The potential of evidence based clinical intake tools to discover or ground prevalence of symptoms using real-life virtual health encounters: a retrospective cohort study

Eden Avnat, Michael Samin, Daniel Ben Joya, Eyal Schneider, Elia Yanko, Dafna Eshel, Shahar Ovadia, Yossi Lev, Daniel Souroujon

Submitted to: Journal of Medical Internet Research
on: June 17, 2023

Table of Contents

The potential of evidence based clinical intake tools to discover or ground prevalence of symptoms using real-life virtual health encounters: a retrospective cohort study

Eden Avnat[1, 2] BMedSci, MPH; Michael Samin[1, 3] BMedSci; Daniel Ben Joya[1, 3] MD; Eyal Schneider[1, 3] MD; Elia Yanko[4] BSc; Dafna Eshel[1] MD; Shahar Ovadia[1] MCS; Yossi Lev[1] BCS; Daniel Souroujon[1, 2] MD, MPH

[1]Kahun Medical Ltd Givatayim IL
[2]Sackler Faculty of Medicine Tel Aviv IL
[3]Department of Otolaryngology-Head and Neck Surgery, Samson Assuta Ashdod University Hospital, Ben Gurion University Faculty of Health Sciences Ashdod IL
[4]The Azrieli Faculty of Medicine, Bar-Ilan University Safed IL

Corresponding Author:
Eden Avnat BMedSci, MPH
Kahun Medical Ltd
4 Ariel sharon St
Givatayim
IL

Abstract

Background: Evidence-based clinical intake tools (EBCITs) are structured assessment tools used to gather information about patients and help healthcare providers make more informed decisions. The growing demand for personalized medicine, along with the big data revolution, have rendered EBCITs a promising solution. EBCITs have the potential to provide comprehensive and individualized assessments of symptoms, enabling accurate and timely diagnosis, while contributing to the grounding of medical care.

Objective: This work examines whether EBCITs cover data concerning disorders and symptoms to a similar extent as physicians, thus can reliably address medical conditions in clinical settings. We also explore the potential of EBCITs to discover and ground real prevalence of symptoms in different disorders thereby expanding medical knowledge and further support medical diagnoses made by physicians.

Methods: Between August 1, 2022, and January 15, 2023, patients who used the services of a virtual healthcare (VH) provider in the USA were first assessed by the Kahun EBCIT. Kahun platform gathered, documented, and analyzed the information from the sessions and its clinical findings. In this study, we estimated the prevalence of patients' symptoms in medical disorders, using two datasets. The first set analyzed symptoms prevalence, as determined by the Kahun's knowledge engine. The second set analyzed symptoms prevalence, relying solely on data from the VH patients gathered by Kahun. The difference in variance between these two prevalence datasets, helped us assess Kahun's ability to incorporate new data, while integrating existing knowledge. To analyze the comprehensiveness of the Kahun's knowledge engine, we compared how well it covers weighted data for the symptoms and disorders found in the 2019 National Ambulatory Medical Care Survey (NAMCS). To assess Kahun's diagnosis accuracy, physicians independently diagnosed 250 of Kahun-VH's sessions. Their diagnoses were compared with Kahun's diagnoses.

Results: As part of this work, 2,550 patients used Kahun to complete a full assessment, among them 1,714 females and 836 males. Kahun collected 314 different chief complaints and proposed 108,523 suggestions related to symptoms during the intake process. At the end of the intake process, 6,496 conditions were presented to the caregiver. Kahun covered 94% (526,157,569/562,150,572) of the weighted symptoms and 91% (1,582,637,476/173,4783,244) of the weighted disorders in NAMCS 2019. In 90% (224/250) of the sessions, at least one identical disorder suggested by both the physicians and Kahun, with total accuracy rate of 72% (367/507). Kahun's engine yielded 519 prevalences while the Kahun-VH cohort yielded 599; 156 prevalences were unique to the latter and 443 prevalences were shared by both databases.

Conclusions: ECBITs, such as Kahun, encompass extensive amounts of knowledge and could serve as a reliable database for inferring medical insights and diagnosis. Using this credible database, potential prevalence of symptoms in different disorders

were discovered or grounded. This highlights the ability of ECBITs to refine the understanding of relationships between disorders and symptoms, which further supports physicians in medical diagnosis.

(JMIR Preprints 17/06/2023:49570)
DOI: https://doi.org/10.2196/preprints.49570

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

The potential of evidence based clinical intake tools to discover or ground prevalence of symptoms using real-life virtual health encounters: a retrospective cohort study

Eden Avnat[1,2*], Michael Samin[1,3], Daniel Ben Joya[1,3], Eyal Schneider[1,4], Elia Yanko[5], Dafna Eshel[1], Shahar Ovadia[1] Yossi Lev[1], Daniel Souroujon[1,2]

[1] Kahun Medical Ltd, Givatayim, IL.

[2] Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, IL.

[3] Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer Sheva, IL.

[4] Department of Otolaryngology-Head and Neck Surgery, Samson Assuta Ashdod University Hospital, Ben Gurion University Faculty of Health Sciences, Ashdod, IL.

[5] The Azrieli Faculty of Medicine, Bar-Ilan University, Safed, IL.

[*] Corresponding author: eden@kahun.com , +972-54-55-33-400

Abstract

Background:

Evidence-based clinical intake tools (EBCITs) are structured assessment tools used to gather information about patients and help healthcare providers make more informed decisions. The growing demand for personalized medicine, along with the big data revolution, have rendered EBCITs a promising solution. EBCITs have the potential to provide comprehensive and individualized assessments of symptoms, enabling accurate and timely diagnosis, while contributing to the grounding of medical care.

Objective:
This work examines whether EBCITs cover data concerning disorders and symptoms to a similar extent as physicians, thus can reliably address medical conditions in clinical settings. We also explore the potential of EBCITs to discover and ground real prevalence of symptoms in different disorders thereby expanding medical knowledge and further support medical diagnoses made by physicians.

Methods:
Between August 1, 2022, and January 15, 2023, patients who used the services of a virtual healthcare (VH) provider in the USA were first assessed by the Kahun EBCIT. Kahun platform gathered, documented, and analyzed the information from the sessions and its clinical findings. In this study, we estimated the prevalence of patients' symptoms in medical disorders, using two datasets. The first set analyzed symptoms prevalence, as determined by the Kahun's knowledge engine. The second set analyzed symptoms prevalence, relying solely on data from the VH patients gathered by Kahun. The difference in variance between these two prevalence datasets, helped us assess Kahun's ability to incorporate new data, while integrating existing knowledge. To analyze the comprehensiveness of the Kahun's knowledge engine, we compared how well it covers weighted data for the symptoms and disorders found in the 2019 National Ambulatory Medical Care Survey (NAMCS). To assess Kahun's diagnosis accuracy, physicians independently diagnosed 250 of Kahun-VH’s sessions. Their diagnoses were compared with Kahun's diagnoses.

Results:
As part of this work, 2,550 patients used Kahun to complete a full assessment, among them 1,714 females and 836 males. Kahun collected 314 different chief complaints and proposed 108,523 suggestions related to symptoms during the intake process. At the end of the intake process, 6,496 conditions were presented to the caregiver. Kahun covered 94% (526,157,569/562,150,572) of the weighted symptoms and 91% (1,582,637,476/173,4783,244) of the weighted disorders in NAMCS 2019. In 90% (224/250) of the sessions, at least one identical disorder suggested by both the physicians and Kahun, with total accuracy rate of 72% (367/507). Kahun’s engine yielded 519 prevalences while the Kahun-VH cohort yielded 599; 156 prevalences were unique to the latter and 443 prevalences were shared by both databases.

Conclusions:
ECBITs, such as Kahun, encompass extensive amounts of knowledge and could serve as a reliable database for inferring medical insights and diagnosis. Using this credible database, potential prevalence of symptoms in different disorders were discovered or grounded. This highlights the ability of ECBITs to refine the understanding of relationships between disorders and symptoms, which further supports physicians in medical diagnosis.

Key Words: clinical intake tool, evidence-based medicine, big data, digital health, symptoms, prevalence

Introduction

Evidence-based medicine (EBM) has become an integral part of modern medical practice [1]. It relies on the use of systematic research to identify up-to-date and reliable data, which serves as a crucial component in making medical decisions for individual patients [2]. The use of EBM can help reduce outdated practices or implicit biases that may disrupt medical decision-making, such as prioritizing caucasian individuals over minorities in emergency care or underestimating diagnosis of coronary heart disease in females [3-5].

The emergence of the big data revolution in the medical world not only reduces some of the bias and keeps physicians updated, but also serves as an important factor in the evolution of EBM [6-8]. By utilizing machine learning tools, implicit insights can be more easily derived from raw data, and instead of focusing on data analysis as a tool to answer questions, it is now used as a tool to find new questions which can lead to new and promising hypotheses [7].

In particular, personalized medicine has benefited from the analysis of abundant and diverse data [9]. Analyzing multi-omics data derived from large-scale cohort and population studies, combined with the study’s conclusions, allows identifying subtle differences in an individual's genetics which may lead to precise and personalized interventions [9]

Today, as the demand for personalized medicine based on the EBM approach is on the rise, health providers are seeking to bridge the gap between the EBM paradigm which formulates generalized conclusions gathered from many studies, and personalized medicine that focus on the individual [10,11]. One solution that bridges that gap, suggests that using evidence-based clinical intake tools (EBCIT), such as Kahun [11,12], can help physicians with more personalized decision-making.

Kahun is an artificial intelligence engine, encompassing more than 150,000 peer-reviewed publications and more than 30,000,000 medical relations and insights that were mapped to a knowledge graph [13]. Using its evidence-based knowledge graph, Kahun asks the patient personalized questions and then uses dynamic reasoning to generate tailored clinical assessments [12,13].

Prior studies have focused primarily on the diagnostic accuracy of Kahun and other diagnostic support tools [12,14-16]. However, one study [17] that examined COVID-19 patients and their specific symptoms, showed that these tools can also provide medical insights such as extending the understanding of symptoms associated with a specific disease such as COVID-19. By collecting patient findings and diagnoses, EBCITs have the potential to become comprehensive real-world databases in their own right. Analyzing these databases using big data methods could contribute to medical knowledge by grounding the real prevalence of symptoms in medical conditions or identifying new ones, while supporting the EBM approach.

This study aims to test whether EBCITs cover symptoms and diagnoses similar to physicians and thus serve as a potential reliable tool for decision making in the clinic. Furthermore, in this study, we demonstrated the ability of EBCITs to identify current or emerging prevalence of symptoms in disorders, thereby enhancing medical knowledge and providing additional support for physicians in clinical practice.

Methods

Study Design and Data Collection:

This study includes the analysis of three data sets:

Kahun's knowledge engine:

As mentioned, Kahun's data is based on more than 30,000,000 medical relations and insights from over than 150,000 peer-reviewed publications that were mapped by medical experts to a knowledge graph [13]. The medical knowledge is represented by the nodes and the edges of the knowledge graph. For example, node1 represents a specific disorder, which is mapped by edge1 to node2, which represents a specific symptom. The data on edge1 represents the relationship between node1 and node2, such as prevalence. Therefore, the triple (node1, edge1, node2) represents the prevalence of the specific symptom in the specific disorder. This data is referred to here as the Kahun's knowledge engine.

Kahun-VH cohort:

This dataset was collected by the Kahun platform which assessed patients who received virtual healthcare(VH) services from a VH provider based in the US, specializing in doctor-patient medical visits that occur using camera-enabled smartphones or computers, between August 1, 2022 and January 15, 2023.
Patients who used the provider's services were first given a link for initial assessment by Kahun. During the Kahun assessment, questions regarding the patient's medical background, chief complaint, symptoms, and risk factors were personally generated based on the Kahun algorithms [12]. Additionally, during each assessment, the algorithm suggested a dynamic list of relevant differential diagnoses (disorders) while computing a matching probability between the findings and each disorder. All the data and metadata regarding questions, answers, and disorders was anonymized and stored in a separate part of the Kahun database, is not relating to Kahun's knowledge engine and is referred to here as the Kahun-VH cohort.

National Ambulatory Medical Care Survey (NAMCS):

The data was obtained from the latest (2019) National Ambulatory Medical Care Survey (NAMCS), which was conducted by the Centers for Disease Control and Prevention [18]. NAMCS 2019 was designed to provide objective information about the provision and use of ambulatory medical care services in the United States [18]. The findings in NAMCS are based on a sample of weighted visits to non-federally employed office-based physicians, who are primarily engaged in direct patient care. The findings included: ‘reason of visit #1-5’ (coded by NAMCS internal method), ‘diagnosis #1-5’ (ICD-10 coded), and more. Since the scope of this current study covered symptoms and disorders, we addressed only NAMCS symptoms related to the ‘SYMPTOM MODULE’ and only those disorders that appeared among the NAMCS’ diagnoses. Findings were excluded where the reason for visit was related to follow-up visits and the prescription of medication or diagnoses related to an encounter for a specific examination.

Study Population:

Kahun-VH cohort included patients aged 16 or older who used the provider's services and completed the assessment by Kahun. Assessments missing information regarding sex, chief complaint, and differential diagnosis were excluded.

Analysis and Variables:

All statistical analyses were performed using R-studio (version 4.2.2).

Categorical variables were represented by percentage while continuous variables were represented by mean and Standard deviation (SD) if distributed normally, and otherwise by median and interquartile range. A ‘positive symptoms ratio’ was determined as the ratio between the total number of new symptoms or refinement of known symptoms the patient confirmed during the assessment, and the total number of symptoms or refinements suggested to the patient during the assessment. ‘Total questions’ included only those questions that were answered by the patients during the assessment.

Kahun’s Coverage Rates:

We calculated how well Kahun covers NAMCS data by multiplying the ratio of symptoms or disorders that were reported in each visit and appear in Kahun's knowledge engine, with the relevant visit weighted score designated by NAMCS. All multiples were summed and then divided by the summed total of all the visits’ weighted scores. We calculated the coverage ratio for the entire set of symptoms and disorders, for the groups of symptoms subset by NAMCS’s ‘SYMPTOM MODULE’, and for groups of disorders classified according to ICD-10 prefixes.

Kahun's Diagnostic Ability:

To evaluate Kahun's diagnostic ability, 250 sessions (10% of all sessions) were randomly selected and blindly assessed by Israeli licensed physicians. The full transcript of Kahun’s sessions, including questions suggested by Kahun and answers provided by the patient, was given to the physicians. The transcript did not include Kahun’s suggestions for differential diagnosis. Then, based on the transcript received, the physicians suggested up to three (non-ordered) suitable diagnoses. The physicians' suggested diagnoses were compared with both the relevant diagnoses and the relevant number of diagnoses suggested by Kahun (for example, if only two diagnoses were suggested by the physicians, only two diagnoses suggested by Kahun were compared). Kahun's accuracy rate was set as the number of matched diagnoses suggested by the physicians and Kahun divided by the total number of diagnoses suggested by Kahun (for example, in a specific session, if Kahun suggested three different disorder and two of them were also suggested by the physician, Kahun's accuracy rate was 67% - 2/3).

Prevalence:

Kahun calculated the prevalence of symptoms in different disorders based on the data from Kahun’s knowledge engine and, independently, based on the Kahun-VH cohort. For each disorder, symptoms that were suggested in more than 29 different assessments were selected. The prevalence of a symptom in a disorder was determined by the ratio of the total occurrences of the confirmed symptom to the total occurrences of the suggested symptom.

Only disorders and symptoms that appeared in both datasets were included. This created two equally dimensioned prevalence matrices: the Kahun prevalence matrix and the Kahun-VH cohort prevalence matrix. The prevalence similarity ratio was determined by dividing a prevalence from Kahun’s prevalence matrix with the corresponding prevalence extracted from Kahun-VH cohort’s prevalence matrix.

Each prevalence matrix underwent hierarchical clustering that clustered disorders by symptoms

using the pheatmap R package [19].

Ethical considerations:

This study analyzed anonymized data from three sources: Kahun's knowledge engine (published peer-reviewed articles), the Kahun-VH cohort, and the NAMCS. The data used in the Kahun-VH cohort was anonymized and de-identified according to HIPAA safe-harbor privacy rules. The data from NAMCS is publicly available and has been anonymized by the Centers for Disease Control and Prevention. As such, this study did not require Institutional Review Board (IRB) approval or ethical review or re-individual informed consent- all the data used was anonymized and de-identified, with no risk to individual privacy. Additionally, no personal or identifying information about participants was accessed or stored during the study. Appropriate measures were taken to ensure compliance with relevant privacy guidelines. No compensation was provided to participants, as this study was a secondary analysis of existing anonymized datasets.

Results

Kahun’s Coverage Rates (Kahun's knowledge engine vs NAMCS):

Kahun covered 94% (526 millions /562 millions) of all weighted symptoms reported in the NAMCS 2019 dataset. In 9 out of 10 different symptom groups, there was coverage for at least 85% weighted symptoms. Notably, there was nearly complete coverage in the Nervous system group (Figure 1). Additionally, Kahun covered 91% (1,582 millions/1,734 millions) of all weighted disorders reported in the NAMCS 2019 dataset, with at least 88% coverage for weighted disorders in 14 out of 17 different disorder groups. It is especially interesting that nearly complete coverage was observed for Diseases of the blood and certain disorders involving the immune mechanism, and for Infectious and parasitic diseases groups (Figure 2).

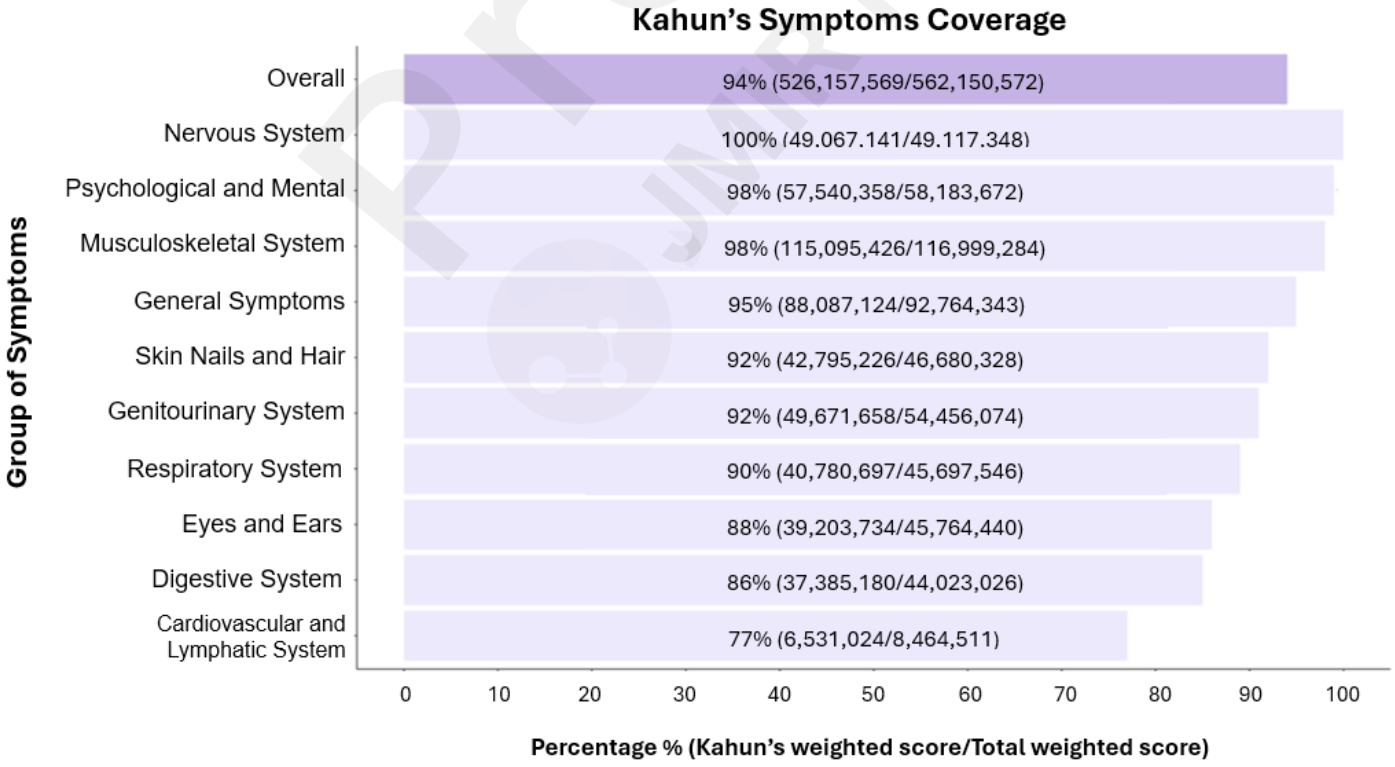


Figure 1: Kahun's coverage rates for weighted symptoms reported in NAMCS 2019. Purple colored bar represents total symptoms while light-purple bars represent different groups of symptoms.

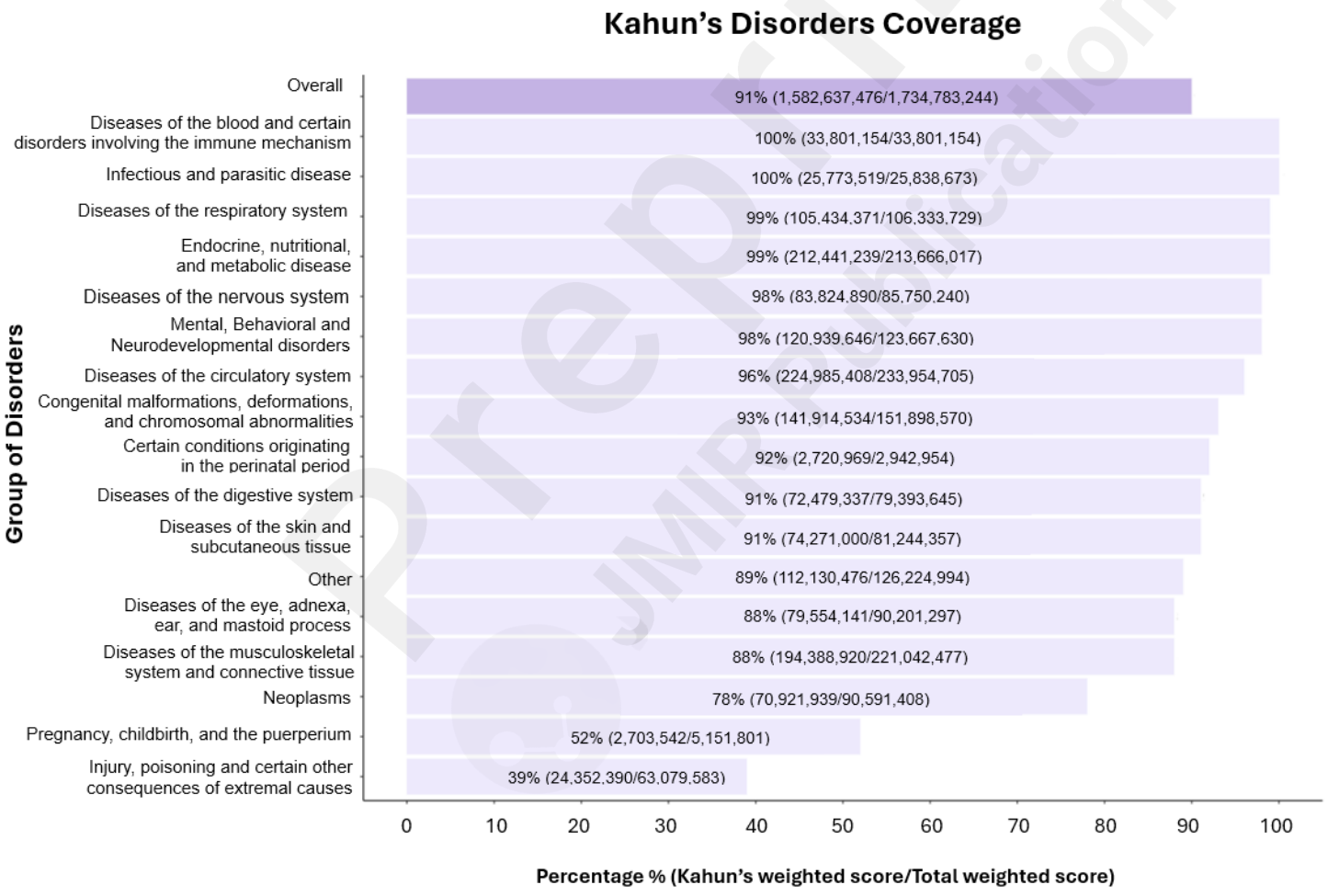


Figure 2: Kahun's coverage rates for weighted disorders reported in NAMCS 2019. Purple colored bar represents total disorders while light-purple bars represent different groups of disorders.

Patients' Characteristics (Kahun-VH cohort):

Kahun-VH cohort included: 1,714 females (67%) and 836 males (33%) with a median age of 35

years old (Table 1). During each session, a median of 34 questions were asked with a positive symptom ratio of 0.3 (Table 1).

During the assessments, 314 unique chief complaints were reported. Among them, the 5 most frequent chief complaints were: anxiety (N=173), sore throat (N=153), cough (N=138), sinus pain (N=114), and headache (N=108). Additionally, the duration of the most frequent chief complaints was between 24 hours and 1 week, for 32.3% (836/2,550) of the complaints (Table 1).

Overall, Kahun generated 108,523 suggestions relating to 905 unique symptoms and provided 6496 disorders (321 unique disorders) as possible diagnoses.

	Overall (N=2550)
Age	
Mean (SD)	38.3 (14.3)
Median [Min, Max]	35.0 [16.0, 90.0]
Sex	
Female	1714 (67.2%)
Male	836 (32.8%)
Duration of Chief Complain	
<24 Hours	325 (12.7%)
>24 Hours and <1 week	824 (32.3%)
>1 week and <12 weeks	481 (18.9%)
>12 weeks	344 (13.5%)
Missing	576 (22.6%)
Total Questions	
Mean (SD)	33.6 (9.20)
Median [Min, Max]	34.0 [11.0, 61.0]
Positive Symptoms Ratio	
Mean (SD)	0.322 (0.201)
Median [Min, Max]	0.297 [0.0244, 1.00]

Table 1: Patients' Characteristics

Kahun's Diagnostic Ability (Kahun-VH cohort):

A random sample of 250 Kahun-VH sessions was selected to assess Kahun's diagnostic ability. The sessions were blindly evaluated by physicians. In 90% (224/250) of the sessions, at least one identical disorder matched between the differential diagnosis suggested by the physicians and the differential diagnosis suggested by Kahun. Additionally, 367 diagnoses that were suggested by Kahun matched the diagnosis suggested by the physicians, resulting in a 72% (367/507) accuracy rate.

Prevalence (Kahun-VH cohort vs Kahun's knowledge engine):

Kahun calculated prevalence of 60 symptoms in 28 disorders. A total of 519 prevalences detected based on Kahun's knowledge engine, whereas 599 prevalences were detected based on the data from the Kahun-VH cohort (Figure 3, Figure 4 respectively). There was no statistically significant difference between the median prevalence value of Kahun's knowledge engine and the Kahun-VH cohort: 21% [5-50%] and 23% [8-47%], respectively, Wilcoxon's *P*= .35.

Out of the prevalences that originated in the Kahun-HV cohort, 159 did not appear in Kahun's knowledge engine. This resulted in a median detection rate of 4 [1-10] new prevalences per disorder. Tonsillitis had the highest detection rate with 24 new prevalences, while 6 disorders had no new prevalences: disorders of the pituitary gland, hypertensive crisis, hypothyroidism, laryngitis, sexually transmitted infection diseases, and urinary tract infection diseases. Additionally, 443 prevalences were identified in both Kahun's knowledge-graph and the VH patient cohort with a median prevalence similarity ratio of 1.04 [0.61-2.27]. Among them, the prevalence similarity ratio of 85 prevalences (19%) ranged from 0.85 to 1.15.

Using hierarchical clustering algorithm, five clusters of disorders were established based on their symptoms prevalence. Disorders with relative similar symptoms prevalences distribution were clustered together. The prevalence of symptoms that were not suggested by Kahun during the sessions in the corresponded disorder, though suggested in other disorders, was not calculated. The median number of disorders per cluster in Kahun's knowledge engine was five, whereas the median number of disorders per cluster was six in the Kahun-VH cohort (Figure 3, Figure 4).

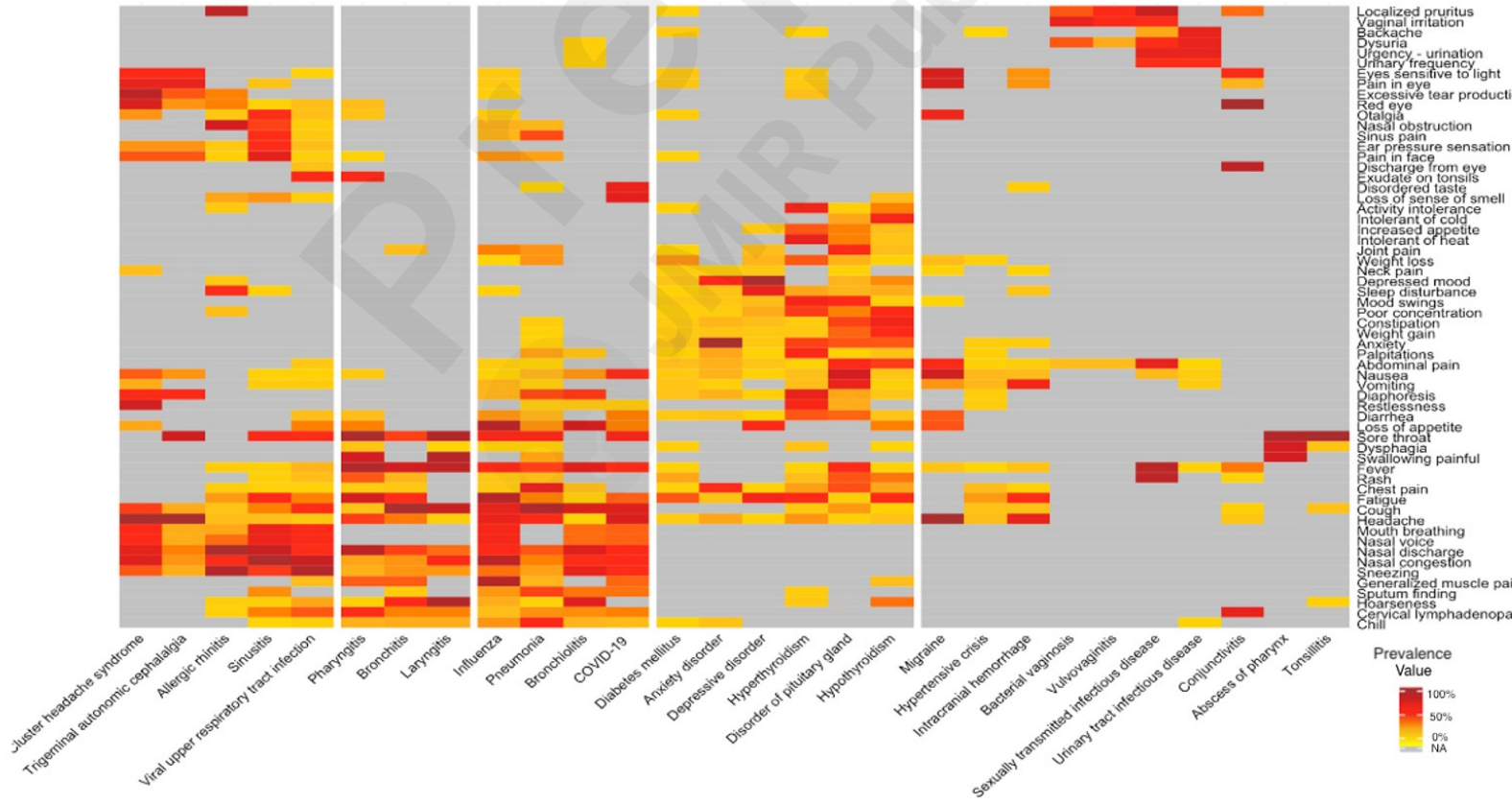


Figure 3: Heatmap of the prevalence matrix of symptoms in disorders based on Kahun's knowledge

engine. The white vertical lines cluster the disorders by their symptoms-prevalences distribution, using the hierarchical clustering algorithm. Dark red colors represent high prevalence while light yellow colors represent low prevalence. Grey squares represent symptoms that were not suggested by Kahun during sessions dealing with the corresponded disorder.

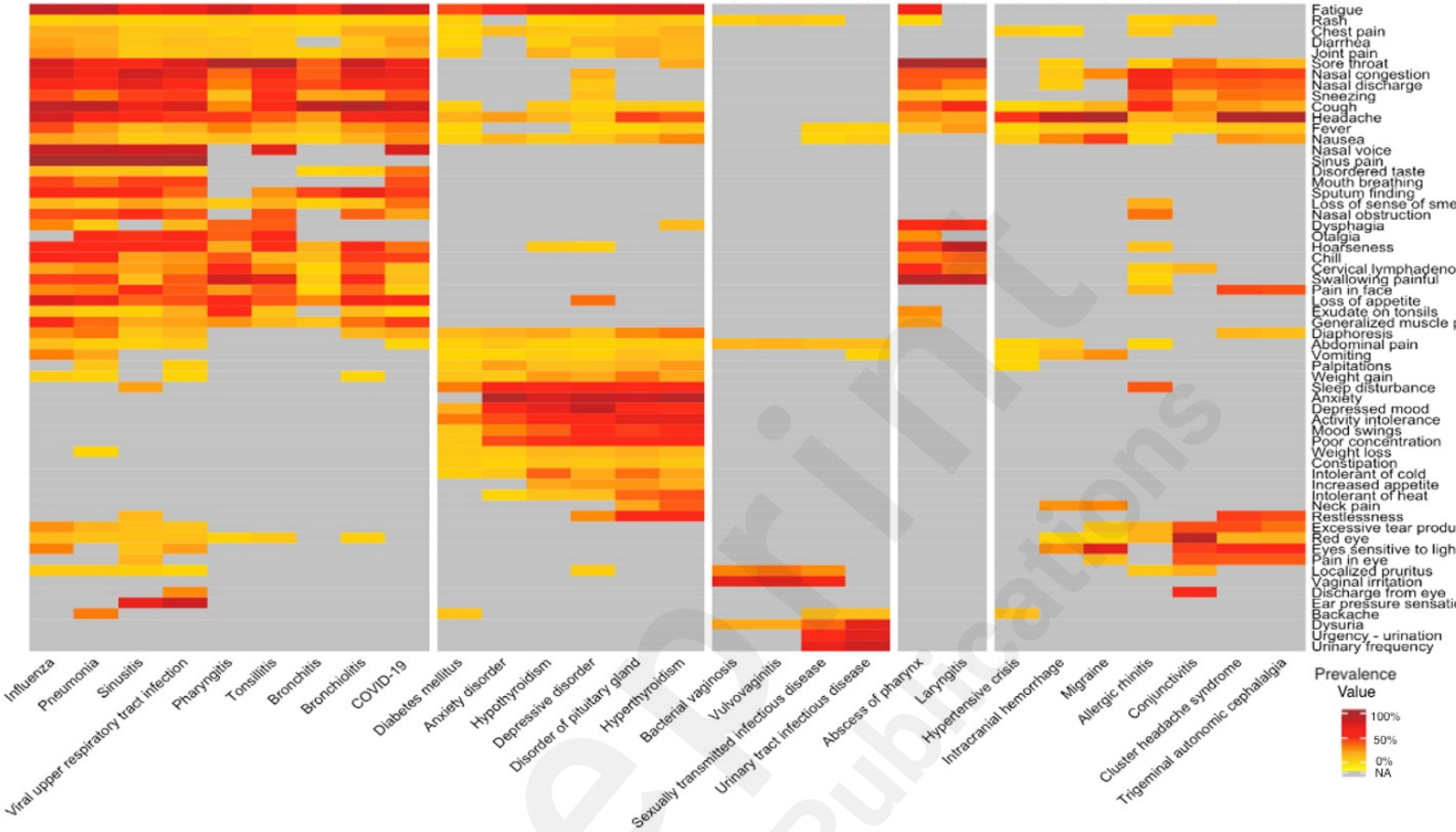


Figure 4: Heatmap of the prevalence matrix of symptoms in disorders, based on Kahun-VH cohort. The white vertical lines cluster the disorders by their symptoms-prevalences distribution, using the hierarchical clustering algorithm. Dark red colors represent high prevalence while light yellow colors represent low prevalence. Grey squares represent symptoms that were not suggested by Kahun during sessions dealing with the corresponded disorder.

Discussion

Principal Results:

This study tested the potential of EBCITs to cover the same medical data assessed by physicians, thus acting as a reliable source for drawing medical conclusions in addition to identifying or grounding possible prevalences of symptoms in different disorders. We used the data in the Kahun EBCIT to examine our hypothesis.

Kahun's relatively comprehensive coverage of the weighted symptoms collected by NAMCS 2019, shows that although Kahun is not a human physician, it can address at least 94% of the actual symptoms that patients report during medical visits. Moreover, when the symptoms are classified into groups, Kahun still covers most of the symptoms in each group. Kahun also generates differential diagnoses while taking into account about 91% (1,582,637,476/1,734,783,244) of the weighted disorders that physicians diagnosed in NAMCS 2019. These encouraging coverage rates

indicate that data gathered by Kahun is almost identical to the data gathered by a human physician and therefore relatively reliable.

Moreover, while compering to human medical licensed physicians, Kahun's decent diagnostic ability indicates that the differential diagnosis suggested by Kahun and therefore it's reasoning and intake process are reliable and trustable. Hence, we can infer that medical insights and assessments based on this data are valid.

To identify prevalences of symptoms in disorders, we analyzed the Kahun-VH cohort. Similarly to other remote-care service users, most of the patients in this cohort were females and relatively young. Additionally, the total number of questions that were asked during the assessment was similar to other clinical intake tools [20]. Comparing between a real physician interview and EBTICs might reveal a difference in the effectiveness of these approaches and should be explored further. While focusing on chief complaints, the study highlighted frequent chief complaints that are considered common among primary physicians, as suggested by others works [21,22]. This supports the claim that the data in the study agrees with real-world data. Additionally, more than 40% of the chief complaints lasted less than a week, presenting the relative acute characteristic of the symptoms, which although not widely studied, are supported by Lee et al [23].

The relatively young age of the patients and the fact that those patients were assessed by the VH's physicians only via virtual encounters, might indicate that some of the typical brick and mortar clinic patients in the traditional setting of primary care are underrepresented in this cohort. Although telemedicine and virtual encounters are gaining more popularity worldwide [24], we believe that evaluating EBCITs in a typical primary clinic setting is also important and should be considered in future research.

Nearly 109,000 suggestions, relating to 905 unique symptoms and approximately 6,500 diagnoses of 321 distinct disorders, were included in this study, demonstrating the abundance of data gathered. Having such diverse data could enable profound research.
Moreover, Kahun-VH cohort included only 2,550 patients. Thus, increasing usage of Kahun, which was shown as having decent diagnostic accuracy [12], has the potential to assist physicians in medical assessment while contributing to extended research by enriching Kahun’s database even further.

In this study, we focused on prevalences of symptoms in disorders. We compared the prevalences generated by Kahun’s knowledge engine with prevalences from the Kahun-VH cohort. We found 156 prevalences, exclusively in the latter group. These unique prevalences emphasize that even after learning and memorizing more that 150,000 medical articles, just as Kahun did [13], medical knowledge can still be improved. Thus, consistently incorporating real-world data into EBCITs like Kahun, could result in more precise and effective diagnosis which may help physicians in medical decision making. An example for this assumption is partly demonstrated by the relative medical-improvement in the hierarchical clustering of the 28 analyzed disorders. The hierarchical clustering algorithm grouped together disorders with similar symptoms-prevalence distributions, which may resemble physicians process of medical reasoning. For example, according to Figure 4, bacterial vaginosis, vulvovaginitis, sexually transmitted infection diseases and urinary tract infection diseases were clustered together as they shared similar symptoms-prevalence distribution and can reasonably clustered as genital related cluster. By comparing the suggested clusters of the disorders, as shown in Figure 3 and Figure 4, sexually transmitted infection diseases and headache-related disorders were re-clustered with more appropriate disorders for each, as noted by certificated physicians. This improvement may indicate that as EBCITs are being used more, they result in better performances,

thus suggesting more relevant diagnoses to the physician.

By assessing patients while taking into account a wide range of differential diagnoses and systematically storing the data, Kahun can ask unorthodox questions and less obvious prevalences may be discovered. These prevalences have the potential to serve as new leads for innovative research or medical trends. For example, a new possible prevalence of hoarseness in pharyngeal abscess was identified by the Kahun-VH cohort, with an estimated prevalence value of 48% (19/40). Literature reviews suggest that such a prevalence exists [25-27]; although, as far as we know, an exact evaluation of this prevalence was not conducted. Using the Kahun-VH cohort, we were able to estimate this prevalence.

A practical example for the potential of EBCIT in grounding prevalences, is the prevalence of nasal congestion in migraines. Although such a prevalence did not appear in Kahun's knowledge engine, according to the data from the Kahun-VH cohort, the prevalence was 28% (9/32, Figure 4). Remarkably, a study by Muehlberger et al. [28] evaluated this prevalence as 25%. These relatively close estimations testify to Kahun’s potential to accurately discover prevalences.

Apart from the novel prevalences that were discovered, 443 prevalences were simultaneously identified in both datasets. Of those prevalences, 81% had relative difference values (prevalence similarity ratio > ±0.15). This crucial finding highlights the importance of grounding prevalences. By combining comprehensive databases, such as the ones EBCITs may provide, with the information from previous studies, these discrepancies could be resolved and more refined prevalences would be computed. A different approach for settling these discrepancies argues that each cohort represents a different population. Hence, each prevalence is valid as long it refers to its original cohort. The latter approach may be helpful in creating personalized medical prevalences.

While compering Kahun-VH's prevalences with Kahun's knowledge engine prevalences, we included only selected symptoms and disorders, mainly to compare reasoning and sample size issues. Other prevalences were identified based on Kahun’s knowledge engine and the Kahun-VH cohort but are not presented in this current analysis. The actual number of prevalences is much greater. Splitting the main cohort into smaller, more homogeneous, sub-cohorts, could yield prevalences that are more personalized.

Limitations:

Our study has some known limitations. The coverage analysis refers to only some of the symptoms and disorders, as explained in the Methods section. On one hand, because Kahun is an EBCIT and presently does not cover physical examination nor does it save personal information for follow-up visits, some valuable information was not covered in this tool by definition. Therefore, the full coverage of the NAMCS is unknown. On the other hand, many other symptoms and disorders that were not included in NAMCS exist in the Kahun's knowledge engine. Therefore, we recommend further coverage analysis of other tools and reference databases. Moreover, NAMCS and VH cohort both cover the US population. Although the US population is diverse, some of the insight from this study is less relevant to other populations.

Although Kahun's diagnostic ability showed encouraging results and improved performance compared to previous studied [13], no such tool can be considered perfect. Therefore, constant improvement of Kahun is advised, and as a result, the grounded prevalences suggested in this study

will probably be refined in the future. Since the focus of this study was not on Kahun's accuracy rate, the sample that was used to determine this accuracy rate included only about 10% of the medical vignettes data (250/2550). Thus, further research that focusing on EBCITs' diagnosis ability and including more data should be considered in the future

An additional limitation of the Kahun-VH cohort is that it consists only of patients who can afford VH services and actively seek for medical care. This fact can potentially lead to selection bias and limit the generalizability of our real-world prevalences.

Moreover, since the setting of this cohort was based on virtual encounters and included younger patients with less comorbidities, there might be an underrepresentation of the older patients with more comorbidities, and the relevant prevalences might be affected. Therefore, future research involving typical primary care setting should be conducted.

Another limitation of this study is that most of the authors are employed by Kahun. This potential bias may have influenced the design and interpretation of the study results. However, we have taken steps to minimize the impact of this bias by using rigorous methods and statistical analyses and by involving other authors which are not employed by Kahun in analyzing some of the data. Despite these efforts, we recognize that the potential for bias remains and we encourage readers to carefully evaluate the evidence presented in this study and consider alternative perspectives.

Comparison with Prior Work:

As far as we know, only one study, conducted by Perlman et al. [17], showed the potential advantage of using clinical intake tools to infer medical insights. Specifically, the study identified relative rates and associations between disorders and symptoms. That in-depth study included 71,619 self-assessments of COVID-19 participants. The study focused only on one disorder: COVID-19. Although it also included only several predetermined symptoms, it managed to identify prominent results such as increased probability of having COVID-19 if a loss of smell or taste was experienced. It is worth noting that the demographic characteristics of the participants in that study, which were mainly females in their 30s, was similar to another study regarding clinical intake tools [14] and similar to our results.

In our view, the pioneering work done by Perlman et al. is extremely significant. Because their study had a relatively narrow scope in terms of coverage, symptoms, and disorders, we believed that at least one study with a wider scope was needed to support our main hypothesis. To the best of our knowledge, our study showed, for the first time, the potential of an EBCIT to generate prevalences of many symptoms in diseases.

Conclusions:

Kahun has the ability to cover most of the symptoms that patients may present, while addressing most of the possible disorders. Although not perfect, Kahun has the potential to serve as a reliable medical database and provide plausible EBM diagnosis. Thus, increased utilization of EBCITs such as Kahun, could help an improvement in EBCITs’ ability to uncover medical insights, including discovering and grounding prevalences of symptoms in disorders which might support physicians in medical diagnosis and expend medical knowledge

Conflicts of Interest:

All authors are salaried employees at Kahun Medical Ltd except from E.Y.

Data Availability:

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Funding Statement:

This study was funded by Kahun Medical Ltd.

Author Contributions:

EA, SO and DS contributed to conceptualization. EA, SO, YL and EY contributed to data curation and Software. MS, EY and DS contributed to investigation. EA, MS, DBJ and EY contributed to methodology and to project administration. DBJ, ES, EY and DE contributed to Validation. EA, MS, YL and DS contributed to writing the original draft. EA, ES, EY and DE contributed to reviewing and editing the paper. All authors reviewed and approved the final version of this manuscript.

List of Abbreviations:

Evidence-based clinical intake tools- EBCITs
Evidence-based medicine- EBM
SARS-CoV-2- COVID-19
National Ambulatory Medical Care Survey- NAMCS
Virtual healthcare- VH
Standard deviation- SD

References:

(1) Bonilla-Velez J, Small M, Urrutia R, Lomberk G. The Enduring Value of Research in Medical Education. *International Journal of Medical Students*. 2017;5(1):37-44. doi:https://doi.org/10.5195/ijms.2017.168
(2) Sackett DL. Evidence-based medicine. Seminars in Perinatology. 1997;21(1):3-5. doi:https://doi.org/10.1016/s0146-0005(97)80013-4
(3) Dehon E, Weiss N, Jones J, Faulconer W, Hinton E, Sterling S. A Systematic Review of the Impact of Physician Implicit Racial Bias on Clinical Decision Making. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine. 2017;24(8):895-904. doi:https://doi.org/10.1111/acem.13214
(4) FitzGerald C, Hurst S. Implicit bias in healthcare professionals: a systematic review. *BMC medical ethics*. 2017;18(1). doi:https://doi.org/10.1186/s12910-017-0179-8
(5) Moore N, Detering KM, Low T, Nolte L, Fraser S, Sellars M. Doctors' perspectives on adhering to advance care directives when making medical decisions for patients: an Australian interview study. *BMJ Open*. 2019;9(10):e032638. doi:https://doi.org/10.1136/bmjopen-2019-032638
(6) Au-Yong-Oliveira M, Pesqueira A, Sousa MJ, Dal Mas F, Soliman M. The Potential of Big Data Research in HealthCare for Medical Doctors' Learning. *Journal of Medical Systems*. 2021;45(1). doi:https://doi.org/10.1007/s10916-020-01691-7
(7) Mayer-Schönberger V, Ingelsson E. Big Data and medicine: a big deal? *Journal of Internal*

Medicine. 2018;283(5):418-429. doi:https://doi.org/10.1111/joim.12721
(8) Ioannidis JPA, Khoury MJ. Evidence-based medicine and big genomic data. Human Molecular Genetics. 2018;27(R1):R2-R7. doi:https://doi.org/10.1093/hmg/ddy065
(9) Cirillo D, Valencia A. Big data analytics for personalized medicine. Current Opinion in Biotechnology. 2019;58:161-167. doi:https://doi.org/10.1016/j.copbio.2019.03.004
(10) Lin JZ, Long JY, Wang AQ, Zheng Y, Zhao HT. Precision medicine: In need of guidance and surveillance. World Journal of Gastroenterology. 2017;23(28):5045. doi:https://doi.org/10.3748/wjg.v23.i28.5045
(11) Afzal M, Riazul Islam SM, Hussain M, Lee S. Precision Medicine Informatics: Principles, Prospects, and Challenges. IEEE Access. 2020;8:13593-13612. doi:https://doi.org/10.1109/ACCESS.2020.2965955
(12) Ben-Shabat N, Sloma A, Weizman T, Kiderman D, Amital H. Assessing the Performance of a New Artificial Intelligence–Driven Diagnostic Support Tool Using Medical Board Exam Simulations: Clinical Vignette Study. JMIR Medical Informatics. 2021;9(11):e32507. doi:https://doi.org/10.2196/32507
(13) Kahun Ltd. URL: https://www.kahun.com/ [accessed 2023-05-24]
(14) Fraser HSF, Cohan G, Koehler C, Anderson J, Lawrence A, Pateña J, et al. Evaluation of Diagnostic and Triage Accuracy and Usability of a Symptom Checker in an Emergency Department: Observational Study. *JMIR mHealth and uHealth*. 2022;10(9):e38364. doi:https://doi.org/10.2196/38364
(15) Gräf M, Knitza J, Leipe J, Krusche M, Welcker M, Kuhn S, et al. Comparison of physician and artificial intelligence-based symptom checker diagnostic accuracy. *Rheumatology International*. Published online September 10, 2022. doi:https://doi.org/10.1007/s00296-022-05202-4
(16) Ceney A, Tolond S, Glowinski A, Marks B, Swift S, Palser T. Accuracy of online symptom checkers and the potential impact on service utilisation. Wilson FA, ed. *PLOS ONE*. 2021;16(7):e0254088. doi:https://doi.org/10.1371/journal.pone.0254088
(17) Perlman A, Vodonos Zilberg A, Bak P, Dreyfuss M, Leventer-Roberts M, Vurembrand Y, et al. Characteristics and Symptoms of App Users Seeking COVID-19–Related Digital Health Information and Remote Services: Retrospective Cohort Study. *Journal of Medical Internet Research*. 2020;22(10):e23197. doi:https://doi.org/10.2196/23197
(18) National Ambulatory Medical Care Survey. URL: https://www.cdc.gov/nchs/ahcd/about_ahcd.htm [accessed 2023-05-24]
(19) Kolde R. Package 'pheatmap'. https://cran.r-project.org/package=pheatmap. 2019.
(20) Ben-Shabat N, Sharvit G, Meimis B, Ben-Joya D, Sloma A, Kiderman D, et al. Assessing data gathering of chatbot based symptom checkers - a clinical vignettes study. International Journal of Medical Informatics. 2022;168:104897. doi:https://doi.org/10.1016/j.ijmedinf.2022.104897
(21) Finley CR, Chan DS, Garrison S,Korownyk C, Kolber MR, Campbell S, et al. What are the most common conditions in primary care? *Canadian Family Physician*. 2018;64(11):832-840. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6234945/
(22) Krüger K, Töpfner N, Berner R, Windfuhr J, Oltrogge JH. Clinical practice guideline: Sore throat. *Deutsches Aerzteblatt Online*. 2021;118(11). doi:https://doi.org/10.3238/arztebl.m2021.0121
(23) Lee HY, Choi SH, Rim JS,Lim HK,Soo-Heo Young,Shin J, et al. Illnesses Encountered during Medical Volunteering in Takeo Province, Cambodia. *Medicina*. 2020;56(1):30. doi:https://doi.org/10.3390/medicina56010030
(24) Meuter LS, Wolf KI, Pacak K. Maintaining professional encounters and enhancing telemedicine interactions with core virtual-clinical values. Endocr Pract. 2021;27(1):77-79. doi: 10.1016/j.eprac.2020.11.006

(25) Christoforidou A, Metallidis S, Kollaras P, Agathangelidis A, Nikolaidis P, Vital V, et al. Tuberculous retropharyngeal abscess as a cause of oropharyngeal dysphagia. *American Journal of Otolaryngology*. 2012;33(2):272-274. doi:https://doi.org/10.1016/j.amjoto.2011.06.006

(26) Lin J, Wu XM, Feng JX, Chen MF. Retropharyngeal abscess presenting as acute airway obstruction in a 66-year-old woman: A case report. *World Journal of Clinical Cases*. 2019;7(22):3838-3843. doi:https://doi.org/10.12998/wjcc.v7.i22.3838

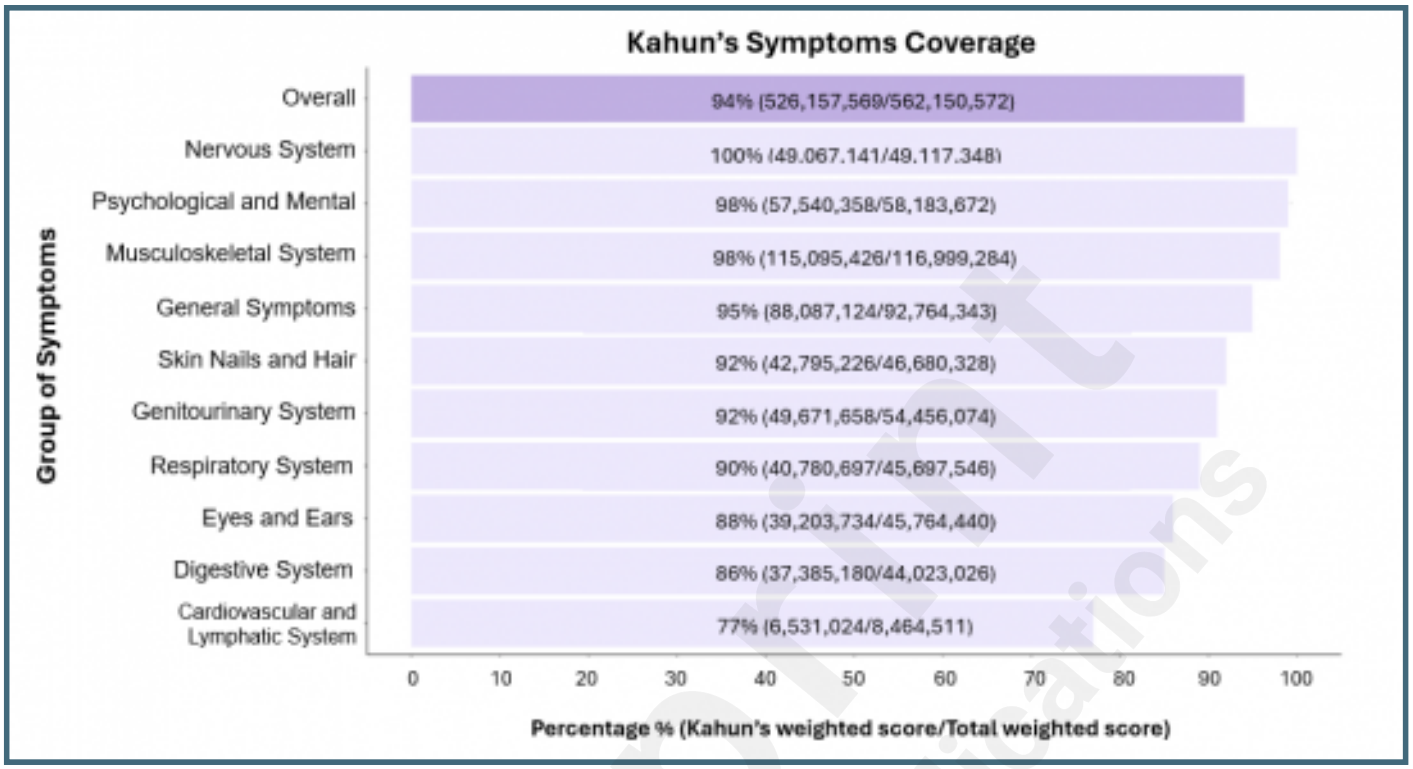
(27) Mishra AK, Anand A, Gupta KK, Saini R. Tuberculous Retropharyngeal Abscess as an Indiscernible Cause of Dysphagia. *Journal of Case Reports*. 2018;8(1):76-79. doi: http://dx.doi.org/10.17659/01.2018.0020

(28) Muehlberger T, Wormald JCR, Hachach-Haram N, Mosahebi A. Migraine: A look down the nose. *Journal of Plastic, Reconstructive & Aesthetic Surgery*. 2017;70(7):914-921. doi:https://doi.org/10.1016/j.bjps.2017.02.009

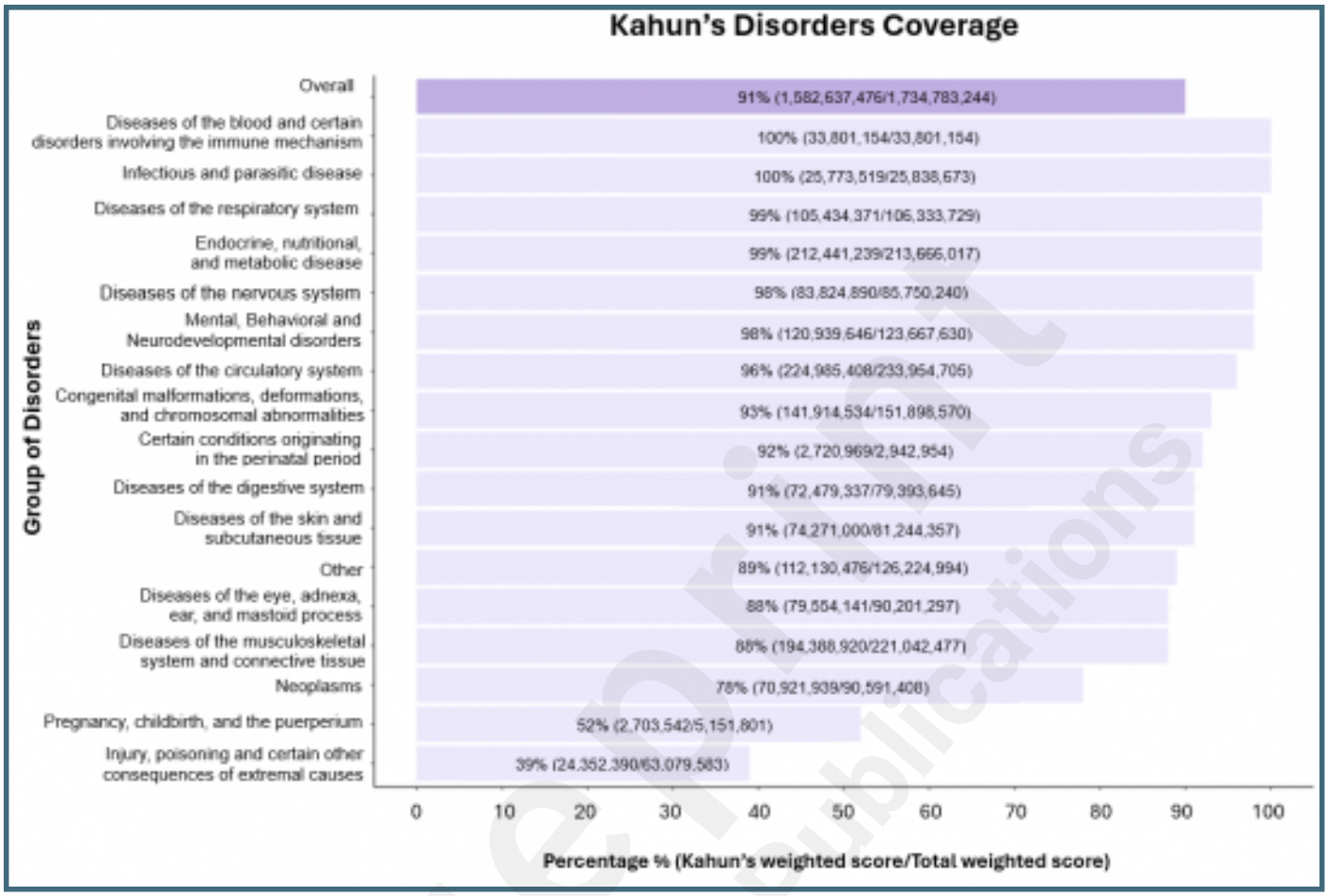
Supplementary Files

Figures

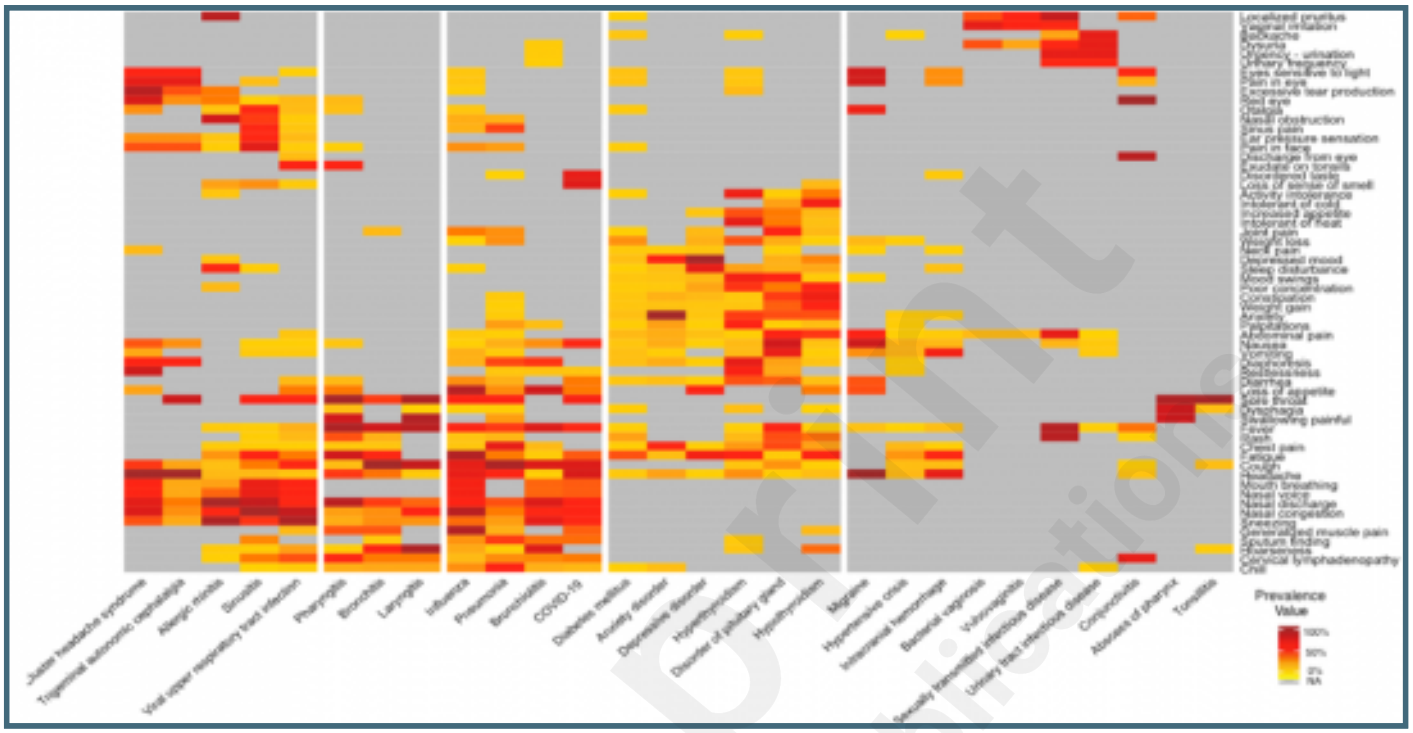
Kahun’s coverage rates for weighted symptoms reported in NAMCS 2019. Purple colored bar represents total symptoms while light-purple bars represent different groups of symptoms.



Kahun’s coverage rates for weighted disorders reported in NAMCS 2019. Purple colored bar represents total disorders while light-purple bars represent different groups of disorders.



Heatmap of the prevalence matrix of symptoms in disorders based on Kahun's knowledge engine. The white vertical lines cluster the disorders by their symptoms-prevalences distribution, using the hierarchical clustering algorithm. Dark red colors represent high prevalence while light yellow colors represent low prevalence. Grey squares represent symptoms that were not suggested by Kahun during sessions dealing with the corresponded disorder.



Heatmap of the prevalence matrix of symptoms in disorders, based on Kahun-VH cohort. The white vertical lines cluster the disorders by their symptoms-prevalences distribution, using the hierarchical clustering algorithm. Dark red colors represent high prevalence while light yellow colors represent low prevalence. Grey squares represent symptoms that were not suggested by Kahun during sessions dealing with the corresponded disorder.