Astro2020 White Paper

Long-term Trends in the Astronomical Workforce: Analysis and Recommendations Based on the Publication Histories of 10,000 US Astronomy PhD Recipients

Principal Author:
Name: Ivelina Momcheva
Institution: Space Telescope Science Institute
Email: imomcheva@stsci.edu
Phone: 410-338-6790

Endorsements:
Kevin Flaherty, WIlliams College
Laura Prichard, STScI

Abstract: Measuring and understanding the long-term trends in the astronomical workforce is crucial for driving policies and decisions that will shape the landscape for new researchers in the next decade. Concerns have been raised about the growing number of astronomy PhDs, the lengthening of the postdoc phase, the number of available tenure-track positions and the scarcity of women and minorities in senior positions. These trends impact career expectations and satisfaction, recruiting, hiring, promotion and retention. In this white paper, we present the summary of an analysis of the publication histories of $>$10,000 US PhD recipients since 1970. We summarize the long-term trends and make recommendations on policies that will be helpful in supporting the community in the next decade.

350-character Summary: We present data on long term career tracks of $>$10,000 US astronomy PhD recipients from 1970 to 2018. We determine cohort membership, size and gender breakdown, utilize publication histories to measure retention in scientific research overall and as a function of gender, and examine trends in the type of institutions where astronomers are employed.

1 Introduction

Every year in the US well over 100 new graduate students receive doctoral degrees in astronomy. As each new class enters the work force, they have to navigate a complex landscape of different career tracks, often equipped with only limited and biased information provided by their mentors and peers. In recent years, concerns have been raised about the growing number of astronomy PhDs, the lengthening of the postdoc phase, the number of available tenure-track positions and the scarcity of women and minorities in senior positions (e.g., [1, 2]). Some of these concerns are new but others are experienced anew by each cohort as they need to navigate their job searches. The lack of available data on the long-term employment trends is the source of much anxiety. We believe that by providing more accurate information about the labor market we can allow people to make better and more informed decisions about their lives and careers, and provide factual information for decisions and policies.

In this white paper, we present data on long term career tracks of >10,000 US astronomy PhD recipients from 1970 to 2018. We use information in online databases to determine cohort membership, size and gender breakdown of PhD recipients over 48 years. We further utilize publication histories to measure long-term retention in scientific research overall and as a function of gender. We finally examine trends in the type of institutions that astronomers are employed.

2 Recommendations

Based on the data analysis presented here, we make the following recommendations:

- Real data on long term employment trends should be used in formulating policies and making decisions that affect future employment opportunities in the field. Planned demographic surveys are crucial and should be supplemented with analysis of historical records, online archives and data collected by departments and universities. Data must be analyzed in the context of national and international funding trends, macro-economic conditions and geopolitical forces in order to understand long-term trends in the field.

- The data presented in this paper does not support claims of a recent over-production of astronomy PhDs in the US. It is therefore unnecessary to limit the number of PhDs awarded at US institutions. The annual number of astronomy PhD recipients has more than doubled over the last 45 years, and the job market has expanded accordingly to absorb this increase, resulting in a consistent retention rate of active researchers.

- In order to increase the representation of women at all post-graduate levels of astronomy beyond the current ceiling of 30%, significant work is needed at the transition from high school to college. No leaks in the pipeline have been identified at the undergraduate level and beyond for all women. However, under-represented minorities, and minority women specifically, are still not represented at the same rates through all professional levels and substantial work is needed to advance their equal participation in astronomy.

- More junior female researches should be considered for and involved in activities where seniority is considered important. Due to the slow and gradual increase of the percentage of

women with PhDs in past graduating classes, there are very few senior female astronomers in many sub-fields of research.

- Mentoring and information on multiple career paths should be a key component in the training of all astronomy PhD recipients and postdoctoral researchers in order to prepare them for the variety of skills needed for professional success inside and outside of academia.

This White Paper is a subset of relevant data and conclusions of a more extensive analysis (Momcheva et al., 2019, in prep.)

3 Data

Data for this analysis was collected using the SAO/NASA Astrophysics Data System (ADS[1]) and specifically the application programming interface (API[2]) it offers. We queried the ADS database for all PhD dissertations in astronomy for each year between 1970 and 2018 which yielded a total of $\sim 20,000$ entries. Cohorts were identified based on the calendar year of their dissertation publication and only recipients of PhDs from US institutions were selected. ADS indexes dissertation data provided by ProQuest[3], the major dissertation publisher in the USA which guarantees that the ADS database is largely complete and accurate for US PhD recipients. The final sample contains $\sim 10,000$ PhD recipients. The search results were verified by comparing cohorts from certain years to classes from several departments were we where able to obtain data from staff or public websites. The ProQuest topic designation is done by the author at the time of submission based on a set of categories provided by ProQuest. We further verified that all resulting dissertations were astronomy themed by spot-checking titles and abstracts.

We identified gender of the PhD recipients using the first name of the dissertation author. We utilized the `genederize.io` API[4] to determine the most likely gender. Considering that $> 20\%$ of US PhD recipients are non-US citizens [3], we did not assume a country of origin when determining the gender. The API returned "male", "female" and "unknown" labels and likelihood percentages. Names where only the first initial is used (N=526) were labeled "unknown".

The publication history of each PhD recipient was determined though a call to the ADS API for refereed publications in astronomy starting at five years prior their graduation and ending in the present year using their name as listed on their dissertation. We separately recorded the number of first author and co-authored publications for each year throughout this period. Note that all PhD recipients have at least one first author peer reviewed publication (their dissertation) and that co-authored papers *include* first-author papers. We determined the longevity of a scientific career by looking for the year of the last published paper, either as a first author or as a co-author. Possible effect that could artificially curtail a publication record (e.g., name change) or extend it (e.g., common name leading to failed disambiguation) were not corrected for in the current analysis.

[1] `https://ui.adsabs.harvard.edu/`
[2] `https://github.com/adsabs/adsabs-dev-api`
[3] `https://www.proquest.com/`
[4] `https://genderize.io/`

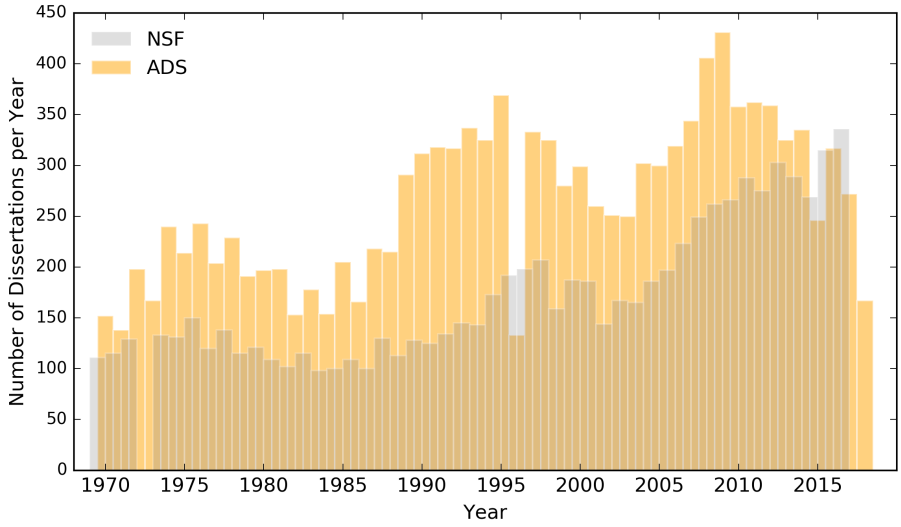

Figure 1: Number of astronomy PhD recipients from US institutions per year from 1970 to 2018 based on the analysis presented here (orange). For comparison we also show the number of astronomy and astrophysics PhDs reported by the NSF Survey of Earned Doctorates (gray, [3]), which uses a different method of classifying degrees into topic areas. The two datasets show similar broad trends but our analysis yields systematically higher cohort sizes.

Our verification and validation methods have found these effects to be small, affecting less that five people per cohort and they are unlikely to significantly change the results.

Finally, we assigned a current institution to each PhD recipient based on the affiliation in their most recent peer-reviewed paper (first author or co-authored). For US institutions, we determined institution type by hand. We grouped the institutions into four different categories: (1) R1 institutions; (2) R2 institutions, liberal arts colleges and other degree-granting institutions; (3) private companies and non-profits; and (4) national labs and observatories. For the first two classes we used the 2018 Carnegie Classification of Institutions of Higher Education[5]. Since this assignment is manual and time-consuming, we only classified institutions for the cohorts of 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005 and 2010.

4 Analysis

4.1 Annual PhD Production Rate

A total of >10,000 PhDs in astronomy have been granted by US institutions between 1970 and 2018. In Figure 1 we show the number of PhD recipients in astronomy from US institutions over these 48 years (orange). The overall trend shows an increase in the annual number of PhD recipients. In the early 1970s, only 150 PhDs were awarded per year while in the late 2000s

[5] https://en.wikipedia.org/wiki/List_of_research_universities_in_the_United_States

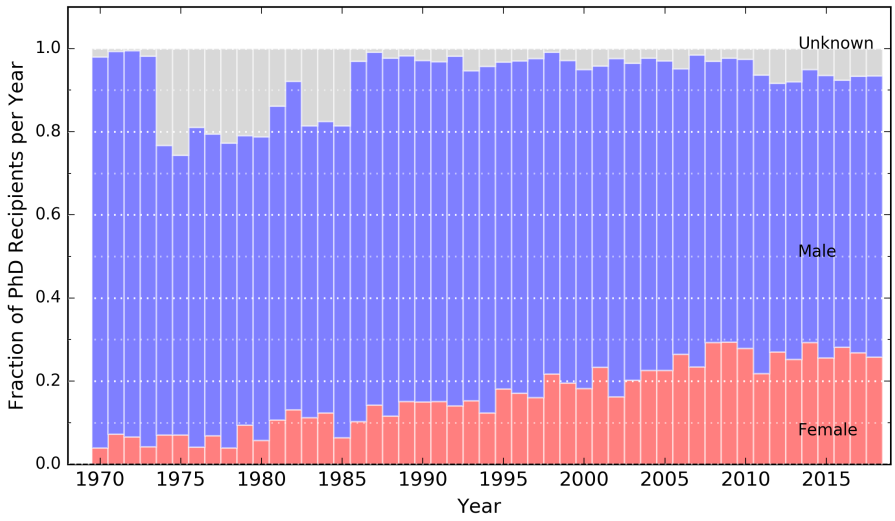

Figure 2: Fraction of PhD recipients identified as male (blue) or female (red) based on their first names for each year from 1970 to 2018. First initials and ambiguous names are considered unknown.

and early 2010s the numbers reached 350 per year with notable peaks in 2009 and 2010 of over 400. The dip in 1996 is due to an incompleteness in the ADS database and the decline in the last three years is likely due to a lag in database updates. Overall, we find a factor of 2.3 increase in the PhD production rate over the period. The trend is clearly non-monotonic and there are large scale increases *and decreases*, likely driven by overall economic factors and changes in academic funding among other complex influences. A better understanding of the trends that drive these changes is important but outside the scope of the current paper.

As a sanity check on the overall numbers, we compare our sample to the data from the NSF Survey of Earned Doctorates (SED, [3]; Figure 1, gray). SED is an annual census conducted since 1957 of all individuals receiving a research doctorate from an accredited U.S. institution in a given academic year. The overall trends between the two samples are quite similar and they exhibit the same temporal variations but the NSF numbers are systematically lower. The label assignment in SED is done on the basis of the degree (PhD in Physics vs. PhD in Astronomy) rather than the topic of research therefore there it is likely that some astronomy PhDs granted from physics departments are not counted as astronomy. This means that the higher PhD production rates from the ADS/ProQuest analysis likely offer a more complete accounting of all PhD recipients who contend for jobs on the astronomy job marketplace and PhD production rates are in fact higher than previously believed.

4.2 Trends with Gender

In Figure 2 we show the fraction of male/female/unknown PhD recipients for each cohort based on the first name analysis. A gradual increase in the fraction of women throughout this period is evident, starting at 5% in the 1970s and peaking at 30% in the late 2010s. In fact, 2008 is the first

year when more than 100 women were granted PhDs from US institutions. However, a notable plateau in the fraction of female graduates is observed in the last 15 years. A comprehensive report on the career tracks of the representation and participation of women in astronomy based on data from the Statistical Research Center at the American Institute of Physics [4] points out that from 2007 to 2017 the percentage of women earning bachelor's degrees in physics and astronomy stagnated at $\sim 20\%$ (their Figure 6). Following a delay of several years, we are likely seeing this stagnation reflected in the fraction of female PhD recipients in astronomy. This suggests that efforts to further increase the representation of women in astronomy must focus on the undergraduate level and on the transition between high-school and undergraduate enrollment in physics and astronomy.

4.3 Active Research Retention Rate

One of the main goals of this analysis is to determine the fraction of astronomy PhD recipients who stay in academia long term. A major concern shared by many recent PhD recipients is that while the number of PhD recipients has increased, the number of available academic positions has seemingly remained constant, therefore increasing competition and increasing the fraction of people who leave the field. Furthermore, concerns have been raised whether female PhD recipients are hired at the same rates as male candidates (e.g., [1, 2]).

We use the publication records of PhD recipients in each cohort to determine the rate at which they leave active research. We determine the year when someone leaves by the date of their last first-author publication or the date of their last co-authored publication. We find that the distribution of last-paper-published dates for each cohort is bi-modal: it exhibits one peak within three years of graduation and a second peak near the present day. The first peak is people leaving active research almost immediately after graduation and the second peak is people who are still actively involved in research. We use these peaks to determine the fraction of people who leave or stay in research. This analysis is limited to the years from 1970 to 2010 because for more recent cohorts we cannot separate the two peaks sufficiently well.

Using first author publications as a measure of active research, we find that $\sim$40% of researchers leave active research within 3 years of receiving a PhD (Figure 3, top left). However, using all co-authored publications as a measure, the fraction is much lower - only $\sim$20% of PhD recipients leave active research within 3 years (Figure 3, lower left). *In total, $\sim$80% of PhD recipients stay in the field at least three years past graduation.* The fraction of PhD recipients who leave active research has stayed strikingly constant over a 40 year stretch. Most notable, the results based on co-authored publications indicate that the fraction of PhDs ceasing research shortly after graduation is only $\sim$20%, with no sharp rise for the most recent cohorts. The results for female PhD recipients show more stochasticity due to the low number statistics but on average they match the trends for male PhD recipients, i.e., women and men leave active research at similar rates.

We also determine the fraction of each cohort that is still scientifically active based on first-authored (Figure 3, top right) or co-authored (Figure 3, lower right) refereed publications within the last three years. Between 20% (1970s cohorts) and 40% (2000s cohorts) of PhD recipients are still involved in research as first authors and between 40% (1970s cohorts) and 70% (2000s cohorts) are still involved in research as co-authors. The increase is gradual and implies that cohorts transition away from active research at a low rate over time: 0.5% of each cohort stops publishing first author paper each year and 0.75% of each cohort stops publishing co-authored papers each year. Again, the trends are similar for male and female PhD recipients and no notable

drop in retention is seen for the most recent cohorts.

Finally, a common analysis for this type of data is to determine the survival rate for cohorts relative to a fixed point in time [6]. In Figure 4 we show the fraction of the PhD recipients who remain scientifically active (measured by all their co-authored papers) as a function of the number of years since PhD for the 1970 to 2010 cohorts in our analysis. We find that all cohorts follow a similar trend with most cohorts preserving more than 50% of their original membership even 20 years post-graduation. A similar analysis performed on the astronomical literature by [6] found a dramatic shortening of careers in astronomy (their Figure 4) which we find no evidence for. The difference is because here we only consider PhD recipients rather than all unique authors. In recent years, programs such as Research Experience for Undergraduates (REU) have allowed many more undergraduates to carry out research that leads to publications (and co-authorship). Very few of them stay in the field long term which is likely to drive the trends observed in [6].

We note that this analysis does not tell us anything about the career tracks taken by people who leave active research. Further research will be needed to understand their trajectories. Some PhD recipients are likely to be in academic careers not captured in this analysis. This is therefore a lower limit on the fraction of people who remain employed within the field. For comparison, a similar analysis was carried out by [2] for a smaller sample of 1063 PhD recipients over 13 years from 28 US departments where the authors did extensive online searches to determine the employment histories of people in their sample. They find that 65% of men and 66% of women find long term employment in astronomy, very much in line with our findings, indicating that we are unlikely to be missing a significant population.

4.4 Where Are They Now?

Combining the two trends observed above - an overall increase in the production of PhDs and a steady fraction of people involved in active research - requires that a substantial number of new positions for astronomy professionals must have become available over time. In this section we examine the current affiliations of PhD recipients who are still scientifically active, based on co-authored publications within the last three years. In Figure 5 we show the break-down of the most recent affiliations of all currently active researchers. As described above, the analysis was done only for the cohorts of 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005 and 2010, and we interpolate between these to fill the gaps.

If, for simplicity, we ignore the movement of people between institutions over time, we can interpret Figure 5 as an annual hiring rate for different institution type. Keep in mind that Figure 5 does not show a sequence in time but rather the present affiliations of **all** currently-active PhD recipients. The total number of PhD recipients in each type of institution is the integral of the corresponding area of the plot. The most significant increases in Figure 5 are in the numbers of researchers being hired at R1 and non-US institutions. In fact, the fraction of each cohort that is currently at R1 institutions has stayed fairly constant over time - approximately 25% of researchers who remain in the field are currently at R1 institutions. The hiring rates of national labs/observatories and non-R1 degree-granting institutions have also increased over time though they represent a smaller share of the job market. The number of positions at all types of institutions has clearly increased over time dispelling beliefs that the number of jobs is constant. The increase in non-US institutions since the mid 1980s is also quite striking and likely driven by the major geopolitical changes in the last three decades, underscoring that the scientific endeavour is very

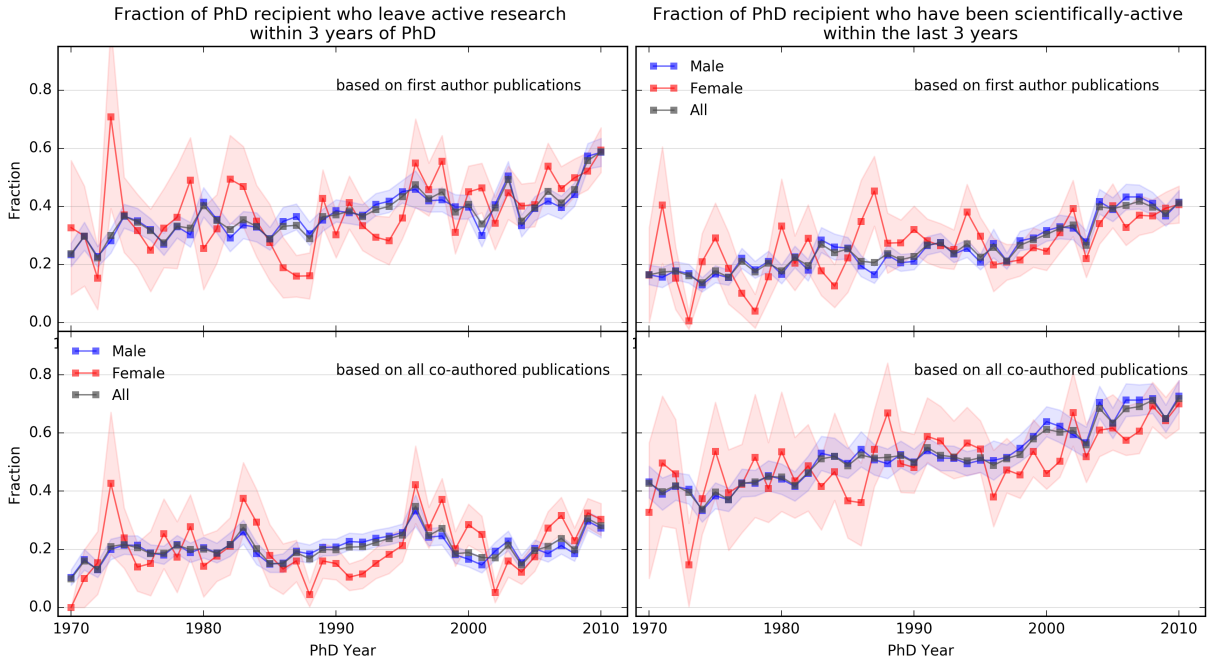

Figure 3: Left: Fraction of PhD recipients who leave active research within three years of PhD based only on first author peer-reviewed publications (top) and based on all peer-reviewed publications (bottom). Right: Fraction of PhD recipients who have been scientifically active within the last three years based only on first author peer-reviewed publications (top) and based on all peer-reviewed publications (bottom).

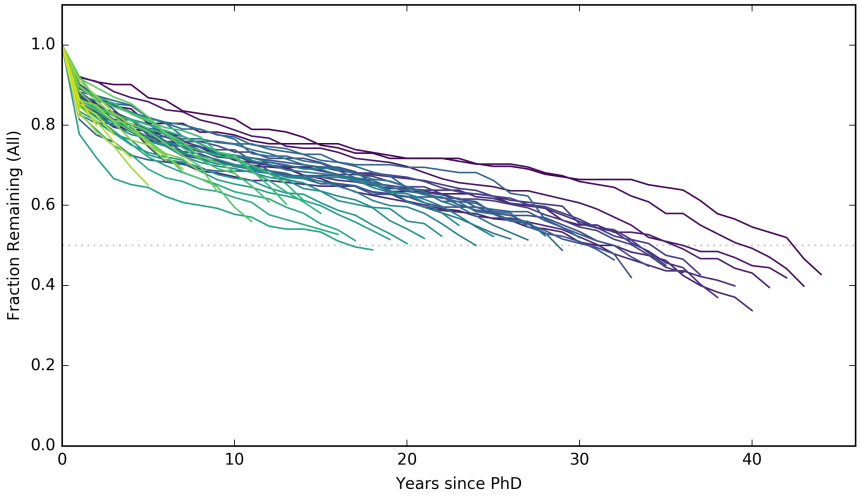

Figure 4: Survival function of all cohorts from 1970 (purple) to 2010 (yellow) as a function of years post PhD. The survival is determined from first-author and co-authored peer-reviewed publications. Most cohorts exhibit very similar trends, with over 50% (dotted line) of the cohort members still in active research 20 years post-PhD, going against suggestions that careers in astronomy have been severely shortened in recent years.

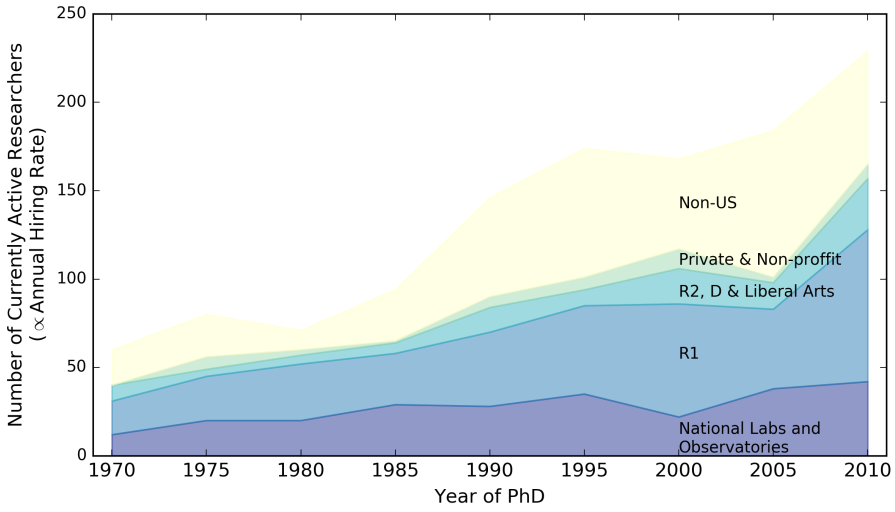

Figure 5: Where are they now? Break down of most recent affiliation type of all currently-active researchers as determined by co-authored publications within the last three years. We group institutions into five categories (from bottom to top): national labs and observatories; R1 universities; R2 universities, liberal arts colleges and other degree granting institutions; private companies including non-profits; and non-US institutions. The analysis was done only for the cohorts of 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005 and 2010 with interpolation between to show trends. We note the increasing number of researchers at R1 and non-US institutions and, to a lesser extent, in non-R1 degree granting institutions and national labs. Non-US hires are not included.

international and opportunities for research and teaching are not limited to the US. Since we only follow US PhD recipients in this analysis, non-US PhD recipients, that are subsequently employed at US institutions, are not captured here and the hiring rates are therefore lower limits. Further work is needed to understand the rate at which non-US PhD recipients are employed at US institutions follow similar patterns.

In summary, PhD recipients who stay in the field find long term employment at a variety of different institutions within the US, including national labs and observatories with functional responsibilities, research-focused universities and teaching-focused institutions. These career require a variety of different soft and hard skills and graduate programs as well as post-doctoral opportunities must focus on developing a range of competencies that would allow graduates to be successful in a range of careers, rather than just focusing on research.

5 Analysis and Summary

The results presented here underscore the need for more comprehensive analyses of employment trends in order to provide data to graduates navigating their job searches and to institutions and policy-making bodies in focusing efforts where they would be most beneficial. Recently, both [1] and [2] have carried out similar analyses, the former focusing on the AAS job rumor mill and the latter on the careers of ~ 1000 PhD recipients over the last 13 years. Both publications were widely

discussed in the field. The extensive and detailed analysis of [4] should receive more attention. Planned demographic surveys, such as those carried out by the AAS, AIP and NSF, are crucial but do not always capture all necessary data and should be supplemented with analysis of historical records, online archives and data collected by departments and universities. Recent analyses of such historical dataset have, for example, shed light on biases in telescope time allocation and citation rates [8, 9]. Other similar datasets should be used to understand trends within the field.

We find that a significant fraction of PhD recipients stay in active research: $\sim$50% of all PhD recipients since 1970 have been an author on a paper in the last 3 years. Focusing on more recent cohorts, 70% of all PhD recipients from the late 2000s, now $\gtrsim$10 years post-PhD, have co-authored a peer-reviewed paper and 40% have a first-author paper in the last 3 years. The fraction of people who leave active research is quite significant but not as high as some have feared. About 30% of PhD recipients from the 2000s are currently in jobs that do not involve academic publishing, many of them likely completely outside of academia. Furthermore, we see that researchers transition into out of research at a rate of 1% per year. Researchers who stay in the field are employed in the US at a wide range of institutions including substantial fractions at R1 universities, R2 universities, liberal arts colleges and national labs and observatories. These statistics underscore the importance of a broad set of skills being taught at the graduate level with a view towards a broad range of careers available for PhD recipients. A white paper presented by [7] focuses on the fact that graduate programs need to teach core competencies required within and outside the field and identifies specific recommendations based on the National Academies of Sciences report on "Graduate STEM Education for the 21st Century". These recommendations are supported by the analysis presented here.

While women represent 30% of recent PhD cohorts they are still only 19.2% of all PhD recipients over the last 48 years. The combination of low representation in past cohorts ($\sim$ 5% to 10% prior to the mid-1980s) and the overall 50% drop out of active research means that the total number of senior women is low. For many cohorts, only five and in some years, none, of the women in the cohort are still actively involved in research. As a result, only one or two senior women may be still active in research in some sub-fields. Therefore, relatively more junior female researches should be considered for and involved in activities where "seniority" is considered important. This provides them with important opportunities for networking and professional growth and ensures appropriate representation on committees, award nominations, invited speaker lists, etc.

We do not find preferential attrition of women at the post-graduate level. This finding is supported by [4] who find no leaky pipeline between undergraduate, PhD and post-graduate employment for women. Furthermore, the Longitudinal Study of Astronomy Graduate Students (LSAGS) [5] also finds that the sex of the survey respondents has no direct effect on their leaving the field but is a second order effect. Therefore, efforts to increase the representation of women in astronomy at the post-graduate level would be futile without substantial effort at the transition from high school to college. Critically, under-represented minorities, and minority women specifically, are still not represented at the same rates through all professional level and substantial work is needed to advance their participation in astronomy [4]. Overall, women still experience slower career progression [4] and work is needed on equalizing opportunities and removing biases from the system.

References

[1] Flaherty, K., The Leaky Pipeline for Postdocs: A study of the time between receiving a PhD and securing a faculty job for male and female astronomers, 2018, https://arxiv.org/abs/1810.01511

[2] Perley, Daniel A., Gender and the Career Outcomes of PhD Astronomers in the United States, 2019, https://arxiv.org/abs/1903.08195

[3] NSF Survey of Earned Doctorates, https://www.nsf.gov/statistics/srvydoctorates/

[4] Anne Marie Porter and Rachel Ivie, "Women in Physics and Astronomy, 2019", https://www.aip.org/statistics/reports/women-physics-and-astronomy-2019

[5] Ivie, Rachel and White, Susan and Chu, Raymond Y., "Women's and men's career choices in astronomy and astrophysics", Phys. Rev. Phys. Educ. Res., 2016, https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.12.020109

[6] Milojević, Staša and Radicchi, Filippo and Walsh, John P., "Changing demographics of scientific careers: The rise of the temporary workforce", PNAS, 2018, https://www.pnas.org/content/115/50/12616

[7] Amaya Moro-Martín, Astro2020 APT White Paper: "Mind the gap": a call to redesign astronomy graduate education, https://arxiv.org/abs/1904.04311

[8] Reid, I. N., Gender-Correlated Systematics in HST Proposal Selection, 2014, , 126, 923

[9] Caplar, N., Tacchella, S., & Birrer, S., Quantitative evaluation of gender bias in astronomical publications from citation counts, 2017, Nature Astronomy, 1, 0141