

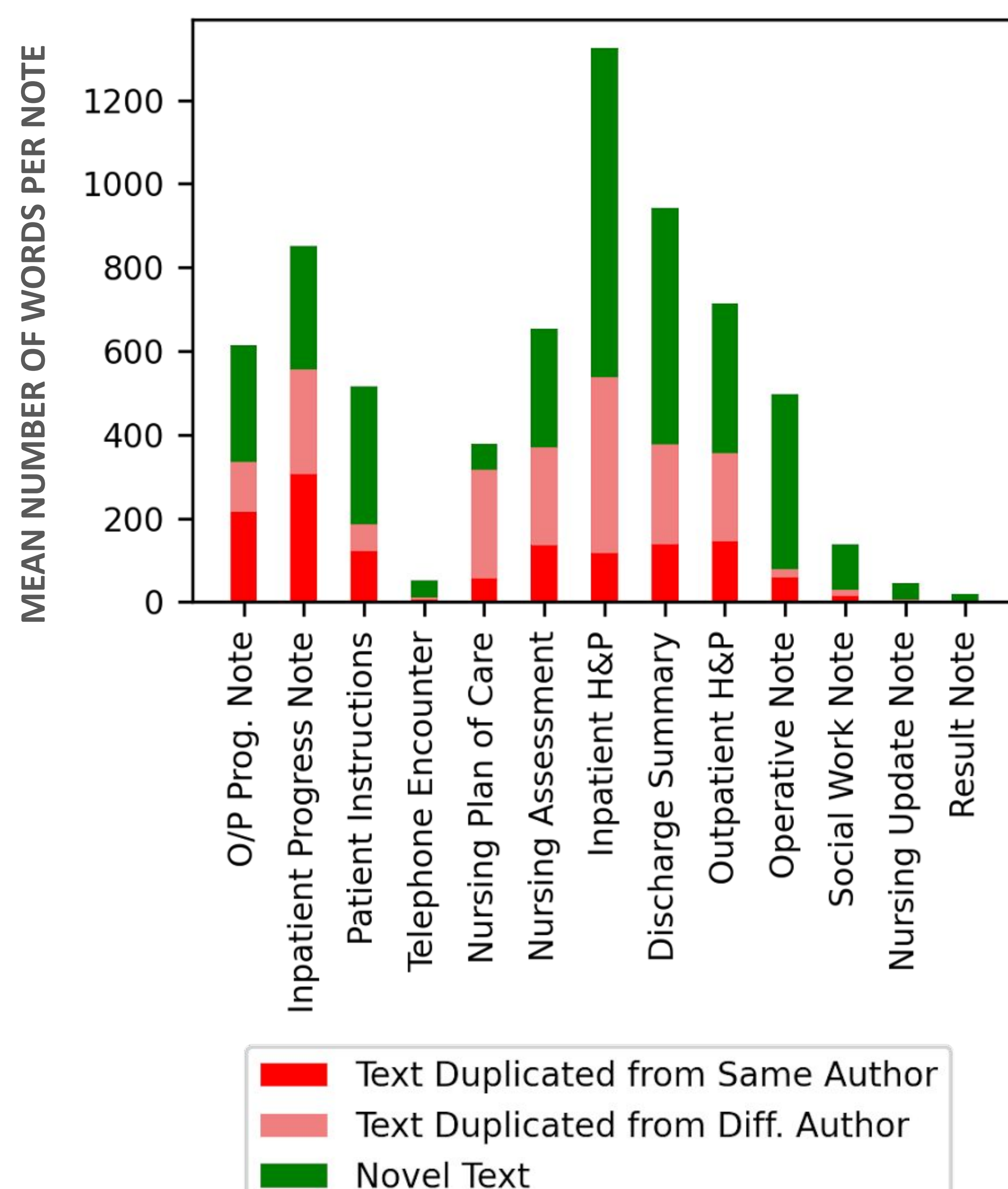
INTRODUCTION

- Duplicated text within electronic medical records leads to wasted clinician time, medical errors, and burnout¹⁻⁶.
- This is the largest study of its kind, examining the prevalence of duplicative data in clinical notes from a large academic health system and determining the factors associated with duplication.

METHODS

- 10-gram sliding window to identify exactly-duplicated spans of text within a patient's charts.
- Examined all inpatient and outpatient notes within the Penn Medicine Health System from 2015 to 2020.
- Quantified (a) text duplicated from a different author vs. (b) text duplicated from the same author.
- Quantified novel and duplicated text per author by note type, author types, and per chart by the number of notes in the chart.
- Examined the relationship between information duplication and information scatter, defined as the inverse of novel text per note, for different note types.

MEAN DUPLICATE TEXT PER NOTE, BY NOTE TYPE



IT'S IN THE CHART!

6 Years
of clinical notes
analyzed

100
Million
NOTES

33
Billion
Words

192 Billion
Characters

8x

the amount
of characters in
all of English
Wikipedia

4,285 WORDS
per chart = 1/6th of hamlet

HAMLET

William
Shakespeare

15 PATIENTS =
READING HAMLET

2X
EVERY DAY

Average
note has
40 words

50% of
all text is
duplicated

HAMLET TWICE IN 40 WORD SEGMENTS

= 1605 SEPARATE PAGES WITH

50% of WORDS COPIED...

BUT YOU DONT KNOW WHICH ONES??

RESULTS

- Analysis included 100 million notes consisting of 33 billion words.
- 50.1% of the total text in charts was duplicated from prior notes written about the same patient.
- Duplication fraction increased year-over-year, from 33.0% for notes written in 2015 to 54.2% for notes written in 2020.
- 54.1% of duplicated text was copied forward from the same author. 45.9% was duplicated from a different author.
- Charts with more notes had more total duplicate text, approaching ~60%.
- Notes with high information scatter had high information overload and vice versa.

DISCUSSION

- More than half of the text in the chart is directly duplicated.
- Our results is likely an underestimate as our algorithm was unable to identify more pernicious forms of duplicate text.
- Both inter- and intra-author duplication are major problems, and charts with more notes have more duplication.
- Alternative non-note documentation paradigms (e.g., a collaborative wiki model) can mitigate duplication.
- Not requiring separate clinicians or teams to create entirely new documents will likely reduce inter-author duplication
- Not requiring new documents to be created for every outpatient encounter or day of hospitalization will likely reduce intra-author duplication.

References

1. Steinkamp J, Kantrowitz J, Sharma A, Bala W. Beyond Notes: Why It Is Time to Abandon an Outdated Documentation Paradigm. *J Med Internet Res.* 2021;23(4):e24179.
2. Beasley JW, Wetterneck TB, Temte J, et al. Information chaos in primary care: implications for physician performance and patient safety. *J Am Board Fam Med.* 2011;24(6):745-751.
3. Bowman S. Impact of electronic health record systems on information integrity: quality and safety implications. *Perspect Health Inf Manag.* 2013;10:1c.
4. Gardner RL, Cooper E, Haskell J, et al. Physician stress and burnout: the impact of health information technology. *J Am Med Inform Assoc.* 2019;26(2):106-114.
5. Babbott S, Manwell LB, Brown R, et al. Electronic medical records and physician stress in primary care: results from the MEMO Study. *J Am Med Inform Assoc.* 2014;21(e1):e100-e106.
6. Arndt BG, Beasley JW, Watkinson MD, et al. Tethered to the EHR: Primary Care Physician Workload Assessment Using EHR Event Log Data and Time-Motion Observations. *Ann Fam Med.* 2017;15(5):419-426.