\text{PhD\_year} = \text{AP\_start\_year}]$; Stage 2: $1[\text{AP\_start\_year} = \text{FP\_start\_year}]$. Reference: 0.
University sector (type)	National; Public; Private; Joint; Missing	Normalized from institution attributes; used as controls depending on stage and specification (doctoral/AP/FP institution). Reference: model-dependent (typically modal sector).

Notes. This table summarizes the covariate blocks described as "path-structure variables" in the main text.

1.2.7 Gender inference and dataset evaluation

Researchers' gender was inferred from personal names, following prior work demonstrating a statistical association between given names and gender categories at the population level (2). Gender inference was implemented using Gemini 3.0 Flash Preview, with gender defined as the category most commonly associated with a name in a statistical sense rather than as an individual's gender identity.

To assess the reliability of name-based inference, we conducted a manual audit on a sample of cases by searching publicly available information on the open web at the individual level. For each audited researcher, we attempted to corroborate the inferred gender using verifiable public cues linked to the researcher (e.g., profile photographs or explicit gender information on publicly accessible platforms). Because many researchers cannot be reliably located through public search or do not provide verifiable gender cues, manual verification was performed only for cases where gender could be determined with sufficient clarity; cases with insufficient information were treated as non-verifiable and excluded from the audit rather than being forced into a classification.

Because confidence scores produced by LLMs are not directly comparable across languages and writing systems (4), and because internal model probabilities do not correspond to posterior probabilities in real-world populations, we did not impose explicit confidence thresholds on model outputs. Instead, we reduced systematic uncertainty in gender inference by controlling the quality of name information at the input stage through a stratified approach (5). Specifically, given that Japanese names convey markedly different amounts of gender-related information across writing systems, we prioritized names written in kanji as inputs for gender inference. Prior linguistic and computational research shows that kanji names typically exhibit greater gender discriminability than names represented solely through phonetic or transliterated forms (6). When kanji representations were unavailable, romanized names were used as an alternative input and treated as a secondary, less informative source.

1.3 Supplementary Note 3: Model specifications and statistical inference

1.3.1 Stage- and discipline-stratified regression specifications

Consistent with the main text, we estimate gender gaps in promotion timelines using ordinary least squares (OLS) models separately by promotion stage and by aggregated discipline group (Humanities, STEM, Interdisciplinary). For each stage $s \in \{\text{Stage } 1, \text{Stage } 2\}$, the outcome is the observed promotion duration in years, constructed from standardized career-history records:

$$Y_i^{(1)} = t_{\text{doc}\to\text{ap},i}, \qquad Y_i^{(2)} = t_{\text{ap}\to\text{fp},i}. \tag{S2}$$

All models are estimated on the corresponding stage-specific *observable-transition* samples defined in Supplementary Note 1 (and Supplementary Table S1). Analyses are restricted to PhD conferral cohorts 1970–2024, grouped into five-year cohort bins, with the 1970–1974 bin serving as the baseline cohort.

Let Female_i be a binary indicator (female=1; male=0). Let $\mathbf{C}_i$ denote a vector of cohort-bin indicators (excluding the baseline bin), and let $\text{Female}_i \times \mathbf{C}_i$ denote the corresponding full set of Female-by-cohort interaction terms. For each stage and discipline group, we report two specifications:

Model 1 (main effects).

$$Y_i^{(s)} = \alpha^{(s)} + \beta^{(s)}\text{Female}_i + \mathbf{C}_i^\top \gamma^{(s)} + \varepsilon_i^{(s)}. \tag{S3}$$

Model 2 (full Female×cohort interactions).

$$Y_i^{(s)} = \alpha^{(s)} + \beta^{(s)}\text{Female}_i + \mathbf{C}_i^\top \gamma^{(s)} + (\text{Female}_i \times \mathbf{C}_i)^\top \theta^{(s)} + \varepsilon_i^{(s)}. \tag{S4}$$

In both models, coefficients are interpreted in year units. Positive values indicate longer observed promotion durations for women (female disadvantage), and negative values indicate shorter observed promotion durations for women (female advantage), conditional on being observed in the corresponding promotion transition.

1.3.2 Covariate-adjusted specifications used in robustness/path-structure analyses

In addition to the baseline cohort models (Eqs. S3–S4), we estimate covariate-adjusted specifications in robustness and auxiliary analyses that incorporate path-structure variables described in the main text (institutional mobility, regional mobility, pre-tenure trajectories, and same-year promotion). The covariate blocks and their stage-specific construction windows are documented in Supplementary Data (`covariate_definitions.csv`) and summarized in Supplementary Table S8.

Formally, the covariate-adjusted interaction model takes the following form within each stage and discipline group:

$$Y_i^{(s)} = \alpha^{(s)} + \beta^{(s)}\text{Female}_i + \mathbf{C}_i^\top \gamma^{(s)} + (\text{Female}_i \times \mathbf{C}_i)^\top \theta^{(s)} + \mathbf{X}_i^{(s)\top} \delta^{(s)} + \varepsilon_i^{(s)}, \tag{S5}$$

where $\mathbf{X}_i^{(s)}$ includes (as applicable) institutional-sector mobility, macro-regional mobility, pre-AP trajectory measures, and a stage-specific same-year promotion indicator (Stage 1: $1[\text{PhD\_year} = \text{AP\_start\_year}]$; Stage 2: $1[\text{AP\_start\_year} = \text{FP\_start\_year}]$).

Categorical covariates include an explicit "Unknown" category where missingness reflects unobserved origin/destination attributes rather than logical impossibility. For specifications that require complete covariate vectors, we use listwise deletion; the resulting regression sample sizes are reported in Supplementary Table S1.

1.3.3 Cell-balanced weighting

To mitigate imbalance in gender composition across cohort bins and discipline groups, the main discipline-stratified regression estimates are computed using weighted least squares (WLS) with *cell-balanced* weights. The target is to equalize the contribution of each cohort-by-gender cell within a stage and discipline group, so that estimated gender effects are not mechanically dominated by large male cells.

For a given stage s and discipline group d, define cells by (c,g) where c indexes the five-year cohort bin and $g \in \{0,1\}$ indexes gender (male=0, female=1). Let $n_{c,g}^{(s,d)}$ denote the observed sample size in cell (c,g). Each observation i in cell (c,g) receives weight

$$w_i^{(s,d)} = \frac{1}{n_{c,g}^{(s,d)}}, \tag{S6}$$

and weights are normalized within each stage–discipline estimation sample to have mean 1 for numerical stability:

$$\tilde{w}_i^{(s,d)} = \frac{w_i^{(s,d)}}{\frac{1}{N^{(s,d)}} \sum_{j=1}^{N^{(s,d)}} w_j^{(s,d)}}. \tag{S7}$$

All reported regression coefficients, marginal effects, and confidence intervals in the discipline-by-cohort figures (Figs. 3–4 in the main text) are derived from WLS estimates using $\tilde{w}_i^{(s,d)}$. Unweighted OLS estimates are provided as robustness checks in Supplementary Note 4.

1.3.4 Statistical inference and standard errors

All hypothesis tests are two-sided. Inference is based on heteroskedasticity-robust standard errors using the HC1 estimator. For each coefficient or linear contrast, we report 95% confidence intervals computed as estimate $\pm 1.96 \times$ SE. No multiple-testing correction is applied in the main regression tables; sensitivity checks focus on the stability of patterns across adjacent cohorts and across discipline groups.

1.3.5 Inference for cohort-specific gender effects (construction of marginal effects)

To visualize cohort dynamics in gender gaps within each discipline and promotion stage, we transform Model 2 interaction coefficients into cohort-specific marginal effects of gender (Female−Male). Under Eq. S4, the cohort-specific gender gap in cohort bin c is defined as the linear contrast:

$$\mathrm{ME}_c^{(s)} = \mathbb{E}\left[Y^{(s)} \mid \text{Female} = 1, c\right] - \mathbb{E}\left[Y^{(s)} \mid \text{Female} = 0, c\right]. \tag{S8}$$

With the 1970–1974 cohort as the baseline, this yields:

$$\mathrm{ME}_{\text{baseline}}^{(s)} = \beta^{(s)}, \qquad \mathrm{ME}_c^{(s)} = \beta^{(s)} + \theta_c^{(s)} \quad \text{for } c \neq \text{baseline}. \tag{S9}$$

Standard errors for $\mathrm{ME}_c^{(s)}$ are computed using the delta method via linear hypothesis tests on the estimated coefficient vector under HC1 robust covariance. Confidence intervals are constructed as $\mathrm{ME}_c^{(s)} \pm 1.96 \times \mathrm{SE}(\mathrm{ME}_c^{(s)})$. These cohort-specific marginal effects correspond to the point estimates and error bars displayed in the cohort-by-discipline marginal-effect plots.

1.4 Supplementary Note 4: Additional results and sensitivity analyses

This note reports robustness checks evaluating whether the cohort-specific gender differences in promotion duration are sensitive to (i) sample-composition reweighting, (ii) the inclusion of the most recent (limited-exposure) cohort, (iii) additional career-path and mobility covariates, and (iv) alternative outcome transformations. Unless otherwise stated, all models are estimated separately by stage and discipline using Model 2 (female main effect, cohort-bin fixed effects, and female$\times$cohort interactions) with HC1 robust standard errors. For all sensitivity checks, marginal female effects are defined as Female$-$Male; positive values indicate longer observed promotion durations for women (female disadvantage), and negative values indicate shorter durations (female advantage).

1.4.1 Sensitivity to cell-balanced weighting

The main-text regressions are estimated using unweighted OLS. To assess whether cohort-specific marginal gender effects depend on the empirical distribution of observations across cohort-by-gender cells, we re-estimated Model 2 using cell-balanced weights within each stage$\times$discipline sample. Specifically, for each cohort bin, observations were weighted inversely proportional to the number of observations in the cohort$\times$gender cell, and weights were normalized to sum to the corresponding sample size. Supplementary Table S5 compares cohort-specific marginal female effects between the baseline (unweighted OLS) and the weighted specification (WLS), using Pearson correlation and absolute differences across matched cohorts. The two specifications yield numerically identical marginal effects across all stage–discipline combinations in this implementation, indicating that the cohort-by-gender interaction estimates are not driven by unbalanced cell sizes.

Supplementary Table S9. Sensitivity to cell-balanced weighting (baseline vs. weighted OLS).

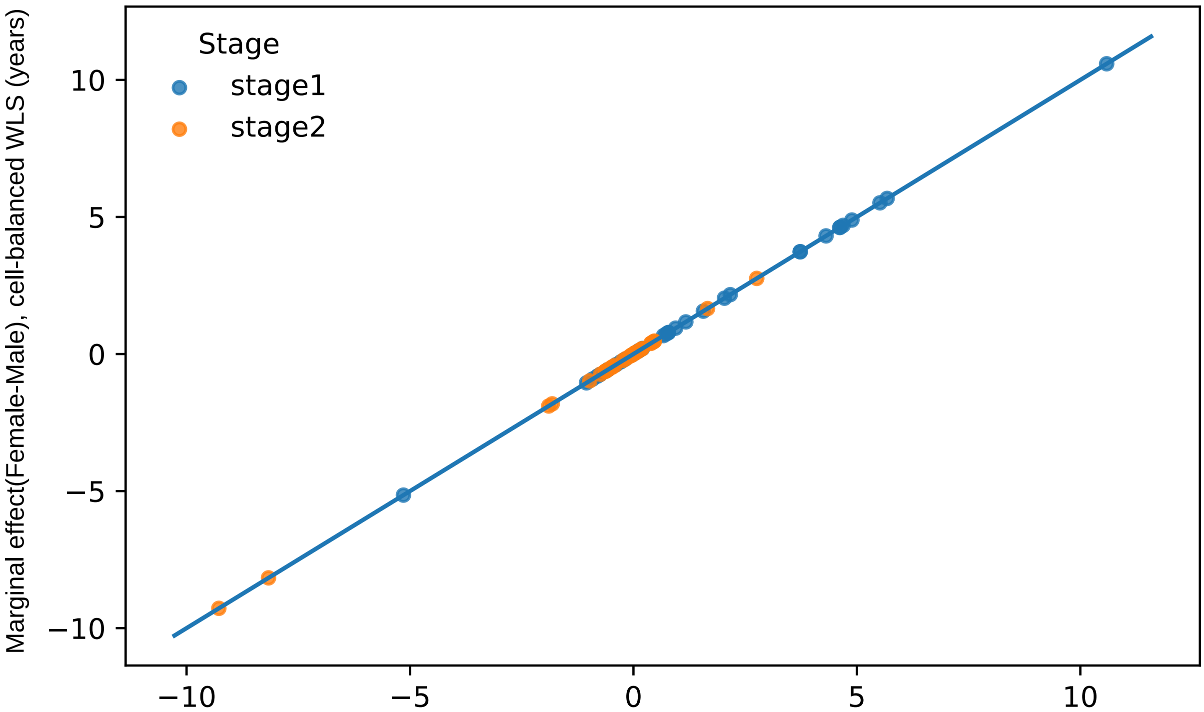
Stage	Discipline	Cohorts	r	Mean $\|\Delta\|$	Max $\|\Delta\|$	Sign flips
Stage 1	Humanities	111.00	0.00	0.00		0
Stage 1	STEM	111.00	0.00	0.00		0
Stage 1	Interdisciplinary	101.00	0.00	0.00		0
Stage 2	Humanities	111.00	0.00	0.00		0
Stage 2	STEM	81.00	0.00	0.00		0
Stage 2	Interdisciplinary	81.00	0.00	0.00		0

Notes. r denotes the Pearson correlation of cohort-specific marginal female effects (Female–Male) between the baseline specification and the alternative specification. Mean $|\Delta|$ and Max $|\Delta|$ report absolute differences in estimated effects (years) across matched cohorts. "Sign flips" counts cohorts where the estimated marginal effect changes sign across specifications.

1.4.2 Sensitivity to cohort window (limited-exposure cohorts)

Cohorts at the end of the observation window (notably 2020–2024) have limited exposure time to complete promotion transitions, which can mechanically attenuate realized transition rates and increase uncertainty in time-to-promotion estimates. To evaluate whether the inclusion of the most recent cohort affects cohort-specific gender effects, we re-estimated the baseline Model 2 on a restricted window (1970–2019), excluding the 2020–2024 cohort. Supplementary Table S6 shows that cohort-specific marginal effects are highly concordant between the 1970–2024 and 1970–2019 windows across stage–discipline combinations, implying that the headline cohort patterns are not an artifact of limited exposure in the most recent cohort.

Supplementary Figure S2. Cohort-window sensitivity (1970–2024 vs. 1970–2019).



Notes: The figure contrasts cohort-specific marginal female effects from the baseline specification estimated on the full window (1970–2024) versus a restricted window excluding the 2020–2024 cohort (1970–2019). All estimates use Model 2 with HC1 robust standard errors and are computed as linear contrasts of interaction coefficients (Female−Male) within each discipline and stage.

Supplementary Table S10. Sensitivity to cohort window (1970–2024 vs. 1970–2019).

Stage	Discipline	Cohorts	r	Mean $\|\Delta\|$	Max $\|\Delta\|$	Sign flips
Stage 1	Humanities	101.00	0.00	0.00		0
Stage 1	STEM	101.00	0.00	0.00		0
Stage 1	Interdisciplinary	91.00	0.00	0.00		0
Stage 2	Humanities	101.00	0.00	0.00		0
Stage 2	STEM	71.00	0.00	0.00		0
Stage 2	Interdisciplinary	71.00	0.00	0.00		0

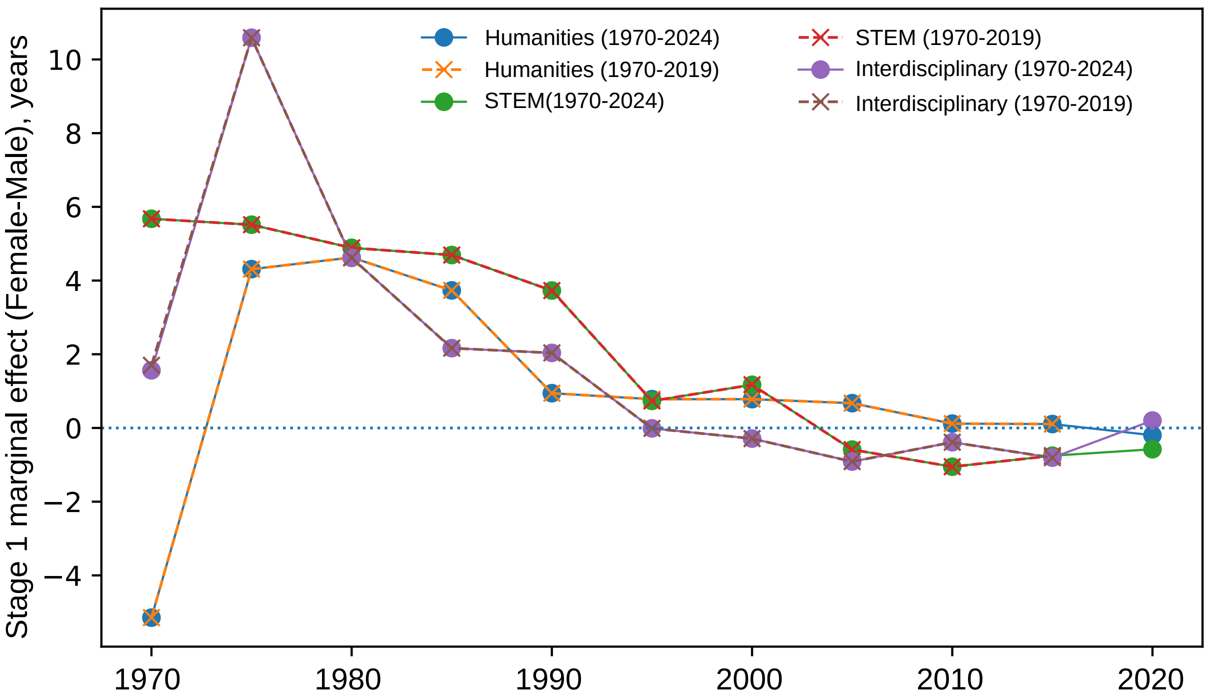
Notes. r denotes the Pearson correlation of cohort-specific marginal female effects (Female–Male) between the two cohort windows. Mean $|\Delta|$ and Max $|\Delta|$ report absolute differences (years) across matched cohorts. "Sign flips" counts cohorts where the estimated marginal effect changes sign across specifications.

1.4.3 Sensitivity to covariate adjustment (career-path and mobility controls)

The baseline Model 2 focuses on cohort structure and does not condition on detailed career-path or mobility features. To assess whether the estimated cohort-specific gender effects are robust to adjustment for career-history and institutional context, we re-estimated Model 2 under an extended specification adding: (i) institutional-sector mobility (same type / down-to-up / up-to-down / unknown), (ii) macro-regional mobility (same region / local-to-metro / metro-to-local / other / unknown), (iii) a same-year promotion indicator (to capture rank skipping or accelerated promotion), and (iv) pre-associate-professor trajectory type (none / postdoctoral / assistant / lecturer / multiple roles). These controls correspond to the main-text path-structure variables and follow the coding rules in Supplementary Note 2.

Supplementary Table S7 summarizes the concordance between cohort-specific marginal female effects from the baseline and extended specifications. For Stage 1, cohort-specific gender effects remain highly stable across all disciplines after adjustment. For Stage 2, STEM and Interdisciplinary estimates are also broadly stable. In Humanities, differences are concentrated in the earliest cohorts where the number of observed female Stage 2 transitions is extremely small; in such sparse cells, adding multiple categorical controls can induce high variance and unstable point estimates. Restricting the comparison to better-populated cohorts increases concordance, indicating that conclusions for adequately populated cohorts are not driven by the covariate set.

Supplementary Figure S3. Covariate sensitivity (baseline vs. extended path controls).



Notes: Comparison of cohort-specific marginal female effects estimated using Model 2 with (i) the baseline covariate set (female, cohort, and female×cohort interactions) and (ii) the extended specification adding institutional mobility, regional mobility, same-year promotion, and pre-AP trajectory controls. Marginal effects are computed as Female−Male within each cohort and discipline; confidence intervals use HC1 robust standard errors.

Supplementary Table S11. Sensitivity to covariate adjustment (baseline vs. extended path controls).

Stage	Discipline	Cohorts	r	Mean $\|\Delta\|$	Max $\|\Delta\|$	Sign flips
Stage 1	Humanities	111.00	0.20	0.39		1
Stage 1	STEM	110.97	0.59	2.02		2
Stage 1	Interdisciplinary	101.00	0.36	0.67		2
Stage 2	Humanities	110.29	1.80	9.79		1
Stage 2	STEM	80.96	0.16	0.44		0
Stage 2	Interdisciplinary	80.95	0.16	0.51		0

Notes. r denotes the Pearson correlation of cohort-specific marginal female effects (Female–Male) between the baseline and extended specifications. Mean $|\Delta|$ and Max $|\Delta|$ report absolute differences (years) across matched cohorts. "Sign flips" counts cohorts where the estimated marginal effect changes sign across specifications. The low concordance for Stage 2 Humanities is driven by extremely sparse earliest cohorts; restricting the analysis to cohorts with at least 10 observed transitions per gender yields substantially higher concordance (see main text for discussion).

1.4.4 Estimator sensitivity (levels vs. log-transformed outcomes)

Promotion durations are right-skewed, and a small number of long trajectories can affect level-based estimates. To evaluate robustness to distributional skew, we re-estimated the baseline Model 2 using a log-transformed outcome, $\log(1+\text{duration})$, and compared cohort-specific marginal female effects against the level-based OLS estimates. For interpretability, log-scale marginal effects are reported as percent differences, computed as $100 \times (\exp(\widehat{ME}) - 1)$ with confidence intervals transformed accordingly. Supplementary Table S8 reports the two estimators side-by-side for each stage, discipline, and cohort bin. Overall, the sign and relative ordering of cohort-specific gender effects are preserved under the log transformation in most cohorts, while effect magnitudes are compressed, as expected under a concave transformation of the outcome.

Supplementary Table S12. Estimator sensitivity: OLS (years) vs. log-OLS (percent difference).

Discipline	Cohort	n	OLS: ME (years) [95% CI]	Log-OLS: ME (%) [95% CI]
Stage 1				
Humanities				
Humanities	1970-1974	19	-5.15 [-7.44, -2.85]	-61.1 [-74.6, -40.3]
Humanities	1975-1979	31	4.31 [-4.39, 13.01]	27.4 [-36.2, 154.5]
Humanities	1980-1984	84	4.63 [1.21, 8.04]	79.4 [27.1, 153.1]
Humanities	1985-1989	135	3.73 [-0.05, 7.52]	49.6 [8.7, 105.9]
Humanities	1990-1994	258	0.94 [-0.66, 2.55]	22.6 [-2.5, 54.3]
Humanities	1995-1999	477	0.78 [-0.14, 1.71]	21.4 [5.0, 40.3]
Humanities	2000-2004	902	0.78 [0.11, 1.46]	14.7 [3.6, 26.9]
Humanities	2005-2009	1303	0.67 [0.16, 1.19]	8.3 [-0.8, 18.2]
Humanities	2010-2014	1273	0.12 [-0.30, 0.54]	-0.3 [-8.1, 8.1]
Humanities	2015-2019	952	0.11 [-0.21, 0.43]	1.9 [-5.9, 10.4]
Humanities	2020-2024	339	-0.20 [-0.54, 0.15]	-10.2 [-22.6, 4.2]
STEM				
STEM	1970-1974	36	5.68 [1.64, 9.71]	61.6 [25.1, 108.6]
STEM	1975-1979	100	5.52 [-1.45, 12.48]	41.6 [-9.9, 122.5]
STEM	1980-1984	322	4.89 [0.99, 8.79]	41.2 [9.7, 81.6]
STEM	1985-1989	635	4.69 [1.71, 7.67]	39.9 [15.4, 69.7]

Continued on next page

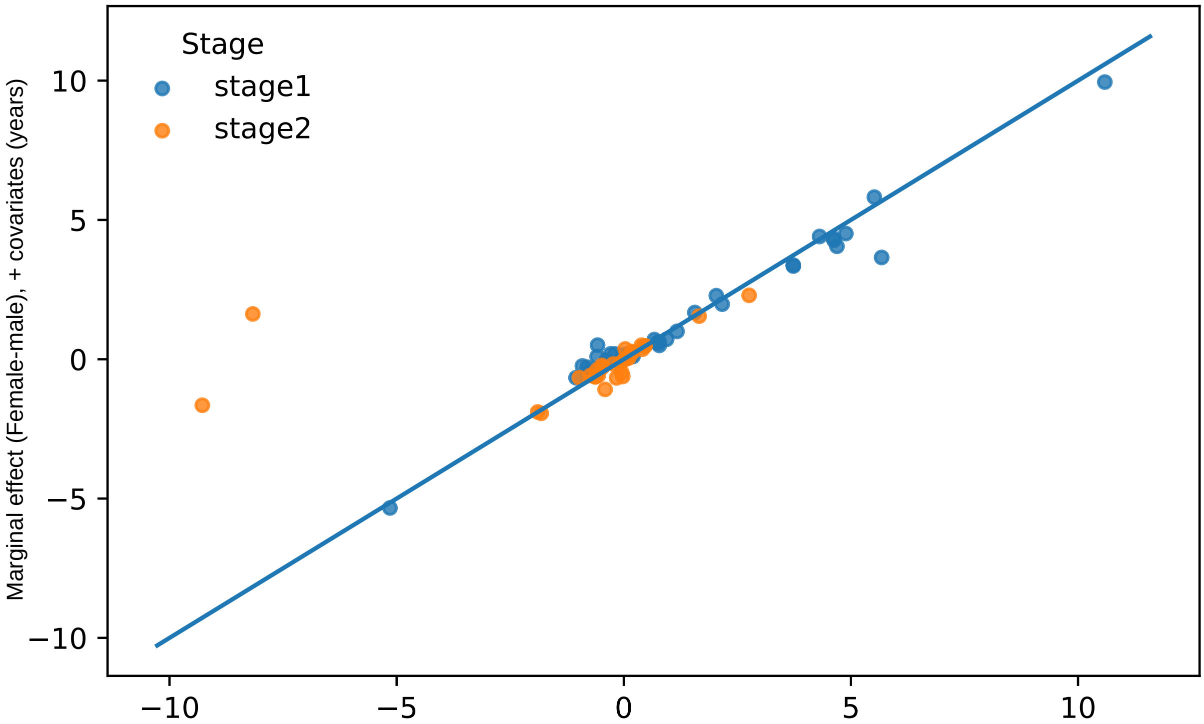
Discipline	Cohort	*n*	OLS: ME (years) [95% CI]	Log-OLS: ME (%) [95% CI]
STEM	1990-1994	935	3.73 [1.72, 5.74]	27.1 [6.6, 51.5]
STEM	1995-1999	1351	0.73 [-0.48, 1.94]	6.2 [-6.0, 20.1]
STEM	2000-2004	1720	1.17 [0.34, 2.00]	8.0 [-1.6, 18.5]
STEM	2005-2009	1634	-0.59 [-1.34, 0.17]	-11.4 [-20.4, -1.4]
STEM	2010-2014	1152	-1.05 [-1.67, -0.44]	-20.9 [-29.3, -11.4]
STEM	2015-2019	659	-0.75 [-1.27, -0.24]	-15.8 [-25.7, -4.5]
STEM	2020-2024	192	-0.58 [-1.06, -0.09]	-22.2 [-36.5, -4.6]
Interdisciplinary				
Interdisciplinary	1975-1979	85	10.59 [2.16, 19.02]	108.6 [33.4, 226.3]
Interdisciplinary	1980-1984	165	4.62 [0.75, 8.48]	48.5 [13.5, 94.4]
Interdisciplinary	1985-1989	313	2.17 [-0.80, 5.14]	18.9 [-8.8, 55.0]
Interdisciplinary	1990-1994	534	2.04 [0.13, 3.95]	28.6 [1.8, 62.6]
Interdisciplinary	1995-1999	819	-0.01 [-1.73, 1.71]	-8.0 [-25.9, 14.3]
Interdisciplinary	2000-2004	1069	-0.29 [-1.32, 0.75]	-5.4 [-17.4, 8.4]
Interdisciplinary	2005-2009	969	-0.91 [-1.74, -0.08]	-13.0 [-23.3, -1.3]
Interdisciplinary	2010-2014	632	-0.39 [-1.23, 0.45]	-8.2 [-20.8, 6.3]
Interdisciplinary	2015-2019	367	-0.81 [-1.47, -0.14]	-17.4 [-30.0, -2.4]
Interdisciplinary	2020-2024	83	0.20 [-0.58, 0.98]	4.7 [-23.6, 43.5]
Stage 2				
Humanities				
Humanities	1970-1974	7	-8.17 [-11.32, -5.01]	-87.9 [-91.7, -82.3]
Humanities	1975-1979	19	-9.28 [-12.19, -6.37]	-86.6 [-91.5, -78.9]
Humanities	1980-1984	38	-0.41 [-3.44, 2.61]	12.4 [-16.2, 50.7]
Humanities	1985-1989	109	2.76 [-0.62, 6.13]	34.6 [3.8, 74.7]
Humanities	1990-1994	221	0.39 [-1.50, 2.28]	0.4 [-15.5, 19.2]
Humanities	1995-1999	354	0.03 [-1.25, 1.31]	-2.0 [-16.6, 15.2]
Humanities	2000-2004	560	0.07 [-0.78, 0.92]	2.0 [-6.6, 11.4]
Humanities	2005-2009	804	0.19 [-0.41, 0.80]	1.4 [-5.6, 8.8]
Humanities	2010-2014	777	0.17 [-0.27, 0.62]	1.4 [-5.1, 8.3]
Humanities	2015-2019	442	0.46 [0.06, 0.85]	7.5 [-0.1, 15.6]
Humanities	2020-2024	58	-0.02 [-0.58, 0.53]	3.4 [-13.9, 24.2]
STEM				
STEM	1985-1989	119	1.66 [-6.29, 9.61]	16.9 [-28.0, 89.9]
STEM	1990-1994	284	0.41 [-2.55, 3.37]	9.3 [-19.4, 48.3]
STEM	1995-1999	530	-0.07 [-1.93, 1.78]	3.5 [-12.8, 22.9]
STEM	2000-2004	706	-0.74 [-2.09, 0.62]	-6.9 [-20.9, 9.6]
STEM	2005-2009	972	-0.47 [-1.35, 0.41]	-7.2 [-17.6, 4.6]
STEM	2010-2014	922	-0.60 [-1.24, 0.05]	-11.0 [-20.4, -0.5]
STEM	2015-2019	618	0.11 [-0.39, 0.61]	2.2 [-7.6, 13.0]
STEM	2020-2024	150	-0.04 [-0.44, 0.36]	3.0 [-9.5, 17.3]
Interdisciplinary				
Interdisciplinary	1985-1989	78	-1.90 [-2.95, -0.84]	-7.5 [-17.7, 3.9]
Interdisciplinary	1990-1994	225	-1.82 [-4.15, 0.52]	-10.6 [-28.2, 11.5]
Interdisciplinary	1995-1999	345	-0.98 [-3.15, 1.19]	-11.8 [-33.5, 17.1]
Interdisciplinary	2000-2004	542	-0.56 [-2.09, 0.97]	-7.4 [-22.4, 10.5]
Interdisciplinary	2005-2009	655	-0.48 [-1.62, 0.66]	-10.3 [-24.0, 5.9]

Continued on next page

Discipline	**Cohort**	n	**OLS: ME (years) [95% CI]**	**Log-OLS: ME (%) [95% CI]**
Interdisciplinary	2010-2014	576	0.47 [-0.28, 1.22]	8.5 [-1.9, 20.0]
Interdisciplinary	2015-2019	340	-0.23 [-0.96, 0.49]	-2.0 [-15.1, 13.1]
Interdisciplinary	2020-2024	76	-0.16 [-0.79, 0.47]	-0.1 [-19.9, 24.7]

Notes. Table S8 compares estimator choices for Model 2 (female, cohort-bin, and female×cohort interactions) estimated within each discipline. "OLS" reports marginal female effects (Female−Male) in years. "Log-OLS" estimates the same model on $\log(1+\text{duration})$ and is reported as a percent difference, computed as $100 \times (\exp(\widehat{ME}) - 1)$ with confidence intervals transformed accordingly. Robust (HC1) standard errors are used throughout.

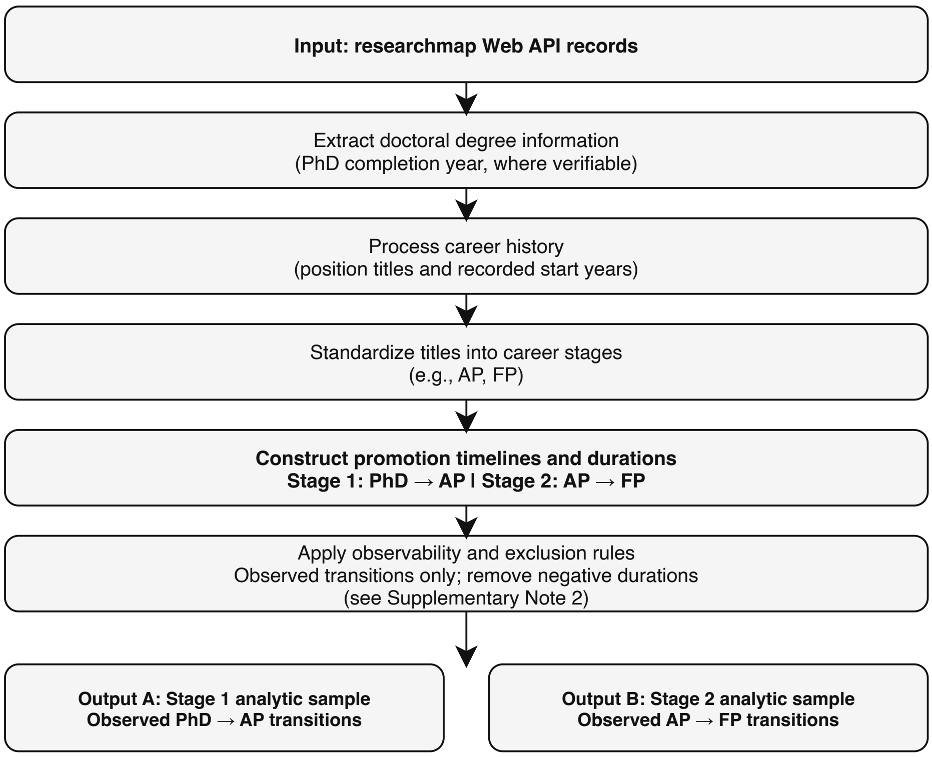
Supplementary Figure S4. Estimator sensitivity (OLS vs. log-OLS).



Notes: The figure compares cohort-specific marginal female effects estimated from level-based OLS (years) to those estimated from log-OLS on $\log(1 + \text{duration})$ (percent difference). Both specifications use Model 2 with HC1 robust standard errors; log-scale effects are exponentiated and expressed as percent differences.

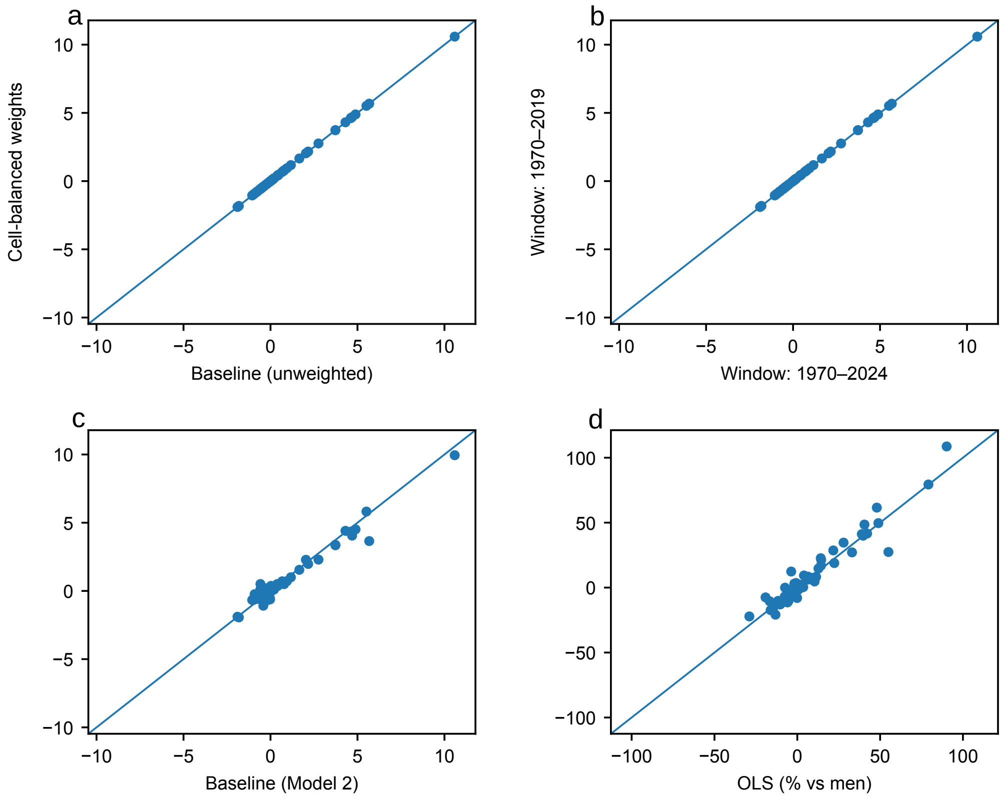
2 Supplementary Figures

Supplementary Figure S5. Data processing workflow and observability rules for constructing promotion durations.



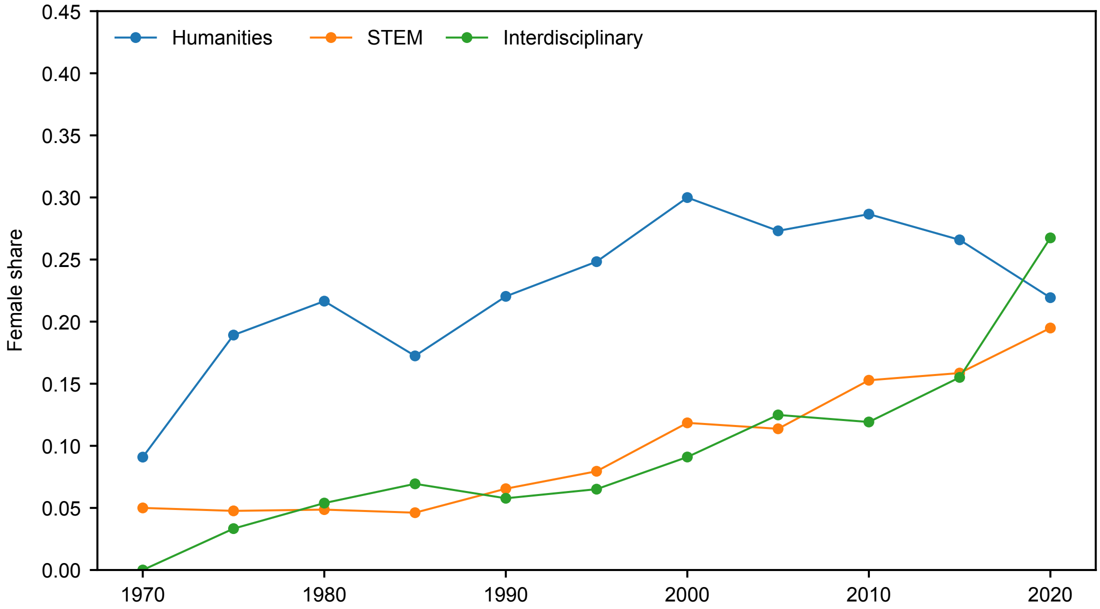
Notes: The figure summarizes the sequential steps used to define stage-specific analytic samples and to compute promotion durations from standardized job-history records extracted via the researchmap Web API. Stage 1 requires an observed (verifiable) PhD completion year followed by an observed AP start year; Stage 2 requires an observed AP start year followed by an observed FP start year. In both stages, we include *observed transitions only* (i.e., the absence of a recorded endpoint is treated as unobserved rather than as non-promotion). Promotion durations are measured as year differences between adjacent recorded start years; logically inconsistent negative durations are excluded.

Supplementary Figure S6. Robustness of cohort-specific gender effects across alternative specifications.



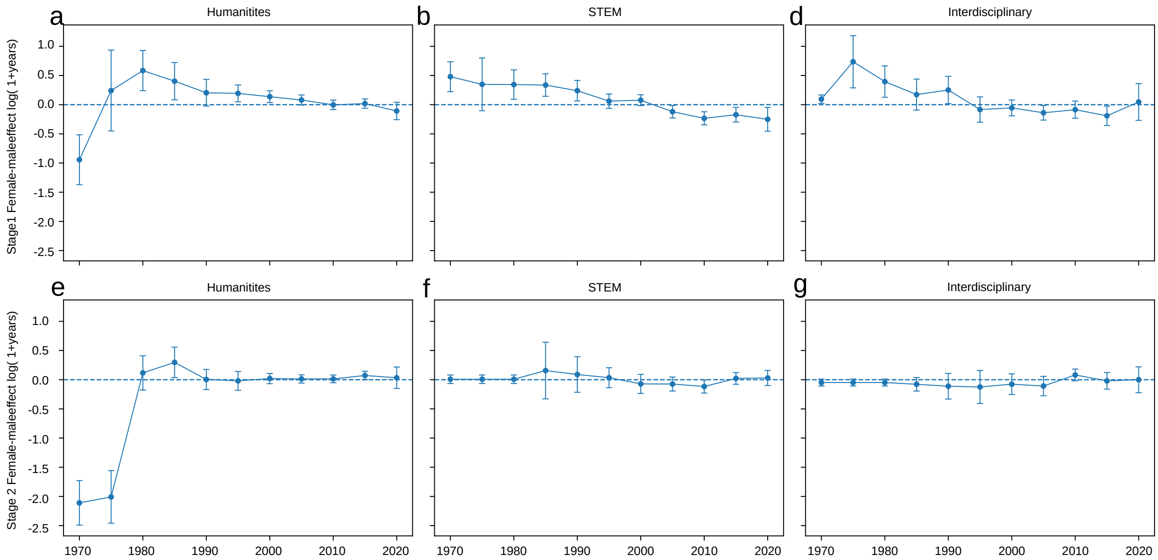
Notes: The figure compares cohort-specific marginal effects of gender (Female–Male) estimated using Model 2 under the baseline setup versus alternative specifications described in Supplementary Note 4. Panel **a** contrasts unweighted OLS and cell-balanced weighting; Panel **b** contrasts the full cohort window (1970–2024) and the restricted window (1970–2019); Panel **c** contrasts the baseline model and covariate-augmented models adding mobility/path controls (including same-year indicators); Panel **d** contrasts effect scales by expressing OLS effects as percentage differences relative to the male mean within each cohort-by-discipline cell and log-OLS effects as percent changes in $(1+\text{duration})$, computed as $100 \times (\exp(\widehat{ME}) - 1) \times \frac{m+1}{m}$, where m is the male mean duration in the corresponding cohort-by-discipline cell. Points correspond to cohort-by-discipline cells; the diagonal line indicates equality between specifications. Open circles denote sparse cells ($n < 20$). Panels **a–b** lie close to the diagonal because cohort-specific female–male contrasts are defined within cohort bins, making them largely invariant to reweighting or to the exclusion of later cohorts when the compared bins are shared.

Supplementary Figure S7. Female representation by PhD cohort across aggregated discipline groups.



Notes: The figure reports the share of women among PhD holders within each five-year cohort bin, stratified by aggregated discipline group (Humanities, STEM, Interdisciplinary). Female shares are computed at the individual level using unique researcher IDs (deduplicated across stages). This cohort composition provides the discipline-resolved denominator for stage-entry and promotion-duration analyses reported in the main text and in Supplementary Notes.

Supplementary Figure S8. Cohort-specific gender effects by discipline for Stage 1 and Stage 2 under the log-OLS specification.



Notes: The figure shows cohort-specific marginal effects of gender (Female−Male) estimated from Model 2 using the log-transformed outcome $\log(1+\text{duration})$, shown separately for Stage 1 and Stage 2 and stratified by discipline (Humanities, STEM, Interdisciplinary). Points indicate estimated marginal effects by five-year cohort bin; vertical bars indicate 95% confidence intervals based on HC1 robust covariance. Although the analytic window spans PhD cohorts from 1970 to 2024, later cohorts are not shown because limited exposure time prevents the estimation of cohort-specific effects for observed Stage 1 or Stage 2 transitions. Log-scale effects are exponentiated and reported as percent changes in $(1+\text{duration})$, computed as $100\times(\exp(\widehat{ME})-1)$ with confidence intervals transformed accordingly.

3 Supplementary Tables

Supplementary Table S13. Promotion durations by gender and discipline for Stage 1 and Stage 2.

	Female			Male		
Discipline group	*n*	Mean (SD)	Median [IQR]	*n*	Mean (SD)	Median [IQR]
Panel A. Stage 1, observed transitions (N = 19,568)						
Overall	2,929	7.446 (5.503)	6.000 [7.000]	16,639	8.449 (5.326)	8.000 [7.000]
Humanities	1,563	6.003 (4.478)	5.000 [5.000]	4,210	5.398 (4.087)	5.000 [6.000]
STEM	883	9.660 (6.389)	9.000 [8.000]	7,853	9.902 (5.426)	9.000 [7.000]
Interdisciplinary	483	8.066 (5.308)	7.000 [7.500]	4,576	8.762 (4.997)	8.000 [7.000]
Panel B. Stage 2, observed transitions (N = 10,638)						
Overall	1,490	8.095 (4.270)	8.000 [5.000]	9,148	8.464 (4.580)	8.000 [6.000]
Humanities	847	8.591 (4.214)	8.000 [5.000]	2,542	8.535 (4.276)	8.000 [5.000]
STEM	406	7.126 (4.274)	6.500 [6.000]	3,954	8.309 (4.749)	8.000 [6.000]
Interdisciplinary	237	7.979 (4.186)	7.000 [5.000]	2,652	8.626 (4.600)	8.000 [6.000]

Notes. Promotion durations are measured in years as differences between adjacent observed start years.Stage 1 includes individuals with an observed PhD completion year followed by an observed AP start year; Stage 2 includes individuals with an observed AP start year followed by an observed FP start year. Only observed transitions are included; the absence of a recorded endpoint is treated as unobserved rather than as non-promotion. Logically inconsistent negative durations are excluded. “Other” is a residual discipline category and is not displayed in the main rows.

Supplementary Table S14. Coverage rates for key variables under the all-researcher denominator and stage-observed subsamples.

Variable	Overall (%)	Humanities (%)	STEM (%)	Interdisciplinary (%)
Panel A. All researchers				
Gender available	100.00	100.00	100.00	100.00
Discipline group available	99.63	100.00	100.00	100.00
Verifiable PhD completion year	100.00	100.00	100.00	100.00
AP start year available	26.74	34.68	21.55	30.70
FP start year available	20.04	24.08	16.14	25.13
Stage 1 duration available	26.74	34.68	21.55	30.70
Stage 2 duration available	13.83	17.91	10.61	17.21
Panel B. Observed Stage 1 transitions				
Gender available	100.00	100.00	100.00	100.00
Discipline group available	99.63	100.00	100.00	100.00
Verifiable PhD completion year	100.00	100.00	100.00	100.00
AP start year available	100.00	100.00	100.00	100.00
FP start year available	51.72	51.63	49.25	56.05
Stage 1 duration available	100.00	100.00	100.00	100.00
Stage 2 duration available	51.72	51.63	49.25	56.05
Panel C. Observed Stage 2 transitions				
Gender available	100.00	100.00	100.00	100.00
Discipline group available	99.57	100.00	100.00	100.00
Verifiable PhD completion year	100.00	100.00	100.00	100.00
AP start year available	100.00	100.00	100.00	100.00
FP start year available	100.00	100.00	100.00	100.00
Stage 1 duration available	100.00	100.00	100.00	100.00
Stage 2 duration available	100.00	100.00	100.00	100.00

Notes. Coverage is defined as the percentage of non-missing values within the panel denominator. Panel A includes all researchers in the API pull. Panel B restricts to researchers with an observed Stage 1 transition. Panel C restricts to researchers with an observed Stage 2 transition. Discipline grouping follows the mapping used throughout the analyses: single-field entries are classified as Humanities or STEM, whereas multi-field entries are classified as Interdisciplinary.

4 Supplementary References

References

[1] Sīle, L., Guns, R., Vandermoere, F., Sivertsen, G., & Engels, T. C. Tracing the context in disciplinary classifications: A bibliometric pairwise comparison of five classifications of journals in the social sciences and humanities. *Quantitative Science Studies* **2**, 65–88 (2021).

[2] Goyanes, M., Demeter, M. & Ochsner, M. Gender detection tools and the gender gap in a large scientific sample. *Scientometrics* (2024). https://doi.org/10.1007/s11192-024-05010-0

[3] Krause, L. *et al.* Confidently Wrong: LMs Are Not Calibrated for Multilingual Classification. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2023). https://aclanthology.org/2023.emnlp-main.230/

[4] Zhang, M. *et al.* Calibrating the Confidence of Large Language Models by Eliciting Fidelity. *Preprint at arXiv:2404.02655* (2024). https://arxiv.org/abs/2404.02655

[5] Dylman, A. S. & Kikutani, M. The role of semantic processing in reading Japanese orthographies: an investigation using a script-switch paradigm. *Reading and Writing* **31**, 503–531 (2018). https://doi.org/10.1007/s11145-017-9796-3

[6] Barešová, M. *et al.* Gender-specific features in contemporary Japanese names. *Names: A Journal of Onomastics* (2024).