

Fig 1. A: Mean probability that an author in a given position is female for primary articles indexed in Pubmed with the MeSH term Biology (black) or Computational Biology (grey). The bio dataset is inclusive of papers in the comp dataset. Error bars represent 95% confidence intervals. B: Mean probability that an author in a given position is female for primary articles indexed in Pubmed with the MeSH term Biology (black) or Computational Biology (grey). The bio dataset is inclusive of papers in the comp dataset. Error bars represent 95% confidence intervals. C: Mean probability that an author is female for publications in a given year. Error bars represent 95% confidence intervals. D: Mean probability that the first (F), second (S), penultimate (P) or other (O) author is female for publications where the last author is male ($P_{female} < 0.2$) or female ($P_{female} > 0.8$). Papers where the gender of the last author was uncertain or could not be determined were excluded. Error bars represent 95% confidence intervals.

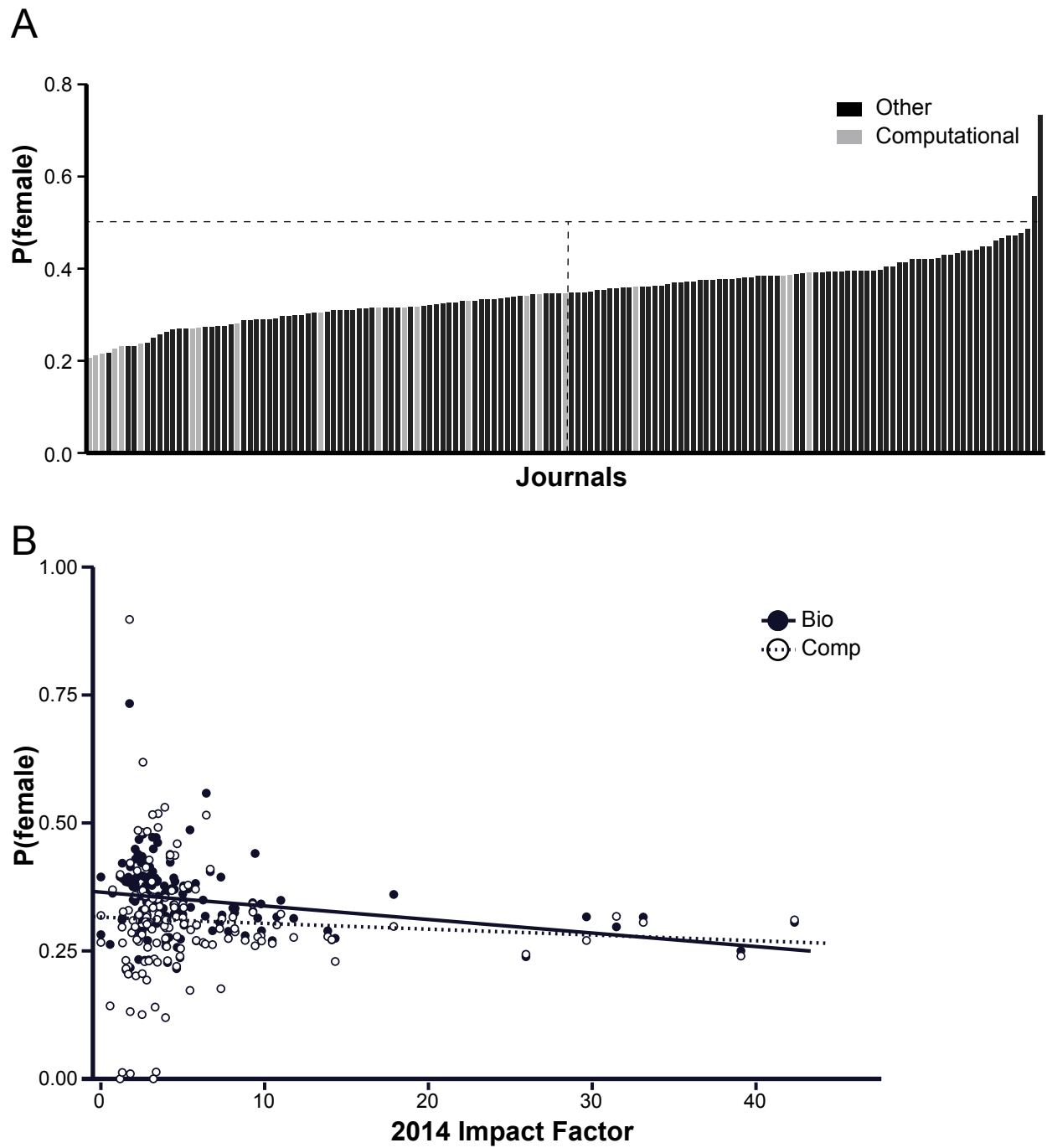
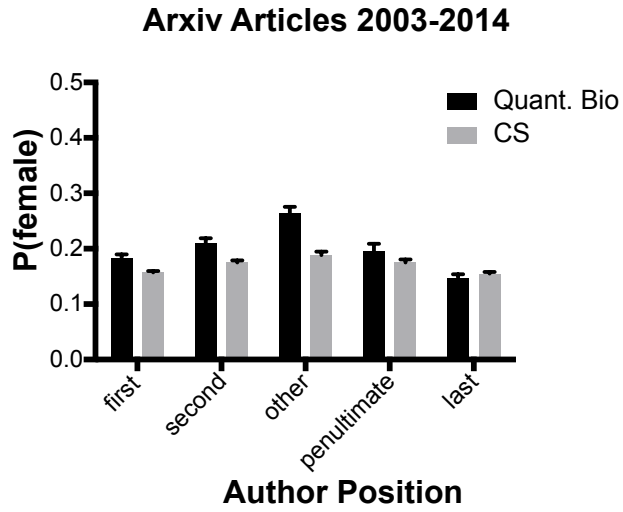


Fig 2. A: Mean probability that an author is female for every journal that had at least 1000 authors in our dataset. Grey bars represent journals that have the words “Bioinformatics,” “Computational,” “Computer,” “System(s),” or “-omic(s)” in their title. Vertical line represents the median for female author representation. See also S1 Table. B: Mean probability that an author is female for articles in the “Bio” dataset (black dot) or in the “Comp” dataset (open square) for each journal that had at least 1000 authors plotted against the journals’ 2014 impact factor. Journals that had computational biology articles are included in both datasets. An ordinary least squares regression was performed for each dataset. Bio: $m = -0.00264$, $P_{Z>|z|} = 0.0022$. Comp: $m = -0.000789$, $P_{Z>|z|} = 0.568$.

A



B

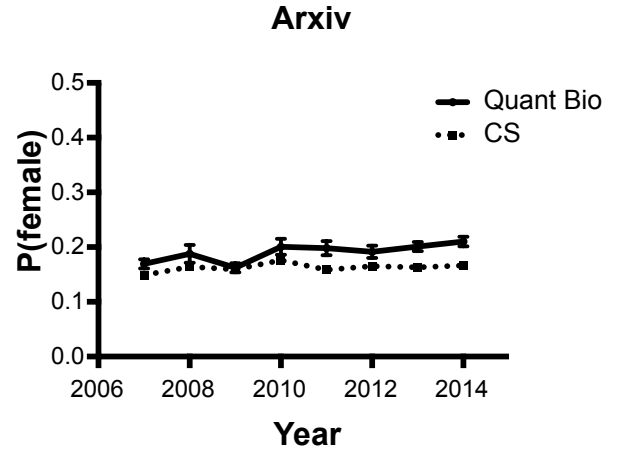
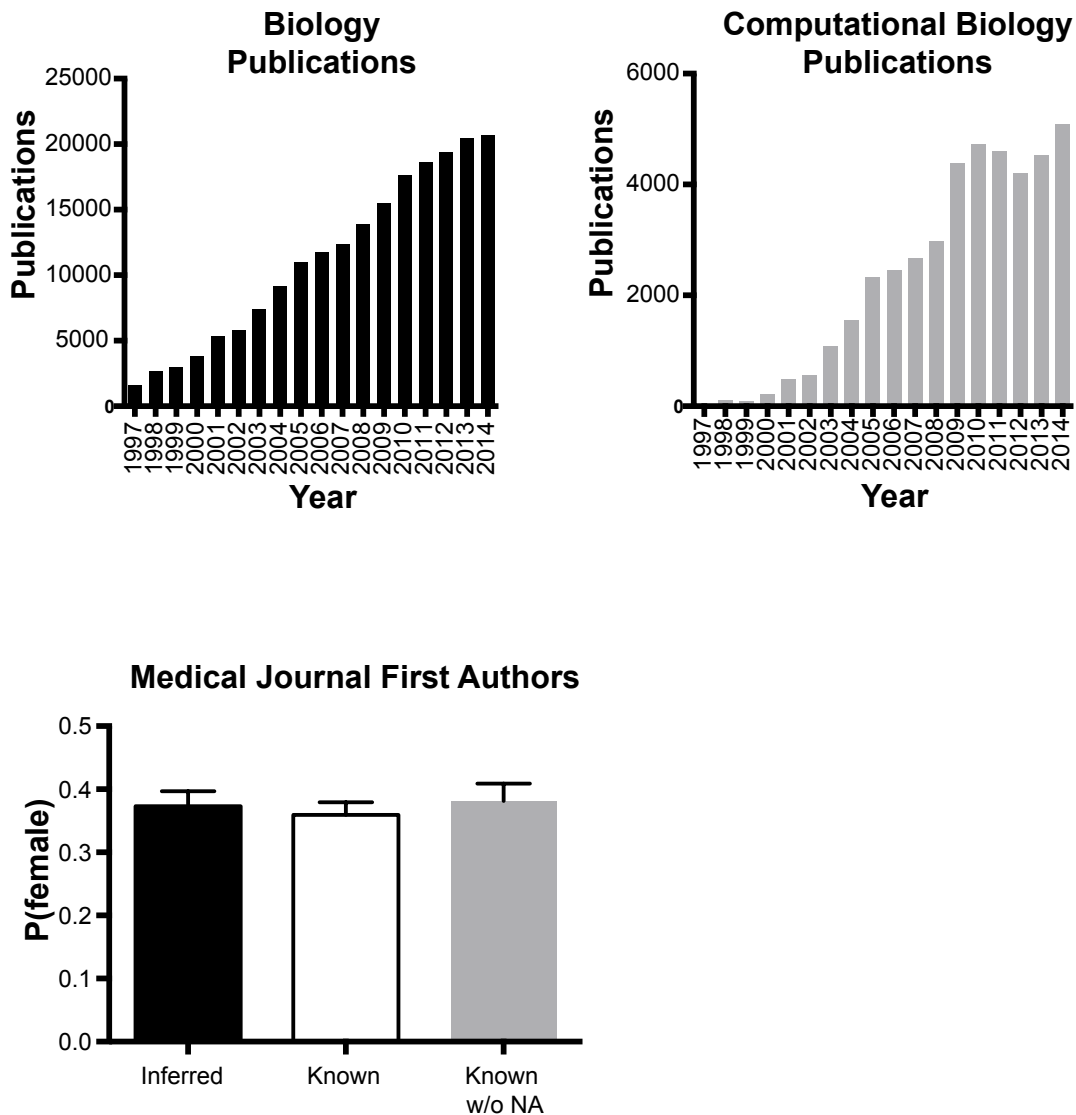
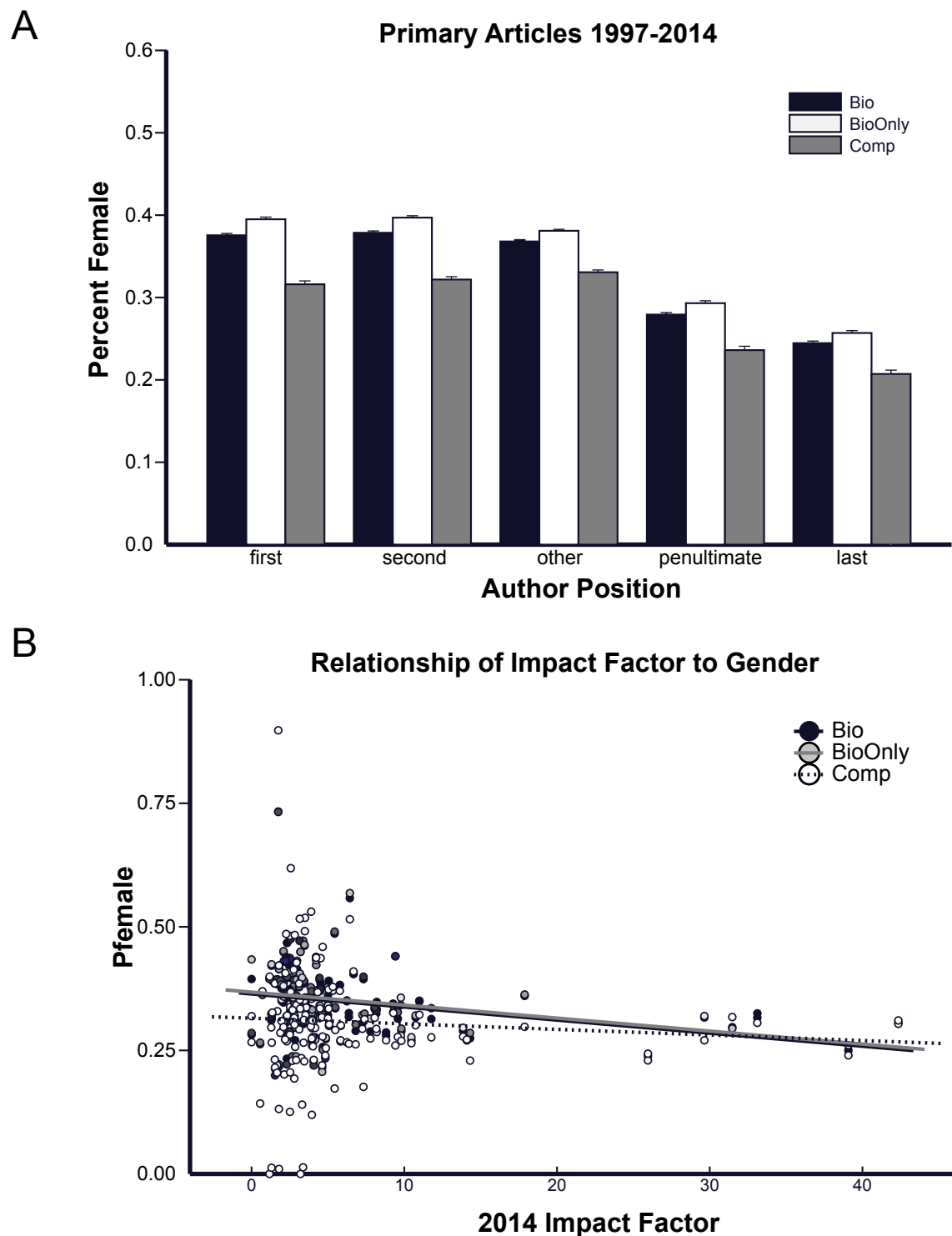


Fig 3. A: Mean probability that an author in a given position is female for all preprints in the arXiv quantitative biology (black) or computer science (grey) categories between 2007 and 2014. Error bars represent 95% confidence intervals. B: Mean probability of authors being female in arXiv preprints in a given year. Error bars represent 95% confidence intervals. Slopes were determined using ordinary least squares regression. The slope for q-bio is slightly positive ($p < 0.05$), but the slope for cs is not.



S1 Fig. A: Number of primary publications per year indexed under the “Biology” MeSH term. B: Number of primary publications per year indexed with “Computational Biology” as a major MeSH term. C: Comparison of computational gender inference (black) with known genders (white) for the dataset from Filardo et. al. [15]. Grey represents the known proportion of female authors when excluding names for which the gender could not be computationally inferred. Error bars represent 95% confidence intervals.



S2 Fig. A: Mean probability that an author in a given position is female for primary articles indexed in Pubmed with the MeSH term Biology (black), Computational Biology (gray) or for those articles with Biology *but not* Computational biology (white). Error bars represent 95% confidence intervals. B: Mean probability that an author is female for articles in the “Bio” dataset (black) in the “Comp” dataset (white), or for articles in the Bio *but not* Comp (gray) for each journal that had at least 1000 authors plotted against the journals’ 2014 impact factor. Excluding computational publications from the biology dataset does not substantially alter the correlation between impact factor and P_{female} .