

# Understanding Primary Care Physicians' Work via Text Analytics on EHR Inbox Messages

Célia Escribe, MS; Stephanie A. Eisenstat, MD; Walter J. O'Donnell, MD; and Retsef Levi, PhD

**E**lectronic health records (EHRs) have transformed the daily work of primary care physicians (PCPs) and are now ubiquitous in most practices. However, in spite of promised benefits, PCPs now find that they spend more than half of their work on the EHR,<sup>1-3</sup> with a significant fraction dedicated to administrative tasks.<sup>1,4</sup> The hope was that more sophisticated EHR systems would reduce work burden on care team members, but this has not been the case: The dissemination of EHR systems reduced the work burden on support staff while at the same time increasing PCPs' work burden.<sup>5</sup> With this, EHR impact on physicians' wellness has become a great concern for most health care organizations.<sup>6-9</sup> Moreover, as more of PCPs' work is now conducted outside well-documented in-person clinic visits, there is limited identification of the scope of tasks they manage and how this varies across PCPs. Partially motivated by these circumstances, there have been time motion studies and studies that leverage EHR log data to better understand various aspects of physicians' EHR work.<sup>1,6,10,11</sup> Core measures of EHR usage have been developed, such as "work outside of work."<sup>12</sup> Studies have also focused on quantifying the work allocation among different tasks (eg, clinical face time with patients).<sup>11</sup>

Given that inbox message management comprises more than 25% of physician work,<sup>1</sup> and these messages have become the primary mode of communication among physicians, team members, and patients,<sup>13</sup> it is important to understand specific drivers of that work at a more granular level, to improve efficiency and quality while decreasing work burden. A few studies have focused on the work component of managing EHR inbox messages, but most were conducted when the dissemination of EHR systems was still evolving and the tasks that they were used to conduct were relatively less complex.<sup>14-16</sup> Other studies focused on quantifying the number of messages received relative to generic categories (eg, system-generated notifications) without considering the specific tasks managed by physicians.<sup>17,18</sup> These studies highlighted that receiving more than the average of system-generated inbox messages was associated with higher probability of burnout.<sup>17</sup> However, these types of statistics do not lend themselves to directly inform related hospital policies. To the best of our knowledge, few studies look

## ABSTRACT

**OBJECTIVES:** To develop a text analytics methodology to analyze in a refined manner the drivers of primary care physicians' (PCPs') electronic health record (EHR) inbox work.

**STUDY DESIGN:** This study used 1 year (2018) of EHR inbox messages obtained from the Epic system for 184 PCPs from 18 practices.

**METHODS:** An advanced text analytics latent Dirichlet allocation model was trained on physicians' inbox message texts to identify the different work themes managed by physicians and their relative share of workload across physicians and clinics.

**RESULTS:** The text analytics model identified 30 different work themes rolled up into 2 categories of medical and administrative tasks. We found that 50.8% [range across physicians, 34.5%-61.9%] of the messages were concerned with medical issues and 34.1% [range, 23.0%-48.9%] focused on administrative matters. More specifically, 13.6% [range, 7.1%-22.6%] of the messages involved ambiguous diagnosis issues, 13.2% [range, 6.9%-18.8%] involved condition management issues, 6.7% [range, 1.9%-13.4%] involved identified symptoms issues, 9.5% [range, 5.2%-28.9%] involved paperwork issues, and 17.6% [range, 9.3%-27.1%] involved scheduling issues. Additionally, there was significant variability among physicians and practices.

**CONCLUSIONS:** This study demonstrated that advanced text analytics provide a reliable data-driven methodology to understand the individual physician's EHR inbox management work with a significantly greater level of detail than previous approaches. This methodology can inform decision makers on appropriate workflow redesign to eliminate unnecessary workload on PCPs and to improve cost and quality of care, as well as staff work satisfaction.

*Am J Manag Care.* 2022;28(1):e24-e30. doi:10.37765/ajmc.2022.88817

at the specific content of inbox messages,<sup>19,20</sup> and none provides an automated scalable methodology to analyze the specific work managed through EHR inbox communications, which is believed to be a critical part of the PCP's clinical work.<sup>13</sup> It is therefore essential to not only quantify the overall level of workload related to inbox messages management by PCPs, but also to understand in a refined manner the specific work themes and related drivers of communications between the PCP and team members and patients.

The purpose of this study is to develop a data-driven scalable approach to identify salient *work themes (topics)* in the EHR messages managed by PCPs. This can help in differentiating clinically complex work of PCPs from work that could be handled by other care team members. Specifically, messages related to the management of complex clinical conditions and patient health need the PCP involvement, whereas messages related to "simpler" clinical tasks (eg, routine preventive work) or administrative tasks (eg, scheduling) are better managed by other team members. The value of this methodological approach is that it allows analyzing physician work composition at a great level of detail. This provides critically important diagnostic tools to help inform policies that address compensation of PCPs, as well as redesign of related workflow processes to improve quality, cost of care, and staff work satisfaction.

One challenge to developing detailed understanding of PCPs' work is that it requires massive manual review of messages or otherwise relying on qualitative reporting by PCPs. This paper leveraged advanced text analytics to develop a data-driven and computer-enabled methodology to conduct such analysis. The methodology and tools developed in this study allow unprecedented understanding of the substance and complexity of work themes managed by PCPs through EHR inbox messages. Moreover, the ability to understand heterogeneity in work composition across PCPs and practices is critical to be able to identify opportunities to improve the quadruple aim: quality, cost, patient satisfaction, and physician satisfaction.

## METHODS

### Study Setting and Population

This work included retrospective analysis of all the EHR messages managed by 184 PCPs who worked in 18 primary care practices at a large academic medical center in Boston, Massachusetts, from March 1, 2018, through March 1, 2019. **Table 1** displays characteristics of this cohort. Note that the academic medical center has been using the Epic EHR system since 2016.

### Data Sources

The analysis used a self-constructed data set that integrates 2 main sources: (1) EHR inbox message data and (2) PCP patient panel data.

## TAKEAWAY POINTS

Electronic health records (EHRs) have transformed the daily work of primary care physicians (PCPs). Decision makers currently have limited ability to understand physicians' actual EHR inbox work composition.

- ▶ This study demonstrates that text analytics allow understanding of specific work themes managed by each PCP through inbox messages with a greater level of detail than previous approaches.
- ▶ A significant part of PCPs' EHR inbox work is in fact concerned with administrative matters and non-clinically complex work.
- ▶ This methodology is a starting point for managers to identify opportunities to redesign workflow processes and practices and to improve quality of care and work satisfaction.

**TABLE 1.** Cohort Characteristics

Characteristics	Physicians, n (%)
Gender	
Female	118 (63)
Male	67 (36)
Years of practice	
≤ 9	29 (16)
10-19	37 (20)
20-29	52 (28)
30-39	36 (19)
≥ 40	14 (8)
Unknown	17 (9)
Clinic type	
Community health centers	47 (25)
Boston downtown	138 (75)
Clinical full-time equivalent	
0.90-1.00	20 (10)
0.75-0.89	21 (11)
0.50-0.74	53 (29)
0.25-0.5	70 (38)
< 0.25	22 (12)

In the academic medical center, the Epic EHR system is the primary messaging system for communication within the practice team and with patients. Inbox messages include messages from patients and care team members, as well as messages generated automatically by the EHR system. Only messages from patients and care team members contain text information that is not protocol driven. More specifically, those inbox messages that are not protocol driven include communications with patients as well as communications among care team members that are initiated by a patient's message, by a patient's phone call, by the need to transfer an electronic note related to a patient, or as part of the coordination of an action required for a given patient. The forthcoming analysis focuses on this specific category of *patient-initiated inbox messages*. Such patient-initiated communications correspond to 36% of inbox messages received by the 184 PCPs included in the study. The remaining EHR protocol-driven messages can be divided into system-generated notifications,

## METHODS

lab results notifications, and prescription-related notifications (authorization and signing), which represent, respectively, 26%, 9%, and 29% of received inbox messages.

The second main data source used in the analysis is an internal registry of patient assignments to the panel of individual PCPs. This defines the specific patient population for which each PCP cares.

Using the PCP patient panel data, each message from the inbox message data regarding a given patient was associated with a PCP. Only inbox messages concerning patient-initiated communication were selected in the final data set. This final data set captured a total of 1,279,712 messages.

### Analysis

**Latent Dirichlet allocation model.** A latent Dirichlet allocation (LDA) model was trained on the text messages.<sup>21</sup> The LDA model is a generative probabilistic Bayesian model for discovering the topics that occur in a corpus of text documents. The number of desired topics to discover is given as input to the model. After training the model on all text messages, each identified topic was characterized by the respective frequency of each word from the corpus vocabulary. The LDA model then identifies, for each message, what specific topics are discussed in the message and what relative portion of the message is dedicated to each topic (eg, the model can determine that a message consists of 60% of topic A and 40% of topic B). Because of uncertainty in the modeling, there can be a small percentage of the message topics mixture that corresponds to a *nonspecific topic* (ie, the part of the message not explained by the LDA-identified topics).

The number of topics given as input to the LDA model was selected manually, using the common practice of expert assessment.<sup>22,23</sup> The detailed methodology can be found in [eAppendix Exhibit 1](#) (eAppendix available at [ajmc.com](#)).

**Work theme (topic) identification.** After selecting the best LDA model, a manual review was performed, with the help of 2 physicians, on each of the topics identified by the LDA model to label the work theme captured in the respective topic. Manual review involved interpretation of the most frequent words within a topic. A general theme was selected based on this review. The methodology used to evaluate external accuracy of the algorithm predictions for a given message is detailed in [eAppendix Exhibit 2](#).

To understand the output of the model and the role of the different topics, some message analysis was performed and is described in detail in [eAppendix Exhibit 3](#).

**Theme (topic) taxonomy.** The overall aim of this methodological step was to provide a hierarchical taxonomy of the LDA-identified topics and their respective work themes, so that it would be possible to analyze the share of PCPs' work concerned with more general work categories. The higher level of the proposed taxonomy included *administrative* and *medical* issues. This fundamentally distinguishes between clinical messages on which the physician's input is important compared with administrative messages, which likely could be handled by other care team members. A second more

granular level was proposed to divide the *administrative* category and the *medical* category into subcategories. After discussion with 2 physicians, the *medical* category was divided into the following subcategories: *identified symptoms*, *ambiguous diagnosis*, *condition management*, *tests and exams*, and *clinical decision-making referral*. Similarly, the *administrative* category was divided into the following subcategories: *paperwork*, *scheduling*, *administrative referral*, and *prescription*. It should be emphasized that other subcategories are possible here, as the purpose of this second level of the taxonomy is only to provide a higher-level analysis of the work composition.

Each previously identified topic was then assigned to one of the categories of the proposed taxonomy described above.

**PCP-level analysis.** All statistics calculated at the PCP level considered all the patient-initiated messages where the PCP was involved either as sender or receiver of the message. The share of messages concerned with a given topic was calculated for each PCP, by summing over all messages the respective percentage that this specific topic represented (ranging from 0%-100%) and then calculating the mean. Following this analysis, it was possible to characterize, for each PCP, what the EHR messages-related workload was across the different work categories from the taxonomy.

**Practice-level analysis.** Practice-level analysis was performed by aggregating the statistics regarding the share of work per topic among all doctors from a given practice.

## RESULTS

### LDA Model

Using human assessment of topic model validity as explained in the Methods section, the best number of topics selected was 30. For models with a smaller number of topics, some topics were missing, and other topics seemed to combine what would preferably be better differentiated themes. For models with more than 30 topics, additional topics were assessed to be less meaningful in terms of discerning a unique theme.

### Work Theme (topic) Identification

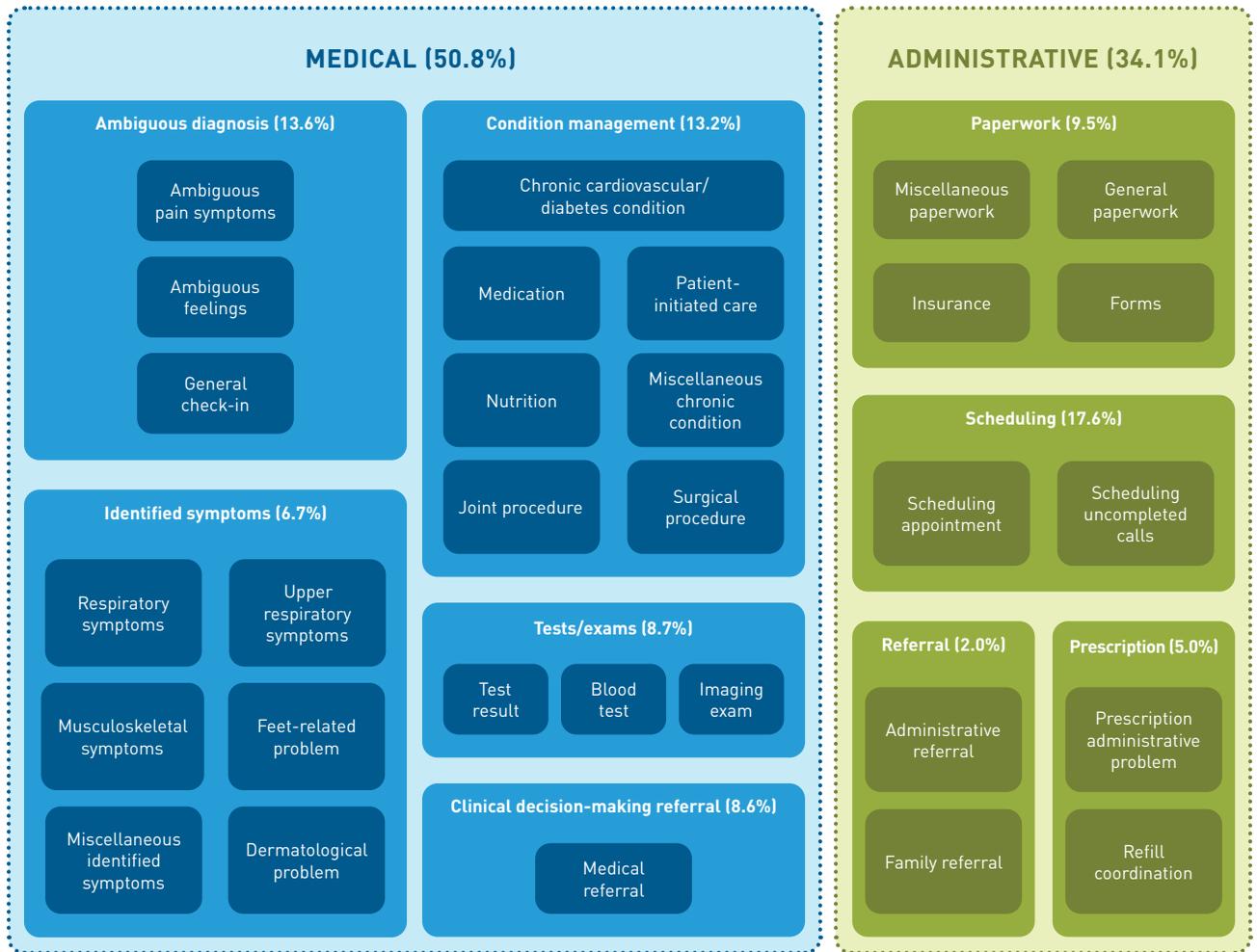
The list of the top 10 most frequent words for all topics is displayed in [eAppendix Exhibit 4](#) and these are displayed as word clouds in [eAppendix Exhibit 5](#) for a subset of topics. For example, the 5 most frequent words for the *nutrition* topic are *food*, *diet*, *weight*, *eat*, and *cholesterol*. Sample messages and their topic mixture are presented in [eAppendix Exhibit 6](#).

The accuracy obtained with external manual validation was equal to 81% (84% and 78% specifically for each of the 2 experts), which corresponds to the average of the cumulative weight of the topics identified correctly. This indicates that most of the topics identified by the algorithm are correct.

### Theme (topic) Taxonomy

Each identified work theme (topic) was first assigned to one of the granular subcategories. For example, the following topics were

**FIGURE 1.** Topic Taxonomy for Patient-Initiated Inbox Messages Sent and Received by PCPs\*



PCP, primary care physician.

\*This figure shows the assignment of the 30 topics into the categories of the high-level taxonomy and the more granular taxonomy. Numbers displayed correspond to the average number of patient-initiated messages sent and received by PCPs concerned with each topic across all PCPs, by summing over all messages the respective percentage that this specific topic represented. *Medical* and *administrative* categories amount to 84.9% of the messages, as the remaining 15.1% correspond to the average percentage of the unclassified topic in each message. The remaining protocol-driven inbox messages are not included in this analysis. The small value for the *prescription* subcategory is explained by the fact that most prescription-related messages handled by the PCPs are protocol-driven ones (signing and authorization) and thus are not included in this analysis. The same holds for the *referral* subcategory.

assigned to the subcategory *condition management: medication, nutrition, patient-initiated care, chronic cardiovascular and diabetes condition, miscellaneous chronic condition, joint procedure, and surgical procedure*. The topic assignment to one of the high-level categories (ie, *administrative* and *medical*) was then determined by the chosen subcategory. The complete resulting assignments of topics to categories are presented in **Figure 1**.

### PCP-Level Analysis

On average, of the patient-initiated messages sent and received by physicians, 34.1% (range, 23.0%-48.9%; SD, 5.3%) involved administrative issues, whereas 50.8% (range, 34.5%-61.9%; SD, 5.3%)

involved medical issues. The large range indicates significant variability among PCPs. The other 15.0% (range, 11.3%-20.5%; SD, 1.7%) were concerned with nonspecific topics (ie, the part of messages not explained by the LDA-identified topics).

Results at the subcategory level are presented in **Table 2**. We found that 13.6% (range, 7.1%-22.6%; SD, 3.1%) of the patient-initiated messages sent and received by physicians involved ambiguous diagnosis issues, 13.2% (range, 6.9%-18.8%; SD, 2.1%) involved condition management issues, 6.7% (range, 1.9%-13.4%; SD, 1.4%) involved identified symptoms issues, 9.5% (range, 5.2%-28.9%; SD, 2.3%) involved paperwork issues, and 17.6% (range, 9.3%-27.1%; SD, 3.2%) involved scheduling issues.

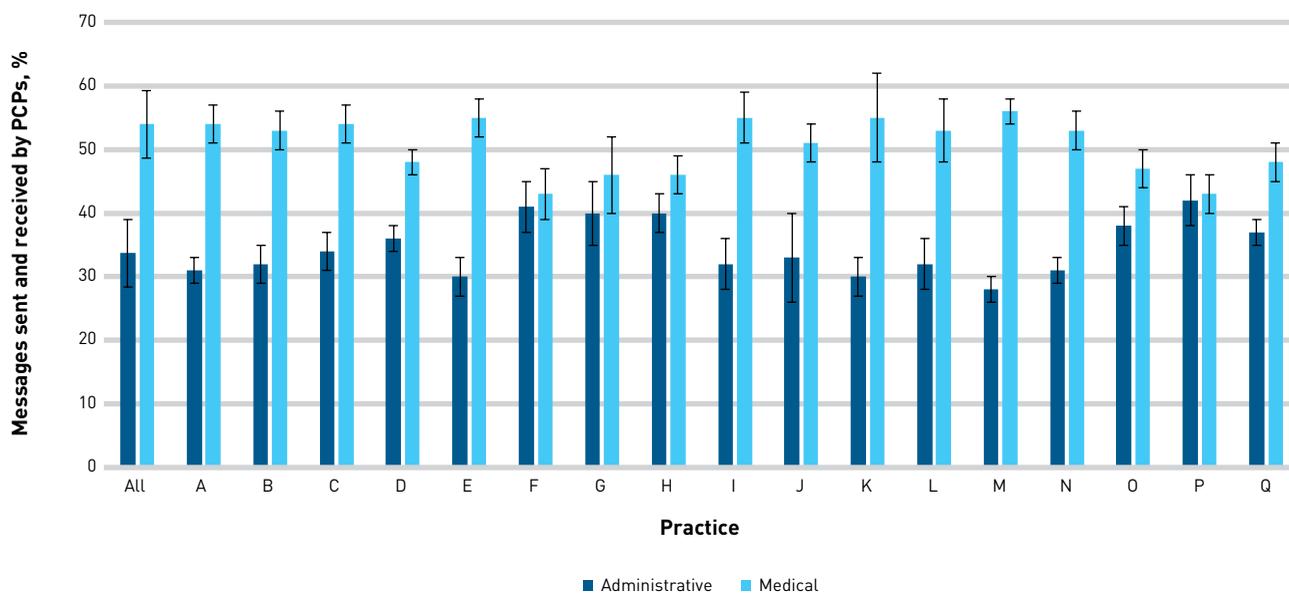
## METHODS

**TABLE 2.** Percentage of Patient-Initiated Inbox Messages Sent and Received by Primary Care Physicians Concerned With Each Subcategory of Topic\*

Topic subcategory	Mean	SD	Interquartile range	Range
Ambiguous diagnosis	13.6	3.1	11.0-16.0	7.1-22.6
Condition management	13.2	2.1	11.8-14.6	6.9-18.8
Identified symptoms	6.7	1.4	5.9-7.6	1.9-13.3
Tests/exams	8.7	1.8	7.8-9.6	1.5-16.3
Clinical decision-making referral	8.6	1.5	7.6-9.3	4.1-15.6
Paperwork	9.5	2.3	8.0-10.8	5.2-28.9
Scheduling	17.6	3.2	15.3-19.6	9.3-27.1
Referral	2.0	0.9	1.3-2.2	0.0-8.1
Prescription	5.0	1.7	3.9-5.8	1.6-12.4

\*Statistics are compiled across all primary care physicians.

**FIGURE 2.** Allocation of Patient-Initiated Inbox Messages Sent and Received by PCPs Between Administrative and Medical Categories Across Practices\*



PCP, primary care physician.

\*Share of inbox messages, with SD, devoted to administrative and medical matters across all PCPs, and across PCPs from each of the 18 considered practices.

### Practice-Level Analysis

Figure 2 reports the mean share of messages concerned with administrative and medical topics across different practices. The mean across all PCPs from a given practice differed greatly from one practice to another; for example, it ranged from 28% to 42% for the administrative topic category.

## DISCUSSION

Demand for non-face-to-face care is increasing for primary care physicians, and handling EHR inbox messages represents a large share of such non-face-time work.<sup>1,13</sup> Moreover, the handling of these messages is a source of significant work burden that has been

identified as being associated with burnout.<sup>17</sup> It is therefore critical to improve the level of understanding of EHR inbox management work burden by exploring the specific drivers of communications between PCPs and their support team and patients. The contributions of this study are 2-fold. First, this study develops an advanced text analytics approach to identify detailed work composition and themes managed by PCPs in their EHR inbox messages that can be implemented at scale with minimal work. Second, the new methodology offers important new insights on PCPs' EHR inbox work composition and how it differs across PCPs and practices.

As already mentioned, the analysis highlights several important concrete insights regarding PCPs' work composition. First, it highlights the great complexity of PCPs' work, which typically

requires simultaneous management of multiple tasks that may be very different in nature from one another, often including a blend of clinical and administrative issues. For example, handling an acute patient problem may require clinical diagnosis of the problem, as well as scheduling an appointment by the doctor. This underscores the importance of understanding the specific work topics managed by PCPs and their corresponding complexity, and not merely monitoring the volume of work.

Related to that, the analysis highlights that of all the time PCPs spent managing EHR inbox messages, a significant fraction was spent handling issues classified as administrative topics. Specifically, more than 34% of the messages managed by the PCPs in the study were concerned with administrative topics. It is likely that some of these administrative messages could be handled by other team members. For example, scheduling and paperwork issues mostly require the involvement of administrative staff, and although signing off on a prescription may require the involvement of a PCP, administrative issues related to prescriptions (eg, prescription sent to wrong pharmacy) could be addressed by nurses or medical assistants. Previous research already highlighted the significant share of administrative and clerical work (eg, related to order entry, billing, and coding)<sup>1</sup> as a possible factor associated with burnout.<sup>7,24</sup> The fact that more than 30% of patient-initiated EHR inbox messages are concerned with administrative work further describes and quantifies the nonclinical work conducted by physicians and raises the question of whether it has any impact on physician burnout. Although this paper does not provide an answer to this question, it provides important data to stimulate further relevant research. In particular, more research should be dedicated to understanding the possible link between this administrative share of messages and physicians' and patients' well-being and, more specifically, what the "optimal" share between clinical and administrative work is.

The broad range and variable complexity of clinical tasks that are managed by PCPs raises another question of whether the PCP is always the best team member to handle them, or whether some of them could also be delegated to nonphysician clinicians. For example, some of the messages related to the *condition management* category (eg, nutrition, routine preventive counseling) could probably be delegated to nurses or medical assistants. This suggests that health care organizations should consider the broad range of work themes currently managed by PCPs when redesigning operational processes. This insight is aligned with the results of other studies that suggested, for example, a shift from physician-centric models to shared-care models by expanding the roles of nurses and medical assistants to meet patients' clinical needs with more efficient use of PCPs' time.<sup>24</sup>

The observed variability across practices regarding the administrative share in the messages managed by PCPs supports the hypothesis that practice-level drivers can have an impact on PCPs' work composition. Practice-level factors such as standardized workflow processes and staffing schemes are modifiable by leaders, and future research could help in determining which of those factors play an important role in affecting work composition.

The previous discussion suggests that the methodology and tools developed in this study enable the first step to go beyond merely monitoring PCPs' work volume and to understand the details of the work currently managed by PCPs. Such diagnostic tools are a starting point for future research to study workflows and their impact on work composition. They can allow health care systems and policy makers to identify opportunities to redesign workflow processes and practices and shape the respective payment schemes accordingly. Moreover, the newly developed approach could be used to accurately measure and assess the impact of these efforts after changes took effect. The complexity and variability of the different work themes handled by the PCPs also suggest that a "one-size-fits-all" solution would probably fall short of success and that different workflow redesigns should be considered.

These are all key enablers to obtain the quadruple aim of improving cost and quality of health care, as well as patient and physician satisfaction.

### Limitations

This study has several limitations. First, the methodology was used on PCPs within a single institution; it would be interesting to observe the generalizability of the obtained results to other hospitals, and to physicians from other specialties. Second, EHR data were extracted from the Epic EHR system. Other EHR systems may yield different insights regarding work composition, because of, for example, different inbox message formats and templates that may affect communication workflows. Third, the algorithm identifying work themes based on text can be sensitive to patient health literacy, as level of education is known to be correlated with how patients use internet-based patient portals.<sup>25,26</sup> Lastly, this study only provides descriptive analysis with respect to the content of managed inbox messages and does not analyze other potentially important factors related to workflow processes (eg, use of pooled email queues or forwarding messages).

## CONCLUSIONS

Handling EHR inbox messages is an important driver of work burden for physicians. This study illustrates that advanced text analytics provide a reliable data-driven methodology to identify the work themes managed by each PCP through their EHR inbox with significantly greater level of detail. This methodology is a critical first step to inform decision makers on the redesign of workflow processes to eliminate unnecessary workload on PCPs and improve cost and quality of care. ■

### Acknowledgments

The funding of the first author was partially supported by the Susan Hockfield–Thomas Byrne Fellow in Clinometrics. The authors would like to thank Joshua Metlay, MD, PhD, chief of the Division of General Internal Medicine, Massachusetts General Hospital, and Kerri Palamara, MD, director of the Center for Physician Well-Being, Massachusetts General Hospital, for their help and consultation on the project. The authors would also like to thank the editor and the anonymous reviewers for their constructive comments that helped in substantially improving the paper substance and exposition.

**Author Affiliations:** Operations Research Center (CE) and Sloan School of Management (RL), Massachusetts Institute of Technology, Cambridge, MA; Division of General Internal Medicine (SAE) and Pulmonary/Critical Care Division (WJO), Massachusetts General Hospital and Harvard Medical School, Boston, MA.

**Source of Funding:** This work was funded by a Massachusetts General Hospital grant (grant no. 220150) through the Susan Hockfield – Thomas Byrne Fellow in Clinometrics.

**Author Disclosures:** Ms Escribe was funded by a grant from Massachusetts General Hospital. The remaining authors report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article.

**Authorship Information:** Concept and design (CE, SAE, WJO, RL); acquisition of data (CE, SAE, WJO, RL); analysis and interpretation of data (CE, SAE, WJO, RL); drafting of the manuscript (CE, SAE, WJO, RL); critical revision of the manuscript for important intellectual content (CE, SAE, WJO, RL); statistical analysis (CE, SAE, RL); and obtaining funding (WJO, RL).

**Address Correspondence to:** Retsef Levi, PhD, Sloan School of Management, Massachusetts Institute of Technology, Operations Research Center, 100 Main St, E62-416, Cambridge, MA 02142. Email: retsef@mit.edu

## REFERENCES

- 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. doi:10.1370/afm.2121
- Young RA, Burge SK, Kumar KA, Wilson JM, Ortiz DF. A time-motion study of primary care physicians' work in the electronic health record era. *Fam Med*. 2018;50(2):91-99. doi:10.22454/FamMed.2018.184803
- Tai-Seale M, Olson CW, Li J, et al. Electronic health record logs indicate that physicians split time evenly between seeing patients and desktop medicine. *Health Aff (Millwood)*. 2017;36(4):655-662. doi:10.1377/hlthaff.2016.0811
- Erickson SM, Rockwern B, Koltov M, McLean RM; Medical Practice and Quality Committee of the American College of Physicians. Putting patients first by reducing administrative tasks in health care: a position paper of the American College of Physicians. *Ann Intern Med*. 2017;166(9):659-661. doi:10.7326/M16-2697
- Howard J, Clark EC, Friedman A, et al. Electronic health record impact on work burden in small, unaffiliated, community-based primary care practices. *J Gen Intern Med*. 2013;28(1):107-113. doi:10.1007/s11606-012-2192-4
- Adler-Milstein J, Zhao W, Willard-Grace R, Knox M, Grumbach K. Electronic health records and burnout: time spent on the electronic health record after hours and message volume associated with exhaustion but not with cynicism among primary care clinicians. *J Am Med Inform Assoc*. 2020;27(4):531-538. doi:10.1093/jamia/oc220
- Shanafelt TD, Dyrbye LN, Sinsky C, et al. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. *Mayo Clin Proc*. 2016;91(7):836-848. doi:10.1016/j.mayocp.2016.05.007
- 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. doi:10.1136/amiainl-2013-001875
- 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. doi:10.1093/jamia/ocy145
- Gilchrist V, McCord G, Schrop SL, et al. Physician activities during time out of the examination room. *Ann Fam Med*. 2005;3(6):494-499. doi:10.1370/afm.391
- Sinsky C, Colligan L, Li L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Ann Intern Med*. 2016;165(11):753-760. doi:10.7326/M16-0961
- Sinsky CA, Rule A, Cohen G, et al. Metrics for assessing physician activity using electronic health record log data. *J Am Med Inform Assoc*. 2020;27(4):639-643. doi:10.1093/jamia/oc223
- McMahon LF Jr, Rize K, Irby-Johnson N, Chopra V. Designed to fail? the future of primary care. *J Gen Intern Med*. 2021;36(2):515-517. doi:10.1007/s11606-020-06077-6
- Murphy DR, Reis B, Sittig DF, Singh H. Notifications received by primary care practitioners in electronic health records: a taxonomy and time analysis. *Am J Med*. 2012;125(2):209.e1-e7. doi:10.1016/j.amjmed.2011.07.029
- Murphy DR, Reis B, Kadiyala H, et al. Electronic health record-based messages to primary care providers: valuable information or just noise? *Arch Intern Med*. 2012;172(3):283-285. doi:10.1001/archinternmed.2011.740
- McDonald CJ, McDonald MH. Electronic medical records and preserving primary care physicians' time: comment on "electronic health record-based messages to primary care providers." *Arch Intern Med*. 2012;172(3):285-287. doi:10.1001/archinternmed.2011.1678
- Tai-Seale M, Dillon EC, Yang Y, et al. Physicians' well-being linked to in-basket messages generated by algorithms in electronic health records. *Health Aff (Millwood)*. 2019;38(7):1073-1078. doi:10.1377/hlthaff.2018.05509
- Murphy DR, Meyer AND, Russo E, Sittig DF, Wei L, Singh H. The burden of inbox notifications in commercial electronic health records. *JAMA Intern Med*. 2016;176(4):559-560. doi:10.1001/jamainternmed.2016.0209
- Heisey-Grove D, Rathert C, McClelland LE, Jackson K, DeShazo JP. Patient and clinician characteristics associated with secure message content: retrospective cohort study. *J Med Internet Res*. 2021;23(8):e26650. doi:10.2196/26650
- Heisey-Grove DM, McClelland LE, Rathert C, Tartaglia A, Jackson K, DeShazo JP. Associations between patient health outcomes and secure message content exchanged between patients and clinicians: retrospective cohort study. *J Med Internet Res*. 2020;22(10):e19477. doi:10.2196/19477
- Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res*. 2003;3:993-1022.
- Hall D, Jurafsky D, Manning CD. Studying the history of ideas using topic models. In: *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics; 2008:363-371. Accessed February 19, 2021. <https://www.aclweb.org/anthology/D08-1038>
- Chang J, Boyd-Graber J, Wang C, Gerrish S, Blei DM. Reading tea leaves: how humans interpret topic models. In: *NIPS '09: Proceedings of the 22nd International Conference on Neural Information Processing Systems*. Citeseer; 2009:288-296.
- Sinsky CA, Willard-Grace R, Schutzbank AM, Sinsky TA, Margolius D, Bodenheimer T. In search of joy in practice: a report of 23 high-functioning primary care practices. *Ann Fam Med*. 2013;11(3):272-278. doi:10.1370/afm.1531
- Sarkar U, Karter AJ, Liu JY, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. *J Am Med Inform Assoc*. 2011;18(3):318-321. doi:10.1136/jamia.2010.006015
- Fareed N, Swoboda CM, Jonnalagadda P, Huerta TR. Persistent digital divide in health-related internet use among cancer survivors: findings from the Health Information National Trends Survey, 2003-2018. *J Cancer Surviv*. 2021;15(1):87-98. doi:10.1007/s11764-020-00913-8

Visit [ajmc.com/link/88817](https://ajmc.com/link/88817) to download PDF and eAppendix

## **eAppendix Exhibit 1: Topic Selection Methodology**

Defining the number of topics by expert assessment is a common practice,<sup>1</sup> as using likelihood-based comparisons of models to optimize the number of topics (e.g., perplexity metric) has been shown to lead to models with less interpretable topics.<sup>2</sup> It was demonstrated that traditional metrics (e.g., perplexity metric) do not capture whether topics are coherent or not, suggesting that practitioners developing topic models should focus on evaluations that depend on real-world task performance rather than optimizing likelihood-based measures.<sup>2</sup> It is therefore more reliable to select the number of topics in the LDA model based on expert assessment rather than by using traditional metrics.

This is what was done in this study. The number of topics given as input to the LDA model was selected manually, retrospectively, among a list of potential number of topics (5, 10, 15, 20, 25, 30, 35, 40). For each specified number of topics, an LDA model was computed. Each topic from the model and its corresponding word distribution were then explored manually by two experts. Topic meaning was subjectively interpreted from the topic-word distribution by these experts (e.g., the words ‘Skin, red, cream, rash, face’ suggest a topic related to skin symptoms). Based on this interpretation, a topic was deemed coherent when its theme was salient and distinguishable (e.g., the words ‘cough, insurance, cholesterol, surgery’ correspond to a non-coherent topic based on expert assessment). The quality of a model was then estimated based on the proportion of coherent topics among all topics. Finally, the LDA model with the highest proportion of coherent topics was selected as the best model, and its corresponding number of topics was selected as the best input number of topics.

## **eAppendix Exhibit 2: Methodology to Evaluate the Accuracy of the LDA Algorithm**

There is no automated way to perform model validation, as this study's approach is an unsupervised approach applied to unlabeled documents. Therefore, out-of-sample manual validation of the mixture of topics was implemented with the help of two physicians for 100 randomly selected messages that were held out from the training set. Each of the two experienced physicians reviewed the 100 messages and selected independently a list of topics representing the wholeness of the message. The accuracy between the LDA-selected themes and the themes selected by each physician is then calculated (see below for the specific methodology). The final accuracy was calculated as the average between the accuracies calculated based on each physician.

The methodology used to evaluate the accuracy of the predictions made by the algorithm for a given message is presented below in more details.

Errors made by the algorithm when predicting a topic are not given the same importance. For example, an error confusing the *Upper-Respiratory* topic with the *Paperwork* topic is considered more significant and is weighted higher than one that confuses the *Test Result Information* topic with the *Blood Test Topic*. Therefore, for each topic, close topics in terms of their main theme are selected among all the topics. For example, the close topics of *Test Result Information* topic will be *Blood Test* topic and *Imaging Exam* topic. The complete methodology used to evaluate the accuracy of the prediction made by the algorithm for a given message is presented below:

1. Initialize accuracy at 0.
2. Initialize the considered topic as the highest-percentage topic of the topic
3. For the considered topic of the message:
  - a. if the topic is considered present in the message by the physician:
    - i. if its associated percentage is considered accurate, the given percentage is added to the accuracy
    - ii. if the associated percentage is considered too large, only the deemed accurate percentage is added to the accuracy.
    - iii. if the associated percentage is considered too small in comparison to the accurate percentage, only the predicted percentage is added to the accuracy.

- b. If the topic is not considered present in the message by the physician:
  - i. if one of its close topics is considered present in the message, then half of the corresponding predicted percentage is added to the accuracy.
  - ii. if none of its close topics are considered present in the message, do nothing.
- 4. Iterate step 2 for all topics present in the message, including the none-specific topic.

### eAppendix Exhibit 3: Message Analysis

#### *Method*

To understand the output of the LDA model and the role of the different topics, it is important to explore if a message is most often described by a single topic, or by a mixture of different topics. To that end, all messages were reviewed by selecting their main (highest percentage) topic and the corresponding percentage. A relatively high percentage for the main topic of a message (e.g., 60%) implies that the message is primarily focused on this topic. On the contrary, a low percentage for the highest-percentage topic of a message (e.g., 20%) implies that the message includes a mixture of equally important topics. The statistics describing the percentage of the main topic across all messages were then calculated. The analysis was also performed at the topic level. For each topic, messages where the main topic corresponds to the given topic were selected, and statistics describing the main percentage were derived similarly.

#### *Results*

The mean percentage of the main topic across all messages was equal to 42.8% (standard deviation (SD) 16.6), and the median was equal to 40%. This implies that while it is uncommon to have a single topic representing a message, the main topic often has a significant weight in the message. The reported standard deviation implies that there is relatively high variability across the messages in terms of weight of the main topic. The same statistics at the topic level are represented in the following table. It should be noted that while some topics (e.g., *General Paperwork*) will often be the primary topic included in a message, others (e.g., *Musculoskeletal Symptoms*) will more often be discussed together other significant topics in a message.

<b>Topic description</b>	<b>Mean of the percentage of the main topic</b>	<b>Standard deviation of the percentage of the main topic</b>
Imaging Exam	0.09	0.13
Miscellaneous Chronic Condition	0.34	0.13
Patient-Initiated Care	0.37	0.13
Nutrition	0.38	0.14
Medical Referral	0.42	0.15

Scheduling Appointment	0.48	0.17
Upper-Respiratory Symptom	0.40	0.13
Skin Symptom	0.37	0.13
Scheduling Uncompleted Calls	0.48	0.17
Ambiguous Feelings	0.40	0.14
Prescription Administrative Problem	0.47	0.17
Family Referral	0.37	0.14
Joint Procedure	0.34	0.12
Musculoskeletal Symptoms	0.37	0.13
Administrative Referral	0.51	0.17
Insurance	0.39	0.14
General Paperwork	0.48	0.15
Medication	0.44	0.15
Blood Test	0.45	0.16
Refill Coordination	0.47	0.16
Miscellaneous Identified Symptoms	0.38	0.13
Respiratory Symptoms	0.45	0.13
Ambiguous Symptoms	0.38	0.13
General Check-In	0.43	0.16
Forms	0.40	0.14
Test Result	0.45	0.17
Feet-Related Problem	0.37	0.13
Miscellaneous Paperwork	0.36	0.12
Chronic Cardiovascular and Diabetes Condition	0.35	0.13
Surgical Procedure	0.36	0.13

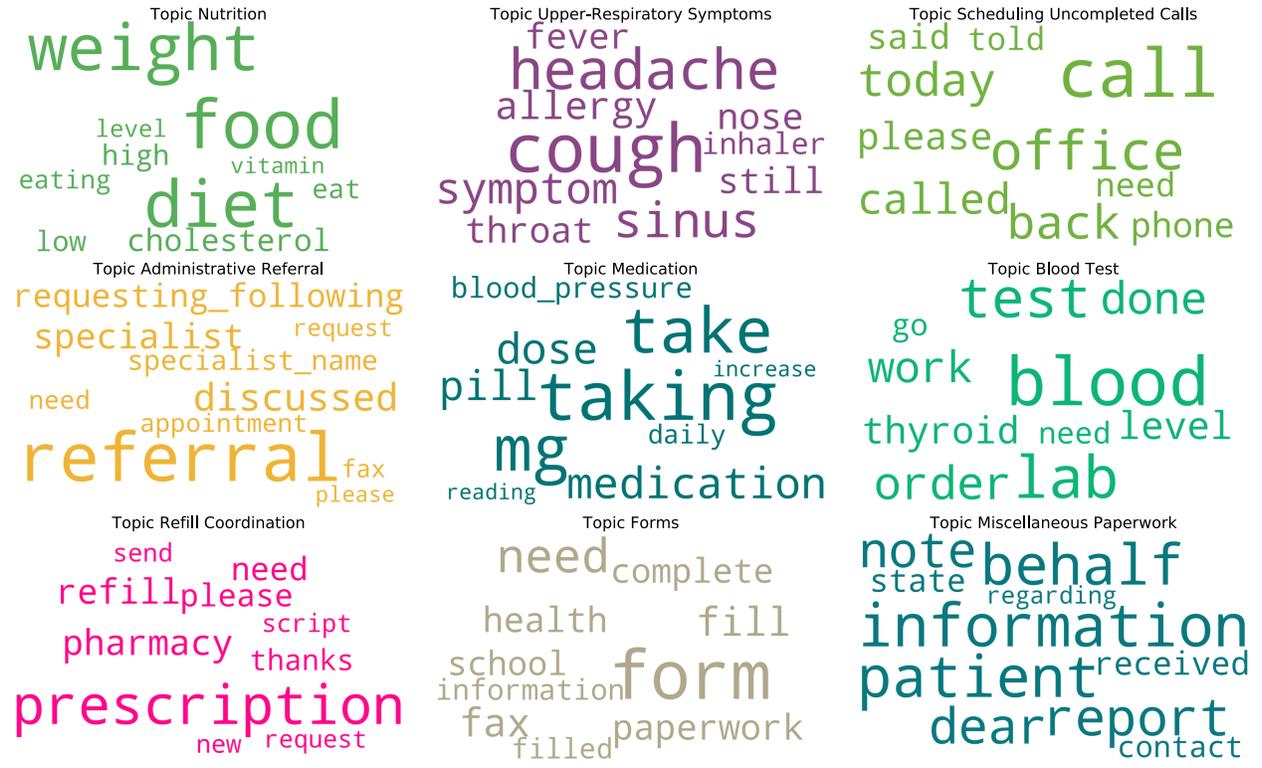
Legend: Description of the Percentage of the Main Topic per Topic.

#### eAppendix Exhibit 4: Topic Labelling and Word Distribution

<b>Topic Description</b>	<b>Top Ten Most Frequent Words</b>
Imaging Exam	Scan, ultrasound, breast, cyst, biopsy, cancer, liver, imaging, nodule, kidney
Miscellaneous Chronic Condition	Physical therapy, cardiologist, stress, heart, condition, change, ongoing, diagnosis, facility, therapy
Patient-Initiated Care	Sleep, cold, try, help, counter, using, take, medication, allergy, night
Nutrition	Food, diet, weight, eat, cholesterol, low, high, level, vitamin, supplement, eating
Medical Referral	See, help, could, someone, think, doctor, need, recommend, specialist, refer
Scheduling Appointment	Appointment, see, schedule, visit, office, call, need, time, appt, follow
Upper-Respiratory Symptom	Cough, headache, sinus, symptom, allergy, nose, throat, fever, inhaler, still
Skin Symptom	Skin, red, cream, rash, face, eye, area, picture, dermatologist, itchy
Scheduling Uncompleted Calls	Call, office, back, today, please, phone, need, told, called, said
Ambiguous Feelings	Think, time, feel, see, going, better, feeling, something, night, try
Prescription Administrative Problem	Mg, example, prescription, day, send, medication, tablet, name, pharmacy, supply
Family Referral	Mother, care, mom, patient, family, told, support, husband, friend, year
Joint Procedure	Knee, hip, surgery, car, right, replacement, injury, exercise, walking, accident
Musculoskeletal Symptoms	Hand, neck, right, arm, shoulder, left, pain, side, finger, nerve
Administrative Referral	Referral, discussed, specialist, requesting following, specialist name, need, appointment, fax, request, please

Insurance	Insurance, shot, vaccine, cover, pay, insurance company, new, cost, coverage, cancer
General Paperwork	Non urgent, letter, need, question, send, email, medical, record, work, mail
Medication	Taking, mg, medication, dose, pill, blood pressure, daily, increase, reading, take
Blood Test	Blood, lab, test, order, done, work, level, thyroid, need, go
Refill Coordination	Prescription, pharmacy, refill, need, thanks, script, send, new, request, please
Miscellaneous Identified Symptoms	Stool, diarrhea, symptom, discharge, white, negative, virus, smear, blood, bacterial
Respiratory Symptoms	Antibiotic, ear, chest, infection, lung, pneumonia, ray, prednisone, lesion, patch
Ambiguous Symptoms	Pain, back, symptom, still, leg, feel, non-urgent, muscle, lower, worse
General Check-In	Well, hope, good, great, weekend, best, nice, thanks, see, time
Forms	Forms, need, fill, fax, paperwork, health, complete, school, information, filled
Test Result	Result, test, question, normal, resulted, urine, blood, sample, lab, negative
Feet-Related Problem	Foot, xray, bone, arthritis, toe, fracture, flare, big, needle, big
Miscellaneous Paperwork	Information, patient, behalf, report, note, received, contact, state, dear, regarding
Chronic Cardiovascular and Diabetes Condition	Risk, dl, period, year, unit, age, insulin, treatment, machine, attack
Surgical Procedure	Surgery, procedure, surgeon, orthopedic, endoscopy, month, post, reflux, finish, need

**eAppendix Exhibit 5: Word Clouds of the Top Ten Most Frequent Words for a Selection of Nine Topics**



## eAppendix Exhibit 6: Sample Messages and Corresponding Topic Mixtures

Sample messages are presented with the topic mixtures and corresponding percentages predicted by the algorithm.

*“Dr. X, I want to keep you informed about my current state. Since Friday I have been taking Celexa 30 mg at dinnertime. I seem to be experiencing periods of 'jitteriness' on and off throughout the day. You have described that feeling as a side effect which should 'pass in a few days'. I'm able to manage the feeling so far but wanted you to know that it was occurring. Thank you for your help and support.”*

Topic mixture: Ambiguous Feelings (60%) + Medication (25%) + none-specific topic (15%)

*“Patient states the pain on her right arm is getting worse. She was taking tylenol max mg/day, 6 times a day but she stopped because she thought she was taking too much of it. Denies swelling. Hurts at all times. It also hurts from above the shoulder, on the right side. She can move the right arm, but it is very painful. Pain is also taking away her appetite.”*

Topic mixture: Ambiguous Pain Symptom (50%) + Musculoskeletal Symptom (27%) + Medication (17%) + none-specific topic (6%)

*"I think I'm getting a sinus infection. I have very thick yellow mucus coming from my nose and I have some sinus pressure under my eyes and above. I was wondering if I should schedule an appointment with you in the coming days. Thanks"*

Topic mixture: Upper-Respiratory Symptoms (64%) + Scheduling Appointment (25%) + Miscellaneous Identified Symptoms (6%) + none-specific topic (5%)

*“Your labs are back, and for the most part normal. Your urinalysis and culture were negative. Your diabetes screen was negative. Your LDL cholesterol level came back elevated. As a reminder, LDL cholesterol, or "bad" cholesterol is affected by animal fat. That includes red meat and dairy products. Cutting back or eliminating these foods, should help lower your cholesterol, as well as increased exercise. Your Vitamin D level did come back low. Ideally your level should be greater than 20. I would recommend a daily supplement of Vitamin D 1000 U. I would*

*recommend a recheck of you level in 6-12 months. Please let me know if you have any questions or concerns regarding your test result(s).*" Topic mixture: Nutrition (50%) + Test Result (26%) + Blood Test (9%) + none-specific topic (15%)

*"Is it possible to send another letter confirming I had my annual exam with the full date of my appointment (MGH Complio rejected the letter you gave me yesterday)? I believe the date of my appointment was December 20, 2017. Thank you!"*

Topic mixture: Scheduling Appointment (65%) + General Paperwork (30%) + none-specific topic (5%)

*"Patient insurance will not cover methocarbamol (Robaxin) 500 mg TID Insurance will cover 750 mg tab. Please advise on possible prescribing of 750 mg "*

Topic mixture: Prescription Administrative Problem (60%) + Insurance (35%) + none-specific topic (5%)

## References

1. Hall D, Jurafsky D, Manning CD. Studying the History of Ideas Using Topic Models. In: *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics; 2008:363-371. Accessed February 19, 2021. <https://www.aclweb.org/anthology/D08-1038>
2. Chang J, Boyd-Graber J, Wang C, Gerrish S, Blei DM. Reading tea leaves: How humans interpret topic models. In: *Neural Information Processing Systems*. Vol 22. Citeseer; 2009:288-296.