<u>Original Paper</u>

A Data-Driven Algorithm to Recommend Initial Clinical Workup for Outpatient Specialty Referral: Algorithm Development and Validation Using Electronic Health Record Data and Expert Surveys

Wui Ip[1], MD; Priya Prahalad[1], MD, PhD; Jonathan Palma[2], MSc, MD; Jonathan H Chen[3,4], MD, PhD

[1]Department of Pediatrics, Stanford University School of Medicine, Palo Alto, CA, United States

[2]Neonatology & Perinatal Medicine, Orlando Health Winnie Palmer Hospital for Women & Babies, Orlando, FL, United States

[3]Department of Medicine, Stanford University School of Medicine, Palo Alto, CA, United States

[4]Stanford Center for Biomedical Informatics Research, Stanford, CA, United States

Corresponding Author:
Wui Ip, MD
Department of Pediatrics
Stanford University School of Medicine
453 Quarry Road, MC 5660
Palo Alto, CA, 94304
United States
Phone: 1 6507234000
Email: wui@stanford.edu

Abstract

Background: Millions of people have limited access to specialty care. The problem is exacerbated by ineffective specialty visits due to incomplete prereferral workup, leading to delays in diagnosis and treatment. Existing processes to guide prereferral diagnostic workup are labor-intensive (ie, building a consensus guideline between primary care doctors and specialists) and require the availability of the specialists (ie, electronic consultation).

Objective: Using pediatric endocrinology as an example, we develop a recommender algorithm to anticipate patients' initial workup needs at the time of specialty referral and compare it to a reference benchmark using the most common workup orders. We also evaluate the clinical appropriateness of the algorithm recommendations.

Methods: Electronic health record data were extracted from 3424 pediatric patients with new outpatient endocrinology referrals at an academic institution from 2015 to 2020. Using item co-occurrence statistics, we predicted the initial workup orders that would be entered by specialists and assessed the recommender's performance in a holdout data set based on what the specialists actually ordered. We surveyed endocrinologists to assess the clinical appropriateness of the predicted orders and to understand the initial workup process.

Results: Specialists (n=12) indicated that <50% of new patient referrals arrive with complete initial workup for common referral reasons. The algorithm achieved an area under the receiver operating characteristic curve of 0.95 (95% CI 0.95-0.96). Compared to a reference benchmark using the most common orders, precision and recall improved from 37% to 48% (*P*<.001) and from 27% to 39% (*P*<.001) for the top 4 recommendations, respectively. The top 4 recommendations generated for common referral conditions (abnormal thyroid studies, obesity, amenorrhea) were considered clinically appropriate the majority of the time by specialists surveyed and practice guidelines reviewed.

Conclusions: An item association–based recommender algorithm can predict appropriate specialists' workup orders with high discriminatory accuracy. This could support future clinical decision support tools to increase effectiveness and access to specialty referrals. Our study demonstrates important first steps toward a data-driven paradigm for outpatient specialty consultation with a tier of automated recommendations that proactively enable initial workup that would otherwise be delayed by awaiting an in-person visit.

XSL•FO
RenderX

KEYWORDS

recommender system; electronic health records; clinical decision support; specialty consultation; machine learning; EHR; algorithm; algorithm development; algorithm validation; automation; prediction; patient needs

Introduction

Background

There is a fundamental and growing gap between the supply and demand of medical expertise, as reflected in the projected shortage of 100,000 physicians by 2030 [1]. The problem is particularly acute for specialty care [2-6], for which over 25 million people in the United States have deficient access [7]. Wait times for in-person specialty visits commonly extend weeks to months after referrals are made [5]. Adding to this problem, essential initial workup is often not completed [8,9], resulting in ineffective visits when the specialists do not have sufficient information to make a definitive diagnosis and treatment recommendations by the time of their first in-person visit. Such inefficiency could lead to care delay, missed opportunity to provide access to more patients, and dissatisfaction of patients and families.

Ideally, referring providers could directly communicate with specialists for their preconsultation advice on an initial recommended clinical workup. However, data show that primary care providers are only able to communicate with specialists half of the time when referring patients [10]. Alternatively, primary care providers and specialists can collaboratively develop guidelines for initial workup [11], but this requires substantial manual effort to produce and maintain up-to-date content as new evidence arises and practice changes over time. Asynchronous electronic consults or synchronous telemedicine consults are emerging approaches for referring providers to solicit specialists' opinions on the need of referral and initial workup [12-16], with potential advantages of streamlining the referral process and empowering primary care providers. However, such consults remain limited in availability, as they still require a human consultant to review and respond to each request [17,18].

A more data-driven approach could boost the capacity of the health system by making initial specialty clinic visits more effective and by sparing the time required by specialists to communicate initial workup needs. Prior studies have shown the efficacy of statistical approaches, including association rules, Bayesian networks, logistic regression, and deep neural networks, for generating clinical order recommendations. The focus of these studies, however, has been primarily in the acute care settings such as inpatient hospitalization and emergency room visits [19-26].

Our aim is to develop a data-driven paradigm for outpatient specialty consultation with a tier of automated recommendations that proactively enable initial workup that would otherwise be delayed by awaiting an in-person visit. Taking advantage of electronic health records that contain thousands of specialist referral visits, we propose a data-driven algorithm inspired by Amazon's "customers who bought A also bought B" [27] to anticipate initial specialty evaluations at the time of referral based on how specialists cared for similar patients in the past. In this study, we chose pediatric endocrinology as a use case because laboratory evaluation is often required to inform specialist treatment recommendations [28-30].

Objective

Using specialty referrals to pediatric endocrinology as an example, we developed a recommender algorithm to anticipate initial workup needs for a variety of endocrine conditions. We compared the performance of the algorithm to a reference benchmark based upon the most common workup orders. We evaluated the need to complete initial workup and the clinical appropriateness of the algorithm recommendations by surveying specialists.

Methods

Recommender Algorithm Development

Deidentified structured electronic health record data from outpatient clinic visits at Stanford Children's Health were extracted from the Stanford Medicine Research Data Repository using the Observational Medical Outcomes Partnership (OMOP) common data model [31]. We include patients younger than 18 years with a pediatric endocrine referral order from any Stanford-affiliated clinic and a subsequent pediatric endocrine visit within 6 months. Between 2015 and 2020, 3424 patients met criteria, whose data yielded >1,150,000 instances of 8263 distinct clinical items.

We used OMOP common data model concepts to define distinct clinical items, including 2966 conditions, 2423 measurements (eg, lab results), 1187 procedures (eg, diagnostic imaging), and 1687 medications. Numeric laboratory results were categorized as "normal," "high," or "low" based upon reference ranges. We excluded clinical items that occurred in fewer than 10 patients.

Based on the timing of pediatric endocrinology referral, we split the patient cohort into a training set (referrals from 2015 to 2019: n=2842 patients) and a test set (referral in 2020: n=582 patients). In the training set, we calculated the co-occurrence statistics of pairs of clinical items to build an item association matrix (Figure 1). We counted duplicate items only once per patient to allow natural interpretation of patient prevalence and diagnostic measures.

Figure 1. Algorithm training and construction of the item co-occurrence matrix.
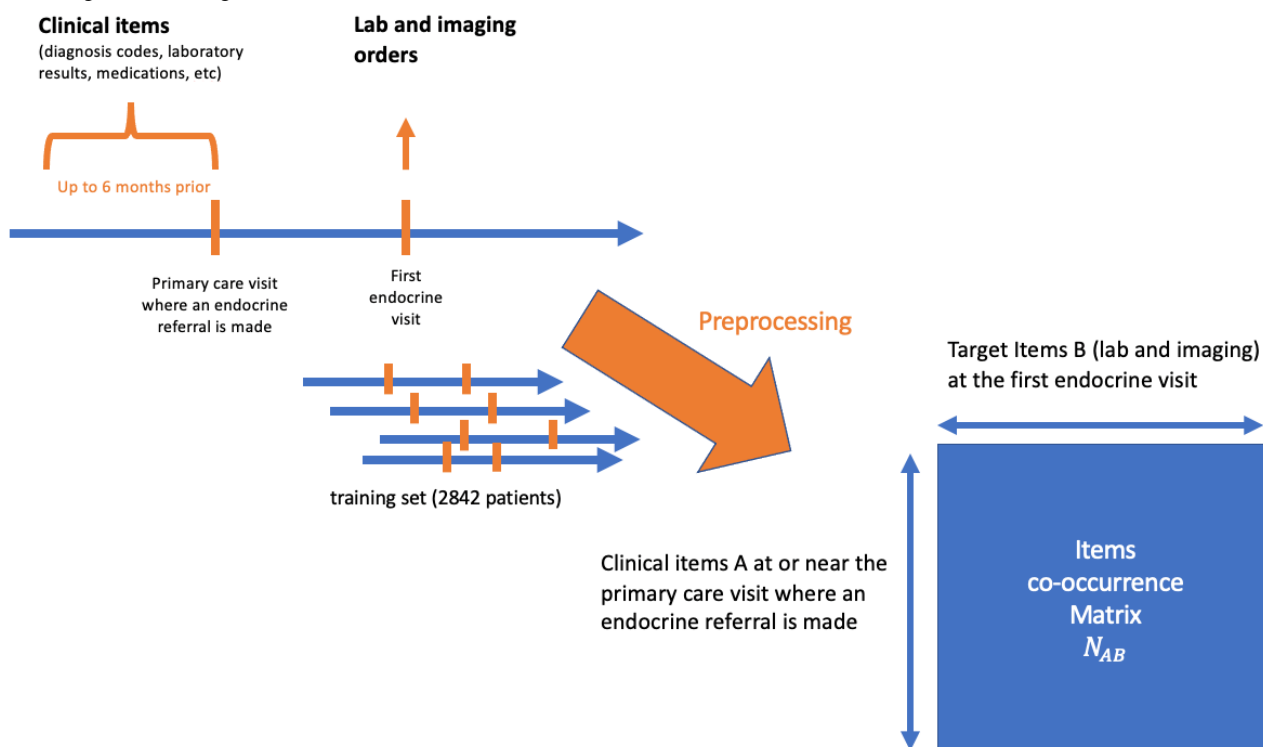
The recommender algorithm is queried with a patient's clinical items (diagnosis, labs, medications, etc) associated with the primary care encounter when the endocrine referral was placed. In addition, we included clinical items associated with the patient in the 6 months prior to the primary care encounter.

Using these clinical items (A_1,..., A_q), the recommender algorithm retrieves scores that resemble posttest probability from the co-occurrence association matrix for all possible target items at the subsequent endocrine visits. We limited the target items to laboratory and imaging orders to focus on diagnostic workup recommendations. For each query item (A), target items (B) are ranked by estimated posttest probability P(B|A), or positive predictive value (PPV), defined as the number of patients who have query item A followed by target item B (N_{AB}) divided by the number of patients with query item A (N_A).

$$PPV = \frac{N_{AB}}{N_A}$$

If a patient has q query items, q separate ranked lists are generated. To aggregate these results, we estimate total pseudo-counts using the following equation that sums across every i-th query item:

$$\frac{1}{\sum_{i=1}^{q} N_{A_i}} \sum_{i=1}^{q} W_{A_i} N_{A_i B}$$

W_A is a weighting factor for the query item. There are several ways one can model W_A. For instance, one can penalize common query items by the following expression:

$$W_A = \frac{1}{N_A}$$

Another method, inspired by a weighting strategy using item clustering based on genres [32], is to weigh a query item based on its relevance to the endocrine referral cohort by using a relative risk term:

$$W_A = RR_A$$

Where:

$$RR_A = \frac{\frac{N_A}{N_{endocrine}}}{\frac{N_{A_{outpt}} - N_A}{N_{outpt} - N_{endocrine}}}$$

The numerator is the prevalence of item A in our endocrine cohort ($N_{endocrine}$ is the total number of patients in our endocrine referral cohort, of which N_A patients have clinical item A). The denominator is the prevalence of item A outside of the endocrine cohort in all outpatient clinics (N_{outpt} is the total number of pediatric patients in all outpatient clinics, of which $N_{A_{outpt}}$ patients have item A).

Using 10-fold cross-validation in our training set, we evaluated these two strategies to model W_A individually and in combination ($W_A = \frac{RR_A}{N_A}$). We selected the W_A that gave the best prediction performance in the training data and subsequently used it in the test set.

The code can be accessed via GitHub [33].

Evaluation Using Electronic Health Record Test Set Data

To evaluate the performance of the recommender algorithm in the test set (Figure 2), we compared the recommended list of orders with the actual workup orders patients received at their first endocrine visit. We calculated the precision (PPV) and

recall (sensitivity) for the top 4 recommendations, and performed the receiver operating characteristics analysis. We chose the top 4 recommendations because 4 is the mean number of workup orders at the first endocrine visit. We calculated 95% CIs using 1000 bootstrap resamples [34].

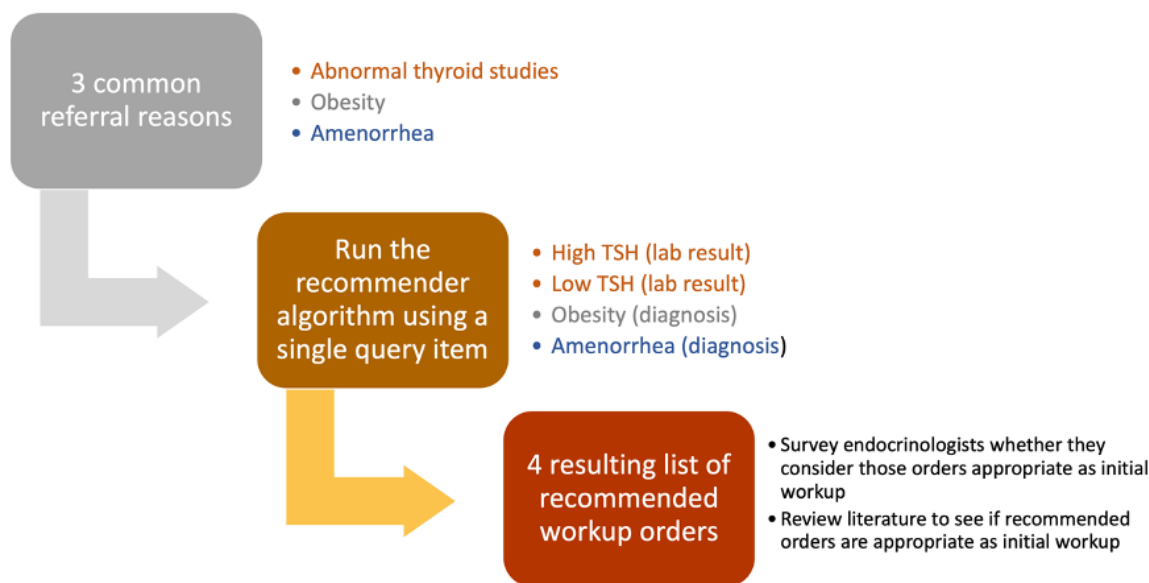**Figure 2.** Algorithm evaluation using the test set.
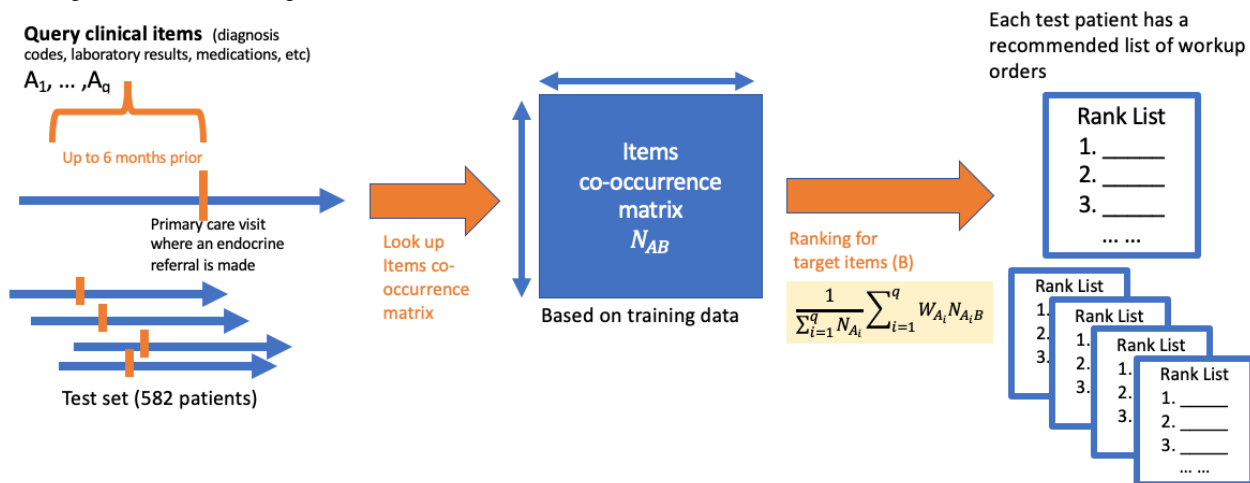
Evaluation Using Expert Surveys

To further understand whether the recommendations would be as clinically appropriate as the initial workup ordered by referring providers, we conducted a survey of all pediatric endocrinologists at Stanford Children's Health on three common referral reasons (abnormal thyroid studies, obesity, and amenorrhea). The survey was approved by the Institutional Review Board at Stanford University. Survey invitations were sent via emails in July 2020, and survey questions and informed consent were included in the supplemental material. For abnormal thyroid studies, we generated two lists of the top recommended workup orders—one queried with high thyroid stimulating hormone (TSH; an abnormal lab result suggesting hypothyroidism) and the other queried with low TSH (an abnormal lab result suggesting hyperthyroidism). For obesity and amenorrhea, we generated a list of the top recommended orders using the diagnosis as a single query item (Figure 3). Subsequently, we asked the endocrinologists to select the orders from the recommended lists that they considered clinically appropriate as initial workup for the corresponding condition. Other than the referral reasons, the endocrinologists received no other information related to the patients. We instructed them to define appropriate workup as workup that gives sufficient information for the endocrinologists to make concrete recommendations at the clinic visits. In addition, for each of the three conditions, we asked them how often initial workup is completed in their practice and how helpful it is if initial workup is completed prior to the first specialty visit. Lastly, we reviewed published literature and consensus guidelines [28,30,35-37] as external validation to assess whether the recommended orders represent a reasonable workup.

Figure 3. Evaluation of algorithm output by practicing endocrinologists. TSH: thyroid stimulating hormone.

Results

Evaluation Using Electronic Health Record Test Set Data

Table 1 compares the performance of the recommender algorithm with a reference benchmark using the most common orders in our endocrine referral cohort (endocrine prevalence).

The recommender algorithm had the best performance with an area under the receiver operating characteristic curve (AUC) of 0.95 (95% CI 0.95-0.96). Comparing with the reference benchmark, precision improved from 37% to 48% ($P<.001$), and recall improved from 27% to 39% ($P<.001$). The recommender algorithm is based on a weighting factor, $W_A = \frac{RR_A}{N_A}$, that resulted in the best cross-validation performance in our training data (Multimedia Appendix 1, Table S1).

Table 1. Recommender algorithm performance in the test set. Precision and recall were calculated at k=4, given that 4 is the average number of workup orders.

	Recommender[a]	Endocrine prevalence[b]	Outpatient prevalence[c]	Random[d]
Precision[e] (%; 95% CI)	48 (45-52)	37 (34-40)	10 (8-12)	2 (2-3)
Recall[f] (%; 95% CI)	39 (36-42)	27 (24-29)	5 (4-6)	2 (1-3)
AUC[g] (95% CI)	0.95 (0.95-0.96)	0.88 (0.87-0.89)	0.64 (0.62-0.66)	0.49 (0.47-0.5)

[a]Recommender: ranking workup orders using the recommender algorithm.

[b]Endocrine prevalence: ranking workup orders using the percentage of patients who had the orders in the endocrine referral cohort ($\frac{N_B}{N_{endocrine}}$) training set.

[c]Outpatient prevalence: ranking workup orders using the percentage of patients who had the orders among all outpatients ($\frac{N_{B_{outpt}}}{N_{outpt}}$).

[d]Random: random ranking of workup orders.

[e]Precision: positive predictive value (proportion of predictions that were correct).

[f]Recall: sensitivity (proportion of correct items that were predicted).

[g]AUC: area under the receiver operating characteristic curve.

Evaluation Using Expert Surveys

Of 14 pediatric endocrinologists at Stanford Children's Health, 12 (86%) responded to our survey on three common referral reasons (abnormal thyroid studies, obesity, and amenorrhea). Table 2 shows less than half of the patients coming to the specialty clinics with appropriate initial workup as estimated by the pediatric endocrinologists for each of the three referral reasons (each endocrinologist provided a value between 0% and 100%). The endocrinologists considered it as moderately to very helpful to have the initial workup completed prior to specialty visits (Table 2).

Table 3 shows the top recommendations based on the recommender algorithm using a single query item as mentioned in Figure 3. Each recommended workup order has a corresponding survey result showing the percentage of endocrinologists who considered the order clinically appropriate as the initial workup. Overall, the majority of the specialists considered the top four recommendations clinically appropriate in each of the lists.

Table 2. Estimated percentages of patients with initial workup completed before specialty visits and mean Likert scale score of how helpful (5: extremely helpful; 1: not helpful at all) it is to have initial workup completed before specialty visits.

	Value, mean (SD)
Estimated percentages of patients with initial workup completed before specialty visits (%)	
Abnormal thyroid studies	49 (21)
Obesity	45 (20)
Amenorrhea	37 (18)
Likert scale score of how helpful it is to have initial workup completed before specialty visits	
Abnormal thyroid studies	4.2 (0.8)
Obesity	3.3 (0.9)
Amenorrhea	4.1 (0.8)

Table 3. Top recommendations for patients referred for high thyroid stimulating hormone (TSH; commonly due to hypothyroidism), low TSH (commonly due to hypothyroidism), obesity, and amenorrhea.

Orders	PPV[a] (%)	Relative ratio[b]	Endocrine prevalence[c] (%)	Outpatient prevalence[d] (%)	Percent of endocrinologist considered appropriate
High TSH[e]					
TSH[f]	60.9	1.0	61.7	17.1	92
Free thyroxine[f]	60.3	1.2	56.8	7.5	100
Thyroglobulin antibody[f]	41.7	3.7	12.8	0.5	92
Thyroperoxidase antibody[f]	39.1	3.2	13.7	1.0	92
Vitamin D level	9.3	0.3	31.4	8.1	0
Serum cortisol	8.6	0.7	11.3	0.7	0
Low TSH[g]					
TSH[h]	61.5	1.0	61.7	17.1	75
Free thyroxine[h]	57.7	1.0	56.8	7.5	100
Thyroglobulin antibody[h]	50.0	4.0	12.8	0.5	67
Thyroperoxidase antibody[h]	46.2	3.4	13.7	1.0	58
Total tri-iodothyronine[h]	42.3	16.1	3.0	0.7	92
Comprehensive metabolic panel	26.9	0.5	53.8	23.1	8
Obesity[i]					
Hemoglobin A_{1c}[j]	40.2	1.9	22.5	12.5	100
TSH	28.0	0.4	61.7	17.1	75
Free thyroxine	25.6	0.4	56.8	7.5	42
Comprehensive metabolic panel[j]	25.6	0.5	53.8	23.1	92
Lipid panel[j]	25.0	1.4	18.3	12.9	100
Vitamin D level	20.7	0.6	31.4	8.1	42
Amenorrhea[k]					
Prolactin[l]	41.4	4.8	8.9	0.4	92
Luteinizing hormone[l]	37.9	2.2	17.5	0.9	100
Follicle stimulating hormone[l]	34.5	2.1	16.5	1.7	100
Estradiol	27.6	4.7	6.1	0.4	100
17 Hydroxy-progesterone	24.1	2.5	9.7	0.2	100
Dehydroepi-androsterone sulfate	17.2	2.2	8.0	0.4	92

[a]PPV: positive predictive value.

[b]Relative ratio: the ratio of the probability of the order given the query item to the probability of the order given the lack of the query item.

[c]Endocrine prevalence: the percentage of patients who had the orders in the endocrine referral cohort ($\frac{N_B}{N_{endocrine}}$).

[d]Outpatient prevalence: the percentage of patients who had the orders among all outpatients ($\frac{N_{B_{outpt}}}{N_{outpt}}$).

[e]The top four orders are considered clinically appropriate by almost all of the endocrinologists and are recommended based on published guidelines. The fifth and sixth recommended items have relatively low PPV.

[f]Recommended based on guidelines [36].

[g]Here, the top five orders are considered clinically appropriate by most endocrinologists and published guidelines.

XSL•FO
RenderX

Discussion

Significance

Using pediatric endocrinology as an example, we developed and evaluated a recommender algorithm to anticipate initial workup needs at the time of specialty referral. The algorithm can predict appropriate specialist's workup orders with high discriminatory accuracy with an AUC>0.9. Our survey shows that, among the three common referral reasons, less than half of the patients typically have appropriate initial workup prior to their initial specialist visit. Most specialists agree that having initial workup completed prior to the first clinic visit is helpful and that our algorithm recommendations for the three referral conditions are clinically appropriate. This supports the potential utility of a data-driven recommender algorithm for referring providers. Although we illustrated 3 common referral conditions in this study, the algorithmic approach is general, and it could be broadly applied to other referral reasons or other specialties, with the benefit of personalization based on individual patient patterns of clinical items, including the combination of multiple conditions.

Although this algorithm is not suitable for full automation given the level of precision and recall, such an algorithm could serve as a clinical decision support tool [38-41] by displaying relevant clinical orders for referring providers to make the referral process more effective. One can imagine coupling this clinical decision support tool with electronic consultation so that specialists can quickly confirm the workup orders in the recommended list, thus augmenting the efficiency of the specialists and increasing their capacity to care for more patients. Advantages of an algorithmic decision support tool compared to building consensus guidelines among specialists [11,42] include scalability to answer unlimited queries on demand, maintainability through automated statistical learning, adaptability to respond to evolving clinical practices [43], and personalizability of individual suggestions with greater accuracy than manually authored checklists [43-45].

Different from our prior recommender algorithm for the inpatient setting [19], we applied a weighting factor to each query item based on its relevance to a specialty and its inverse frequency. The motivation is that inpatient clinical items are often related to acute reasons of hospitalization, while outpatient clinical items vary in scope, ranging from health maintenance or chronic disease management to treatment of urgent care issues. We show that differentially weighting query items significantly improves the performance of the recommender algorithm in both precision and recall. This makes intuitive sense because common clinical items seen in primary care clinics that are irrelevant to endocrinology likely provide less predictive power. A similar weighting scheme could be applied

to other recommender algorithms when the clinical use case is specialty specific.

The association rule mining methods shown here are relatively simple to implement with interpretable results and associated statistics. Other approaches including Bayesian networks [21] and deep machine learning [46] are computationally more complex with less interpretable results. Although future research should compare these different methods, our focus primarily is to demonstrate the applicability of a data-driven approach in workup recommendations for specialty referral.

Although we ranked the recommended workup items based on PPV as shown in Table 3, we have also provided alternative metrics such as relative ratio, which could be used to look for less common but more specific or "interesting" items for a given query. For instance, in Table 3, total tri-iodothyronine had a relative ratio of 16.1, suggesting this is highly specific for patients with low TSH (indicating hyperthyroidism). In comparison, free thyroxine ranks higher based on its PPV but has a relative ratio of 1.0, suggesting this is not specific for patients with low TSH. Indeed, we observed free thyroxine also appeared in the list of recommendations for patients with high TSH (Table 3).

For a crowdsourcing clinical decision support solution like recommender algorithms, a typical concern is that recommendations drawn from common practices do not necessarily imply clinical appropriateness. To address this concern, we solicited specialist opinions on the algorithm outputs. Overall, the majority of the top recommendations were considered clinically appropriate as initial workup by the specialists. We also performed external validation by reviewing relevant guidelines, which revealed general agreement with the specialists' assessments.

Limitations

Limitations in this study include that the algorithm was developed at a single institution, requiring future work to expand to other institutions to evaluate generalizability. However, the algorithmic framework is general, as we used a common data model with data schema and features that were not institution specific. Second, in recommender systems such as ours, there is a well-known cold start problem when there is a lack of clinical items. Our algorithm starts with a generic "best seller list" by using the cohort item prevalence, but the algorithm could rapidly bootstrap itself with even just a couple of clinical items such as diagnosis codes or laboratory results to dynamically converge on recommendations specific to the patient scenario. Third, our cohort definition relied on referral orders placed in the electronic health records, potentially failing to capture patients who were referred to specialty clinics by other means (eg, fax or phone communication). Additionally,

structured data in the electronic health records such as diagnosis codes or problem lists are often optimized for billing purpose and may be incomplete. Future research should investigate whether using unstructured text and leveraging natural language processing in clinical notes could further optimize the algorithm performance [47]. Fourth, our survey results are limited to three common referral conditions; further validation on other less common clinical conditions with more specialists from other institutions are needed. Future work should also include a prospective study to assess the effectiveness of the recommender algorithm in the specialty referral workflow. Lastly, this study did not include an analysis on the potential cost benefits of this recommender algorithm. Future research should compare the cost of additional visits due to incomplete workup with the cost of unnecessary labs if ordered based on algorithm recommendations.

Conclusion

An item association–based recommender algorithm can predict appropriate specialist's workup orders with high discriminatory accuracy. This could support future clinical decision support tools to increase effectiveness and access to specialty referrals. Our study demonstrates important first steps toward a data-driven paradigm for outpatient specialty consultation with a tier of automated recommendations that proactively enable initial workup that would otherwise be delayed by awaiting an in-person visit.

Acknowledgments

Authors' Contributions

WI conducted extraction and analysis of the data. WI and JHC have verified the underlying data. WI and PP conducted the survey study with specialists. WI drafted the manuscript. WI, PP, JP, and JHC contributed to the study concept and design, interpretation of data, and critical revision of the manuscript.

Conflicts of Interest

JHC is the cofounder of Reaction Explorer LLC that develops and licenses organic chemistry education software. He also received paid consulting or speaker fees from the National Institute of Drug Abuse Clinical Trials Network, Tuolc Inc, Roche Inc, and Younker Hyde MacFarlane PLLC. WI serves as Medical Director of Healthcare Data at nference and receives compensation. Other authors have no conflicts of interest to report.

Multimedia Appendix 1

Supplemental material and table.
[PDF File (Adobe PDF File), 215 KB-Multimedia Appendix 1]

References

1. 2017 update: the complexities of physician supply and demand: projections from 2015 to 2030: final report. IHS Markit. URL: https://aamc-black.global.ssl.fastly.net/production/media/filer_public/a5/c3/a5c3d565-14ec-48fb-974b-99fafaeecb00/aamc_projections_update_2017.pdf [accessed 2020-12-04]
2. Mehrotra A, Forrest CB, Lin CY. Dropping the baton: specialty referrals in the United States. Milbank Q 2011 Mar;89(1):39-68 [FREE Full text] [doi: 10.1111/j.1468-0009.2011.00619.x] [Medline: 21418312]
3. Ray KN, Bogen DL, Bertolet M, Forrest CB, Mehrotra A. Supply and utilization of pediatric subspecialists in the United States. Pediatrics 2014 Jun;133(6):1061-1069. [doi: 10.1542/peds.2013-3466] [Medline: 24799548]
4. Jaakkimainen L, Glazier R, Barnsley J, Salkeld E, Lu H, Tu K. Waiting to see the specialist: patient and provider characteristics of wait times from primary to specialty care. BMC Fam Pract 2014 Jan 25;15:16 [FREE Full text] [doi: 10.1186/1471-2296-15-16] [Medline: 24460619]
5. Bisgaier J, Rhodes KV. Auditing access to specialty care for children with public insurance. N Engl J Med 2011 Jun 16;364(24):2324-2333. [doi: 10.1056/NEJMsa1013285] [Medline: 21675891]
6. Mayer ML. Are we there yet? Distance to care and relative supply among pediatric medical subspecialties. Pediatrics 2006 Dec;118(6):2313-2321. [doi: 10.1542/peds.2006-1570] [Medline: 17142513]

7. Woolhandler S, Himmelstein DU. The relationship of health insurance and mortality: is lack of insurance deadly? Ann Intern Med 2017 Sep 19;167(6):424-431 [FREE Full text] [doi: 10.7326/M17-1403] [Medline: 28655034]

8. Hendrickson CD, Saini S, Pothuloori A, Mecchella JN. Assessing referrals and improving information availability for consultations in an academic endocrinology clinic. Endocr Pract 2017 Feb;23(2):190-198. [doi: 10.4158/EP161514.OR] [Medline: 27849384]

9. Hendrickson CD, Lacourciere SL, Zanetti CA, Donaldson PC, Larson RJ. Interventions to improve the quality of outpatient specialty referral requests: a systematic review. Am J Med Qual 2016 Sep;31(5):454-462. [doi: 10.1177/1062860615587741] [Medline: 26013165]

10. Stille CJ, McLaughlin TJ, Primack WA, Mazor KM, Wasserman RC. Determinants and impact of generalist-specialist communication about pediatric outpatient referrals. Pediatrics 2006 Oct;118(4):1341-1349. [doi: 10.1542/peds.2005-3010] [Medline: 17015522]

11. Ho CK, Boscardin CK, Gleason N, Collado D, Terdiman J, Terrault NA, et al. Optimizing the pre-referral workup for gastroenterology and hepatology specialty care: consensus using the Delphi method. J Eval Clin Pract 2016 Feb;22(1):46-52 [FREE Full text] [doi: 10.1111/jep.12429] [Medline: 26223584]

12. Chen AH, Murphy EJ, Yee HF. eReferral--a new model for integrated care. N Engl J Med 2013 Jun 27;368(26):2450-2453. [doi: 10.1056/NEJMp1215594] [Medline: 23802515]

13. Joschko J, Keely E, Grant R, Moroz I, Graveline M, Drimer N, et al. Electronic consultation services worldwide: environmental scan. J Med Internet Res 2018 Dec 21;20(12):e11112 [FREE Full text] [doi: 10.2196/11112] [Medline: 30578187]

14. Osman MA, Schick-Makaroff K, Thompson S, Bialy L, Featherstone R, Kurzawa J, et al. Barriers and facilitators for implementation of electronic consultations (eConsult) to enhance access to specialist care: a scoping review. BMJ Glob Health 2019;4(5):e001629 [FREE Full text] [doi: 10.1136/bmjgh-2019-001629] [Medline: 31565409]

15. Vimalananda VG, Orlander JD, Afable MK, Fincke BG, Solch AK, Rinne ST, et al. Electronic consultations (E-consults) and their outcomes: a systematic review. J Am Med Inform Assoc 2020 Mar 01;27(3):471-479 [FREE Full text] [doi: 10.1093/jamia/ocz185] [Medline: 31621847]

16. Kern C, Fu DJ, Kortuem K, Huemer J, Barker D, Davis A, et al. Implementation of a cloud-based referral platform in ophthalmology: making telemedicine services a reality in eye care. Br J Ophthalmol 2020 Mar;104(3):312-317 [FREE Full text] [doi: 10.1136/bjophthalmol-2019-314161] [Medline: 31320383]

17. Nabelsi V, Lévesque-Chouinard A, Liddy C, Dumas Pilon M. Improving the referral process, timeliness, effectiveness, and equity of access to specialist medical services through electronic consultation: pilot study. JMIR Med Inform 2019 Jul 10;7(3):e13354 [FREE Full text] [doi: 10.2196/13354] [Medline: 31293239]

18. Chang Y, Carsen S, Keely E, Liddy C, Kontio K, Smit K. Electronic consultation systems: impact on pediatric orthopaedic care. J Pediatr Orthop 2020 Oct;40(9):531-535. [doi: 10.1097/BPO.0000000000001607] [Medline: 32931692]

19. Chen JH, Podchiyska T, Altman R. OrderRex: clinical order decision support and outcome predictions by data-mining electronic medical records. J Am Med Inform Assoc 2016 Mar;23(2):339-348 [FREE Full text] [doi: 10.1093/jamia/ocv091] [Medline: 26198303]

20. Zhang Y, Padman R, Levin JE. Paving the COWpath: data-driven design of pediatric order sets. J Am Med Inform Assoc 2014 Oct;21(e2):e304-e311 [FREE Full text] [doi: 10.1136/amiajnl-2013-002316] [Medline: 24674844]

21. Klann JG, Szolovits P, Downs SM, Schadow G. Decision support from local data: creating adaptive order menus from past clinician behavior. J Biomed Inform 2014 Apr;48:84-93 [FREE Full text] [doi: 10.1016/j.jbi.2013.12.005] [Medline: 24355978]

22. Klann J, Schadow G, Downs S. A method to compute treatment suggestions from local order entry data. AMIA Annu Symp Proc 2010 Nov 13;2010:387-391 [FREE Full text] [Medline: 21347006]

23. Klann J, Schadow G, McCoy J. A recommendation algorithm for automating corollary order generation. AMIA Annu Symp Proc 2009 Nov 14;2009:333-337 [FREE Full text] [Medline: 20351875]

24. Hunter-Zinck HS, Peck J, Strout T, Gaehde SA. Predicting emergency department orders with multilabel machine learning techniques and simulating effects on length of stay. J Am Med Inform Assoc 2019 Dec 01;26(12):1427-1436 [FREE Full text] [doi: 10.1093/jamia/ocz171] [Medline: 31578568]

25. Wright A, Chen ES, Maloney FL. An automated technique for identifying associations between medications, laboratory results and problems. J Biomed Inform 2010 Dec;43(6):891-901 [FREE Full text] [doi: 10.1016/j.jbi.2010.09.009] [Medline: 20884377]

26. Wang JX, Sullivan D, Wells A, Chen JH. ClinicNet: machine learning for personalized clinical order set recommendations. JAMIA Open 2020 Jul;3(2):216-224 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa021] [Medline: 32734162]

27. Smith B, Linden G. Two decades of recommender systems at Amazon.com. IEEE Internet Computing 2017 May;21(3):12-18. [doi: 10.1109/mic.2017.72]

28. Hanley P, Lord K, Bauer AJ. Thyroid disorders in children and adolescents: a review. JAMA Pediatr 2016 Oct 01;170(10):1008-1019. [doi: 10.1001/jamapediatrics.2016.0486] [Medline: 27571216]

29. Cuda SE, Censani M. Pediatric obesity algorithm: a practical approach to obesity diagnosis and management. Front Pediatr 2018;6:431. [doi: 10.3389/fped.2018.00431] [Medline: 30729102]

30. Klein DA, Poth MA. Amenorrhea: an approach to diagnosis and management. Am Fam Physician 2013 Jun 01;87(11):781-788 [FREE Full text] [Medline: 23939500]

31. Datta S, Posada J, Olson G, Li W, O'Reilly C, Balraj D, et al. A new paradigm for accelerating clinical data science at Stanford Medicine. arXiv. Preprint posted online on March 17, 2020 [FREE Full text]

32. Frémal S, Lecron F. Weighting strategies for a recommender system using item clustering based on genres. Expert Syst Applications 2017 Jul;77:105-113. [doi: 10.1016/j.eswa.2017.01.031]

33. HealthRex / CDSS. GitHub. URL: https://github.com/HealthRex/CDSS/tree/master/scripts/specialty_recommender [accessed 2022-02-22]

34. Davison AC, Hinkley DV. Bootstrap Methods and Their Application. Cambridge: Cambridge University Press; 1997.

35. Styne DM, Arslanian S, Connor E, Farooqi IS, Murad MH, Silverstein JH, et al. Pediatric obesity-assessment, treatment, and prevention: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 2017 Mar 01;102(3):709-757 [FREE Full text] [doi: 10.1210/jc.2016-2573] [Medline: 28359099]

36. Child with suspected acquired hypothyroidism. Pediatric Endocrine Society. URL: https://pedsendo.org/clinical-resource/child-with-suspected-acquired-hypothyroidism/ [accessed 2021-04-16]

37. Child with suspected hyperthyroidism. Pediatric Endocrine Society. URL: https://pedsendo.org/clinical-resource/child-with-suspected-hyperthyroidism/ [accessed 2021-04-16]

38. Ostropolets A, Zhang L, Hripcsak G. A scoping review of clinical decision support tools that generate new knowledge to support decision making in real time. J Am Med Inform Assoc 2020 Dec 09;27(12):1968-1976 [FREE Full text] [doi: 10.1093/jamia/ocaa200] [Medline: 33120430]

39. Middleton B, Sittig DF, Wright A. Clinical decision support: a 25 year retrospective and a 25 year vision. Yearb Med Inform 2016 Aug 02;Suppl 1:S103-S116 [FREE Full text] [doi: 10.15265/IYS-2016-s034] [Medline: 27488402]

40. Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, et al. Grand challenges in clinical decision support. J Biomed Inform 2008 Apr;41(2):387-392 [FREE Full text] [doi: 10.1016/j.jbi.2007.09.003] [Medline: 18029232]

41. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. J Am Med Inform Assoc 2007;14(2):141-145 [FREE Full text] [doi: 10.1197/jamia.M2334] [Medline: 17213487]

42. Cornell E, Chandhok L, Rubin K. Implementation of referral guidelines at the interface between pediatric primary and subspecialty care. Healthc (Amst) 2015 Jun;3(2):74-79. [doi: 10.1016/j.hjdsi.2015.02.003] [Medline: 26179727]

43. Chen JH, Alagappan M, Goldstein MK, Asch SM, Altman RB. Decaying relevance of clinical data towards future decisions in data-driven inpatient clinical order sets. Int J Med Inform 2017 Jun;102:71-79 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.03.006] [Medline: 28495350]

44. Chen JH, Goldstein M, Asch S, Mackey L, Altman RB. Predicting inpatient clinical order patterns with probabilistic topic models vs conventional order sets. J Am Med Inform Assoc 2017 May 01;24(3):472-480 [FREE Full text] [doi: 10.1093/jamia/ocw136] [Medline: 27655861]

45. Wang JK, Hom J, Balasubramanian S, Schuler A, Shah NH, Goldstein MK, et al. An evaluation of clinical order patterns machine-learned from clinician cohorts stratified by patient mortality outcomes. J Biomed Inform 2018 Oct;86:109-119 [FREE Full text] [doi: 10.1016/j.jbi.2018.09.005] [Medline: 30195660]

46. Islam MM, Yang H, Poly TN, Li YJ. Development of an artificial intelligence-based automated recommendation system for clinical laboratory tests: retrospective analysis of the National Health Insurance Database. JMIR Med Inform 2020 Nov 18;8(11):e24163 [FREE Full text] [doi: 10.2196/24163] [Medline: 33206057]

47. Lee H, Kang J, Yeo J. Medical specialty recommendations by an artificial intelligence chatbot on a smartphone: development and deployment. J Med Internet Res 2021 May 06;23(5):e27460 [FREE Full text] [doi: 10.2196/27460] [Medline: 33882012]

Abbreviations

AUC: area under the receiver operating characteristic curve
OMOP: Observational Medical Outcomes Partnership
PPV: positive predictive value
STARR: Stanford Medicine Research Data Repository
TSH: thyroid stimulating hormone

XSL•FO
RenderX