This folder contains all the data and program files required to reproduce the results in the main body of the manuscript, *"Death of the Salesman, or the Salesforce?"*

Important note: GitHub does not allow us to upload large data files. Use this dropbox link to download the data, This may turn out to not be stable. If you cannot download the data this way, e-mail us at pazoulay@mit.edu.

xtqmlp is a modification of the built-in Stata command, **xtpoisson, fe** that allows the clustering of the standard errors at a coarser level than the panel id variable (in the case of our manuscript, we cluster the errors around each star scientist, uniquely identified by the variable `setnb`). The reader should place `xtqmlp.ado` in the personal subfolder within his/her own ado folder.[1]

Our replication archive contains six folders named `data`, `do-files`, `estimates`, `figures`, `raw_data`, and `tables`.

Let us begin in the dataset contained in the data folder. The Stata file **afterlife_dd_dataset.dta** contains the data at the article-year level, which can be used to replicate all of the results in the paper that leverage our difference-in-differences research design (Tables 3, 4, 5, 6, and 9; Figure 1, 3, and 4). It is created by a series of do files that combine publicly available data (e.g., PubMed), commercial data (e.g., Web of Science), and confidential data (specifically, the AAMC Faculty Roster) to build the final analytic sample. A complete data dictionary is included at the end of this document.

Note that we cannot make publicly available all the raw and intermediate files that combined allow us to construct **afterlife_dd_dataset.dta**, for two reasons. First, we assembled the superstar sample (N=13,426) from original CVs guaranteeing anonymity to our respondents. Second, we have licensed the Faculty Roster of the Association of American Medical Colleges, the main industry group for US Medical Schools under strict conditions of confidentiality. The public release dataset uses the AAMC data only in very aggregate form. Any researcher wishing to replicate our results from the ground up, however, should contact Dr. Hershel Alexander, Director of Data Operations and Services (e-mail: halexander@aamc.org). The licensing of AAMC data to other researchers, for the purposes of replication or original research, is entirely at the discretion of AAMC. However, researchers interested in examining our code should contact us.

The file **afterlife_dead_stars_memories.dta** is necessary to produce the descriptive statistics in Table 1. The file **afterlife_mediation_analysis.dta** contains all the data at the article level to reproduce Tables 7 and 8, as well as Figure 5.

---

[1] Our do-files make use of a number of packages available on the statistical software component archive: Ben Jann's `coefplot` to generate figures from estimated coefficients, as well as his `estout` series of packages (which also include `esttab`, `estpost`, and `estadd`) to prepare table output; Austin Nichols' `byhist` package to produce interlaced histograms; Gary Longton and Nicholas Cox's `distinct` to report the number of distinct observations.

Finally, the files **afterlife_prdct_cites_invst_level.dta** and **afterlife_prdct_cites_pmid_level.dta** contain the predicted posthumous citations at the scientist and the investigator level. These files can be generated by running the do file `cmpt_inxss_cites.do`. This is not strictly required to replicate the results however, since the file **afterlife_mediation_analysis.dta** contains the covariates corresponding to these predictions already. We chose to make these available so that interested readers could inspect (and possibly improve upon) the penalized Poisson procedure that generates the predictions.

In addition to the data, this folder contains the following subfolders: (1) `do-files`, which contains a set of Stata do-files needed to replicate every table or figure[2] in the paper (one do-file per table or figure); (2) `estimates`, which contains, for each statistical model, the stored estimates in a .ster file (convenient because the fe Poisson models sometimes take a long time to run); (3) `figures`, which contains the Stata graphs generated by some of the do-files; and (4) `tables`, which contains the tables generated by the do-files.

Finally, the subfolder `raw_data` contains three excel spreadsheets. `dead_stars.xls` contains the information presented in Appendix F, i.e., the comprehensive list of 720 deceased superstars. `dead_stars_memrlzd.xlsx` contains the comprehensive set[3] of memory events for the 720 prematurely deceased superstars. `academic_recognition_events.xls` contains the set of 5,850 academic recognition events for the deceased as well as the 8,326 control stars.

Note that every do-file references a working directory. Interested researchers should of course substitute the directory path that works on their machine (by providing it as a value for the global macro `{F10}` in their profile.ado file). We have made use of forward slashes in the directory structure, which should make these files portable to Windows, MacOS, and Linux operating systems.

[2] An exception is Figure 1A. Replication here necessitates access to the entire list of 14,326 superstars. We cannot post this list publicly. However, interested researchers should contact us if they would like to replicate this particular figure.

[3] By comprehensive we mean academic as well as non-academic memories, e.g. *Wikipedia* pages, *New York Times* obits, etc.

Data Dictionary

afterlife_dd_dataset.dta [unique key is srce_pmid/death_year/setnb/year, 13,087,941 article-years]

Variable Name	Definition
id	Article identifier (srce_pmid/death_year/setnb)
xid	Strata identifier
srce_pmid	Source article pmid
srce_pubyear	Source article year of publications
year	Focal calendar year
yr	Focal calendar year, slightly coarsened (the years 1950-1955 are collapsed)
setnb	Star identifier
star_yob	Star's year of birth
star_yod	Star's year of death
female	Star's gender
deg	Star's degree
deg_year	Star's highest degree year
aad	Birth age at death/counterfactual death
age_at_death	Career age at death/counterfactual death
nas	Star is a member of the National Academy of Sciences in year of death/counterfactual death
stk_inv_pubs	Star's cumulative nb. of publications in year of death
stk_inv_cites	Star's cumulative nb. of citations in year of death
stk_inv_nih	Star's cumulative NIH funding in year of death
stk_trainees	Star's cumulative nb. of trainees in year of death
stk_nwghtd_collabs	Star's cumulative nb. of coauthors (excluding trainees) in year of death
frac1_slf_yprmt_setnb	Star's fraction of unrelated self-citations as a % of all citations (self-promotion index), at time of death
index_pubs	Star's cumulative nb. of pubs., 20 ventiles
index_amount	Star's cumulative NIH funding, 20 ventiles
index cites	Star's cumulative nb. of citations, 20 ventiles
index_stk_trainees	Star's cumulative nb. of trainees, 20 ventiles
index_stk_collabs	Star's cumulative nb. of coauthors (excluding trainees), 20 ventiles
index_slf_prmtr1_setnb	Star's self-promotion index, 20 ventiles
cod	Cause of death

journal_name	Source article journal
jif	Source article journal impact factor
nbauthors	Source article nb. of authors
treat	Article associated with deceased star
death_year	Year of death (or counterfactual death)
after_death	After death for treated articles
after_death_cmn	After death for treated and control articles
artcl_age	Article age (in years) in focal year
nb_cites	Nb. of citations in focal year (excluding self-citations)
nb_cites_ycoauthor	Nb. of citations from coauthors in focal year (excluding self-citations)
nb_cites_ncoauthor	Nb. of citations from non-coauthors in focal year (excluding self-citations)
nb_cites_yfield	Nb. of citations in same subfield as the source in focal year (excluding self-citations)
nb_cites_nfield	Nb. of citations not in same subfield as the source in focal year (excluding self-citations)
nb_cites_ycoloc	Nb. of citations from same institution in focal year (excluding self-citations)
nb_cites_ncoloc	Nb. of citations not from same institution in focal year (excluding self-citations)
nb_cites_ncmmrlzr	Nb. of citations in focal year, excluding those from coauthors and memorializers
nb_cites_ncmmrlzr5	Nb. of citations in focal year, excluding those from coauthors and memorializers and 5 years post-death window
rltv_rank	Source article citation rank within the star's corpus at time of death
yrltv_btm10pct	Source article is in the bottom 10% of citations within the star's corpus at time of death
yrltv_mdl25_75pct	Source article is in the middle quartiles of citations within the star's corpus at time of death
yrltv_top10pct	Source article is in the top 10% of citations within the star's corpus at time of death
nrltv_top5pct	Source article is in the top 5% of cites in the *PubMed*/*WoS* universe at time of death
nrltv_top1pct	Source article is in the top 1% of cites in the *PubMed*/*WoS* universe at time of death

afterlife_dd_dataset.dta [unique key is setnb, 720 deceased scientists]

Variable Name	Definition
setnb	Star identifier
itrmrl_nih	Star is an NIH intramural scientist
total_mem	Total number of memory events for the deceased star
academic_mem	Total number of academic memory events for the deceased star (whether or not PubMed indexed)
nyt_obit	New York Times wrote an obit for the star
wikipedia_page	Star has a Wikipedia page
named_award	Star has an eponymous award
festschrift	A festschrift was organized for the star
xpctd_lasso_all_ntrt_cites	Cumulative predicted posthumous citations for the star (using penalized Poisson procedure)
inxss_lasso_all_ntrt_cites	Cumulative "excess" posthumous citations for the star (using penalized Poisson procedure)

afterlife_mediation_analysis.dta [unique key is id=pmid/death_year/setnb, 481,746 articles]

Variable Name	Definition
id	Article identifier
xid	Article strata
pmid	PubMed identifier for article (can be repeated because an article can serve as control in distinct death years)
setnb	Star identifier
star_yob	Star's year of birth
star_yod	Star's year of death
srce_pubyear	Source article year of publications
treat	Article associated with deceased (treated) star
death_year	Year of death (or counterfactual death)
cod	Cause of death
deg	Star's degree
deg_year	Star's highest degree year
female	Star's gender

stk_inv_pubs	Star's cumulative nb. of publications in year of death
stk_inv_cites	Star's cumulative nb. of citations in year of death
stk_inv_nih	Star's cumulative NIH funding in year of death
itrmrl_nih	Star is an NIH intramural scientist
log_cites	Star's cumulative nb. of citations in year of death (logged)
log_pubs	Star's cumulative nb. of publications in year of death (logged)
log_funding	Star's cumulative NIH funding in year of death (logged); zero if no funding.
no_funding	Indicator variable if the star has no NIH funding and is not an intramural NIH scientist
nas	Star is a member of the National Academy of Sciences in year of death/counterfactual death
nb_trainees	Star's cumulative nb. of trainees in year of death
nb_coauthors	Star's cumulative nb. of coauthors in year of death
ln_nb_trainees	Star's cumulative nb. of trainees in year of death (logged)
ln_nb_coauthors	Star's cumulative nb. of coauthors in year of death (logged)
ln_nb_coauthors_nt	Star's cumulative nb. of coauthors—excluding trainees—in year of death (logged)
zero_trainees	Star does not have any recorded trainee
zero_collabs	Star does not have any recorded coauthor (possible because only AAMC coauthors are counted)
zero_collabs_nt	Star does not have any recorded non-trainee coauthor (possible because only AAMC coauthors are counted)
acd_mem_ttl	Total number of recognition events
acd_mem_vdplus	Total number of recognition events, post-death/counterfactual death
acd_mem_vdwthn5	Total number of recognition events, within a window of five years after death/counterfactual death
sudden_death	Star death was sudden
antcpt_death	Star death was anticipated
nknwn_death	Star's cause of death is unknown
frac1_slf_yprmt_setnb	Star's fraction of unrelated self-citations as a % of all citations (self-promotion index), at time of death
index_slf_yprmtr1_setnb	Star's self-promotion index, 20 ventiles
top_prmtr1_setnb	Star has a self-promotion index above the median
xpctd_lasso_all_ntrt_cites	Article predicted posthumous citations for the star
inxss_lasso_all_ntrt_cites	Article "excess" posthumous citations for the star
xpctd_lasso_nwndw5_ntrt_cites	Article predicted posthumous citations for the star, excluding five years post-death window
inxss_lasso_nwndw5_ntrt_cites	Article "excess" posthumous citations for the star, excluding five years post-death window
xpctd_lasso_ncmmrlzr_ntrt_cites	Article predicted posthumous citations for the star, excluding those from coauthors and memorializers
inxss_lasso_ncmmrlzr_ntrt_cites	Article "excess" posthumous citations for the star, excluding those from coauthors and memorializers
xpctd_lasso_ncmmrlzr5_ntrt_cites	Article predicted posthumous citations for the star, excluding both
inxss_lasso_ncmmrlzr5_ntrt_cites	Article "excess" posthumous citations for the star, excluding both

afterlife_prdct_cites_pmid_level.dta [unique key is id=pmid/death_year/setnb, 481,746 articles]

Variable Name	Definition
id	Article identifier
pmid	PubMed identifier for article (can be repeated because an article can serve as control in distinct death years)
setnb	Star identifier
death_year	Year of death (or counterfactual death)
treat	Article associated with deceased (treated) star
death_year	Year of death (or counterfactual death)
nb_cites_all	Article actual posthumous citations for the star
xpctd_lasso_all_ntrt_cites	Article predicted posthumous citations for the star
inxss_lasso_all_ntrt_cites	Article "excess" posthumous citations for the star
nb_cites_nwndw5	Article actual posthumous citations for the star, excluding five years post-death window
xpctd_lasso_nwndw5_ntrt_cites	Article predicted posthumous citations for the star, excluding five years post-death window
inxss_lasso_nwndw5_ntrt_cites	Article "excess" posthumous citations for the star, excluding five years post-death window
nb_cites_ncmmrlzr	Article actual posthumous citations for the star, excluding those from coauthors and memorializers
xpctd_lasso_ncmmrlzr_ntrt_cites	Article predicted posthumous citations for the star, excluding those from coauthors and memorializers
inxss_lasso_ncmmrlzr_ntrt_cites	Article "excess" posthumous citations for the star, excluding those from coauthors and memorializers
nb_cites_ncmmrlzr5	Article actual posthumous citations for the star, excluding both
xpctd_lasso_ncmmrlzr5_ntrt_cites	Article predicted posthumous citations for the star, excluding both
inxss_lasso_ncmmrlzr5_ntrt_cites	Article "excess" posthumous citations for the star, excluding both

afterlife_prdct_cites_invst_level.dta [unique key is setnb, 720 deceased scientists]

Variable Name	Definition
setnb	Star identifier
death_year	Year of death (or counterfactual death)
nb_cites_all	Cumulative actual posthumous citations for the star
xpctd_lasso_all_ntrt_cites	Cumulative predicted posthumous citations for the star
inxss_lasso_all_ntrt_cites	Cumulative "excess" posthumous citations for the star
nb_cites_nwndw5	Cumulative actual posthumous citations for the star, excluding five years post-death window
xpctd_lasso_nwndw5_ntrt_cites	Cumulative predicted posthumous citations for the star, excluding five years post-death window
inxss_lasso_nwndw5_ntrt_cites	Cumulative "excess" posthumous citations for the star, excluding five years post-death window
nb_cites_ncmmrlzr	Cumulative actual posthumous citations for the star, excluding those from coauthors and memorializers
xpctd_lasso_ncmmrlzr_ntrt_cites	Cumulative predicted posthumous citations for the star, excluding those from coauthors and memorializers
inxss_lasso_ncmmrlzr_ntrt_cites	Cumulative "excess" posthumous citations for the star, excluding those from coauthors and memorializers
nb_cites_ncmmrlzr5	Cumulative actual posthumous citations for the star, excluding both
xpctd_lasso_ncmmrlzr5_ntrt_cites	Cumulative predicted posthumous citations for the star, excluding both
inxss_lasso_ncmmrlzr5_ntrt_cites	Cumulative "excess" posthumous citations for the star, excluding both