The Value of Real Time Labor Market Information for Monitoring Health Workforce Demand:
A Case Study Examining Employer Demand for Health Information Technology Skills

February 2017

Benjamin A. Stubbs, MPH, Bianca K. Frogner, PhD, Susan M. Skillman, MS

CONTENTS:

KEY FINDINGS

We used data provided by a job search engine company, LinkUp, which provided an opportunity to peer inside the black box of data coding structures otherwise provided as a paid service by other RT-LMI vendors. The goal of this study was to understand the value and limitations of RT-LMI for monitoring health workforce demand, including allied health professions.

The following were key study findings:

- Over 1.4 million records had one or more of the occupations from our designated healthcare occupation terms, and approximately half had a job description that could be used to search for skills required by the employer.
- The percentage of records with a job title and a job description that referenced a specific HIT skill varied greatly by occupation, with most occupations having fewer than 10% of records containing a HIT skill from our list of search terms.
- The 5 occupations with the highest percentage of job ads that referenced a specific HIT skill were: medical records and health information technicians, 60.4% of records; health educators, 19.5%; medical and clinical laboratory technologists, 17.0%; podiatrists and optometrists, 13.0%; medical assistants, 12.1%.
- Among our seven domains of HIT skills searched, the "health IT (general)" domain, comprised of search terms such as health information, health information technology, IT, or information technology, was most commonly identified (37.7% of records), and "privacy and security" (e.g., data security, cyber security, and risk analysis) was the least common domain (0.5% of records).

The patterns we found suggest that healthcare employers are requesting a range of HIT skills across occupations. Use of these data requires some caution and work to refine the data mining process. While continuing work is needed to improve the use of these data, knowing how RT-LMI best informs health workforce planning is valuable to ensure that the current and future health workforce have the training and education they need to succeed.

CENTER FOR HEALTH WORKFORCE STUDIES
UNIVERSITY of WASHINGTON

The Value of Real Time Labor Market Information for Monitoring Health Workforce Demand:
A Case Study Examining Employer Demand for Health Information Technology Skills

INTRODUCTION

As the United States healthcare system undergoes practice transformation, workforce planners and educators need up-to-date information on the skills employers expect of the future health workforce to ensure a properly trained pipeline of workers. Several administrative and survey data sources such as the American Community Survey (ACS) and Current Population Survey (CPS) are available to monitor the supply, distribution, and characteristics of the health workforce.[1] National surveys such as the Bureau of Labor Statistics (BLS) Job Openings and Labor Turnover Survey (JOLTS) provide estimates of the size of employer demand. These data have limitations, however, in their ability to inform workforce planners and educators about what employers require of the future workforce.

An emerging source of data, Real Time Labor Market Information (RT-LMI), is being used to monitor employer demand by extracting information from online job ads. Although RT-LMI is increasingly being used to track the skills in demand from the general labor market,[2] this type of data is a relatively new source for tracking changes in the health workforce. Understanding how frequently, and for which occupations, specific skills are being requested by healthcare employers is important information to guide education and training programs. This study examines the value of RT-LMI for this purpose: to describe the extent to which skills related to health information technology (HIT) are in demand by employers. Tracking changes in workforce skills and roles has been difficult using traditional sources of health workforce demand data. To explore this new data source, we used data provided by a job search engine company called LinkUp. Other than removing duplicate records and limiting the dataset to job ads they classified as falling into a "health and medical" category, LinkUp did not impose any additional restrictions or pre-code the data, which provided an opportunity to peer inside the black box of data coding structures typically provided as a paid service by other RT-LMI vendors. The goal of this study was to understand the value and limitations of RT-LMI for monitoring health workforce demand, including allied health professions.

MONITORING HEALTH WORKFORCE DEMAND CHANGES BROUGHT ON BY HEALTH INFORMATION TECHNOLOGY

It is known that implementation of HIT, particularly electronic health records (EHRs), affects the roles and skills required of current workers, in many cases requiring significant retraining.[3] Sufficient and appropriate staffing, training, and workflow issues have been consistent barriers to the adoption and effective implementation of EHRs.[4,5] In a recent survey of rural primary care providers, almost half of the respondents reported that a lack of qualified applicants was a barrier to the successful use of HIT even though primary care practices in rural areas have been adopting EHR technology at rates similar to urban practices.[6] Additionally, one-quarter to one-third of respondents reported the absence of community college or baccalaureate training programs as a barrier.[6] The Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 invested $116 million in community colleges and universities to train and educate the workforce to overcome these barriers and assist with the adoption of EHRs across the United States health system.[7]

While studies have identified a lack of HIT training and education as barriers to the successful adoption of EHRs, little is known about the specific skills and qualifications needed, primarily due to a paucity of data in this realm. Instead, studies mostly focus on how EHR adoption affects staffing levels, staff configurations, and the resulting productivity of the health workforce.[8,9] But given the limitations of available data, these studies are not able to identify the changing roles and shifting tasks that the health workforce experiences with the adoption of EHRs. Workforce planners, including those involved in industry training and the education of healthcare occupations, need up-to-date information about the skills required of healthcare workers in order to keep up with the rapidly changing technology.

Currently available data resources are not able to answer these questions. Labor market surveys are periodically conducted for specific occupations, but are often costly and have long lags between data collection and reporting. For example, the Occupational Information Network (O*NET) conducted by the U.S. Department of Labor Employment and Training Administration provides a warehouse of detailed job descriptions for approximately 1,000 occupations, including the knowledge, skills and abilities required of each occupation, based on regular surveys of employers. The limitations of this data source include its small sample size, costliness due to the complex survey design, and two- to three-year lag in reporting.[10] RT-LMI data, however, has the potential to be a new source of information about the changing skills that employers demand from the health workforce. This new source of workforce demand data has the potential to provide information about a large number of occupations, provide near real-time reporting, and is available at a relatively low cost.

REAL TIME LABOR MARKET INFORMATION (RT-LMI)

DATA AVAILABLE IN RT-LMI

Real Time Labor Market Information (RT-LMI) refers to the extraction of data on a regular basis from online job ads using an automated process called web crawling or "spidering." A vendor of RT-LMI searches a pre-determined set of websites, which is monitored and updated frequently to ensure complete and accurate data acquisition. This near real-time updating of activity in the labor market is particularly appealing to users compared with employer surveys and other forms of workforce demand data collection that result in significant lags between data collection and reporting.

Many vendors now aggregate and re-sell RT-LMI to interested stakeholders such as employers, educators, government agencies, researchers, and other workforce planners.[2,11] Vendors often compete based on their stated ability to "de-duplicate" job ads based on their own proprietary algorithms to identify job ads posted on multiple websites and their ability to account for unfilled or re-posted positions. Vendors vary in how they address obstacles to extracting these data, such as secured websites requiring logins to access job ads and job ads managed by contracting agencies that hinder the ability to discern the originating employer.

The information extracted from RT-LMI is limited to what is listed in a job ad, but at a minimum it tends to include job title, company, geographic information (e.g., ZIP code, city, state, and country), and job description. Vendors may take steps to code the job ad information in order to categorize job titles into nationally accepted structures such as the Standard Occupational Classification (SOC) System, and companies into the North American Industry Classification System (NAICS). These steps allow RT-LMI users to compare data from job ads to other data sources such as ACS, CPS, and JOLTS. Vendors may also parse the job description into searchable keywords such as skills, and education and training requirements. Additional information that vendors may glean from job ads is the length of job opening based on when a job is posted and when it was removed from the website, which is often interpreted as length of a job vacancy. Wage rates may also be extracted, though this information is not always listed.

THE USE AND CHALLENGES OF RT-LMI

Employers are a primary user group of RT-LMI. They use RT-LMI to understand which of the many online job advertisers would be most effective in providing qualified applicants (e.g., targeted local sites compared to large, national job boards) or to identify the phrases they should include to make their ads more searchable and effective in attracting applicants with the skills and experience they desire. Recently, researchers, educators and policymakers have started to use RT-LMI to complement the data contained in more traditional, survey-based labor market data sources. Several studies have used RT-LMI data to explore the effect of macroeconomic measures, such as unemployment rates, on hiring difficulty (as measured by the time that internet job ads were left open) or employee education and experience required by employers.[12,13] Another study examined the supply of graduates in biomedical research fields in ten major metropolitan areas compared to the demand nationally, with a focus on the skills that graduates should develop to become more competitive in the job market after graduation.[14]

In all the studies mentioned above, the authors supplemented the data found in traditional labor market sources, such as educational supply data, surveys of employers or data provided by the BLS, with information found in RT-LMI data. Most of these studies compared trends across different geographic locations in the United States, which was aided by location information associated with RT-LMI and the high volume of job ads from all areas of the country. In general, these studies chose to use RT-LMI data because it allowed the authors to investigate variables that are not available in other data sources or because RT-LMI data are more current. Almost all of these studies pointed out the importance of understanding the strengths and weaknesses of RT-LMI, and that RT-LMI should be used as a complement to traditional workforce demand data sources rather than as a replacement.

RT-LMI is not considered a replacement for traditional labor market data or other occupation/industry specific surveys due, in part, to several challenges that RT-LMI vendors face in their data coding process. Carnevale and colleagues conducted extensive quality control testing of data from one vendor (Burning Glass Technologies) and found that accuracy was generally high (above 80%) for location, occupation title, 2-digit occupation code, and skills.[15] In the same study, accuracy was 73% for 6-digit occupation code, 76% for major industry designation (2-digit NAICS codes) and "declined considerably in identifying detailed industries." In addition to the accuracy of coding, studies using RT-LMI may also be limited by the high rate of missing data for some variables. For example, in a study using a different subset of Burning Glass data, Rothwell noted that 55% of records contained education requirements, 52% contained experience requirements and only 7% contained salary information.[12]

Subject matter expertise is also needed to correctly map the information contained in job ads to NAICS and SOC, to identify relevant keywords, and to understand an industry's localized hiring practices (e.g., keeping job ads open for positions with high turnover, recruiting by word-of-mouth, or hiring internally). Carnevale and colleagues found that RT-LMI tends to bias towards large job announcement sites and geographical areas given that many smaller service-related businesses and those in rural areas may not post jobs online.[15] Even for industries that commonly use online job ads, there may be biases in the types of jobs that are advertised. In Carnevale and colleagues' review of RT-LMI data, they found an over-representation of high-skilled and high-wage jobs, and underrepresentation of less-skilled and low-wage jobs.[15] The review further estimated that 60% of all job openings are posted online, versus 80% of all jobs requiring a bachelor's degree or greater.[15]

RT-LMI TO UNDERSTAND HEALTH WORKFORCE DEMAND

In most studies using RT-LMI, healthcare has been one of multiple industries examined in analyses of the entire labor market. To date, only one peer-reviewed study has explicitly focused on the healthcare sector: Morgan and colleagues used RT-LMI to quantify employer demand for physician assistants in primary care settings compared with specialty practices.[16] Yet in spite of the paucity research on the health workforce using RT-LMI data, one study found that healthcare and social assistance jobs are generally overrepresented in these data when compared to JOLTS job openings,[15] which should make RT-LMI data an appealing source of information for health workforce planners and job developers.

DATA AND METHODS

DATA SOURCE

We obtained a dataset of online job ads from the job search engine company "LinkUp" for the mutually agreed upon purpose of this study.[17] LinkUp extracts information from company and government websites only (i.e., it does not search job aggregation websites), and claims to have one of the best de-duplication algorithms in the industry. For 2016, each day of data contains approximately three million active jobs from 50,000 employers across many industries. A recent report reviewed LinkUp data favorably because over half of the companies in the LinkUp data were present in a list of 3,000 publicly traded companies and because the percentage of LinkUp companies in each industry category was similar to the breakdown of industry categories in the list of publicly traded companies.[18]

We obtained LinkUp data for job ads posted in the fifty states and the District of Columbia during the 2015 calendar year. We requested data for jobs that LinkUp categorized as "health and medical" as defined by their proprietary algorithm (for example, a subcategory of health and medical jobs was "CNAs, aides, MAs, home health", and another subcategory was "healthcare support services"). For every job ad, LinkUp provided the text information identifying the following fields: unique job identifier, employer/company name, job title, city, state, zip code, county, date posted, date created, date checked by LinkUp, job ad website url, and job description. The job descriptions were delivered as unstructured text strings that required additional coding to identify keywords of interest.

Our study team developed a coding and parsing process to define the key variables of interest - occupation and HIT skills - rather than rely on algorithms developed by vendors such as LinkUp or Burning Glass. This allowed our team to precisely identify occupations of interest, and create a level of transparency that may not otherwise be available with vendor-constructed RT-LMI.

DEFINING OCCUPATIONS

Although LinkUp used the job title field as part of their algorithm to classify health and medical jobs, we did not use LinkUp's assigned job titles because they were not structured using the SOC system. Instead, our team developed a list of healthcare occupation titles derived from a detailed list of the 2010 SOC system[19] and then used string matching algorithms to code an occupation title derived from the job title field of each ad. We focused on occupations that fell under the two major groups of the SOC that are related to healthcare: 1) 29-0000 Healthcare Practitioners and Technical Occupations, and 2) 39-000 Personal Care and Service Occupations. Given that these occupations do not define the universe of healthcare occupations, we looked across the SOC to identify other relevant healthcare occupations including: 1) 21-1000 Counselors, Social Workers, and Other Community and Social Service Specialists, 2) 39-9020 Personal Care Aides, 3) 43-6013 Medical Secretaries, and 4) 51-9080 Medical, Dental, and Ophthalmic Laboratory Technicians. Occupations were grouped at the detailed level according to the SOC hierarchical system.

SOC 29-2070 Medical Record and Health Information Technicians was the only healthcare occupation that directly relates to HIT. We excluded medical and health services manager, medical scientists, computer, engineering, business, financial, office and administrative occupations that may have a role in using HIT and work in the healthcare industry. We excluded these occupations because our focus was on occupations that have direct involvement in patient care or directly support activities essential to patient care (such as review of the electronic health record, transcription of provider records or coding of records for reimbursement) to examine the extent to which employers were demanding HIT skills among occupations that have not traditionally been expected to have these skills.

As part of our iterative process for mapping job titles to occupations within the SOC, we examined a sample of LinkUp job titles that did not match to our occupation list. We found that several relevant occupations were missed by our initial use of the SOC classification due to the colloquial terms used for occupations such as "doctor" versus "physician" per SOC. We updated our occupation coding scheme to include these terms. We also noted that occupations in SOC incorporate qualifications into job titles such as "certified medical assistant," while job ads generally list the shortened title of "medical assistant" within the job title and list qualification details within the job description. In these cases, our final classification scheme should be understood to focus on the specific occupation and not the credential/qualification. We generally assumed fully delineated occupations rather than abbreviations with a few exceptions such as RN (registered nurse), LPN (licensed practical nurse), NP (nurse practitioner) and EMT (emergency medical technician), among others. In these few cases, both the abbreviation and the fully delineated term were included in the final classification scheme.

Occupation titles were not case-sensitive except in the use of abbreviations. We took steps to avoid erroneous categorization of job ads in which one occupation title is a subset of another occupation title (for example, where "physician" may be identified within the full occupation title of "physician assistant") by comparing the character position of the occupation title within the text string. We also allowed for minor variations in occupation titles not captured by SOC such as "nurses aide" versus "nurse aide" per SOC. Several emerging occupations not yet adopted by SOC such as patient navigator and care coordinator were not included in the final coding scheme for this study but will be included in a future study. A full list of the healthcare occupations used for this analysis is available upon request.

DEFINING HEALTH INFORMATION TECHNOLOGY SKILLS

Our study team, with input from a panel of outside experts, developed a list of HIT skills commonly employed within healthcare occupations for use in identifying the HIT-related skills mentioned as part of the healthcare job ads. We consulted the literature and online tools such as the Occupational Information Network (O*NET).[6, 20] We created a list of specific HIT skills, and categorized them into seven domains: health IT (general), application support, hardware and network support, database management, analytics, informatics, and privacy and security (see Appendix A for list of specific HIT skills and domains). We accounted for differences in the way HIT skills were advertised by including variants of search terms, such as "predictive analytics" and "predictive analysis." In some cases, it was necessary to account for capitalization and the context in which the search term was found. For example, "IT" could be interpreted as the word "it" or could be part of a larger word such as "additional." For this and other terms such as "EHR," "HIE" or "HIT" we specified that all the letters had to be capitalized and spaces had to follow the first and last letters. We did not impose capitalization or context rules for non-abbreviated search terms.

DATA ANALYSIS AND CODING VALIDATION

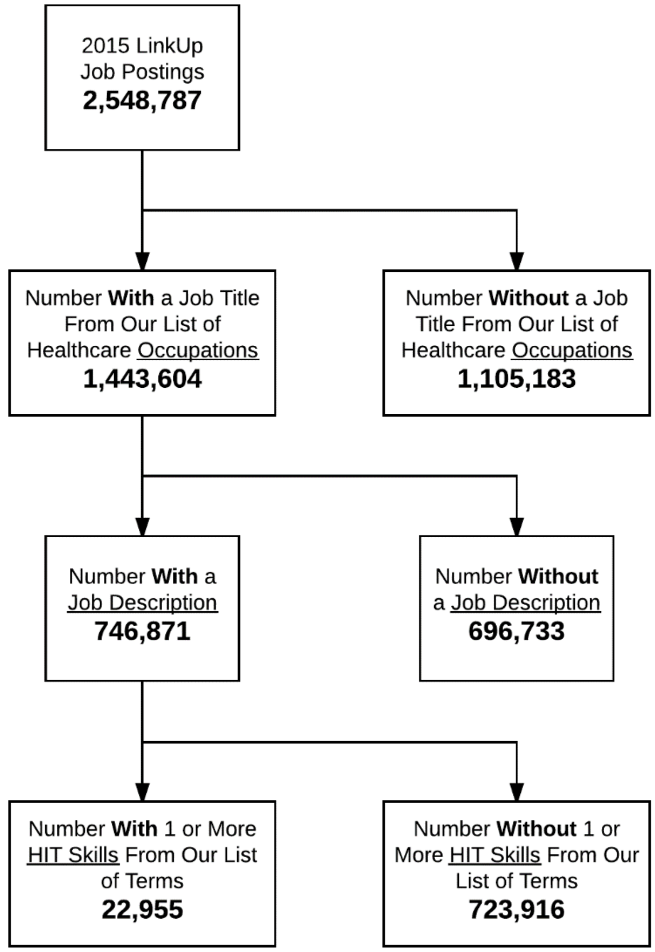
Once the coding schemes for occupations and HIT skills were finalized, we applied basic natural language processing algorithms using string matching to search for the terms within the fields provided by LinkUp. We searched for our detailed occupation list within the job title field and these identified occupations became the subset of job ads that were used to search for specific HIT skills. We searched unstructured job description text for the terms in our list of specific HIT skills. Multiple occupations and multiple HIT skills were allowed per job ad. Exact matches were required for occupation terms and HIT skill terms (case insensitive except as described above for abbreviations). We report the percent of job ads per occupation, the percent of job ads with at least one HIT skill as well as the distribution of HIT skills across the seven coding domains.

One member of our study team reviewed the full job description text for records from selected occupations (i.e., medical records and health information technicians, health educators, medical and clinical laboratory technologists, medical and clinical laboratory technicians, dispensing opticians, and nurse practitioners). A random set of 63 records was selected for review to assess the accuracy of the occupation title and HIT skills coding.

RESULTS

The 2015 LinkUp dataset of "health and medical" job ads contained 2,538,787 records. Of these, 1,443,604 records (56.9%) had one or more of the occupations from our list of search terms in the job title and of these, approximately half (746,871 records or 29.4% of the entire dataset) had a job description. Of the records with a defined healthcare occupation and a job description, 3.1% (22,955 records or 0.9% of the LinkUp dataset) contained one or more of our HIT search terms (Figure 1).

Figure 1: Number of Records with Selected Healthcare Occupations and Health Information Technology Skill in 2015



Of the approximately 1.4 million job ads with one of our designated occupation terms, registered nurses (RNs) was the most common occupation advertised at 43.6% of these job ads (Table 1). The high frequency of RNs is on par with the fact that RNs are the most common single occupation within the health workforce.[21] Some occupations that appeared at higher frequencies were generally more common healthcare occupations such as physicians/surgeons, licensed practical/vocational nurses, medical assistants, and nursing assistants. Occupations that appeared at relatively higher frequency among the job ads than they are represented in the current health workforce supply, such as pharmacy technicians and physical therapists, most likely reflect high relative demand for the occupation.

Although we found 746,871 unique job ads matching our occupation list, some ads listed multiple occupations within a job title, multiple specific HIT skills within a job description or both. We considered each of the multiple occupations and HIT skills as a unique record. As a result, our dataset expanded to 873,209 records for our analysis of specific HIT skills. Of the 873,209 records, 3.3% (28,594 records) contained one of our specific HIT skills (see Table 1). The percentage of records that referenced a specific HIT skill varied greatly by occupation, with most occupations having fewer than 10% of records containing a HIT skill from our list of search terms. Exceptions included: medical records and health information technicians, 60.4% of records; health educators, 19.5%; medical and clinical laboratory technologists, 17.0%; podiatrists and optometrists, 13.0%; medical assistants, 12.1%; medical transcriptionists, 11.2%; and physicians and surgeons, 10.4%. There were 9 out of 67 occupations with no HIT search terms, although some of these occupations, such as medical secretaries, ophthalmic medical technicians and occupational therapy aides, had relatively few records in the dataset to begin with.

Of the 28,594 records with HIT skills terms, "health IT (general)" domain was the most common at 38.0% (10,782 records) (see Appendix B for count information). This domain is comprised of search terms such as health information, health information

Table 1: Occupations for Which Health Information Technology Skills Were Identified

Occupation	Records with at Least 1 HIT Skill	Total Numberof Records with Occupation Title	Percent of Records With At Least 1 HIT Skill*	Percent of Job Ads Identifying This Occupation**
Medical Records and Health Information Technicians	197	326	60.4	0.03
Health Educators	17	87	19.5	0.01
Medical and Clinical Laboratory Technologists	148	873	17	0.10
Podiatrists	11	84	13.1	0.27
Optometrists	35	270	13	0.04
Medical Assistants	4,048	33,471	12.1	4.01
Medical Transcriptionists	61	547	11.2	0.06
Physicians and Surgeons	3,132	29,995	10.4	4.06
Healthcare Social Workers	1,283	15,151	8.5	2.04
Medical and Clinical Laboratory Technicians	36	544	6.6	0.06
Mental Health Counselors	49	741	6.6	0.11
Dental Assistants	285	4,675	6.1	0.64
Pharmacists	1,143	19,323	5.9	2.44
Opticians, Dispensing	6	103	5.8	0.01
Nurse Practitioners	1.056	19,759	5.3	2.73
Surgical Technologists	169	3,397	5	0.36
Physician Assistants	355	7,709	4.6	0.98
Community Health Workers	17	393	4.3	0.07
Nuclear Medicine Technologists	22	524	4.2	0.07
Substance Abuse and Behavioral Disorder Counselors	13	327	4	0.04
Medical Equipment Preparers	34	863	3.9	0.12
Dietitians and Nutritionists	214	5,481	3.9	0.77
Counselors	219	5,910	3.7	0.71
Nurse Midwives	7	221	3.2	0.02
Diagnostic Medical Sonographers	156	5,008	3.1	0.54
Emergency Medical Technicians and Paramedics	156	5,607	2.8	0.62
Exercise Physiologists	8	298	2.7	0.04
Athletic Trainers	4	162	2.5	0.02
Dental Hygienists	33	1,301	2.5	0.19
Respiratory Therapists	208	8,186	2.5	0.99
Registered Nurses	9,429	371,318	2.5	43.57
Radiation Therapists	18	753	2.4	0.09
Magnetic Resonance Imaging Technologists	86	3,648	2.4	0.40
Pharmacy Technicians	1,282	55,586	2.3	5.50
Licensed Practical and Licensed Vocational Nurses	1,839	78,665	2.3	7.36
Dentists	43	1,976	2.2	0.27
Audiologists	7	335	2.1	0.05
Physical Therapists	837	39,127	2.1	4.47
Nurse Anesthetists	14	703	2	0.12
Cardiovascular Technologists and Technicians	7	369	1.9	0.04
Radiologic Technologists	3	164	1.8	0.03
Phlebotomists	258	14,072	1.8	1.52
Orthotists and Prosthetists	6	384	1.6	0.03
Psychiatric Technicians	63	4,246	1.5	0.50
Nursing Assistants	1,140	74,841	1.5	7.10
Health Diagnosing and Treating Practitioners, All Other	2	141	1.4	0.01
Pharmacy Aides	14	1,237	1.1	0.14
Home Health Aides	42	4,162	1	0.47
Physical Therapist Aides	3	319	0.9	0.04
Orderlies	3	347	0.9	0.03
Occupational Therapists	153	16,991	0.9	2.00
Speech-Language Pathologists	108	12,157	0.9	1.57
Physical Therapist Assistants	56	6,638	0.8	0.89
Recreational Therapists	2	328	0.6	0.04
Dietetic Technicians	30	5,327	0.6	0.55
Health Technologists and Technicians, All Other	2	425	0.5	0.05
Occupational Therapy Assistants	21	4,291	0.5	0.61
Massage Therapists	4	2,744	0.1	0.30
Medical Secretaries	0	1	0	<0.01
Ophthalmic Medical Technicians	0	14	0	<0.01
Occupational Therapy Aides	0	16	0	<0.01
Social and Human Service Assistants	0	28	0	<0.01
Medical Appliance Technicians	0	29	0	<0.01
Psychiatric Aides	0	31	0	0.01
Respiratory Therapy Technicians	0	94	0	0.01
Dental Laboratory Technicians	0	123	0	0.01
Personal Care Aides	0	243	0	0.03
Total	28,594	873,209	3.3	100.00

*Denominator is the number of job ads with a job title and a job description. Each job ad could contain multiple occupations or multiple HIT skills (N = 873,209)
**Denominator is the number of job ads with a matching occupation in the job title (N = 1,443,604)

technology, IT, or information technology (see Appendix A for details). The second most common domain (appearing in 27.0% of the records with HIT skills terms) was "application support" (skills primarily relating to EHR software). The "privacy and security" domain (e.g., data security, cyber security, and risk analysis) was the least-represented domain (0.5%). Of the 58 occupations with records that referenced a HIT skill from our list of search terms, 3 occupations referenced skills that belonged to 1 domain, 8 occupations referenced skills that belonged to 2 domains, 10 occupations referenced skills that belonged to 3 domains, 7 occupations referenced skills that belonged to 4 domains, 12 occupations referenced skills that belonged to 5 domains, 7 occupations referenced skills that belonged to 6 domains and 11 occupations referenced skills that belonged to all 7 domains.

Figure 2 shows, for records containing one or more HIT search terms, the percentage of records falling within each HIT domain, by occupation. Across occupations, the specific HIT skills requested by employers most often fell into the "health IT (general)" domain. Two occupations—medical and clinical laboratory technologists, and dental assistants—were exceptions in that HIT skills most often fell into the "analytics" domain, representing approximately half of the records requesting HIT skills within these occupations. Within the "application support" domain, the following occupations had relatively high percentages of job ads requesting this set of skills: optometrists, medical assistants, dental assistants, nurse practitioners, physician assistants, dentists, nursing assistants and physical therapist assistants. Among other notable findings, approximately half of the records for the occupations of healthcare social workers, mental health counselors and magnetic resonance imaging technologists requested skills in the "database management" domain. Also, approximately half of the records for surgical technologists and phlebotomists requested skills that fell into the "hardware and network support" domain.

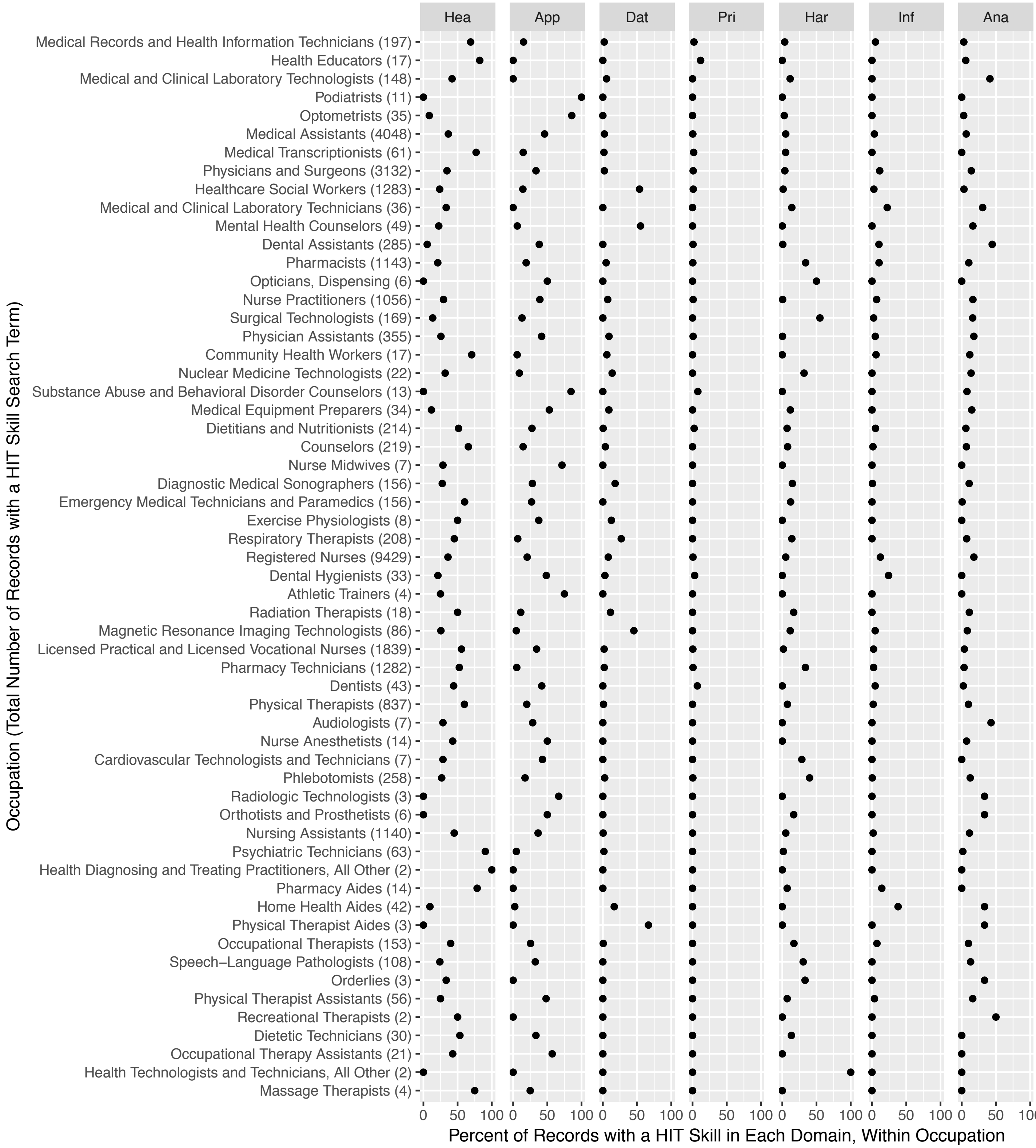
We reviewed 63 records to assess the accuracy of our final coding schemes. We found that 98% of the records (62 out of 63) had the correct occupation coded. In addition, 70% of all records reviewed had the correct coding for the HIT skill required in the job description, but the accuracy varied by occupation. For example, 33% (3 of 9) of records reviewed for dispensing opticians were correctly coded compared to 50% (5 of 10) of the records reviewed for medical and clinical laboratory technicians, and 100% (all 10) of records reviewed for medical records and health information technicians.

DISCUSSION

RT-LMI has value for monitoring the skills requested by employers of the health workforce, but use of these data requires some caution and work to refine the data mining process. This study was successful in using RT-LMI from one major job search engine, LinkUp, to examine over 2.5 million job ads from the health and medical field posted in one year across the country. From these records, we identified 1.4 million job ads that mapped to nearly all of the occupations in our SOC-derived detailed occupation list. Even though the number of job ads that included specific health HIT skills matching our list of search terms was relatively small compared to the overall number of 2015 job ads in the LinkUp dataset, the resulting set of 22,955 job ads provides a large pool of records representing a diverse set of healthcare occupations. The patterns we found suggest that healthcare employers are requesting a range of HIT skills across occupations.

Given constraints on our time and resources for this study, we used a convenience sample of job descriptions using a subset of jobs identified as "health and medical" jobs per LinkUp's proprietary algorithm. In the future, the detailed list of health occupations developed in this study could be searched against LinkUp's full database of job ads. It is hard to determine the extent to which this convenience sample led to any bias towards particular occupations as we had no a priori hypothesis about the distribution of specific healthcare occupations found in RT-LMI. Given prior studies of RT-LMI, however, we assume a slight bias may also exist towards job ads from larger health systems that may rely more heavily on online job ads to recruit healthcare workers, so generalizability is limited.

Figure 2: For Records Containing HIT Search Terms, Percentage of Records Falling Within Each HIT Domain, by Occupation



Excludes occupations with no records with HIT search terms. Hea = health IT (general), App = application support, Dat = database management, Pri = privacy and security, Har = hardware and network support, Inf = informatics, Ana = Analytics.

While in aggregate, across healthcare occupations and HIT skills, we found a relatively large number of records requiring HIT skills from potential employees, the number was small within specific occupations and within specific HIT skill domains. We hypothesize that job advertisements do not always list the full range of HIT skills required by employers. Using medical records and health information technicians as an example, if the job description were to include all responsibilities of the position, we would have expected that all of the job ads would have at least one HIT skill listed in the job description. Using another example, two primary tasks that O*NET lists for RNs are to "record patients' medical information and vital signs" and "monitor, record, and report symptoms or changes in patients' conditions."[22] In addition, O*NET lists fifteen different EHR systems that RNs listed among required technical skills, indicating that the patient monitoring tasks are primarily performed using EHRs. And yet only 2.5% of the records for RNs in our study included one or more of our HIT skills, and of those, only 20% fell into the "application support" domain (which mostly includes terms related to the use of EHRs). Therefore, it is likely that for some occupations, employers do not routinely advertise preferences for HIT skills because these skills are understood to be part of the worker's day-to-day activities. This represents a limitation to using RT-LMI data from the healthcare industry to track demand for HIT-related skills among the health workforce, at least for occupations for which the demand is well established.

A considerable challenge in using RT-LMI related to a specific industry is having the subject matter expertise to establish the initial coding structure required to correctly classify records. The coding process requires several iterations to ensure that the coding structure has content and face validity. For example, in our study, we found that medical records and health information technicians had the highest percentage of records with matching HIT skills, which is what we would have expected. Alternatively, in a detailed review of job descriptions for medical and clinical laboratory technologists, we found that, in many cases, the term "analytics" was used to refer to the analysis of biological samples (e.g., blood and tissue sample) rather than analysis of information technology. Similarly, in a review of job descriptions for health educators, we found the term "health information" within the terms "private health information," indicating that the job applicant would be expected to understand how to protect sensitive health documents, and "provide public health information [to patients]," which does not necessarily imply the use of information technology. While in these cases we correctly identified skills from our vetted list of relevant terms, the context was not consistent with our study's intent and resulted in some level of noise in our analysis. Despite our careful iterative process to develop an accurate coding scheme, these misclassification issues remained, indicating that further refinement of our search terms or employing more advanced search methods that account for the context in which search terms are found may have led to more accurate or more complete classification of the HIT-related terms in these job ads. In our coding validation, 70% of the records we reviewed were coded correctly for the advertised HIT skill, indicating a good level of overall accuracy, with some occupations found to have been coded more accurately than others.

An additional limitation in the text classification approach used in this study was the duplicate count of occupations in some job ads. However, the percentage of affected records was low – multiple occupations were identified within the job title in less than 2% of the total number of job ads. We treated each of these job titles as an independent observation because it was not clear without detailed review of the job description which of the duplicated job titles was primary or if the ad was for multiple occupations. Additionally, each job ad could contain more than one HIT skill from our list of search terms. As such, we counted as separate records: 1) job ads with more than one of our occupation search terms in the job title, and 2) job ads with more than one specific HIT skill in the job description.

AREAS FOR FUTURE HEALTH WORKFORCE RESEARCH USING RT-LMI

In our text classification strategy, we performed basic string searches established using rules regarding punctuation, capitalization, and word positioning to identify occupations and required HIT skills within online job ads. With this work as a foundation, as well as additional work classifying text relating to other topics often found in job ads (such as experience, education and job setting), supervised machine learning (ML) – in which pre-classified text is used to automatically build computer algorithms to search and classify large quantities of unclassified text – could be used to identify other relevant terms and to more accurately capture the context in which terms are used.[23] For example, ML may provide information about skills that are listed as being required versus preferred by the employer. Also, with ML, one may be able to identify the degrees, certifications, and other related expertise (e.g., specific EHR software such as Epic or experience with meaningful use criteria) that are not explicit HIT skills but often appear in conjunction with these skills and could help refine understanding of the HIT-related aspects of the job postings.

Identifying the setting in which HIT skills are requested by occupation needs further work. Educators may want to know whether physical therapists, for example, need specialized HIT skills training to work in a hospital versus an ambulatory care setting. Setting and industry are often not explicitly described within a job ad. Although company name is typically provided, company names tend not to be descriptive. Our study team developed a list of settings such as hospitals, clinics, and nursing homes, and attempted to search for these terms within the company name field, job title field, and job description. In a review of the identified settings, over half appeared to be erroneous. As an example, the job description field may describe a job that is located in a clinic that is part of a larger health system that networks with other ambulatory care and long-term care settings. Our search criteria did not accurately classify the setting as a clinic compared to an ambulatory or long-term care facility. Similarly, geography is also a challenge given that some healthcare settings span across multiple cities and states, making it difficult to identify the particular city and state of the job position.

Another area that requires future attention is using RT-LMI to monitor vacancies over time among healthcare occupations, and the extent to which vacancies exist due to a lack of unqualified applicants with the necessary HIT skills. RT-LMI provides information about when a job is posted and when it is pulled from the website by the vendor. The difference in dates provides a proxy for the duration of a job opening. Several authors who have studied the ability of RT-LMI data to estimate job openings and hires have found that RT-LMI data, while representing a clear undercount, are correlated with job openings or hires from other sources.[15,24] The causal link between job openings, hirings, and changes in unemployment is not entirely clear, according to one report.[15] Detailed work to investigate the extent to which job vacancies correlate to actual job market activity within healthcare is needed, and particularly valuable given that available data sources such as JOLTS do not provide a specific lens on vacancies in healthcare.

CONCLUSIONS

In conclusion, our study finds value in using RT-LMI data as an addition to the toolkit for monitoring health workforce demand. Caution is warranted in interpreting data from RT-LMI given that accuracy of this information relies on the parsing and coding practices of the vendors that aggregate online data as well as the extent to which the goals of employers writing job postings correspond to the goals of the study using these data. Effective use of the information within job ads requires knowledge about the recruitment practices of employers, which varies across industries. RT-LMI may be a useful complement to other available data sources for monitoring health workforce demand trends, but requires further comparisons across data sources to understand the relationships among these resources. In particularly, work is needed to identify which occupations may be over- or underestimated in each data source, particularly among low- and middle-skilled occupations that are known to be in high demand, but for which less comparison and historical data are available. While continuing work is needed to improve our use of these data, knowing how RT-LMI best informs health workforce planning is valuable to ensure that the current and future health workforce have the training and education they need to succeed.

AUTHORS

Benjamin A. Stubbs, MPH, Center for Health Workforce Studies, University of Washington
Bianca K. Frogner, PhD, Center for Health Workforce Studies, University of Washington
Susan M. Skillman, MS, Center for Health Workforce Studies, University of Washington

FUNDING

This study was supported by the National Center for Health Workforce Analysis (NCHWA), Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services (HHS) under cooperative agreement #U81HP27844. The information, conclusions and opinions expressed in this report are those of the authors and no endorsement by NCHWA, HRSA or HHS is intended or should be inferred.

ACKNOWLEDGEMENTS

We appreciate the data and related in-kind assistance provided by Toby Dayton (CEO), Rochelle Dickinson, Brad Squibb, and Lee Buermann at LinkUp. We also appreciate the data management provided by Steve Senter, RN, consultant at the University of Washington (UW) Institute of Translational Health Sciences, and data analysis provided by Nikki Gurley, UW graduate student. We thank Mary McCaskill and Ann Lefebvre, MSW, CPHQ at North Carolina Area Health Education Centers Program, and Rachel Machta from University of North Carolina for input into the list of health IT skills. We also thank Susan Fenton, PhD, RHI, FAHIMA, Associate Professor and Associate Dean for Academic Affairs at The University of Texas Health Science Center at Houston, School of Biomedical Informatics for feedback on results from this report. Lastly, we thank Anne Basye for her support in editing.

SUGGESTED CITATION

Stubbs BA, Frogner BK, Skillman SM. *The Value of Real Time Labor Market Information for Monitoring the Allied Health Workforce: A Case Study of Demand for Health Information Technology Skills*. Center for Health Workforce Studies, University of Washington, Feb 2017.

REFERENCES

1. Skillman SM, Dahal A, Frogner BK, Stubbs BA. Leveraging Data to Monitor the Allied Health Workforce: National Supply Estimates Using Different Data Sources. Center for Health Workforce Studies, University of Washington. Dec 2016.
2. Maher & Maher. Real-Time Labor Market Information: An Environmental Scan of Vendors and Workforce Development Users. Report in collaboration with Jobs for the Future & New York City Labor Market Information Service. September 2014. http://www.jff.org/sites/default/files/publications/materials/Real%20Time%20Labor%20Market%20Information.pdf. Accessed December 8, 2016.
3. Masys DR. Effects of Current and Future Information Technologies on the Health Care Workforce. *Health Affairs*. 2002;21(5): 33-41.
4. Gabriel MH, Jones EB, Samy L, King J. Progress and Challenges: Implementation and Use of Health Information Technology among Critical-Access Hospitals. *Health Affairs* 2014;33(7): 1262-1270.
5. Heisey-Grove D, Danehy LN, Consolazio M, Lynch K, Mostashari F. A National Study of Challenge to Electronic Health Record Adoption and Meaningful Use. *Medical Care*. 2014; 52(2): 144-148.
6. Skillman SM, Andrilla CHA, Patterson DG, Fenton SH, Ostergard SJ. Health Information Technology Workforce Needs of Rural Primary Care Practices. *Journal of Rural Health*. 2015;31(1):58-66.
7. US Department of Health and Human Services, Office of the National Coordinator for Health Information Technology. Get the Facts about Health IT Workforce Development Program. https://www.healthit.gov/sites/default/files/get_the_facts_workforce_development.pdf. Accessed March 6, 2017.
8. Frogner BK, Wu X, Ku L, Pittman P, and Masselink LE. Do Years of Experience with Electronic Health Records Matter for Productivity in Community Health Centers? *Journal of Ambulatory Care Management*. 2017;40(1):36-47.

9. Frogner BK, Wu X, Park J, and Pittman P. The Association of Electronic Health Record Adoption with Staffing Mix in Community Health Centers. *Health Services Research*. 2017; 52(S1): 407-421. DOI: 10.1111/1475-6773.12648.
10. O*NET Resource Center. http://www.onetcenter.org/overview.html. Accessed December 8, 2016..
11. Dorrer J, Milfort M. Vendor Product Review: A Consumer's Guide to Real-Time Labor Market Information. Jobs for the Future. April 2012. http://www.jff.org/sites/default/files/publications/VendorProductReview_041712.pdf. Accessed December 8, 2016.
12. Rothwell J. Education, Job Openings, and Unemployment in Metropolitan America. Brooking Institution. August 2012. https://www.brookings.edu/wp-content/uploads/2016/06/29-education-gap-rothwell.pdf. Accessed December 8, 2016.
13. Modestino AS, Shoag D, Ballance J. Downskilling: Changes in Employer Skill Requirements Over the Business Cycle. *Labour Economics*. 2016;41:333-347.
14. Mason JL, Johnston E, Berndt S, Segal K, Lei M, Wiest JS. Labor and Skills Gap Analysis of the Biomedical Research Workforce. *FASEB J*. 2016;30(8):2673-83.
15. Carnevale A, Jayasundera T, Repnikov D. Understanding Online Job Ads Data: A Technical Report. April 2014. https://cew.georgetown.edu/wp-content/uploads/2014/11/OCLM.Tech_.Web_.pdf. Accessed December 8, 2016.
16. Morgan P, Himmerick KA, Leach B, Dieter P, Everett C. Scarcity of Primary Care Positions May Divert Physician Assistants Into Specialty Practice. *Medical Care Research and Review*. 2016 Feb 4 [ePub] pii: 1077558715627555..
17. LinkUp Job Search Engine. http://www.linkup.com. Accessed December 8, 2016.
18. Jussa J, Zao G, Luo Y, Alvarez MA, Wang S, Rohal G, Wang A, Elledge D, Webster K. Macro and Micro JobEnomics: Gleaning Alpha and Macro Insights from Job Postings. Deutsche Bank Markets Research. May 2015.
19. Bureau of Labor Statistics, U.S. Department of Labor. 2010 SOC Downloadable Materials: Direct Match Title File, 2010 SOC. https://www.bls.gov/soc/soc_2010_direct_match_title_file.xls. Accessed December 8, 2016.
20. Fenton SH, Joost E, Gongora J, Patterson DG, Andrilla CHA, Skillman SM. Health Information Technology Employer Needs Survey: An Assessment Instrument for Workforce Planning. *Educ Perspect Health Inform Inf Manage*. 2013;Winter:1-36.
21. Bureau of Labor Statistics, U.S. Department of Labor. Occupational Outlook Handbook: Registered Nurses: Work Environment. https://www.bls.gov/ooh/healthcare/registered-nurses.htm#tab-3. Accessed December 8, 2016.
22. U.S. Department of Labor Employment and Training Administration O*NET Online. Summary Report for: 29-1141.00 - Registered Nurses. https://www.onetonline.org/link/summary/29-1141.00. Accessed March 17, 2016.
23. Sebastiani F. Machine Learning in Automated Text Categorization. *ACM Computin Surv*. 2002;34(1):1-47.
24. Templin T, Hirsch L. Do Online Job Ads Predict Hiring? New York City Labor Market Information Services. February 2013 https://www.gc.cuny.edu/CUNY_GC/media/365-Images/Uploads%20for%20LMIS/Reports%20and%20Briefs/NYCLMIS-RESEARCH-BRIEF-Do-Online-Ads-Predict-Hiring.pdf. Accessed December 8, 2016.

APPENDIX A: HEALTH INFORMATION TECHNOLOGY DOMAINS AND CORRESPONDING SEARCH TERMS

HIT Skills Domain	Search Term
Hea = Health IT (General)	Health information
	Health information technology
	Health information management
	Health IT
	HIT*
	Information technology
	IT*
App = Application Support	Application specialist
	Electronic health record
	EHR*
	Personal health record
	Software support
	application manager
	application analyst
	implementation specialist
	EHR specialist
	computerized physician order entry
	clinical decision support
Dat = Database Management	Data management
	Data aggregation
Pri = Privacy and Security	Data security
	Cyber security
	Information security
	risk analysis
Har = Hardware & Network Support	Technical support
Inf = Informatics	Informatics
	Population health
	Report builder
	Health information exchange
	HIE*
	Certified professional in health information and management systems
	Certified associate in health information and management systems
	HL7
	Health level 7
	continuity of care document
	direct messaging
	patient portal implementation
	patient portal administration

*All letters capitalized and spaces before and after exact phrase

APPENDIX A: *continued*

HIT Skills Domain	Search Term
Ana = Analytics	Analytics
	Data analysis
	Data analytics
	Projective analysis
	Projective analytics
	Predictive analysis
	Predictive analytics

*All letters capitalized and spaces before and after exact phrase

APPENDIX B: FOR JOB ADS WITH A JOB TITLE FROM OUR LIST OF HEALTHCARE OCCUPATIONS AND A JOB DESCRIPTION, NUMBER OF RECORDS FALLING WITHIN EACH HIT DOMAIN, BY OCCUPATION

Occupation	Ana	App	Dat	Har	Hea	Inf	Pri	None	Total
Medical Records and Health Information Technicians	6	30	4	7	136	10	4	129	326
Health Educators	1	0	0	0	14	0	2	70	87
Medical and Clinical Laboratory Technologists	61	0	8	17	62	0	0	725	873
Podiatrists	0	11	0	0	0	0	0	73	84
Optometrists	1	30	0	1	3	0	0	235	270
Medical Assistants	265	1,862	91	202	1,475	139	14	29,423	33,471
Medical Transcriptionists	0	9	1	3	47	0	1	486	547
Physicians and Surgeons	438	1,047	70	117	1,079	353	28	26,863	29,995
Healthcare Social Workers	42	183	688	16	306	35	13	13,868	15,151
Medical and Clinical Laboratory Technicians	11	0	0	5	12	8	0	508	544
Mental Health Counselors	8	3	27	0	11	0	0	692	741
Dental Assistants	127	109	0	2	16	29	2	4,390	4,675
Pharmacists	118	218	55	390	240	118	4	18,180	19,323
Opticians, Dispensing	0	3	0	3	0	0	0	97	103
Nurse Practitioners	172	413	75	5	310	72	9	18,703	19,759
Surgical Technologists	27	22	0	93	23	4	0	3,228	3,397
Physician Assistants	63	148	32	1	91	17	3	7,354	7,709
Community Health Workers	2	1	1	0	12	1	0	376	393
Nuclear Medicine Technologists	3	2	3	7	7	0	0	502	524
Substance Abuse and Behavioral Disorder Counselors	1	11	0	0	0	0	1	314	327
Medical Equipment Preparers	5	18	3	4	4	0	0	829	863
Dietitians and Nutritionists	13	59	1	15	110	11	5	5,267	5,481
Counselors	15	32	8	17	144	3	0	5,691	5,910
Nurse Midwives	0	5	0	0	2	0	0	214	221
Diagnostic Medical Sonographers	17	44	28	23	43	1	0	4,852	5,008
Emergency Medical Technicians and Paramedics	1	42	0	19	94	0	0	5,451	5,607
Exercise Physiologists	0	3	1	0	4	0	0	290	298
Athletic Trainers	0	3	0	0	1	0	0	158	162
Dental Hygienists	0	16	1	0	7	8	1	1,268	1,301
Respiratory Therapists	15	14	56	29	94	0	0	7,978	8,186
Registered Nurses	1,674	1,934	748	475	3,389	1,152	57	361,889	371,318
Radiation Therapists	2	2	2	3	9	0	0	735	753
Magnetic Resonance Imaging Technologists	7	4	39	10	22	4	0	3,562	3,648

Job ads could have multiple occupations or multiple HIT skills. Hea = health IT (general), App = application support, Dat = database management, Pri = privacy and security, Har = hardware and network support, Inf = informatics, Ana = Analytics.

APPENDIX B: *continued*

Occupation	Ana	App	Dat	Har	Hea	Inf	Pri	None	Total
Pharmacy Technicians	44	68	26	433	673	29	9	54,304	55,586
Licensed Practical and Licensed Vocational Nurses	71	631	34	30	1,024	42	7	76,826	78,665
Dentists	1	18	0	0	19	2	3	1,933	1,976
Audiologists	3	2	0	0	2	0	0	328	335
Physical Therapists	82	167	8	63	502	15	0	38,290	39,127
Nurse Anesthetists	1	7	0	0	6	0	0	689	703
Cardiovascular Technologists and Technicians	0	3	0	2	2	0	0	362	369
Radiologic Technologists	1	2	0	0	0	0	0	161	164
Phlebotomists	32	45	7	103	69	1	1	13,814	14,072
Orthotists and Prosthetists	2	3	0	1	0	0	0	378	384
Psychiatric Technicians	1	3	1	1	57	0	0	4,183	4,246
Nursing Assistants	129	416	5	58	513	19	0	73,701	74,841
Health Diagnosing and Treating Practitioners, All Other	0	0	0	0	2	0	0	139	141
Pharmacy Aides	0	0	0	1	11	2	0	1,223	1,237
Home Health Aides	14	1	7	0	4	16	0	4,120	4,162
Physical Therapist Aides	1	0	2	0	0	0	0	316	319
Orderlies	1	0	0	1	1	0	0	344	347
Occupational Therapists	15	39	1	26	61	11	0	16,838	16,991
Speech-Language Pathologists	14	35	0	33	26	0	0	12,049	12,157
Physical Therapist Assistants	9	27	0	4	14	2	0	6,582	6,638
Recreational Therapists	1	0	0	0	1	0	0	326	328
Dietetic Technicians	0	10	0	4	16	0	0	5,297	5,327
Health Technologists and Technicians, All Other	0	0	0	2	0	0	0	423	425
Occupational Therapy Assistants	0	12	0	0	9	0	0	4,270	4,291
Massage Therapists	0	1	0	0	3	0	0	2,740	2,744
Medical Secretaries	0	0	0	0	0	0	0	1	1
Ophthalmic Medical Technicians	0	0	0	0	0	0	0	14	14
Occupational Therapy Aides	0	0	0	0	0	0	0	16	16
Social and Human Service Assistants	0	0	0	0	0	0	0	28	28
Medical Appliance Technicians	0	0	0	0	0	0	0	29	29
Psychiatric Aides	0	0	0	0	0	0	0	31	31
Respiratory Therapy Technicians	0	0	0	0	0	0	0	94	94
Dental Laboratory Technicians	0	0	0	0	0	0	0	123	123
Personal Care Aides	0	0	0	0	0	0	0	243	243
Total	3,517	7,768	2,033	2,226	10,782	2,104	164	844.615	873,209

Job ads could have multiple occupations or multiple HIT skills. Hea = health IT (general), App = application support, Dat = database management, Pri = privacy and security, Har = hardware and network support, Inf = informatics, Ana = Analytics.