Original Paper

Artificial Intelligence–Based Prediction of Lung Cancer Risk Using Nonimaging Electronic Medical Records: Deep Learning Approach

Marvin Chia-Han Yeh[1,2], MD, PhD; Yu-Hsiang Wang[3], MD; Hsuan-Chia Yang[4,5], PhD; Kuan-Jen Bai[6,7,8], MD; Hsiao-Han Wang[1,2,4,9*], MD; Yu-Chuan (Jack) Li[1,2,4,5,9*], MD, PhD

[1]Department of Dermatology, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan
[2]Research Center of Big Data and Meta-analysis, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan
[3]School of Medicine, Taipei Medical University, Taipei, Taiwan
[4]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan
[5]International Center for Health Information Technology, Taipei Medical University, Taipei, Taiwan
[6]Division of Pulmonary Medicine, Department of Internal Medicine, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan
[7]School of Respiratory Therapy, College of Medicine, Taipei Medical University, Taipei, Taiwan
[8]Pulmonary Research Center, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan
[9]Department of Dermatology, School of Medicine, Taipei Medical University, Taipei, Taiwan
[*]these authors contributed equally

Corresponding Author:
Yu-Chuan (Jack) Li, MD, PhD
Department of Dermatology
Wan Fang Hospital
Taipei Medical University
No 111, Section 3, Xinglong Road
Wenshan District
Taipei, 116
Taiwan
Phone: 886 29307930 ext 2980
Email: jaak88@gmail.com

Abstract

Background: Artificial intelligence approaches can integrate complex features and can be used to predict a patient's risk of developing lung cancer, thereby decreasing the need for unnecessary and expensive diagnostic interventions.

Objective: The aim of this study was to use electronic medical records to prescreen patients who are at risk of developing lung cancer.

Methods: We randomly selected 2 million participants from the Taiwan National Health Insurance Research Database who received care between 1999 and 2013. We built a predictive lung cancer screening model with neural networks that were trained and validated using pre-2012 data, and we tested the model prospectively on post-2012 data. An age- and gender-matched subgroup that was 10 times larger than the original lung cancer group was used to assess the predictive power of the electronic medical record. Discrimination (area under the receiver operating characteristic curve [AUC]) and calibration analyses were performed.

Results: The analysis included 11,617 patients with lung cancer and 1,423,154 control patients. The model achieved AUCs of 0.90 for the overall population and 0.87 in patients ≥55 years of age. The AUC in the matched subgroup was 0.82. The positive predictive value was highest (14.3%) among people aged ≥55 years with a pre-existing history of lung disease.

Conclusions: Our model achieved excellent performance in predicting lung cancer within 1 year and has potential to be deployed for digital patient screening. Convolution neural networks facilitate the effective use of EMRs to identify individuals at high risk for developing lung cancer.

(J Med Internet Res 2021;23(8):e26256) doi: 10.2196/26256

KEYWORDS

artificial intelligence; lung cancer screening; electronic medical record

Introduction

Lung cancer is a leading cause of cancer death worldwide, and to reduce its mortality, early detection is crucial. The National Lung Cancer Screening Trial (NLST) revealed that screening with low-dose computed tomography (LDCT) can reduce the mortality associated with lung cancer by 20% [1]. Likewise, the Dutch-Belgian Randomized Lung Cancer Screening Trial (NELSON study) recently revealed that screening with LDCT resulted in a 24% decrease in the 10-year cumulative mortality for men and a 33% decrease for women [2]. Multiple organizations have recommended LDCT screening for lung cancer to be used on target populations [3,4]. Given the potential harm due to radiation exposure, false-positive results, and costs associated with LDCT, most organizations only recommend annual screening that targets high-risk individuals; this group is largely identified by epidemiological factors, including age and smoking/cessation history [5]. Furthermore, due to the potential harm associated with false-positive results, the cost-effectiveness of implementing annual LDCT screening remains controversial [6]. Multiple research groups have attempted to overcome this problem by developing risk prediction models to identify patients who might benefit from LDCT screening and to generate criteria that are superior to those introduced by the NLST and related studies [7-14]. These models frequently include self-reported information, such as family history, BMI, socioeconomic status, and smoking/cessation history, and they use conventional regression models for the final risk analysis.

In the era of digital medicine, the use of artificial intelligence has resulted in good performance for predicting image-related tasks, specifically the use of convolutional neural networks (CNNs). In lung cancer research, CNNs have been applied to LDCT and chest radiographic images to facilitate detection and classification of pulmonary nodules; these models demonstrate performance that is comparable to that achieved by human experts [15-19]. The prediction performance is largely based on high-level feature extraction and nonlinear prediction via the use of CNNs. Given proper data conversion, using CNN methodologies to generate predictions using other nonimaging medical data may be possible. Our group recently described a risk prediction model for nonmelanoma skin cancer that was generated using data extracted from electronic medical records (EMRs) [20].

In predicting lung cancer risk, the EMR should be suited to the task of identifying high-risk individuals [21]. In this study, our goal is to develop a risk model for the prediction of lung cancer using data from EMRs. As such, we applied established CNN algorithms to the large data set available in EMRs to identify important patterns associated with the development of lung cancer. In contrast with methods used for traditional regression analysis, we attempted to include evolving sequential information found in EMRs to generate our prediction model. Our goal was to generate a model that facilitated the prospective identification of individuals at higher risk for developing lung cancer; these individuals might then undergo further follow-up examinations, including LDCT. The use of a predictive model to identify individuals at high risk could serve to limit unnecessary radiation exposure and reduce costs associated with LDCT and related interventions.

Methods

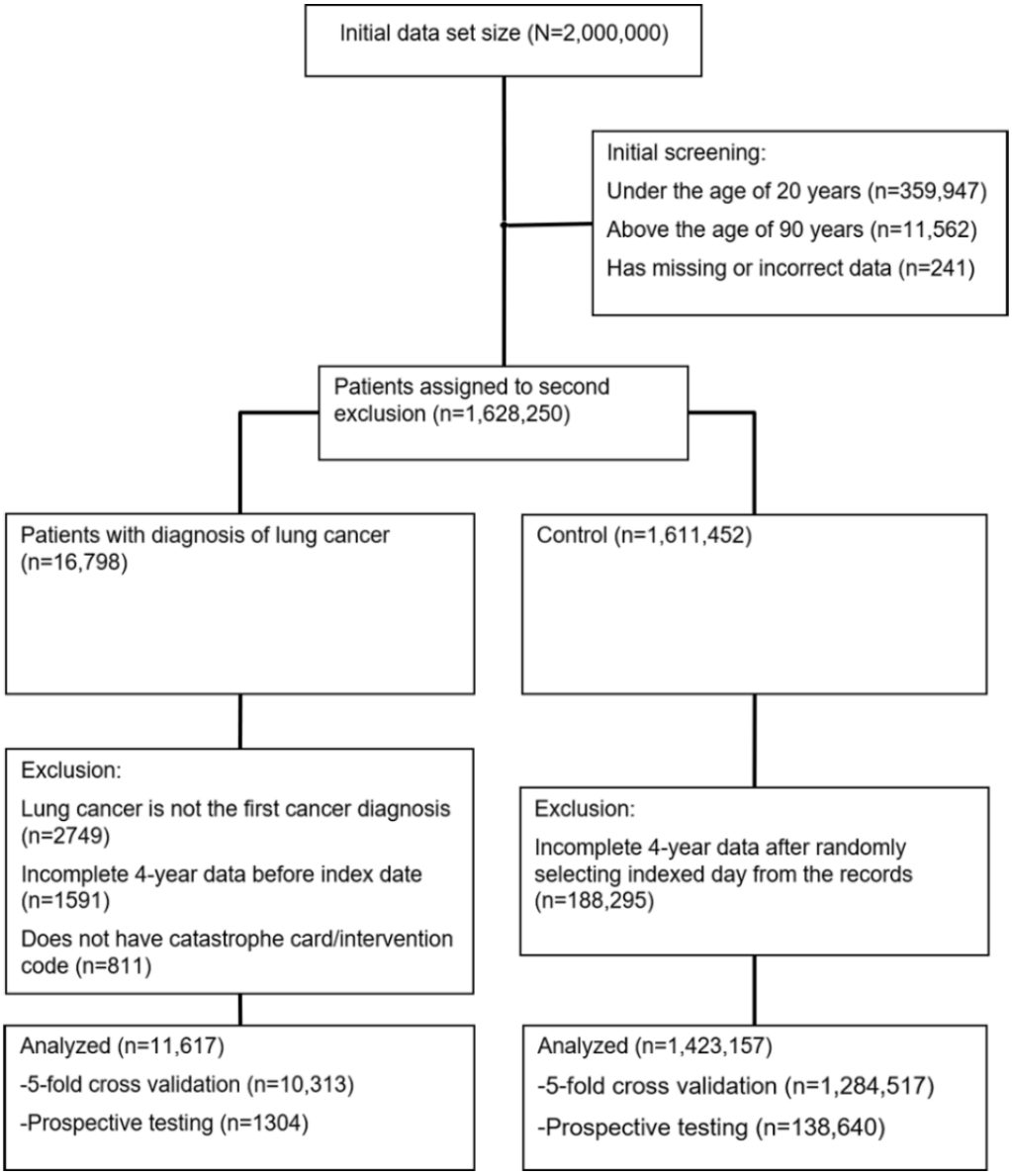
Study Population

Deidentified EMRs of 2 million patients who received care between January 01, 1999, and December 31, 2013, were initially sampled from the Taiwan National Health Insurance Research Database (NHIRD). These EMRs included the demographic information, diagnoses, and procedure codes from the *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD-9-CM*) and prescriptions from both outpatient clinical declaration files and in-hospital declaration files. This study included participants between the ages of 20 and 90 years who had at least 4 years of medical records on file. Participants with missing data were excluded. These criteria yielded 1,628,250 EMRs with over 300 million record entries for evaluation and analysis. This study was approved by the Taipei Medical University Institutional Review Board; informed patient consent was waived, as all data were anonymous and deidentified before analysis [22].

Data Preprocessing

Previous validation studies that focused on lung cancer using the NHIRD have shown a positive predictive value (PPV) of 95% [23]. In this study, we provide further validation of the diagnosis of lung cancer using intervention codes (eg, thoracic surgery, subsequent radiotherapy, or chemotherapy) and national catastrophic illness cards (which require definite pathologic proof of a cancer diagnosis). The inclusion and exclusion criteria used in this study are indicated in Figure 1.

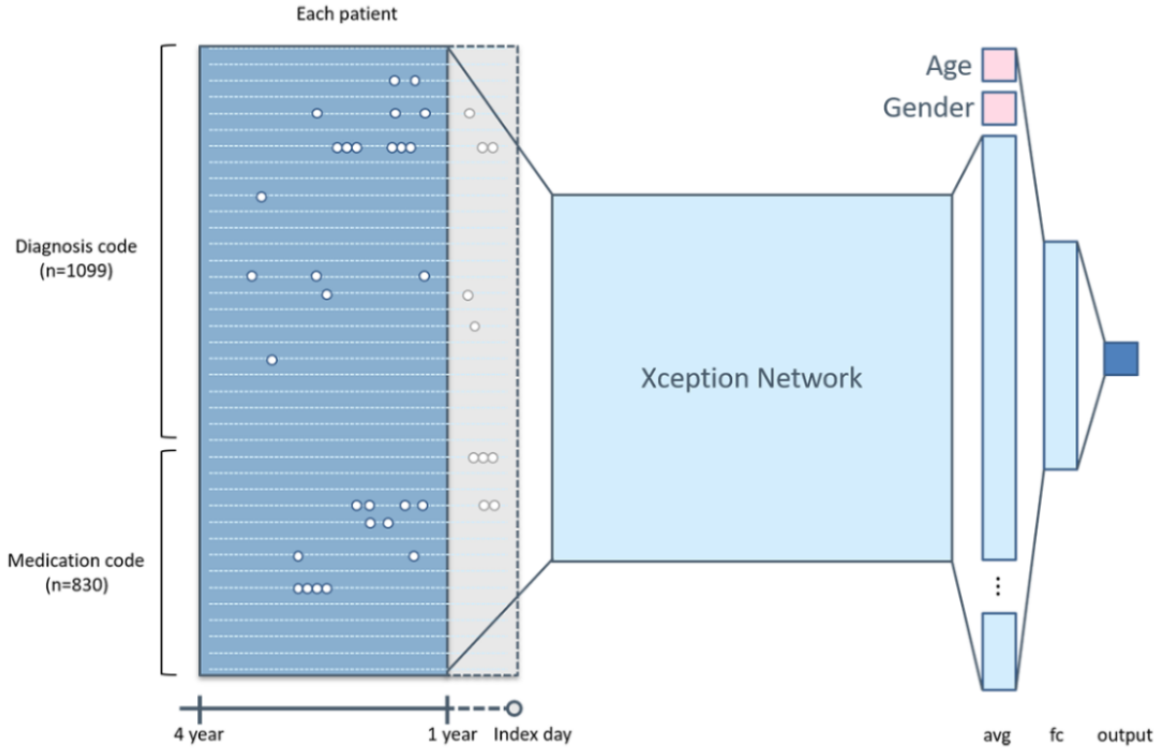
Figure 1. Inclusion and exclusion criteria for the study.



The index date for patients with lung cancer was defined as the date of first diagnosis. For the control patients, the index dates were randomly selected from their medical history. *ICD-9-CM* diagnosis codes and World Health Organization-Anatomical Therapeutic Chemical (WHO-ATC) prescription codes were collected from each case for preprocessing; the date 1 year prior to the index date was used to define the prediction window. The observation window included the 3 years prior to the date included in the prediction window. Thus, we used 3 years of patient medical information to predict the risk of new-onset lung cancer at or within 1 year later (Figure 2). The *ICD-9-CM* and WHO-ATC codes were preprocessed as described in our previous study [20]. Briefly, the EMRs were classified into 1099 *ICD-9-CM* code groups and 830 WHO-ATC drug groups. Together, 1929 features were recorded weekly for 157 weeks. For each patient, the diagnoses and medications prescribed at each visit were recorded and converted to an image-like array that preserved temporal information associated with both diagnosis and medication history.

The inputs included age, gender, and an image representing the patient's 3-year history of diagnosis and medication. The image was input into Xception, a 126-layer neural network, in which feature extraction was performed. The final layer of the Xception network was connected to an average pooling layer and then connected to a fully connected layer with the patient's age and gender.

Figure 2. Visualization of the hidden layer of the model using t-stochastic neighbor embedding. Avg: average; fc: fully connected layer.



We performed 3 subgroup analyses to investigate the performance of the model in different populations. According to the age criteria used in previous trials focused on lung cancer screening [1], patients above and below 55 years of age were included among the subgroups. We also examined patients both with and without previous lung disease [24], including subgroups of patients diagnosed with asbestosis, bronchiectasis, chronic bronchitis, chronic obstructive pulmonary disease (COPD), emphysema, fibrosis, pneumonia, sarcoidosis, silicosis, and tuberculosis. Finally, to focus on the discriminative power of the diagnosis and medication without the confounding effects of age, a subgroup of age- and gender-matched controls was identified.

Model Construction and Evaluation

All patient data were split into training, validation, and testing sets based on their respective index dates. Data with index dates prior to December 31, 2012, were used for training and internal validation, and data with index dates after that date were used for prospective testing. The patients' age, gender, and image-like arrays described above were used as inputs to generate the model (Figure 2).

Lung cancer risk prediction was treated as a binary classification task using supervised learning. The model was trained to determine whether a given patient was likely to develop lung cancer within 1 year. The Xception architecture [25], which includes a 126-layer CNN-based neural network with a moderate number of parameters, was used for feature extraction. The detailed model structure is shown in Figure 2; the model construction and hyperparameters are listed in Section S1 in Multimedia Appendix 1. During training, class weights based on the population size were set to address data imbalance. To ensure the robustness of the model, a 5-fold cross validation was performed on the model. The performance of the model was assessed by its sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). Model calibration was assessed using a reliability curve and the median absolute error.

To understand the model prediction, occlusion sensitivity analysis was performed by iteratively masking information from a single diagnosis or medication followed by evaluating any changes in the model prediction [26]. In addition, a dimensional reduction technique, t-distributed stochastic neighbor embedding (t-SNE), was performed on the fully connected hidden layer output of the final testing data. We randomly selected 1000 lung cancer patients and 9000 control patients for visualization. The model construction, data preprocessing, model training, and statistical processing were performed using the Python programming language, version 3.6.

Results

Baseline Demographics

A total of 11,617 lung cancer patients and 1,423,154 control patients were identified in our data set. The mean age of the lung cancer group was 66.62 years (SD 14.01); the overall data set included 856,558 (59.7%) men and 578,213 (40.3%) women. The baseline demographics of this patient cohort and the assigned subgroups are summarized in Table 1 and Tables S1-S10 in Multimedia Appendix 1.

Table 1. Demographics of the patients with lung cancer and control patients (N=1,434,771).

Group	Patients, n	Age (years), mean (SD)	Male gender, n (%)	Mean diagnosis record count (SD), n	Mean medication record count (SD), n
Whole population					
Lung cancer	11,617	66.62 (14.01)	6931 (59.7)	121.62 (113.19)	202.68 (208.97)
Control	1,423,154	44.95 (16.32)	683,375 (48.0)	66.09 (76.60)	105.99 (135.54)
Age and gender match (1:10)					
Lung cancer	11,617	66.62 (14.01)	6931 (59.7)	121.62 (113.19)	202.68 (208.97)
Control	116,169	66.62 (14.01)	69,310 (59.7)	117.99 (113.67)	190.22 (196.78)
Age ≥55 years					
Lung cancer	9261	71.99 (9.46)	5673 (61.3)	135.12 (116.31)	227.81 (218.12)
Control	385,052	66.57 (9.04)	56,730 (48.6)	114.23 (106.76)	184.50 (189.50)
Age <55 years					
Lung cancer	2356	45.50 (7.55)	1258 (53.4)	68.58 (80.42)	103.90 (126.71)
Control	1,038,102	36.93 (9.85)	496,256 (47.8)	48.23 (51.36)	76.87 (93.45)
History of lung disease[a]					
Lung cancer	3565	70.79 (12.73)	2244 (63.0)	175.12 (134.36)	297.56 (245.55)
Control	182,098	53.01 (18.09)	85,070(46.7)	125.17 (114.53)	204.85 (204.66)
No history of lung disease					
Lung cancer	8052	64.77 (14.16)	4687 (58.2)	97.94 (93.08)	160.67 (174.80)
Control	1,270,651	43.77 (15.70)	598,305 (48.2)	57.42 (64.94)	91.48 (115.23)

[a]Lung diseases included asbestosis, bronchiectasis, chronic bronchitis, chronic obstructive pulmonary disease, emphysema, fibrosis, pneumonia, sarcoidosis, silicosis, and tuberculosis. More information is provided in Table S11 in Multimedia Appendix 1.

Model Performance

For all patients, the model revealed an AUC of 0.821 when the input image-like array included sequential diagnostic information only. By contrast, the AUC was 0.894 when the input features included sequential medication information only; when the sequential diagnostic and medication information was simplified to binary variables, the model performance decreased (AUC=0.827). When both sequential diagnostic and medication information were integrated, the model reached an AUC of 0.902 on prospective testing, with a sensitivity of 0.804 and specificity of 0.837 (Table S12 in Multimedia Appendix 1). The calibration of the model showed a median expected error of 0.125; the reliability curve is shown in Figure S1 in Multimedia Appendix 1.

The model performance at different age cutoffs was then investigated. Screening using an age cutoff of 55 years revealed a superior AUC of 0.871 compared to those obtained when cutoffs of 50 or 60 years were used (0.866 and 0.863, respectively) (Table S13, Multimedia Appendix 1).

Subgroup Analysis

Analyses of the subgroups included one that was both age-and-gender-matched, those at ages above and below 55 years, and those with or without lung disease were performed. For this analysis, we identified an age- and gender-matched control subgroup that was 10 times larger than the original lung cancer subgroup. This model revealed an AUC of 0.818 (SD 0.005) with a sensitivity of 0.647 (SD 0.017) and a specificity of 0.873 (0.023 SD), as shown in Table 2 and in Table S14 in Multimedia Appendix 1. For patients above 55 years of age, the model revealed an AUC of 0.869 (SD 0.005) with a sensitivity of 0.784 (SD 0.011) and a specificity of 0.785 (SD 0.016). The PPV in this subgroup was 0.081% (SD 0.005%), and the negative predictive value was 0.993% (SD 0.000%). The performance of the model was inferior in patients below the age of 55 years; however, it still achieved an AUC of 0.815 (SD 0.007). The discriminatory powers of these models were both excellent among patients with and without a history of lung disease; the AUCs for these subgroups were 0.914 (SD 0.003) and 0.887 (SD 0.002), respectively. Among all the subgroups, the model had the weakest performance in patients below 55 years of age who had no history of lung disease; the AUC for this subgroup was only 0.797 (SD 0.008) for the one-year prospective prediction. By contrast, the model provided the strongest performance for individuals above the age of 55 years with a history of lung disease, which revealed the highest PPV of 14.3% (SD 2.3%). The model exhibited the lowest PPV of 1.0% (SD 0.2%) for individuals less than 55 years of age with no history of lung disease (Table 2). The receiver operating characteristic curves associated with each of these subgroups are summarized in sections S2.1-S2.9 in Multimedia Appendix 1.

Table 2. Discrimination performance (testing set) of the model in the subgroups.

Subgroup	Lung cancer group, n	Control, n	Testing AUC[a] (SD)	Testing sensitivity (SD)	Testing specificity (SD)	PPV[b] (SD), %	NPV[c] (SD), %
Whole population	1304	138,640	0.898 (0.002)	0.805 (0.015)	0.825 (0.018)	4.2 (0.3)	99.8 (0)
Matching age and gender	1304	13,040	0.818 (0.005)	0.647 (0.017)	0.873 (0.023)	34.6 (0.4)	96.0 (0.1)
Age ≥55 years	1046	43,328	0.869 (0.002)	0.784 (0.011)	0.785 (0.016)	8.1 (0.5)	99.3 (0)
Age <55 years	258	95,312	0.815 (0.007)	0.620 (0.080)	0.838 (0.054)	1.1 (0.2)	99.9 (0)
History of lung disease	361	16,596	*0.914 (0.003)*[d]	0.829 (0.021)	0.816 (0.021)	9.0 (0.8)	0.995 (0.1)
No history of lung disease	943	122,044	0.887 (0.002)	0.781 (0.025)	0.827 (0.026)	3.4 (0.5)	99.8 (0.0)
Age ≥55 years with history of lung disease	318	8184	0.875 (0.005)	0.755 (0.047)	0.819 (0.044)	*14.3* (*2.3*)	98.9 (0.2)
Age ≥55 years with no history of lung disease	728	35,144	0.865 (0.003)	0.775 (0.019)	0.786 (0.018)	7.0 (0.4)	99.4 (0.0)
Age <55 years with history of lung disease	43	8,412	0.909 (0.006)	0.777 (0.054)	0.891 (0.036)	3.8 (1.0)	99.9 (0.0)
Age <55 years with no history of lung disease	215	86,900	0.797 (0.008)	0.533 (0.048)	0.865 (0.026)	*1.0* (*0.2*)	99.9 (0.0)

[a]AUC: area under the curve.

[b]PPV: positive predictive value.

[c]NPV: negative predictive value.

[d]Italic text indicates the best performance for the parameter.

Table 3 summarizes the age, gender, diagnosis, and medications associated with both the correctly and incorrectly classified groups from the testing data set. The mean age of the true-positive group was similar to that of the false-positive group and somewhat greater than that of the false-negative group. This tendency was also observed in other subgroups; overall, our results suggest that age and sex are important predictive factors. This is consistent with the t-SNE analysis, in which patients with lung cancer and control patients over 55 years of age were clustered centrally, as compared to the other patients, who were located at the periphery (Figure 3).

The model's hidden layer outputs of 1000 patients with cancer (red dots) and 9000 control patients (green dots) were visualized using t-SNE (Figure 3). Dark green and red represent old age control patients and patients with cancer, respectively. As shown in the left image, most patients with cancer can be clustered away from the control patients. Some dark red dots are mixed with dark green dots in the upper area. These are the patients that were wrongly predicted to be controls by the model. The center images shows that patients aged ≥55 years were clustered in the center of the graph, with the patients with cancer were successfully clustered in the tip area. The right image shows that patients aged <55 years were clustered at the periphery of the graph. Some patients with cancer were also clustered in the tip area, whereas the others were scattered with the control patients.

Occlusion sensitivity analysis further revealed that the specific diagnosis and medication factors were associated with an increased risk of developing lung cancer. Interestingly, "other noninfectious gastroenteritis and colitis" and "other agents for local oral treatment" were associated with the highest risks of developing lung cancer with respect to patient diagnosis and medication, respectively. The top 20 factors identified in the analysis are summarized in Table 4.

Table 3. Prediction analysis of the prospective testing data set (N=139,944).

Group	Patients, n	Age (years), mean (SD)	Male gender, n (%)	Mean diagnosis count (SD), n	Mean medication count (SD), n
All patients					
True positive	1052	69.91 (11.58)	617 (58.65)	141.75 (113.31)	210.7 (186.32)
False positive	22,624	69.19 (12.48)	12,641 (55.87)	114.96 (111.04)	159.14 (171.74)
True negative	116,016	41.94 (13.14)	53,671 (46.26)	63.08 (67.53)	81.46 (101.84)
False negative	252	50.96 (10.79)	134 (53.17)	81.37 (95.67)	104.03 (139.98)
Patients aged ≥55 years					
True positive	851	72.86 (9.25)	510 (59.93)	146.32 (110.84)	217.88 (181.04)
False positive	10,989	74.88 (9.66)	6640 (60.42)	124.11 (119.27)	170.8 (179.15)
True negative	32,339	63.28 (6.58)	13,871 (42.89)	110.24 (97.26)	152.69 (154.96)
False negative	195	64.62 (6.63)	106 (54.36)	125.98 (132.09)	185.08 (216.55)
Patients aged <55 years					
True positive	209	47.87 (6.07)	113 (54.07)	83.3 (87.98)	106.48 (128.64)
False positive	32,765	46.78 (6.58)	18,422 (56.22)	59.4 (63.22)	74.38 (92.27)
True negative	62,547	32.45 (7.43)	27,379 (43.77)	48.67 (48.88)	60.74 (71.36)
False negative	49	36.22 (5.82)	22 (44.90)	63.98 (63.75)	83.88 (115.66)
Patients with a history of lung disease					
True positive	300	72.86 (11.18)	182 (60.67)	184.91 (118.07)	278.71 (194.81)
False positive	2791	75.41 (11.97)	1750 (62.70)	180.66 (140.56)	253.68 (214.05)
True negative	13,805	49.34 (15.6)	5876 (42.56)	119.33 (102.8)	162.24 (162.85)
False negative	61	61.41 (12.11)	34(55.74)	171.72 (155.81)	246.79 (226.86)
Patients with no history of lung disease					
True positive	757	68.45 (11.4)	442 (58.39)	120.97 (104.28)	177.03 (172.5)
False positive	23,328	66.54 (12.25)	12,881 (55.22)	95.23 (94.24)	130.24 (146.34)
True negative	98,716	40.39 (12.27)	45,805 (46.40)	56.19 (59.51)	71.56 (88.63)
False negative	186	48.19 (10.32)	93 (50.00)	65.08 (66.98)	81.69 (101.83)
Patients aged ≥55 years with a history of lung disease					
True positive	255	74.89 (9.03)	160 (62.75)	188.33 (119.58)	284.4 (193.99)
False positive	1778	78.53 (9.16)	1205 (67.77)	188.16 (142.99)	263 (215.97)
True negative	6406	66.38 (7.88)	2669 (41.66)	169.82 (121.41)	239.26 (195.71)
False negative	63	70.44 (7.81)	35 (55.56)	203.87 (148.87)	308.17 (221.29)
Patients aged ≥55 years with no history of lung disease					
True positive	587	71.76 (9.24)	347(59.11)	126.04 (102.89)	185.01 (166.72)
False positive	8958	73.86 (9.69)	5,281(58.95)	104.85 (103.3)	142.56 (154.72)
True negative	26,186	62.73 (6.27)	11,356(43.37)	98.04 (87.47)	135.09 (139.76)
False negative	141	63.47 (6.25)	74(52.48)	100.89 (103.77)	148.73 (195.18)
Patients aged <55 years with lung diseases					
True positive	37	48.89 (6.08)	18 (48.65)	120.46 (100.27)	157.62 (173.25)
False positive	1080	46.56 (7.56)	653 (60.46)	85.56 (72.24)	109.78 (108.74)
True negative	7332	37.7 (9.58)	3099 (42.27)	86.84 (75.16)	113.06 (116.51)
False negative	6	43.33 (9.24)	3 (50.00)	103.67 (98.36)	149.83 (152.85)
Patients aged <55 years with no history of lung disease					

Group	Patients, n	Age (years), mean (SD)	Male gender, n (%)	Mean diagnosis count (SD), n	Mean medication count (SD), n
True positive	172	47.55 (6.07)	95(55.23)	74.94 (83.33)	94.44 (114.72)
False positive	30,982	46.56 (6.56)	17,478(56.41)	55.1 (58.63)	68.47 (84.96)
True negative	55,918	32.06 (7.25)	24,571(43.94)	45.68 (45.68)	56.64 (65.81)
False negative	43	35.65 (5.54)	19(44.19)	59.88 (56.98)	78.84 (108.63)

Figure 3. Visualization of the hidden layer of the model using t-stochastic neighbor embedding.

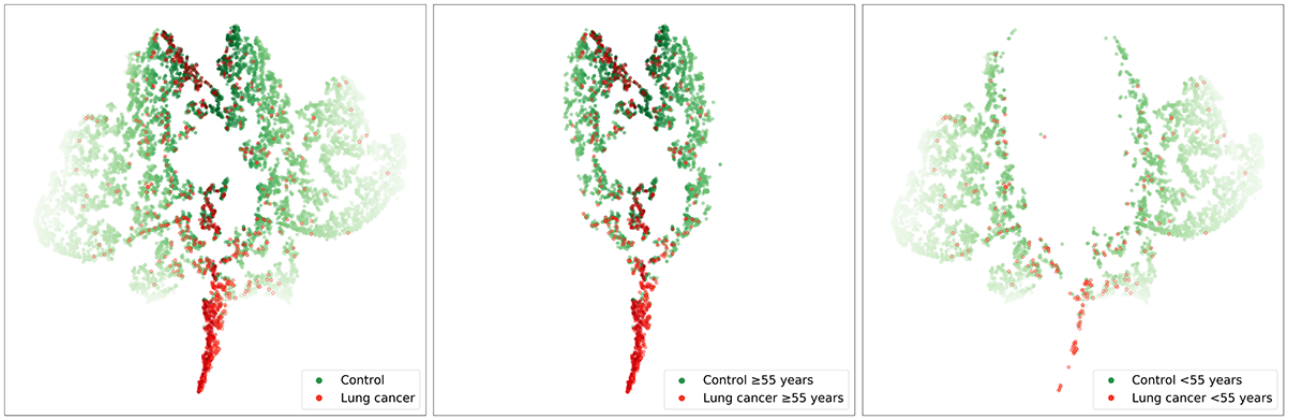


Table 4. Top 20 factors related to lung cancer learned by the model.

Rank	Factor	Lung cancer risk increase (%), mean (SD)
1	Other noninfectious gastroenteritis and colitis	1.85 (1.01)
2	Other congenital anomalies of the circulatory system	1.84 (2.21)
3	Other agents for local oral treatment	1.76 (1.02)
4	Antidotes	1.69 (1.55)
5	Postinflammatory pulmonary fibrosis	1.69 (1.43)
6	Metronidazole	1.69 (1.29)
7	Acariasis	1.65 (1.73)
8	Antiviral drugs	1.57 (1.03)
9	Orchitis and epididymitis	1.57 (1.48)
10	Pneumococcal pneumonia	1.52 (0.93)
11	Buflomedil	1.44 (1.76)
12	Danazol	1.42 (1.41)
13	Calcineurin inhibitors	1.42 (1.29)
14	Other disorders of the urethra and urinary tract	1.37 (1.34)
15	Angina pectoris	1.35 (1.44)
16	Other nonorganic psychoses	1.35 (1.99)
17	Respiratory conditions due to other and unspecified external agents	1.33 (1.33)
18	Open wound of back	1.33 (2.46)
19	Hydrazinophthalazine derivatives	1.31 (1.57)
20	Insulin	1.30 (1.51)

Discussion

Principal Findings

In this study, we explored the possibility of predicting lung cancer using a CNN with diagnosis and medication history extracted from EMRs as a data source. Unlike other proposed lung cancer risk models, our model does not rely on self-reported parameters such as smoking/cessation history, family history, socioeconomic status, or BMI. This model could be readily deployed as a means to evaluate centralized health care databases and perform efficient population-based screening. Such an approach has potential to improve the accuracy of current screening methods, as it can identify those most likely to benefit from interventions [21]. In addition, we attempted to include time-related sequential information as reflected in the medical histories as a means to evaluate lung cancer risk. This approach is different from those used in traditional regression analysis, in which personal history is often simplified and limited to binary or categorical variables. We found that the integration of temporal aspects resulted in improvements in the model performance (Table S12 in Multimedia Appendix 1). The capacity for complex integration of multiple variables is one of the strengths of deep neural networks. To generate this model, we used an established computer vision model (Xception) to extract high-level features from the array representing individual clinical case histories; this ensured that the high-level features associated with the clinical information were effectively extracted for risk prediction.

Related Work

Lung cancer prediction models are under investigation with the goal of identifying high-risk populations that might benefit from LDCT screening. A variety of parameters have been used for prediction, including epidemiologic factors (eg, socioeconomic status, BMI, and smoking history), clinical history (eg, family history and individual history of lung disease history), and results of clinical examinations (eg, blood tests, genetic analysis, and imaging results). The PLCOm2012 model is the most widely validated, with AUCs of 0.78 to 0.82 [27-30]. Likewise, the Bach model exhibited AUCs of 0.66 to 0.75 on external validation [5,31]. Other models include the Haggart model, which exhibited AUCs of 0.71 to 0.84 [5,9], the Liverpool Lung Project model, with AUCs of 0.67 to 0.82 [32], and the Lung Cancer Risk Assessment Tool, which achieved AUCs of 0.77 to 0.78 [5,33]. Some models used information extracted from patient EMRs. The model proposed by Iyen-Omofoman et al [10] used lung-associated clinical symptoms and social-epidemiologic factors from a general practice database, and they achieved an AUC of 0.88; likewise, Wang et al [13] included 33,788 clinical features from clinical histories and laboratory tests evaluated in an extreme gradient boosting (XGBoost) model to achieve an AUC of 0.88. With these previous studies in mind, our model featured a deep learning approach and achieved a prospective prediction AUC of 0.87 in patients older than 55 years and 0.90 for the entire patient cohort. It is possible to test other machine learning models (eg, support vector machine or random forest) on our data set. However, this study serves as a proof of concept of using CNN with nonimaging medical records. Comparing the performance of this model to that of different machine learning models of practical interest would be an interesting approach for future studies.

We recognize that direct comparisons between models may not be fully appropriate, as the target populations and predicted outcomes can vary. Previous reports suggested that the performance of models is inflated when nonsmokers and younger subjects (<55 years of age) are included in the study groups [34]. Our findings confirm this point, as can be observed from the higher AUCs associated with the younger age cutoffs (Table S3, Multimedia Appendix 1). Although our data set did not directly include reports of smoking history or cessation, we did include a history of lung diseases (eg, chronic bronchitis, COPD, and emphysema) among our parameters; these could easily be considered as surrogate factors for smoking history. Further analysis of this patient subgroup may help us understand and mitigate the possibility of performance inflation.

In the original NLST trial, the PPV for the LDCT was determined to be 3.4% [1]. The high false-positive rate associated with this intervention remains a major concern with respect to LDCT screening. In this study, the highest PPV (14.5%) was observed in patients ≥55 years of age with a history of lung disease. As noted above, an increase in cancer diagnoses might be expected in this patient subgroup, as a history of lung disease may be a direct consequence of smoking. As such, this finding suggested that individuals in this subgroup are suitable candidates for model prescreening in an effort to avoid unnecessary radiation exposure and costs associated with LDCT. In addition, we found that the 55-year age cutoff selected in the original NLST trial was also appropriate for our model, as the predictive performance was higher with this age cutoff compared to that observed at cutoffs at age 50 or 60 years (Table S3, Multimedia Appendix 1).

Predictive Factor Analysis

The inclusion of an age- and gender-matched subgroup was necessary to explore the roles of clinical diagnosis and medication history in the predictions generated by our model; evaluation of this subgroup prevented the confounding effects of age and its correlations to clinical history (eg, older people are typically prescribed more chronic disease-related medications). With this consideration, our model achieved an AUC of 0.818. These findings can be compared to the model proposed by Spitz et al [12], which included gender-, age-, and smoking status–matched patients and achieved an AUC of 0.63 in former smokers. Although the models generated from matched populations tended to display weaker performance than those from nonmatched populations and may not be clinically useful, this result provided us with a more clear-cut evaluation of the specific parameters included in this model. Taken together, our findings suggest that our model is capable of identifying factors that are useful for predicting lung cancer using clinical information available 1 year before the clinical diagnosis is made.

Our model demonstrated the worst performance in young patients without pre-existing lung diseases. This finding suggests that identifying high-risk patients among young and asymptomatic patients is still the most challenging task. Further

studies are required to assess the performance of the model in patients with different staging. One of the major concerns with respect to the use of lung cancer prediction models is that they tend to select individuals who are older and who have multiple comorbidities [35], thus reducing the overall benefit gained from the screening process [36]. This tendency was also observed in our model. This phenomenon cannot be fully avoided, as it simply reflects the high percentage of older patients in the population who are diagnosed with lung cancer. However, when focused on patients younger than 55 years of age, our model maintained excellent discriminative power (the AUC was 0.82, with a mean age of true positives of 47.8 years). With the current model, the inclusion of younger individuals remains possible; multiple age-stratified thresholds for lung cancer risk could further optimize the clinical benefits of the predictions from this model.

Although deep learning is often considered a "black box," and it is often challenging to explain the reasoning behind the outcomes, our study used t-SNE and occlusion sensitivity analysis to identify the most critical of the contributing parameters. Our occlusion sensitivity analysis revealed that many of the important factors were those associated with a history of preexisting lung conditions (eg, postinflammatory pulmonary fibrosis and pneumococcal pneumonia) and medications used to treat smoking-related diseases (eg, buflomedil for peripheral arterial disease and angina pectoris, and insulin for insulin resistance of diabetes mellitus) with increased cancer risk (eg, congenital anomalies of the circulatory system [37] and periodontal conditions [38]), and paraneoplastic phenomena (eg, noninfectious gastroenteritis and colitis [39]). This information must be interpreted carefully, as these findings do not imply a causal relationship. For example, the model may predict an increased likelihood of future lung cancer in patients with pre-existing lung disease simply because these patients receive frequent medical attention; thus, there is a higher likelihood that cancer will be detected incidentally. In addition, the sensitivity analysis in this study is only capable of evaluating one factor at a time; this is a major limitation of the explainability of the model, given the fact that our model was designed to integrate complex, high-level features. Finally, we could not explain some of the predictive features identified by this model, such as the associations with terms including *antidote*, *orchitis*, and *epididymitis*. More studies will be required to decode the findings from the CNN and to elucidate the interactions between age, sex, previous diagnoses, and medications.

Although our model achieved excellent discriminative performance, poor calibration was noted, together with the fact that direct numeric output would overestimate the actual risk. This is a known phenomenon associated with modern neural networks [40]. Unlike the traditional logistic regression models, which perform well in calibration because they directly minimize the loss of calibration, modern neural networks tend to perform suboptimally in this regard. This is likely due to the regularization methods (eg, dropout and batch normalization) and the multiple deep layers applied as components of the model architecture [40]. In our study, poor calibration did not limit the use of the model, as individuals were selected based on a predefined threshold identified in the validation data set rather than on the numerical output of the model. As a result, the increased rates reported in Table 4 do not represent the actual cancer risk.

Our model used nonimaging medical information from EMRs; however, we still used CNN as the model backbone. The study design and aims are different from other lung cancer studies that used CNN to analyze computed tomography (CT) scans and determine if a pulmonary nodule is malignant. Their models were used to automatically identify suspicious nodules from CT scans, which were already present, whereas our model attempted to identify patients with high risk of developing lung cancer in the future.

Limitations

There are several limitations to this study. First, the data collection was limited to the NHIRD database of Taiwan; the patient records do not include tissue histology or lung cancer staging data. Patients with small cell lung cancer and mutation-rich non–small cell lung cancer (eg, epidermal growth factor receptor, anaplastic lymphoma kinase, ROS-1) could not be separated. These specific types may have different disease courses and risk factors; therefore, they were usually not included in the traditional screening, and the benefit of receiving screening is undetermined. Our subgroup analysis did include only patients with pre-existing lung diseases, but this did not mitigate the issue entirely. Similarly, the NHIRD database does not include information on patients' lifestyles or any genetic or laboratory data. A subgroup analysis of patients with lung cancer based on tissue histology and staging might help to develop a prediction model that was tailored to different risk groups. Second, the data set did not contain any information on smoking status, which is clearly an important risk factor associated with lung cancer development. This limitation restricted the external validation and the comparisons that could be made between our model and those described in earlier published studies. The authors believe that self-reported information, such as family history, smoking/cessation history, and duration of symptoms, are valuable pieces of information for lung cancer prediction that are very important and can further improve prediction accuracy. In our study, a history of lung diseases (eg, COPD and emphysema) was used as a proxy for a smoking history; our model performed with excellent discriminative power with respect to this subgroup. Finally, the NHIRD includes primarily Taiwanese people; as such, the target population was fairly homogeneous, with limited ethnic diversity. The identified risk factors may not apply to other populations with other ethnicities. Nonetheless, the methodology used here could be easily applied to other medical databases with more diverse patient populations.

Conclusion

Our CNN model exhibited robust performance with respect to the 1-year prospective prediction of the risk of developing lung cancer. As our model included sequential data on clinical diagnoses and medication history, it was capable of capturing features associated with evolving clinical conditions and as such was able to identify patients at higher risk of developing lung cancer. With appropriate ethical regulation, this model may be

deployed as a means to analyze medical databases, thus paving the way for efficient population-based screening and digital precision medicine. A future randomized controlled trial will be required to explore the clinical benefit of this model in diverse populations.

Acknowledgments

This research was funded in part by Ministry of Education (MOE) grants MOE 109-6604-001-400 and DP2-110-21121-01-A-01.

Authors' Contributions

MCHY contributed to the data analysis, model construction, interpretation of results, drafting of the manuscript, and literature review. YHW and HCY contributed to the data curation and data preprocessing. KJB contributed to the investigation and the interpretation of the results. HHW contributed to the interpretation of results, conceptualization, supervision, and manuscript editing. YCL contributed to the conceptualization, supervision, manuscript editing, and interpretation of the results. HHW and YCL contributed equally to this article. The corresponding author, YCL, affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary tables and figures.
[DOCX File , 770 KB-Multimedia Appendix 1]

References

1. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011 Aug 04;365(5):395-409 [FREE Full text] [doi: 10.1056/NEJMoa1102873] [Medline: 21714641]
2. de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. N Engl J Med 2020 Feb 06;382(6):503-513. [doi: 10.1056/NEJMoa1911793] [Medline: 31995683]
3. Moyer VA, U.S. Preventive Services Task Force. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2014 Mar 04;160(5):330-338 [FREE Full text] [doi: 10.7326/M13-2771] [Medline: 24378917]
4. Wood DE, Kazerooni EA, Baum SL, Eapen GA, Ettinger DS, Hou L, et al. Lung cancer screening, version 3.2018, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw 2018 Apr;16(4):412-441 [FREE Full text] [doi: 10.6004/jnccn.2018.0020] [Medline: 29632061]
5. Katki HA, Kovalchik SA, Petito LC, Cheung LC, Jacobs E, Jemal A, et al. Implications of nine risk prediction models for selecting ever-smokers for computed tomography lung cancer screening. Ann Intern Med 2018 Jul 03;169(1):10-19 [FREE Full text] [doi: 10.7326/M17-2701] [Medline: 29800127]
6. Ruano-Ravina A, Pérez-Ríos M, Casàn-Clará P, Provencio-Pulla M. Low-dose CT for lung cancer screening. Lancet Oncol 2018 Mar;19(3):e131-e132. [doi: 10.1016/S1470-2045(18)30121-9] [Medline: 29508756]
7. Bach PB, Kattan MW, Thornquist MD, Kris MG, Tate RC, Barnett MJ, et al. Variations in lung cancer risk among smokers. J Natl Cancer Inst 2003 Mar 19;95(6):470-478. [doi: 10.1093/jnci/95.6.470] [Medline: 12644540]
8. Cassidy A, Myles JP, van Tongeren M, Page RD, Liloglou T, Duffy SW, et al. The LLP risk model: an individual risk prediction model for lung cancer. Br J Cancer 2008 Jan 29;98(2):270-276 [FREE Full text] [doi: 10.1038/sj.bjc.6604158] [Medline: 18087271]
9. Hoggart C, Brennan P, Tjonneland A, Vogel U, Overvad K, Østergaard JN, et al. A risk model for lung cancer incidence. Cancer Prev Res (Phila) 2012 Jun;5(6):834-846 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-11-0237] [Medline: 22496387]
10. Iyen-Omofoman B, Tata LJ, Baldwin DR, Smith CJP, Hubbard RB. Using socio-demographic and early clinical features in general practice to identify people with lung cancer earlier. Thorax 2013 May;68(5):451-459. [doi: 10.1136/thoraxjnl-2012-202348] [Medline: 23321602]
11. Marcus MW, Chen Y, Raji OY, Duffy SW, Field JK. LLPi: Liverpool Lung Project risk prediction model for lung cancer incidence. Cancer Prev Res (Phila) 2015 Jun;8(6):570-575 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-14-0438] [Medline: 25873368]
12. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, et al. A risk model for prediction of lung cancer. J Natl Cancer Inst 2007 May 02;99(9):715-726. [doi: 10.1093/jnci/djk153] [Medline: 17470739]

13. Wang X, Zhang Y, Hao S, Zheng L, Liao J, Ye C, et al. Prediction of the 1-year risk of incident lung cancer: prospective study using electronic health records from the state of Maine. J Med Internet Res 2019 May 16;21(5):e13260 [FREE Full text] [doi: 10.2196/13260] [Medline: 31099339]
14. Wilson DO, Weissfeld J. A simple model for predicting lung cancer occurrence in a lung cancer screening program: The Pittsburgh Predictor. Lung Cancer 2015 Jul;89(1):31-37 [FREE Full text] [doi: 10.1016/j.lungcan.2015.03.021] [Medline: 25863905]
15. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 2019 Jun;25(6):954-961. [doi: 10.1038/s41591-019-0447-x] [Medline: 31110349]
16. Jena S, George S. Morphological feature extraction and KNG‐CNN classification of CT images for early lung cancer detection. Int J Imaging Syst Technol 2020 Jun 09;30(4):1324-1336 [FREE Full text] [doi: 10.1002/ima.22445]
17. Priya MMA, Jawhar SJ. Advanced lung cancer classification approach adopting modified graph clustering and whale optimisation‐based feature selection technique accompanied by a hybrid ensemble classifier. IET Image Process 2020 Jul 08;14(10):2204-2215 [FREE Full text] [doi: 10.1049/iet-ipr.2019.0178]
18. Manikandan T, Bharathi N. Lung cancer detection using fuzzy auto-seed cluster means morphological segmentation and SVM classifier. J Med Syst 2016 Jul;40(7):181. [doi: 10.1007/s10916-016-0539-9] [Medline: 27299354]
19. Gao N, Tian S, Li X, Huang J, Wang J, Chen S, et al. Three-dimensional texture feature analysis of pulmonary nodules in CT images: lung cancer predictive models based on Support Vector Machine Classifier. J Digit Imaging 2020 Apr;33(2):414-422 [FREE Full text] [doi: 10.1007/s10278-019-00238-8] [Medline: 31529236]
20. Wang H, Wang Y, Liang C, Li Y. Assessment of deep learning using nonimaging information and sequential medical records to develop a prediction model for nonmelanoma skin cancer. JAMA Dermatol 2019 Nov 01;155(11):1277-1283 [FREE Full text] [doi: 10.1001/jamadermatol.2019.2335] [Medline: 31483437]
21. Wender RC, Brawley OW, Fedewa SA, Gansler T, Smith RA. A blueprint for cancer screening and early detection: advancing screening's contribution to cancer control. CA Cancer J Clin 2019 Jan;69(1):50-79 [FREE Full text] [doi: 10.3322/caac.21550] [Medline: 30452086]
22. Hsieh C, Su C, Shao S, Sung S, Lin S, Kao Yang Y, et al. Taiwan's National Health Insurance Research Database: past and future. Clin Epidemiol 2019;11:349-358 [FREE Full text] [doi: 10.2147/CLEP.S196293] [Medline: 31118821]
23. Kao W, Hong J, See L, Yu H, Hsu J, Chou I, et al. Validity of cancer diagnosis in the National Health Insurance database compared with the linked National Cancer Registry in Taiwan. Pharmacoepidemiol Drug Saf 2018 Oct;27(10):1060-1066. [doi: 10.1002/pds.4267] [Medline: 28815803]
24. Kovalchik SA, Tammemagi M, Berg CD, Caporaso NE, Riley TL, Korch M, et al. Targeting of low-dose CT screening according to the risk of lung-cancer death. N Engl J Med 2013 Jul 18;369(3):245-254 [FREE Full text] [doi: 10.1056/NEJMoa1301851] [Medline: 23863051]
25. Chollet F. Xception: deep learning with depthwise separable convolutions. 2017 Apr 04 Presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); July 22-25, 2017; Honolulu, HI. [doi: 10.1109/cvpr.2017.195]
26. Zeiler M, Fergus R. Visualizing and understanding convolutional networks. In: Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8689. 2014 Presented at: European Conference on Computer Vision 2014; September 6-12, 2014; Zurich, Switzerland p. 818-833. [doi: 10.1007/978-3-319-10590-1_53]
27. Li K, Hüsing A, Sookthai D, Bergmann M, Boeing H, Becker N, et al. Selecting high-risk individuals for lung cancer screening: a prospective evaluation of existing risk models and eligibility criteria in the German EPIC cohort. Cancer Prev Res (Phila) 2015 Sep;8(9):777-785 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-14-0424] [Medline: 26076698]
28. Tammemagi MC, Schmidt H, Martel S, McWilliams A, Goffin JR, Johnston MR, PanCan Study Team. Participant selection for lung cancer screening by risk modelling (the Pan-Canadian Early Detection of Lung Cancer [PanCan] study): a single-arm, prospective study. Lancet Oncol 2017 Nov;18(11):1523-1531. [doi: 10.1016/S1470-2045(17)30597-1] [Medline: 29055736]
29. Weber M, Yap S, Goldsbury D, Manners D, Tammemagi M, Marshall H, et al. Identifying high risk individuals for targeted lung cancer screening: independent validation of the PLCO risk prediction tool. Int J Cancer 2017 Jul 15;141(2):242-253 [FREE Full text] [doi: 10.1002/ijc.30673] [Medline: 28249359]
30. Tammemägi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection criteria for lung-cancer screening. N Engl J Med 2013 Feb 21;368(8):728-736 [FREE Full text] [doi: 10.1056/NEJMoa1211776] [Medline: 23425165]
31. D'Amelio AM, Cassidy A, Asomaning K, Raji OY, Duffy SW, Field JK, et al. Comparison of discriminatory power and accuracy of three lung cancer risk models. Br J Cancer 2010 Jul 27;103(3):423-429 [FREE Full text] [doi: 10.1038/sj.bjc.6605759] [Medline: 20588271]
32. Raji OY, Duffy SW, Agbaje OF, Baker SG, Christiani DC, Cassidy A, et al. Predictive accuracy of the Liverpool Lung Project risk model for stratifying patients for computed tomography screening for lung cancer: a case-control and cohort validation study. Ann Intern Med 2012 Aug 21;157(4):242-250 [FREE Full text] [doi: 10.7326/0003-4819-157-4-201208210-00004] [Medline: 22910935]
33. Katki HA, Kovalchik SA, Berg CD, Cheung LC, Chaturvedi AK. Development and validation of risk models to select ever-smokers for CT lung cancer screening. JAMA 2016 Jun 07;315(21):2300-2311 [FREE Full text] [doi: 10.1001/jama.2016.6255] [Medline: 27179989]

34. Tammemägi MC. Application of risk prediction models to lung cancer screening: a review. J Thorac Imaging 2015 Mar;30(2):88-100. [doi: 10.1097/RTI.0000000000000142] [Medline: 25692785]
35. Mazzone PJ, Silvestri GA, Patel S, Kanne JP, Kinsinger LS, Wiener RS, et al. Screening for lung cancer: CHEST guideline and expert panel report. Chest 2018 Apr;153(4):954-985. [doi: 10.1016/j.chest.2018.01.016] [Medline: 29374513]
36. Cheung LC, Berg CD, Castle PE, Katki HA, Chaturvedi AK. Life-gained-based versus risk-based selection of smokers for lung cancer screening. Ann Intern Med 2019 Nov 05;171(9):623-632 [FREE Full text] [doi: 10.7326/M19-1263] [Medline: 31634914]
37. Mandalenakis Z, Karazisi C, Skoglund K, Rosengren A, Lappas G, Eriksson P, et al. Risk of cancer among children and young adults with congenital heart disease compared with healthy controls. JAMA Netw Open 2019 Jul 03;2(7):e196762 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.6762] [Medline: 31276179]
38. Mai X, LaMonte MJ, Hovey KM, Nwizu N, Freudenheim JL, Tezal M, et al. History of periodontal disease diagnosis and lung cancer incidence in the Women's Health Initiative Observational Study. Cancer Causes Control 2014 Aug;25(8):1045-1053 [FREE Full text] [doi: 10.1007/s10552-014-0405-3] [Medline: 24913780]
39. Schmidt-Hieber M, Bierwirth J, Buchheidt D, Cornely OA, Hentrich M, Maschmeyer G, AGIHO Working Group. Diagnosis and management of gastrointestinal complications in adult cancer patients: 2017 updated evidence-based guidelines of the Infectious Diseases Working Party (AGIHO) of the German Society of Hematology and Medical Oncology (DGHO). Ann Hematol 2018 Jan;97(1):31-49 [FREE Full text] [doi: 10.1007/s00277-017-3183-7] [Medline: 29177551]
40. Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. In: Proceedings of the 34th International Conference on Machine Learning. 2017 Presented at: 34th International Conference on Machine Learning; August 6-11, 2017; Sydney, Australia p. 1321-1330.

Abbreviations

ATC: Anatomical Therapeutic Chemical
AUC: area under the receiver operating characteristic curve
CNN: convolutional neural network
COPD: chronic obstructive pulmonary disease
CT: computed tomography
EMR: electronic medical record
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
LDCT: low-dose computed tomography
MOE: Ministry of Education
NHIRD: National Health Insurance Research Database
NLST: National Lung Cancer Screening Trial
PPV: positive predictive value
t-SNE: t-distributed stochastic neighbor embedding
WHO: World Health Organization
XGBoost: extreme gradient boosting

Edited by R Kukafka; submitted 04.12.20; peer-reviewed by KL Ong, V Montmirail, JA Benítez-Andrades; comments to author 13.02.21; revised version received 03.04.21; accepted 04.05.21; published 03.08.21

Please cite as:
Yeh MCH, Wang YH, Yang HC, Bai KJ, Wang HH, Li YC(
Artificial Intelligence–Based Prediction of Lung Cancer Risk Using Nonimaging Electronic Medical Records: Deep Learning Approach
J Med Internet Res 2021;23(8):e26256
URL: https://www.jmir.org/2021/8/e26256
doi: 10.2196/26256
PMID:

Original Paper

Obesity and BMI Cut Points for Associated Comorbidities: Electronic Health Record Study

Natalie Liu[1], MD; Jen Birstler[2], MSc; Manasa Venkatesh[1], MSc; Lawrence Hanrahan[3], MSc, PhD; Guanhua Chen[2], PhD; Luke Funk[1,4], MD, MPH

[1]Department of Surgery, University of Wisconsin-Madison, Madison, WI, United States

[2]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, United States

[3]Department of Family Medicine and Community Health, University of Wisconsin-Madison, Madison, WI, United States

[4]Department of Surgery, William S. Middleton Memorial VA, Madison, WI, United States

Corresponding Author:
Luke Funk, MD, MPH
Department of Surgery
University of Wisconsin-Madison
600 Highland Ave
Madison, WI, 53792
United States
Phone: 1 6082631036
Fax: 1 6082520942
Email: funk@surgery.wisc.edu

Abstract

Background: Studies have found associations between increasing BMIs and the development of various chronic health conditions. The BMI cut points, or thresholds beyond which comorbidity incidence can be accurately detected, are unknown.

Objective: The aim of this study is to identify whether BMI cut points exist for 11 obesity-related comorbidities.

Methods: US adults aged 18-75 years who had ≥3 health care visits at an academic medical center from 2008 to 2016 were identified from eHealth records. Pregnant patients, patients with cancer, and patients who had undergone bariatric surgery were excluded. Quantile regression, with BMI as the outcome, was used to evaluate the associations between BMI and disease incidence. A comorbidity was determined to have a cut point if the area under the receiver operating curve was >0.6. The cut point was defined as the BMI value that maximized the Youden index.

Results: We included 243,332 patients in the study cohort. The mean age and BMI were 46.8 (SD 15.3) years and 29.1 kg/m^2, respectively. We found statistically significant associations between increasing BMIs and the incidence of all comorbidities except anxiety and cerebrovascular disease. Cut points were identified for hyperlipidemia (27.1 kg/m^2), coronary artery disease (27.7 kg/m^2), hypertension (28.4 kg/m^2), osteoarthritis (28.7 kg/m^2), obstructive sleep apnea (30.1 kg/m^2), and type 2 diabetes (30.9 kg/m^2).

Conclusions: The BMI cut points that accurately predicted the risks of developing 6 obesity-related comorbidities occurred when patients were overweight or barely met the criteria for class 1 obesity. Further studies using national, longitudinal data are needed to determine whether screening guidelines for appropriate comorbidities may need to be revised.

(J Med Internet Res 2021;23(8):e24017) doi: 10.2196/24017

KEYWORDS

obesity; body mass index (BMI); risk factors; screening; health services; chronic disease

Introduction

Background

Obesity (BMI ≥30.0 kg/m^2) is a global public health problem. The highest rates of obesity occur in the United States, where over one-third of adults have obesity [1]. In 1998, the World Health Organization created international standardized BMI classifications for adults who are overweight and have obesity based on risks of obesity-related diseases for European adults [2]. These classifications were based on the risks of obesity-related diseases in European adults with varied BMI values [3]. On the basis of these classifications, overweight and obesity were defined as having a BMI between 25.0 and 29.9 kg/m^2 and a BMI≥30.0 kg/m^2, respectively. However, studies have demonstrated that the risks of obesity-related comorbidities differ based on sex and race or ethnicity. Female Asian patients have been shown to develop comorbidities at lower BMIs, suggesting that BMI thresholds for overweight and obesity should be lower for these groups [2,4-7].

Study Significance

Obesity is associated with numerous comorbidities, including hypertension, hyperlipidemia, type 2 diabetes mellitus (T2DM), and coronary artery disease (CAD) [8-10]. The cross-sectional study by Pantalone et al [8], which used electronic health record (EHR) data, showed that patients with higher BMIs had a higher prevalence of T2DM, hypertension, and CAD. However, studies have not addressed whether specific BMI cut points exist for US adults. BMI cut points are defined as the thresholds beyond which disease incidence can be accurately detected. In addition, no studies have evaluated cut points by using EHR data that provide patient-level information for large, multiethnic cohorts. Studies have concluded that it is feasible to use EHR analysis to study chronic diseases such as obesity, diabetes, and hypertension [11,12].

Objective

The objective of this study is to examine EHR data from a large health care system in the United States to determine whether BMI cut points exist for 11 common comorbidities associated with obesity and being overweight. We also evaluate whether cut points varied with sex and race or ethnicity. We hypothesize that most cut points would occur in the class 1 obesity category.

Methods

Data Source

We used data from the University of Wisconsin Hospital and Clinics EHR over a 10-year period (June 1, 2008, to December 31, 2018). All patient data and analyses were stored on a secure server managed through the University of Wisconsin Health Information Services and the Institute for Clinical and Translational Research. The Epic Clarity Database was used as the data source for all patients. This study was approved by the University of Wisconsin Minimal Risk institutional review board (protocol #2017-0443), and the need for informed consent was waived. We followed the Strengthening the Reporting of Observational Studies in Epidemiology guidelines within the Enhancing the Quality and Transparency of Health Research network in the methodology and reporting of this study (Multimedia Appendix 1 contains the full Strengthening the Reporting of Observational Studies in Epidemiology checklist) [13].

Data Validation and Cleaning

All recorded heights and weights in the EHR were cleaned to reduce the inclusion of incorrect heights and weights because of errors in data entry. Similar to our previous study using EHR data, we used the methodology proposed by Cheng et al [14] to remove biologically implausible heights and weights [15]. All heights >90 inches, <44 inches, and >1 SD from the mean height when SD was >2.5% of the mean were removed. All weights >1000 pounds, <55 pounds, >70% of the range from the mean when the range ≥50 pounds, and >1 SD from the mean when the SD was >20% of the mean were removed. Missing height data were imputed with the most recent previous nonmissing valid height. Any remaining missing height was replaced with the most recent subsequent nonmissing valid height. BMI values were calculated using the valid heights and weights. No patients were excluded from the study because of the data cleaning process.

Study Population

We included all patients between the ages of 18 and 75 years who had ≥3 in-person clinical visits over a minimum of 2 years documented in the EHR during the study period. All included patients had an *index visit* with a valid BMI measurement, another visit at least 1 year before the index visit, and an additional visit 1 year after their index visit. The minimum 1-year period between the index visit and the previous visit was used to identify patients who had each disease of interest versus those who did not. The 1-year period between the index visit and the subsequent visit was used to calculate 1-year incidence rates for patients who did not have the disease before the index visit but were later diagnosed with the disease. Patients with multiple intervals of ≥3 clinical visits had an interval selected at random.

Patients with a pregnancy or cancer diagnosis at any time before or during the study period were excluded using the International Classification of Disease (ICD)-9 and ICD-10 codes. Patients who had undergone bariatric surgery were identified from our institutional bariatric surgery registry and excluded.

Study Variables

Baseline BMI (BMI at the index visit), age (at the index visit), sex (male or female), race or ethnicity (White, non-Hispanic; Black, non-Hispanic; Asian, non-Hispanic; Native American, non-Hispanic; Hispanic; or other or unspecified), insurance type (commercial or private insurance, Medicare, Medicaid, or other or unspecified), and smoking status (at the index visit; active smoker, former smoker, passive smoker [defined as an individual who has had exposure to tobacco smoke but has never smoked themselves], or nonsmoker) were identified from the EHR. Insurance type was defined as the insurance type used during or before the index visit.

Through a literature review, we identified 11 common obesity-related comorbidities that were included in this study: anxiety, CAD, cerebrovascular disease, chronic pain, depression, gastroesophageal reflux disease, hyperlipidemia, hypertension, obstructive sleep apnea (OSA), osteoarthritis, and T2DM [8,9,16,17]. Incident cases were defined as patients who did not have the disease before the index visit and subsequently developed the disease after the index visit. The 1-year incidence rates (defined per 100 person-years) were calculated based on the occurrence of an ICD-9 or ICD-10 code (Multimedia Appendix 2 contains the full list of ICD-9 and ICD-10 codes) during the 1-year period following the index visit for patients who did not have a diagnosis before the index visit. Prevalent cases were defined as patients who had a diagnosis of comorbidity at or before the index visit and identified using the occurrence of an ICD-9 or ICD-10 code during this time.

Statistical Analysis

We used quantile regression with BMI as the outcome to identify differences in the median BMIs between incident cases of each comorbidity and those who did not develop each comorbidity. Two models were fit for each comorbidity to evaluate the associations between BMI and disease incidence—an unadjusted model with disease incidence as the only independent variable and an adjusted model accounting for baseline age, sex, race or ethnicity, and smoking status. We used quantile regression because we were unable to meet the assumptions of the linear model. Quantile regression also allowed for the evaluation of differences in BMI distributions among patients who developed each comorbidity versus those who did not, which is more informative than differences in single mean values [18]. The difference in median BMIs (the median BMI of incident cases minus the median BMI of patients who did not develop the disease) was the outcome of the quantile model.

We conducted cut point analyses with BMI as a screening test for the incidence of each obesity-related comorbidity. Sensitivity and specificity were calculated for continuous BMI values. A comorbidity had a BMI cut point if the area under the receiver operating curve (AUROC) was >0.6. We chose an AUROC>0.6 to ensure that cut points had significant diagnostic value. Although there is no gold standard method, other investigators have used AUROC thresholds that range from >0.5 to >0.7 to determine cut points [6]. For all comorbidities with an AUROC >0.6, the cut point was defined as the BMI value that maximized the Youden index (sensitivity+specificity-1). BMI cut points were also calculated by sex and race or ethnicity and compared using the bootstrap method with 1000 resamplings. The overall incidence rates above and below each cut point were calculated. For any comorbidities that had an identifiable cut point, baseline characteristics and prevalence of any concurrent comorbidities were compared between patients who developed the comorbidity and those who did not develop the comorbidity.

All statistical analyses were conducted using R version 3.6.3 (R Foundation for Statistical Computing).

Incidence Versus Prevalence Cut Point Analysis

Studies have identified cut points for diseases such as diabetes, hypertension, and hyperlipidemia using both incidence and prevalence [6,19]. As there is no standardized method to determine cut points, we analyzed cut point differences between prevalent and incident cases. For any comorbidities that had an identifiable cut point, we used the bootstrap method with 1000 resamplings to determine cut points and *P* values comparing incident and prevalent cases.

Results

Patient Characteristics

Over 300,000 patients had at least three clinical visits during the study period. After applying exclusion criteria, 243,332 patients met inclusion criteria (Figure 1). The mean age was 46.8 (SD 15.3) years (Table 1). Of the patients, 54.9% (133,654/243,332) of the patients were female, and 88.7% (215,950/243,332) patients were White and non-Hispanic. The mean BMI was 29.1 (SD 7.0) kg/m^2, and 36.8% (89,660/243,332) of patients had a BMI ≥30 kg/m^2. In our study cohort, 57.7% (139,753/243,332) of patients had never smoked or used tobacco products, whereas 14.1% (34,328/243,332) of patients were active smokers. Hyperlipidemia and hypertension were the most common comorbidities, affecting 24.3% (59,097/243,332) patients and 21.5% (52,365/243,332) of the study population, respectively (Table 1).

Figure 1. Study cohort creation (Strengthening the Reporting of Observational Studies in Epidemiology diagram). EHR: electronic health record.

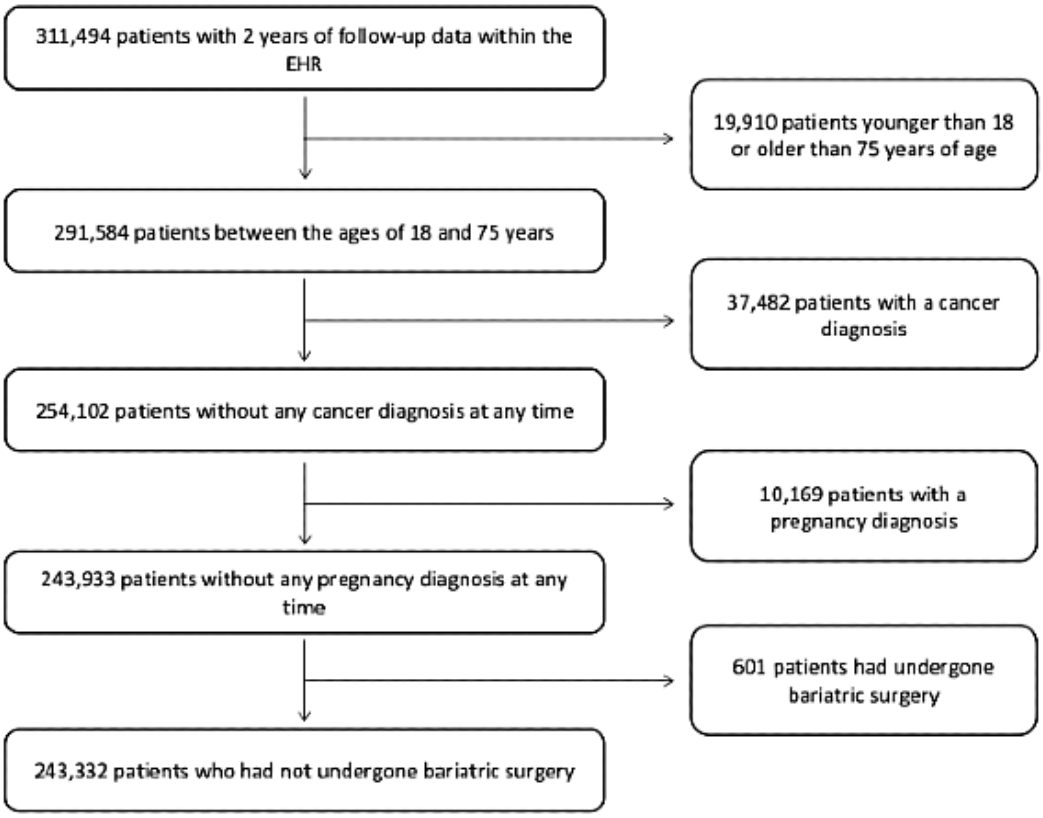


Table 1. Baseline demographics and patient characteristics (N=243,332).

Characteristics	Values
Age (years), mean (SD)	46.8 (15.3)
Sex, n (%)	
Male	109,678 (45.1)
Female	133,654 (54.9)
Race or ethnicity, n (%)	
White, non-Hispanic	215,950 (88.7)
Black, non-Hispanic	9463 (3.9)
Asian, non-Hispanic	6621 (2.7)
Native American, non-Hispanic	1161 (0.5)
Hispanic	7730 (3)
Other or unspecified	2767 (1.1)
Baseline BMI category (kg/m^2)[a], n (%)	
Underweight (<18.5)	3000 (1.2)
Normal (18.5-24.9)	72,803 (29.9)
Overweight (25.0-29.9)	77,869 (32)
Class 1 obesity (30.0-34.9)	48,213 (19.8)
Class 2 obesity (35.0-39.9)	23,371 (9.6)
Class 3 obesity (>40)	18,076 (7.4)
Insurance type, n (%)	
Commercial	191,697 (78.8)
Medicare	31,778 (5.7)
Medicaid	6032 (2.5)
Other or unspecified	13,825 (5.7)
Prevalence of comorbidities, n (%)	
Anxiety	33,984 (14)
Coronary artery disease	9543 (3.9)
Cerebrovascular disease	3076 (1.3)
Chronic pain	14,479 (6)
Depression	32,210 (13.2)
Gastroesophageal reflux	29,512 (12.1)
Hyperlipidemia	59,097 (24.3)
Hypertension	52,365 (21.5)
Obstructive sleep apnea	13,746 (5.6)
Osteoarthritis	21,408 (8.8)
Type 2 diabetes mellitus	18,182 (7.5)
Smoking status, n (%)	
Active smoker	34,328 (14.1)
Former smoker	64,331 (26.4)
Passive smoker	2746 (1.1)
Nonsmoker	139,753 (57.4)

[a]The mean baseline BMI was 29.1 kg/m^2 (SD 7.0 kg/m^2).

Incidence of 11 Comorbidities and Their Associations With BMI

The highest 1-year incidence rates were for hyperlipidemia (4.0 cases per 100 person-years) and hypertension (3.6 cases per 100 person-years; Multimedia Appendix 3 contains the full table of 1-year incidence rates). CAD and cerebrovascular disease had the lowest 1-year incidence rates (0.9 and 0.4 cases per 100-person-years, respectively).

In quantile regression, when comparing the median BMI of those who developed each comorbidity (incident group) versus the median BMI of those who did not, we found statistically significant differences in the median BMIs for all obesity-related comorbidities (Multimedia Appendix 4 contains the full table of the quantile regression analysis evaluating associations between BMI and comorbidity incidence). The median BMIs of the incident groups were higher for all comorbidities except for anxiety (–0.6 kg/m^2; 95% CI –0.8 to –0.4).

After adjusting for age, sex, race or ethnicity, and smoking status, we found statistically significant differences in the median BMIs for all comorbidities except anxiety and cerebrovascular disease (Multimedia Appendix 4 contains the full table of the quantile regression analysis evaluating associations between BMI and comorbidity incidence). The adjusted median BMIs of the incident groups were higher for all comorbidities. The greatest differences in adjusted median BMI were for OSA (6.0 kg/m^2; 95% CI 5.7-6.4) and T2DM (5.0 kg/m^2; 95% CI 4.6-5.4).

BMI Cut Points for All Study Patients

Six comorbidities had BMI cut points: CAD, hyperlipidemia, hypertension, OSA, osteoarthritis, and T2DM (Table 2). Hyperlipidemia had the lowest cut point (27.1 kg/m^2; sensitivity=68.8%; specificity=52.1%), followed by CAD (27.7 kg/m^2; sensitivity=66.5%; specificity=50.5%), hypertension (28.4 kg/m^2; sensitivity=62.3%; specificity=60.7%), osteoarthritis (28.7 kg/m^2; sensitivity=58.7%; specificity=51.7%), OSA (30.1 kg/m^2; sensitivity=72%; specificity=66.6%), and T2DM (30.9 kg/m^2; sensitivity=63.3%; specificity=70.9%).

The 1-year incidence rates above the cut point were higher than the rates below the cut point for the six comorbidities that had identified cut points (Figure 2). The greatest differences were for OSA (0.7 cases per 100 person-years below vs 3.4 cases per 100 person-years above the cut point) and T2DM (0.6 cases per 100 person-years below vs 2.5 cases per 100 person-years above the cut point).

When comparing baseline demographics for the comorbidities with an identifiable cut point (CAD, hyperlipidemia, hypertension, OSA, osteoarthritis, and T2DM), we found that patients who developed each disease were older and more likely to be male than those who did not develop each disease for all six comorbidities (Multimedia Appendices 5-10 contain tables comparing baseline characteristics of patients who developed each comorbidity vs those who did not for all six comorbidities with a cut point). Patients who developed each comorbidity had a higher prevalence of each of the other five comorbidities with an identifiable cut point. For example, patients who developed hypertension had a higher prevalence of CAD, hyperlipidemia, OSA, osteoarthritis, and T2DM.

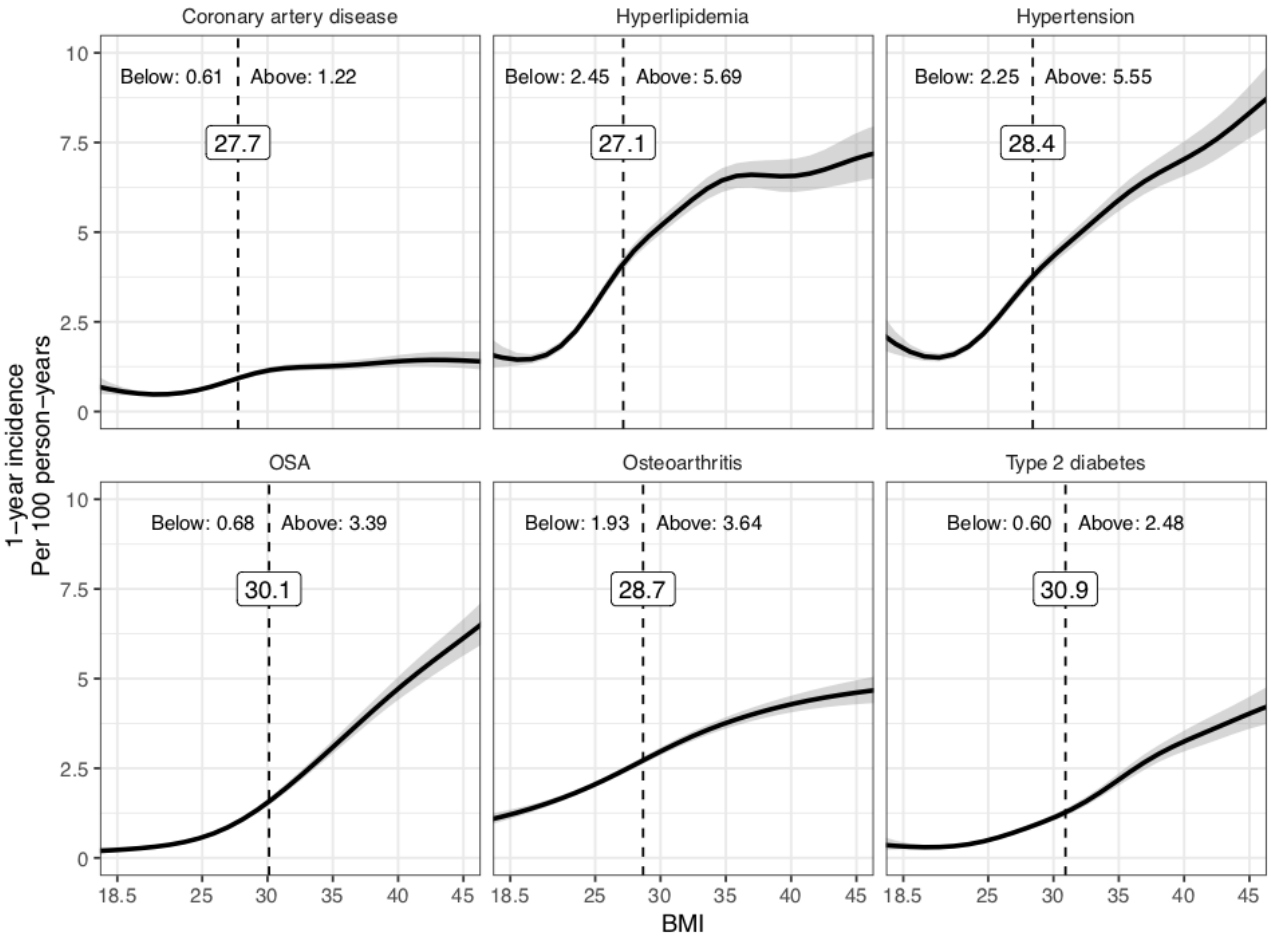
Table 2. Cut points for comorbidities.

Comorbidity	AUROC[a]	Youden index	Sensitivity, %	Specificity, %	Cut point (kg/m^2)
Anxiety	0.477	N/A[b]	N/A	N/A	N/A
Coronary artery disease	0.603	0.170	66.5	50.5	27.7
Cerebrovascular disease	0.561	N/A	N/A	N/A	N/A
Chronic pain	0.559	N/A	N/A	N/A	N/A
Depression	0.521	N/A	N/A	N/A	N/A
Gastroesophageal reflux	0.555	N/A	N/A	N/A	N/A
Hyperlipidemia	0.637	0.209	68.8	52.1	27.1
Hypertension	0.653	0.230	62.3	60.7	28.4
Obstructive sleep apnea	0.754	0.386	72	66.6	30.1
Osteoarthritis	0.606	0.161	58.7	51.7	28.7
Type 2 diabetes mellitus	0.725	0.341	63.3	70.9	30.9

[a]AUROC: area under the receiver operating curve.

[b]N/A: not applicable.

Figure 2. Cut points and comorbidity incidence. Gray shaded areas represent 95% CIs. The dotted line and the values in the box represent BMI cut points. "Below" corresponds to overall disease incidence (per 100 person-years) for all patients with a BMI that is less than the cut point. "Above" corresponds to overall disease incidence (per 100 person-years) for all patients with a BMI that is greater than the cut point. OSA: obstructive sleep apnea.



BMI Cut Points by Sex

Both male and female patients had cut points for hyperlipidemia, hypertension, OSA, and T2DM, but only female patients had cut points for CAD and osteoarthritis (Table 3; Multimedia Appendix 11 contains the full table of AUROC, Youden index, sensitivity, and specificity values for cut points by sex and race or ethnicity). Female patients had a statistically significant lower cut point for T2DM (29.9 vs 32.1 kg/m^2; *P*=.02). There were no differences in other cut points between the male and female patients.

Table 3. Cut points by sex.

Comorbidity	Male cut point (kg/m^2)	Female cut point (kg/m^2)	*P* value[a]
Coronary artery disease	N/A[b]	27.8	N/A
Hyperlipidemia	28.3	28.6	.78
Hypertension	28.8	28.5	.84
Obstructive sleep apnea	31.3	30.2	.74
Osteoarthritis	N/A	29.2	N/A
Type 2 diabetes mellitus	32.1	29.9	.02

[a]*P* value indicates the comparison of cut points between male and female patients.

[b]N/A: not applicable.

BMI Cut Points by Race or Ethnicity

When evaluating cut points by race or ethnicity, Black patients had higher cut points for hypertension (30.3 vs 28.7 kg/m^2 for White patients; P<.001) and OSA (35.1 vs 30.1 kg/m^2; P=.005; Table 4; Multimedia Appendix 11 contains the full table of AUROC, Youden index, sensitivity, and specificity values for cut points by sex and race or ethnicity). Asian patients had lower cut points for hyperlipidemia (24.1 vs 26.5 kg/m^2 for White patients; P=.02), OSA (29.0 vs 30.1 kg/m^2; P=.02), and T2DM (27.5 vs 31.3 kg/m^2; P=.04). Native American patients had lower cut points for hypertension (26.0 vs 28.7 kg/m^2 for White patients) and T2DM (29.3 vs 31.3 kg/m^2) and a higher cut point for hyperlipidemia (28.8 vs 26.5 kg/m^2), but these differences were not statistically significant. For Hispanic patients, we only identified a cut point for OSA (31.3 kg/m^2; sensitivity=69.2%; specificity=70.4%).

Table 4. Cut points by race or ethnicity.

Comorbidity	White, non-Hispanic	Black, non-Hispanic		Asian, non-Hispanic		Native American, non-Hispanic		Hispanic	
	Cut point (kg/m^2)	Cut point (kg/m^2)	P value[a]	Cut point (kg/m^2)	P value	Cut point (kg/m^2)	P value	Cut point (kg/m^2)	P value
Coronary artery disease	27.4	N/A[b]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Hyperlipidemia	26.5	N/A	N/A	24.1	.02	28.8	.41	N/A	N/A
Hypertension	28.7	30.3	.001	25.0	.23	26.0	.45	N/A	N/A
Obstructive sleep apnea	30.1	35.1	.005	29.0	.02	N/A	N/A	31.3	.08
Osteoarthritis	28.7	31.0	.55	N/A	N/A	N/A	N/A	N/A	N/A
Type 2 diabetes mellitus	31.3	31.3	.91	27.5	.04	29.3	.55	N/A	N/A

[a]P value indicates the comparison to cut points for White, non-Hispanic patients.

[b]N/A: not applicable.

Incidence Versus Prevalence Cut Point Analysis

For the six comorbidities that had BMI cut points, we found no statistically significant differences in cut points between the incident and prevalent cases for CAD, hypertension, OSA, and osteoarthritis (Multimedia Appendix 12 contains the full table of incidence vs prevalence cut points). There were statistically significant differences between incidence and prevalence cut points for hyperlipidemia (27.5 vs 27.0 kg/m^2; P=.02) and T2DM (30.7 vs 30.0 kg/m^2; P<.001).

Discussion

Principal Findings

Our findings suggest that the BMI cut points or thresholds beyond which disease incidence can be accurately detected for developing six obesity-related comorbidities occur when patients are overweight or barely meet the criteria for class 1 obesity. The cut points for developing CAD, hyperlipidemia, hypertension, and osteoarthritis were in the overweight category, while the cut points for OSA and T2DM occurred at the transition between overweight and class 1 obesity. In our study cohort, female patients had lower cut points for T2DM. Asian patients had lower cut points for hyperlipidemia, OSA, and T2DM, while Black patients had higher cut points for hypertension and OSA.

Most cut points identified in our study were within the overweight BMI category. Published studies are currently mixed with regard to the association between being overweight and the development of *obesity-related* comorbidities. The meta-analysis by Guh et al [9] found that the relative risks for comorbidities, such as T2DM and CAD, increased when patients were overweight but increased most when patients were obese. Other studies, such as the cross-sectional study by Nguyen et al [16], which used National Health and Nutrition Examination Survey (NHANES) data, demonstrated that higher BMIs were associated with an increased risk of these diseases. In contrast, a retrospective cohort study of Swiss adults by Faeh et al [10] showed increased mortality rates in patients with obesity because of CAD but not in patients who were overweight. Despite numerous studies identifying associations between these chronic diseases and obesity, no studies have identified these cut points in multiracial or ethnic populations.

We found that female patients had a lower cut point for T2DM than male patients. The literature is inconclusive regarding the association between sex and the development of obesity-related comorbidities. The retrospective study by Chu et al [6] found lower cut points for both hypertension and T2DM in Taiwanese women than men. A large cohort study evaluating the incidence of hypertension in Japanese adults with obesity showed that the relationship between BMI and hypertension was influenced by sex, with female patients experiencing a greater risk of developing hypertension [20]. In contrast, a retrospective study by Ong et al [21] of US adults using data from NHANES showed no difference in the risk of hypertension between men and women. Although our results showed no differences in hyperlipidemia cut points between male and female patients, a retrospective cohort study by Tseng et al [19] demonstrated a lower cut point for hyperlipidemia in Taiwanese women than men.

Our study found that compared with White patients, Black patients had higher cut points for hypertension and OSA. The cross-sectional study by Fontaine et al [22] using NHANES data found that Black patients experienced obesity-related morbidity, such as reduction in lifespan, at higher BMIs than White patients. In a review, Wagner and Heyward [23] hypothesized that differences in the development of obesity-related comorbidities between Black patients and those of different racial or ethnic backgrounds stemmed from variations in body composition; Black patients typically have higher BMIs than White patients despite having similar levels of body fat.

We also found that Asian patients had lower cut points. This is supported by the Expert Committee of the World Health Organization, which concluded that Asian populations have different associations between BMI and obesity-related diseases and that the cut points of obesity-related comorbidities in Asians varied between 22.0-25.0 kg/m^2 [4,7]. The population-based cross-sectional study by Cheong et al [24] of Malaysian adults identified BMI cut points for predicting the presence of diabetes, hypertension, and hyperlipidemia to be between 23.3-24.1 kg/m^2 in men and 24.0-25.4 kg/m^2 in women. A prospective study by Chan et al [25] of Chinese adults diagnosed with CAD identified a BMI cut point of 27.3 kg/m^2 for the development of OSA. The lower cut points in Asian patients have been attributed to a multitude of genetic and metabolic differences between Asian and White patients, such as different associations between BMI and body fat percentage in Asian versus White populations [4,7]. In addition, there may be differences among the various Asian subgroups. A secondary analysis by Jih et al [7] of the California Health Interview Survey found the highest rates of overweight or obesity and diabetes in Filipino populations, suggesting that genetic, lifestyle, and dietary factors may account for variations in cut points and disease risk.

Study Implications

Our results suggest that although some current screening guidelines incorporating BMI have appropriate cut points, others may need to be revised. For example, the United States Preventative Services Task Force (USPSTF) recommends screening for T2DM [26] and hypertension [27] at a BMI cut point of 25 kg/m^2. Our BMI cut points of 30.9 kg/m^2 and 28.4 kg/m^2 for T2DM and hypertension, respectively, support these guideline cut points.

In contrast, guidelines for OSA screening vary. The American Academy of Sleep Medicine recommends OSA screening for adults with a BMI ≥30 kg/m^2 [28]. The American Federal Aviation Administration and the US Federal Motor Carrier Safety Administration suggest that pilots with BMI ≥40 kg/m^2 and drivers with BMI ≥35 kg/m^2, respectively, should be screened for OSA [29,30]. The American Academy of Sleep Medicine BMI cut point of 30 kg/m^2 and our cut point of 30.1 kg/m^2 suggest that the Federal Aviation Administration and US Federal Motor Carrier Safety Administration screening cut points for OSA may be too high.

BMI is not included in the current screening recommendations for hyperlipidemia, CAD, or osteoarthritis. Although the USPSTF and American College of Cardiology/American Heart Association have guidelines for hyperlipidemia screening and statin use for some patients who meet age and cardiovascular disease risk criteria, BMI is not one of those criteria [31,32]. This study identified a cut point of 27.1 kg/m^2 for hyperlipidemia risk, indicating that the inclusion of BMI as a risk factor may be warranted. The USPSTF does not recommend screening for CAD but suggests that clinicians offer or refer adults with a BMI ≥30 kg/m^2 for behavioral weight loss therapy to prevent CAD development [33]. We are not aware of any USPSTF or professional society screening recommendations for osteoarthritis. Screening questionnaires for osteoarthritis exist [34] and could be provided to patients who exceed the BMI cut point of 28.7 kg/m^2. We also identified sex and race or ethnicity differences that may need to be considered when screening adults for obesity-related comorbidities.

Our previous EHR publication found that our patient population was demographically similar to the US adult population [35]; thus, our findings may be generalizable to US adults. However, further investigation of the BMI cut points identified in this study using multi-institutional EHR data sets would further elucidate whether our findings are generalizable. If the BMI cut points are similar within multi-institutional EHR data sets, screening recommendations for some comorbidities may need to be re-evaluated to help guide health care providers on when to screen patients for obesity-related comorbidities.

Limitations

First, although our methodology using the Youden index is established in the literature [6,19], there is no *gold standard* method for determining optimal cut points for continuous data, such as BMI. Some investigators have used disease prevalence rather than incidence to establish cut points. Our analysis comparing cut point calculations using incidence versus prevalence identified no clinically significant differences. We believe that cut points determined with incidence have more clinical utility because incidence evaluates the development of disease, whereas prevalence describes a disease that has already been diagnosed. Second, most Youden indices, sensitivities, and specificities were low, which suggests that BMI is not a perfect screening tool for these diseases. However, it has significant clinical use because it is recorded for most patients in the EHR, whereas other markers, such as waist circumference and biomarkers, are not. In addition, the AUROCs were >0.6, indicating that our analyses were able to discriminate between those with and without the disease. Third, there may be selection bias, given that all patients were required to have data in our EHR spanning at least 2 years. For example, our EHR had a lower percentage of Medicaid patients than the national estimates. Fourth, our study was observational, so no inferences can be made about causation. Finally, there may be inaccuracies in our data set because of errors in data entry by health care providers. We removed biologically implausible values using our BMI algorithm, but coding inaccuracies in the ICD-9 and ICD-10 entries may still exist.

Conclusions

The BMI cut points that accurately predict the risks of developing six obesity-related comorbidities (CAD, hyperlipidemia, hypertension, OSA, osteoarthritis, and T2DM) occurred when patients were overweight or barely met the criteria for class 1 obesity. Weight loss counseling for these patients is important because they are at an increased risk of morbidity and mortality related to obesity. Further studies using longitudinal, national data are needed to determine whether screening guidelines for CAD, hyperlipidemia, OSA, and osteoarthritis should be reconsidered.

Acknowledgments

Effort on this study and manuscript was made possible by an American College of Surgeons George HA Clowes Career Development Award, and a Veterans Affair (VA) Career Development Award to LF (CDA 015-060). The views represented in this study represent those of the authors and not those of the Department of Veteran Affairs or the US government. The project described was also supported by the Clinical and Translational Science Award program through the National Institutes of Health (NIH) National Center for Advancing Translational Sciences (grant UL1TR002373). Further funding was obtained through the NIH T32 Surgical Oncology Research Training Program (grant T32 CA090217-17). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. GC was partially funded through the Patient‐Centered Outcomes Research Institute Awards (ME-2018C2-13180). The views in this publication are solely the responsibility of the authors and do not necessarily represent the views of the Patient‐Centered Outcomes Research Institute, its Board of Governors, or Methodology Committee. The authors would like to acknowledge the internal fund from the Department of Family Medicine and Community Health for supporting GC's effort on the collaboration between the Department of Family Medicine and Community Health and the Department of Biostatistics and Medical Informatics. This study was presented as an oral presentation at the American College of Surgeons Clinical Congress 2019 on October 29, 2019, in San Francisco, California, and as an oral presentation at the 2019 Wisconsin Surgical Society meeting on November 8, 2019, in Kohler, Wisconsin.

Authors' Contributions

LF, NL, JB, MV, LH, and GC contributed to the study design. NL, JB, MV, and GC contributed to data collection and analysis. NL, JB, and LF contributed to manuscript composition. All coauthors participated in data interpretation and manuscript revision. All coauthors approved the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work were appropriately investigated and resolved.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Strengthening the Reporting of Observational Studies in Epidemiology statement checklist.
[DOCX File , 32 KB-Multimedia Appendix 1]

Multimedia Appendix 2

International Classification of Disease-9 and International Classification of Disease-10 codes used to identify comorbidities.
[DOCX File , 14 KB-Multimedia Appendix 2]

Multimedia Appendix 3

The 1-year incidence rates of comorbidities.
[DOCX File , 14 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Quantile regression analysis of the associations between the incidence of comorbidities and median BMIs.
[DOCX File , 14 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Comparison of baseline characteristics between patients who developed coronary artery disease and those who did not.
[DOCX File , 17 KB-Multimedia Appendix 5]

Multimedia Appendix 6

Comparison of baseline characteristics between patients who developed hyperlipidemia and those who did not.

[DOCX File , 16 KB-Multimedia Appendix 6]

Multimedia Appendix 7

Comparison of baseline characteristics between patients who developed hypertension and those who did not.
[DOCX File , 16 KB-Multimedia Appendix 7]

Multimedia Appendix 8

Comparison of baseline characteristics between patients who developed obstructive sleep apnea and those who did not.
[DOCX File , 16 KB-Multimedia Appendix 8]

Multimedia Appendix 9

Comparison of baseline characteristics between patients who developed osteoarthritis and those who did not.
[DOCX File , 16 KB-Multimedia Appendix 9]

Multimedia Appendix 10

Comparison of baseline characteristics between patients who developed type 2 diabetes mellitus and those who did not.
[DOCX File , 16 KB-Multimedia Appendix 10]

Multimedia Appendix 11

Area under the receiving operating curve, Youden index, and sensitivity or specificity of sex and race or ethnicity-specific cut points.
[DOCX File , 21 KB-Multimedia Appendix 11]

Multimedia Appendix 12

Incidence versus prevalence cut points.
[DOCX File , 13 KB-Multimedia Appendix 12]

References

1. Ward ZJ, Bleich SN, Cradock AL, Barrett JL, Giles CM, Flax C, et al. Projected U.S. state-level prevalence of adult obesity and severe obesity. N Engl J Med 2019 Dec 19;381(25):2440-2450. [doi: 10.1056/nejmsa1909301]
2. Mascie-Taylor CGN, Goto R. Human variation and body mass index: a review of the universality of BMI cut-offs, gender and urban-rural differences, and secular changes. J Physiol Anthropol 2007 Mar;26(2):109-112 [FREE Full text] [doi: 10.2114/jpa2.26.109] [Medline: 17435352]
3. Obesity : Preventing and Managing the Global Epidemic : Report of a WHO Consultation on Obesity. Geneva: World Health Organization; 1998:1-178.
4. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004 Jan;363(9403):157-163. [doi: 10.1016/s0140-6736(03)15268-3]
5. Arroyo-Johnson C, Mincey K. Obesity epidemiology worldwide. Gastroenterol Clin North Am 2016 Dec;45(4):571-579 [FREE Full text] [doi: 10.1016/j.gtc.2016.07.012] [Medline: 27837773]
6. Chu F, Hsu C, Jeng C. Lowered cutoff points of obesity indicators are better predictors of hypertension and diabetes mellitus in premenopausal Taiwanese women. Obes Res Clin Pract 2015 Jul;9(4):328-335. [doi: 10.1016/j.orcp.2014.08.001] [Medline: 25154649]
7. Jih J, Mukherjea A, Vittinghoff E, Nguyen TT, Tsoh JY, Fukuoka Y, et al. Using appropriate body mass index cut points for overweight and obesity among Asian Americans. Prev Med 2014 Aug;65:1-6 [FREE Full text] [doi: 10.1016/j.ypmed.2014.04.010] [Medline: 24736092]
8. Pantalone KM, Hobbs TM, Chagin KM, Kong SX, Wells BJ, Kattan MW, et al. Prevalence and recognition of obesity and its associated comorbidities: cross-sectional analysis of electronic health record data from a large US integrated health system. BMJ Open 2017 Nov 16;7(11):e017583 [FREE Full text] [doi: 10.1136/bmjopen-2017-017583] [Medline: 29150468]
9. Guh DP, Zhang W, Bansback N, Amarsi Z, Birmingham CL, Anis AH. The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. BMC Public Health 2009 Mar 25;9(1):88 [FREE Full text] [doi: 10.1186/1471-2458-9-88] [Medline: 19320986]
10. Faeh D, Braun J, Tarnutzer S, Bopp M. Obesity but not overweight is associated with increased mortality risk. Eur J Epidemiol 2011 Aug 17;26(8):647-655. [doi: 10.1007/s10654-011-9593-2] [Medline: 21681546]
11. Klompas M, Cocoros NM, Menchaca JT, Erani D, Hafer E, Herrick B, et al. State and local chronic disease surveillance using electronic health record systems. Am J Public Health 2017 Sep;107(9):1406-1412. [doi: 10.2105/ajph.2017.303874]

12. Fildes A, Charlton J, Rudisill C, Littlejohns P, Prevost AT, Gulliford MC. Probability of an obese person attaining normal body weight: cohort study using electronic health records. Am J Public Health 2015 Sep;105(9):54-59. [doi: 10.2105/ajph.2015.302773]
13. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008] [Medline: 18313558]
14. Cheng FW, Gao X, Mitchell DC, Wood C, Still CD, Rolston D, et al. Body mass index and all-cause mortality among older adults. Obesity (Silver Spring) 2016 Oct 29;24(10):2232-2239 [FREE Full text] [doi: 10.1002/oby.21612] [Medline: 27570944]
15. Liu N, Birstler J, Venkatesh M, Hanrahan L, Chen G, Funk L. Weight loss for patients with obesity: an analysis of long-term electronic health record data. Med Care 2020 Mar;58(3):265-272 [FREE Full text] [doi: 10.1097/MLR.0000000000001277] [Medline: 31876663]
16. Nguyen NT, Magno CP, Lane KT, Hinojosa MW, Lane JS. Association of hypertension, diabetes, dyslipidemia, and metabolic syndrome with obesity: findings from the National Health and Nutrition Examination Survey, 1999 to 2004. J Am Coll Surg 2008 Dec;207(6):928-934. [doi: 10.1016/j.jamcollsurg.2008.08.022] [Medline: 19183541]
17. Zhao G, Ford ES, Dhingra S, Li C, Strine TW, Mokdad AH. Depression and anxiety among US adults: associations with body mass index. Int J Obes (Lond) 2009 Feb 6;33(2):257-266. [doi: 10.1038/ijo.2008.268] [Medline: 19125163]
18. Koenker R, Hallock KF. Quantile regression. J Econ Perspect 2001 Nov 01;15(4):143-156. [doi: 10.1257/jep.15.4.143]
19. Tseng C, Chong C, Chan T, Bai C, You S, Chiou H, et al. Optimal anthropometric factor cutoffs for hyperglycemia, hypertension and dyslipidemia for the Taiwanese population. Atherosclerosis 2010 Jun;210(2):585-589. [doi: 10.1016/j.atherosclerosis.2009.12.015] [Medline: 20053403]
20. Fujita M, Hata A. Sex and age differences in the effect of obesity on incidence of hypertension in the Japanese population: a large historical cohort study. J Am Soc Hypertens 2014 Jan;8(1):64-70. [doi: 10.1016/j.jash.2013.08.001] [Medline: 24113321]
21. Ong KL, Tso AW, Lam KS, Cheung BM. Gender difference in blood pressure control and cardiovascular risk factors in Americans with diagnosed hypertension. Hypertension 2008 Apr;51(4):1142-1148. [doi: 10.1161/hypertensionaha.107.105205]
22. Fontaine KR, Redden DT, Wang C, Westfall AO, Allison DB. Years of life lost due to obesity. J Am Med Assoc 2003 Jan 08;289(2):187-193. [doi: 10.1001/jama.289.2.187] [Medline: 12517229]
23. Wagner D, Heyward V. Measures of body composition in blacks and whites: a comparative review. Am J Clin Nutr 2000 Jun;71(6):1392-1402. [doi: 10.1093/ajcn/71.6.1392] [Medline: 10837277]
24. Cheong KC, Yusoff AF, Ghazali SM, Lim KH, Selvarajah S, Haniff J, et al. Optimal BMI cut-off values for predicting diabetes, hypertension and hypercholesterolaemia in a multi-ethnic population. Public Health Nutr 2012 Jun 01;16(3):453-459. [doi: 10.1017/s1368980012002911]
25. Chan P, Tai B, Loo G, Koo C, Ong T, Yeo T, et al. Optimal body mass index cut-offs for identification of patients with coronary artery disease at high risk of obstructive sleep apnoea. Heart Lung Circ 2016 Aug;25(8):847-854. [doi: 10.1016/j.hlc.2016.02.010] [Medline: 27067667]
26. Siu AL. Screening for abnormal blood glucose and type 2 diabetes mellitus: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2015 Oct 27;163(11):861. [doi: 10.7326/m15-2345]
27. Siu AL. Screening for high blood pressure in adults: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2015 Oct 13;163(10):778. [doi: 10.7326/m15-2223]
28. Aurora RN, Quan SF. Quality measure for screening for adult obstructive sleep apnea by primary care physicians. J Clin Sleep Med 2016 Aug 15;12(8):1185-1187 [FREE Full text] [doi: 10.5664/jcsm.6064] [Medline: 27397659]
29. Squire P. Obstructive sleep apnea. In: Goldschneider K, Davidson A, Wittkugel E, Skinner A, editors. Clinical Pediatric Anesthesia: A Case-Based Handbook. Oxford, United Kingdom: Oxford University Press; 2013.
30. Final report: Obstructive sleep apnea (Task 11-05). Federal Motor Carrier Safety Administration. 2005. URL: https://www.fmcsa.dot.gov/advisory-committees/mcsac/final-report-obstructive-sleep-apnea-task-11-05 [accessed 2019-08-01]
31. US Preventive Services Task Force. Screening adults for lipid disorders: recommendations and rationale. Am J Prev Med 2001 Apr;20(3 Suppl):73-76. [Medline: 11306235]
32. Sniderman AD. Did the ACC/AHA/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA cholesterol guidelines get apoB right? J Clin Lipidol 2019 May;13(3):360-366. [doi: 10.1016/j.jacl.2019.05.010] [Medline: 31229023]
33. US Preventive Services Task Force, Curry SJ, Krist AH, Owens DK, Barry MJ, Caughey AB, et al. Behavioral weight loss interventions to prevent obesity-related morbidity and mortality in adults: US Preventive Services Task Force recommendation statement. J Am Med Assoc 2018 Sep 18;320(11):1163-1171. [doi: 10.1001/jama.2018.13022] [Medline: 30326502]
34. Roux CH, Saraux A, Mazieres B, Pouchot J, Morvan J, Fautrel B, KHOALA Osteoarthritis Group. Screening for hip and knee osteoarthritis in the general population: predictive value of a questionnaire and prevalence estimates. Ann Rheum Dis 2008 Oct 01;67(10):1406-1411. [doi: 10.1136/ard.2007.075952] [Medline: 18077540]

35. Funk L, Shan Y, Voils C, Kloke J, Hanrahan L. Electronic health record data versus the National Health and Nutrition Examination Survey (NHANES): a comparison of overweight and obesity rates. Med Care 2017 Jun;55(6):598-605 [FREE Full text] [doi: 10.1097/MLR.0000000000000693] [Medline: 28079710]

Abbreviations

AUROC: area under the receiver operating curve
CAD: coronary artery disease
EHR: electronic health record
ICD: International Classification of Disease
NHANES: National Health and Nutrition Examination Survey
NIH: National Institutes of Health
OSA: obstructive sleep apnea
T2DM: type 2 diabetes mellitus
USPSTF: United States Preventative Services Task Force

Edited by R Kukafka, G Eysenbach; submitted 31.08.20; peer-reviewed by A Jo, J Lau; comments to author 25.09.20; revised version received 01.11.20; accepted 21.06.21; published 09.08.21

Please cite as:
Liu N, Birstler J, Venkatesh M, Hanrahan L, Chen G, Funk L
Obesity and BMI Cut Points for Associated Comorbidities: Electronic Health Record Study
J Med Internet Res 2021;23(8):e24017
URL: https://www.jmir.org/2021/8/e24017
doi: 10.2196/24017
PMID:

Original Paper

The Effect of Collaborative Reviews of Electronic Patient-Reported Outcomes on the Congruence of Patient- and Clinician-Reported Toxicity in Cancer Patients Receiving Systemic Therapy: Prospective, Multicenter, Observational Clinical Trial

Andreas Trojan[1*], MD; Nicolas Leuthold[2*], MMed; Christoph Thomssen[3], MD; Achim Rody[4], MD; Thomas Winder[5,6], MD, PhD; Andreas Jakob[7], MD; Claudine Egger[8], MD; Ulrike Held[9], PhD; Christian Jackisch[10], MD

[1]OnkoZentrum Zürich, Zurich, Switzerland
[2]Clinic for Clinical Pharmacology and Toxicology, University Hospital Zurich, Zurich, Switzerland
[3]Universitätsklinikum Halle (Saale), Halle, Germany
[4]Klinik für Frauenheilkunde und Geburtshilfe, Universitätsklinikum Schleswig-Holstein, Lübeck, Germany
[5]Department of Internal Medicine II, Academic Teaching Hospital Feldkirch, Feldkirch, Austria
[6]University of Zurich, Zurich, Switzerland
[7]Tumor Zentrum Aarau, Hirslanden Medical Center, Aarau, Switzerland
[8]Spital Limmattal, Schlieren, Switzerland
[9]Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland
[10]Sana Klinikum Offenbach, Offenbach, Germany
[*]these authors contributed equally

Corresponding Author:
Andreas Trojan, MD
OnkoZentrum Zürich
Seestrasse 259
Zurich, 8038
Switzerland
Phone: 41 43 344 33 33
Email: trojan@1st.ch

Abstract

Background: Electronic patient-reported outcomes (ePRO) are a relatively novel form of data and have the potential to improve clinical practice for cancer patients. In this prospective, multicenter, observational clinical trial, efforts were made to demonstrate the reliability of patient-reported symptoms.

Objective: The primary objective of this study was to assess the level of agreement κ between symptom ratings by physicians and patients via a shared review process in order to determine the future reliability and utility of self-reported electronic symptom monitoring.

Methods: Patients receiving systemic therapy in a (neo-)adjuvant or noncurative intention setting captured ePRO for 52 symptoms over an observational period of 90 days. At 3-week intervals, randomly selected symptoms were reviewed between the patient and physician for congruency on severity of the grading of adverse events according to the Common Terminology Criteria of Adverse Events (CTCAE). The patient-physician agreement for the symptom review was assessed via Cohen kappa (κ), through which the interrater reliability was calculated. Chi-square tests were used to determine whether the patient-reported outcome was different among symptoms, types of cancer, demographics, and physicians' experience.

Results: Among the 181 patients (158 women and 23 men; median age 54.4 years), there was a fair scoring agreement (κ=0.24; 95% CI 0.16-0.33) for symptoms that were entered 2 to 4 weeks before the intended review (first rating) and a moderate agreement (κ=0.41; 95% CI 0.34-0.48) for symptoms that were entered within 1 week of the intended review (second rating). However, the level of agreement increased from moderate (first rating, κ=0.43) to substantial (second rating, κ=0.68) for common symptoms of pain, fever, diarrhea, obstipation, nausea, vomiting, and stomatitis. Similar congruency levels of ratings were found for the most frequently entered symptoms (first rating: κ=0.42; second rating: κ=0.65). The symptom with the lowest agreement was

hair loss (κ=–0.05). With regard to the latency of symptom entry into the review, hardly any difference was demonstrated between symptoms that were entered from days 1 to 3 and from days 4 to 7 before the intended review (κ=0.40 vs κ=0.39, respectively). In contrast, for symptoms that were entered 15 to 21 days before the intended review, no congruency was demonstrated (κ=–0.15). Congruency levels seemed to be unrelated to the type of cancer, demographics, and physicians' review experience.

Conclusions: The shared monitoring and review of symptoms between patients and clinicians has the potential to improve the understanding of patient self-reporting. Our data indicate that the integration of ePRO into oncological clinical research and continuous clinical practice provides reliable information for self-empowerment and the timely intervention of symptoms.

Trial Registration: ClinicalTrials.gov NCT03578731; https://clinicaltrials.gov/ct2/show/NCT03578731

(J Med Internet Res 2021;23(8):e29271) doi: 10.2196/29271

KEYWORDS

cancer; consilium; app; eHealth; ePRO; CTCAE; congruency; patient-reported; symptoms

Introduction

Patient-reported outcomes (PRO), such as symptoms and functional status, are commonly measured in clinical trials. There is growing interest in integrating electronic PRO (ePRO) into routine clinical practice during chemotherapeutic and immunotherapeutic interventions. Most cancer patients are motivated to spend time and effort documenting symptoms during their consultation for shared reporting with physicians. Patients' self-empowerment and self-reporting should also improve patient-clinician communication, symptom detection, and symptom control [1]. As patient experience has gained importance in regulatory decision-making, patient-reported data are increasingly being used for quality assessment and comparative effectiveness research. Mobile health solutions have the potential to improve electronic symptom documentation, and when the collection of such PRO is widely used, it facilitates communication among stakeholders [1,2]. Several apps have been designed and tested with input from patients, nurses, and physicians. These apps have gained attention and quality with respect to improving the efficacy and safety data in oncology trials and drug discovery [3-5]. Their benefits in real-world digital patient monitoring during cancer immunotherapy have been demonstrated in terms of more accurate symptom assessment, better patient-physician communication, and reduced need for telephone consultations [6-8]. As oncologists intend to share information on symptom grading with their patients, as defined by the Common Terminology Criteria for Adverse Events (CTCAE) standards, reliable information on PRO should not only improve symptom management but also allow for the reduction of emergency admissions and improve patients' quality of life. However, early responsiveness to symptoms and presumably longer continuation of chemotherapy, as well as a potential benefit of follow-up integration of ePRO for symptom monitoring during routine cancer care, frequently involve patient-physician or patient–nurse specialist communication [9,10]. In addition, compliance rates and the use of symptom alerts seem to be enhanced by structured graphic displays on outcome reporting [3,4]. Several digital platforms are currently implementing the capture of ePRO to allow for the sharing of data with treatment teams or to apply automatic algorithms for alert notifications in a timely and structured manner if symptoms worsen [2,11,12]. The consilium care smartphone app continuously allows oncologists to monitor the progress of patients' symptoms through visualized progression charts based on structured patient entries. In the case of severe symptoms that exceed a determined threshold, the app notifies the patient to contact the treatment center. Previous published breast cancer studies showed the potential of the app to stabilize daily functional activity and well-being of patients in collaboration with the physician [1]. In addition, more distinct symptom entries were received from those users who shared reporting with their physicians. The functionality and utility of 2 comparable app versions for collecting ePRO have also demonstrated that the request for a collaborative review of ePRO for shared reporting increases the number of data entries and potentially affects the ability to deal with the symptoms of illness [13]. Since clinical oncology strives for a standardized recording of adverse events, the congruence between doctor and patient should serve as an important indicator that patients' self-reporting can enhance the quality of outcome data for the accuracy of clinician ratings and safety. This has the potential to reduce the problem of patients reporting high symptom severity while their clinicians note low toxicity grades. Further, it has the potential to identify challenges in effective patient-clinician communication regarding symptom experience, to stimulate the processes of recording and reviewing patient-reported symptoms, to facilitate consultation with oncologists, and to provide self-care algorithms for real-time interventions that reduce symptom severity [13].

In this study, we evaluated the efforts being made using the consilium care app in a cohort of patients with breast, colon, lung, or prostate cancer, as well as those with hematological malignancies, to demonstrate the reliability of electronically captured patient-reported symptom entries for shared reporting with the physician to detect critical symptoms in routine cancer care. For this study, we intended to demonstrate that a collaborative review of randomly selected patient-reported symptoms improves congruency of patient- and clinician-reported toxicity in patients receiving systemic anticancer therapy. In particular, we examined whether important and frequent symptoms, such as pain, fever, diarrhea, obstipation, nausea, vomiting, and stomatitis, can be described appropriately according to the CTCAE in order to potentially implement recommendations for mitigation.

Methods

Study Design

We conducted a multicenter, observational, noninterventional study. The protocol was approved by the competent regulatory ethics committee (KEK-ZH:2017-02028) and registered on ClinicalTrials.gov (NCT03578731). Patients with breast, colon, prostate, or lung cancer, as well as those with hematological malignancies, aged 18 years and older, and initiating adjuvant or neoadjuvant systemic therapy were eligible to participate after providing written informed consent. In addition, participants had to speak German and own a smartphone. Eligible participants were recruited consecutively and without preselection according to the recommendation of the local tumor boards in centers in Switzerland, Germany, and Austria.

Objective

The primary objective of this study was to assess the level of agreement, κ, between symptom ratings by physicians at the time of the regular consultation and the ratings derived from the daily PRO between consultations. The level of agreement was analyzed in order to determine the reliability and utility of self-reported electronic symptom monitoring.

Mobile App

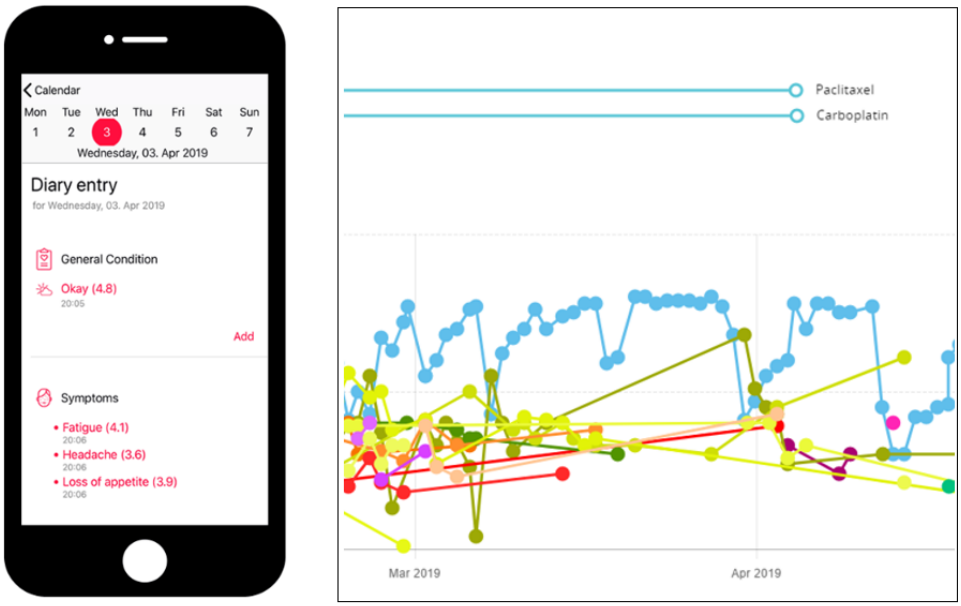
To begin, patients downloaded the consilium care app (available for iOS and Android) and connected themselves via a quick response code to their study centers. For the patients' convenience, a summary of diagnostic workup, treatment medication, and contact information of the respective treatment center was entered into the consilium care web app—the treatment team's counterpart to the smartphone app.

The app (Figure 1) facilitated the selection of well-being, symptoms, medication, and private notes. Symptoms, which were structured in groups according to organ systems, could be selected. The symptom entry display (52 distinct symptoms were available for which severity, onset, and duration could be indicated) was equipped with date and time stamps. Symptom severity, with descriptions based on the CTCAE, could be selected via a slider. The symptom history was displayed on a timeline with individual colors for each symptom. In addition, diary entries and information on diagnosis and therapy were indicated separately.

Patients were encouraged to capture data on well-being and symptoms on a daily basis. Recording usually started on the day of the therapy's initiation or the change in therapy and continued through an observational period of 12 weeks. The app allowed the continuous recording of well-being and symptoms based on the CTCAE through use of virtual analogue scales. Definitions for CTCAE grades were displayed above the slider, with which the grade of the entry could be selected via the virtual analogue scales. The severity level of a symptom, as rated by physicians and patients, was measured on an ordinal scale, with 0 indicating the lowest possible degree of severity and 4 indicating the highest possible one. The history of recorded data was displayed and visualized in the form of a symptom progression chart. In the case of severe symptoms, patients were encouraged by push notifications to seek medical advice. In addition, patients recorded their well-being according to the Eastern Cooperative Oncology Group (ECOG) performance status via a slider, with possible impairments in daily functional activities being displayed. Information for self-care (derived from the Swiss Cancer League and the Sächsische Krebsgesellschaft) was provided to them via the app depending on the severity of symptoms upon data entry.

Functional data security was ensured by identification being made only possible through the patient's ID. The data on the patient's device were encapsulated in the app, and data exchange was encrypted with the patient ID. At the study center, personal data were kept strictly separate from the data collected by the app. Data matching was performed by using the patient ID.

Figure 1. Entrance screen and a representative symptom history chart with indication of medication, well-being (blue graph), and various symptoms presented in different colors.



Collaborative Symptom Reviews

Patients were assigned to medical oncology visits every 3 weeks and invited for shared reporting and intended symptom review, which were preferably scheduled on days of therapeutic intervention. Some exceptions were made for reviews to be carried out over the phone. At the scheduled visit, the app was triggered to randomly select 2 patient-reported symptoms from the past 20 days. A first measurement of congruence (symptom 1) was performed on a symptom that was entered 2 to 3 weeks (14 to 21 days) before the actual consultation, whereas a second measurement (symptom 2) was performed on a symptom that was entered within the previous week (1 to 7 days). Patients and physicians were then prompted to perform a detailed, shared review of these symptoms in order to focus on the collection and appropriate interpretation for symptom severity grading. Up to 4 such reviews were planned per patient, including 2 electronic symptom entries per review.

Questionnaire

At the end of the observational period, participants were asked to complete a questionnaire on paper regarding the usability and usefulness of the app to clarify quality of care and the relationship between the patient and physician during the course of treatment. To this end, a 5-point Likert scale was used, with a rating from 1 (disagree) to 5 (agree very strongly).

Sample Size

We calculated the sample size on a 5% significance level to test the level of agreement, κ=0.5, between 2 raters (ie, fair to good agreement) with a precision of 0.1 on each side of 156 patients. In order to estimate κ with the necessary precision within these subgroups, we included at least 170 patients with breast cancer and 170 patients with colon cancer. We anticipated a difficultly in recruiting the same number of patients with lung cancer or prostate cancer due to their lower prevalence. Thus, the aim was to include 130 patients with either lung cancer (not fewer than 50) or prostate cancer and 130 patients with hematological malignancies. We planned to enroll a total of 600 patients, as we expected 15% to 20% of enrolled patients to discontinue participation (dropout) early.

The originally planned study population size for the entire study cohort was 600. The study duration was estimated to be about 3 years, starting in March 2018. In autumn 2020, only about one-third of the planned study patients were recruited, and the sponsor decided to prematurely terminate the study on October 11, 2020, due to insufficient recruitment. Despite the continuous opening of many study sites beginning 2018, due to the present recruitment rate and the ongoing COVID-19 situation, the planned number of 600 patients was unachievable.

Statistical Analysis

Descriptive statistics included mean and SD for continuous variables, and numbers and percentages of total for categorical variables. For statistical analysis, the associations between physicians' and patients' ratings were visualized by plots. Multiple ratings for patients were included and accounted for by the analysis. For the quantification of levels of agreement, Cohen kappa (κ) values were calculated with squared weights. κ values are reported with 95% CIs. These CIs were based on 1000 bootstrap samples. According to Landis and Koch [14], values for κ were characterized as follows: <0, no agreement; 0 to 0.20, slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; and 0.81 to 1, almost perfect agreement. All analyses were carried out with R version 4.0.2 (The R Foundation for Statistical Computing) [15], and Excel R Markdown was used for dynamic reporting.

Results

Baseline Characteristics

Between February 2018 and October 2020, 223 patients (190 female and 33 male) with cancer (170 breast, 19 lung, 15 colon, 7 prostate, and 12 hematological [B cell] malignancies) were included using the consilium care app. Among them, 181 patients (158 women and 23 men; age at therapy start: mean 54.4 years, SD 12.1) had performed at least 1 validated review with the treating physician. About half of the 181 patients who used the consilium care app were treated in an adjuvant setting (vs neoadjuvant). Fewer than one-third (51/181, 28.2%) of the patients received treatment for advanced disease with noncurative intention. In total, 27 distinct chemotherapeutic agents in 17 different chemotherapy regimens were administered, including antihormones, CDK4/6 inhibitors, and immunotherapies.

Due to the lack of appropriate accrual within the context of the COVID-19 pandemic, premature closing of the study, and other issues, 42 patients included could not perform a minimum of 1 intended review. In addition to this, 7 patients were not evaluable due to the premature study termination, 10 patients did not enter a sufficient number of symptoms, and another 14 patients were not evaluable due to technical issues. Only 3 patients withdrew their informed consent. Baseline characteristics are displayed in Table 1, and an overview flow chart of the patient enrollment is available in Multimedia Appendix 1.

Table 1. Baseline characteristics.

Characteristic	Value (N=181)
Primary tumor, n (%)	
Hematological	9 (5.0)
Breast	142 (78.5)
Colon	11 (6.1)
Lung	13 (7.2)
Prostate	6 (3.3)
Sex, n (%)	
Female	157 (86.7)
Male	23 (12.7)
N/A[a]	1 (0.6)
Age at start, mean (SD)	54.4 (12.1)

[a]N/A: not applicable.

Agreement Levels

A total of 181 patients underwent at least 1 intended symptom review for this analysis. From a subset of 110 patients (60.8%), more than 2 collaborative symptom reviews of patients with their physicians were available for analysis. For the analysis of the first symptom agreement levels (across all multiple ratings per patient), there were 497 (first rating) reviews available for analysis, while for the second symptom agreement levels, 483 reviews (second rating) were available.

An estimation of general agreement levels between physicians' and patients' observations in the first symptom (defined as recorded 14 to 21 days before the review) revealed a fair congruency of κ=0.24 (95% CI 0.16-0.33), while for the second most recent symptom (defined as being recorded 1 to 7 days before), the value rose to κ=0.41 (95% CI 0.34-0.48; Figure 2).

Analysis of the levels of agreement in subgroups of the specific symptoms, including pain, fever, diarrhea, obstipation, nausea, vomiting, and stomatitis, revealed a higher congruency between the patient and physician estimate (symptom 1: κ=0.43, 95% CI 0.21-0.62; symptom 2: κ =0.68, 95% CI 0.54-0.77; Figure 3). Whether this observation was due to a different perception of clinical relevance and frequency of these symptoms or to a clearer description, as 5 of the 7 symptoms were associated with objectifiable values in their definition (eg, fewer than 4 loose stools per day) at some point, remains unclear.

Next, we evaluated the levels of agreement in the subgroup of physicians with at least 10 ratings. The distribution of rating frequencies revealed large differences; of the 29 participating in this study, 9 physicians performed 10 or more ratings. These were considered experienced raters and were included in the subsequent assessments. For the analysis, there were 417 observations for symptom 1 (first rating) and 405 observations for symptom 2 (second rating). Again, multiple ratings per patient were included. As shown in Figure 4, a fair congruency between patient and physician estimates was present for those considered experienced (≥10 ratings; symptom 1: κ =0.25, 95% CI 0.17-0.34; symptom 2: κ=0.41, 95% CI 0.33-0.49). Compared to all physicians' (experienced and less experienced) ratings for symptom 1 (κ=0.24) and symptom 2 (κ=0.41), the agreement levels hardly differed, indicating that congruency was more likely affected by timing and symptom description than the physicians' particular skills.

Similar results of congruency as those seen in the specific symptoms displayed in Figure 2 were obtained for the most frequent symptom as rated by experienced physicians (>10 ratings; symptom 1: κ=0.42, 95% CI 0.18-0.62; symptom 2: κ =0.65, 95% CI 0.5-0.75; Figure 5). The most frequently captured symptoms were fatigue, hot flashes, sleep disorder, headache, and taste disorder.

The levels of agreement with respect to time intervals between the date of collaborative review and the date of symptom entry within the previous week did not reveal a significant difference (days 1-3: κ=0.40; days 4-7: κ=0.39; overall days 1-7: κ=0.41). For the rating of symptoms entered 15 to 21 days prior to the review, a significant lack of congruency was noted (κ=–0.15). This finding indicated that patients recalled symptoms and their severity much better if they occurred more recently. For future studies, a collaborative review of a symptom from the recent past may be considered sufficient to demonstrate the accuracy of the electronic symptom recording in general, particularly for distinct and frequently occurring symptoms. Although this observation might require confirmation in a subsequent study, the idea of recent-past symptom validation (less than 7 days) might be applicable in real-world cancer care, clinical trials, or pay-for-performance models [16]. Furthermore, we noted a moderate increase of congruence between ratings from week 3 (first rating) to week 9 (third rating) in our approach (symptom 1: κ=0.23 vs κ=0.29; symptom 2: κ=0.36 vs κ=0.41), indicating a potential training effect in patients and physicians. The quality of ratings neither appeared differently with regard to light or moderate symptoms (CTCAE grade ≤ 2) nor in comparison to severe symptoms, defined as CTCAE grade >2 (κ=0.13 vs κ=0.11), which is important in cases of early-intervention clinical practice. Congruency of symptom reporting according to the review of the second symptom was similar for breast (396

reviews; κ=0.39), lung (30 reviews; κ=0.45), and colon cancer (23 reviews; κ=0.51), as well as hematological malignancies (20 reviews; κ=0.49). For prostate cancer, there was an almost perfect congruency (12 reviews; κ=0.82) although the low number of reviews had to be considered with regard to statistical significance. The subgroup analysis for age and gender showed overall congruency levels of κ=0.50 for older (>65 years; 99 reviews; κ=0.50;) and younger patients (<65 years; 380 reviews; κ=0.38), as well as for female (435 reviews; κ=0.40) and male (44 reviews; κ=0.49 for) patients.

Figure 2. Estimations of agreement levels between physicians' and patients' observations for the first and second symptom. diarr: diarrhea; fev: fever; obstip: obstipation; stomat: stomatatis; vomit: vomiting.

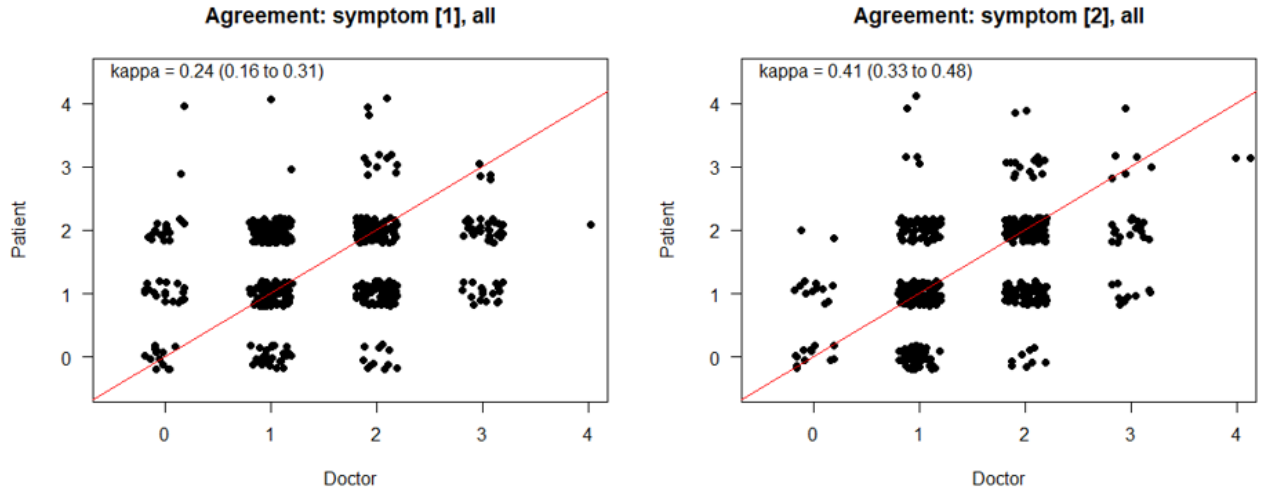


Figure 3. Estimations of agreement levels between physicians' and patients' observations for specific symptoms.

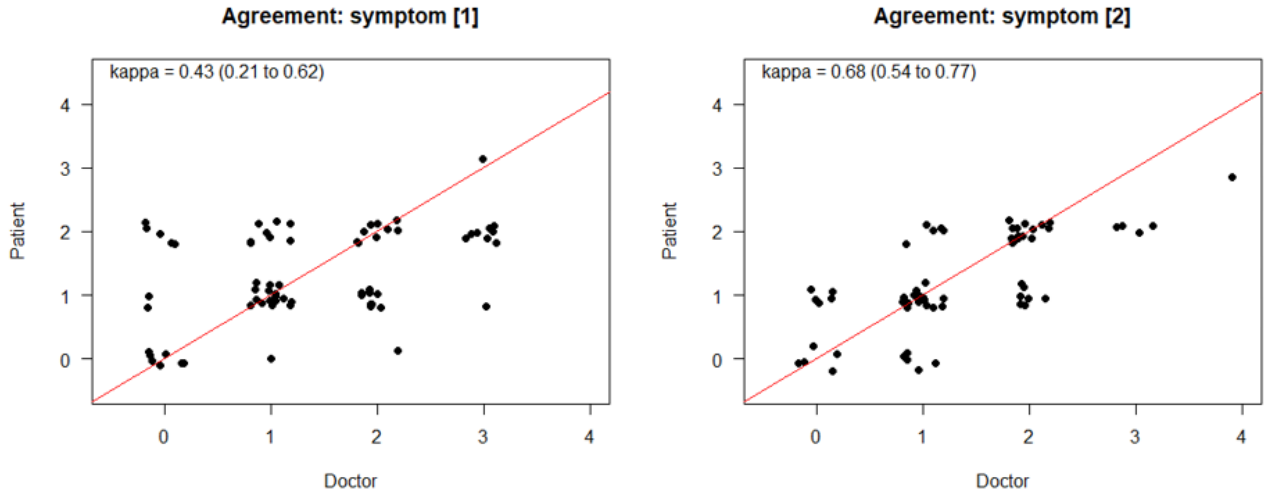


Figure 4. Estimations of agreement levels between physicians' and patients' observations for experienced physicians.

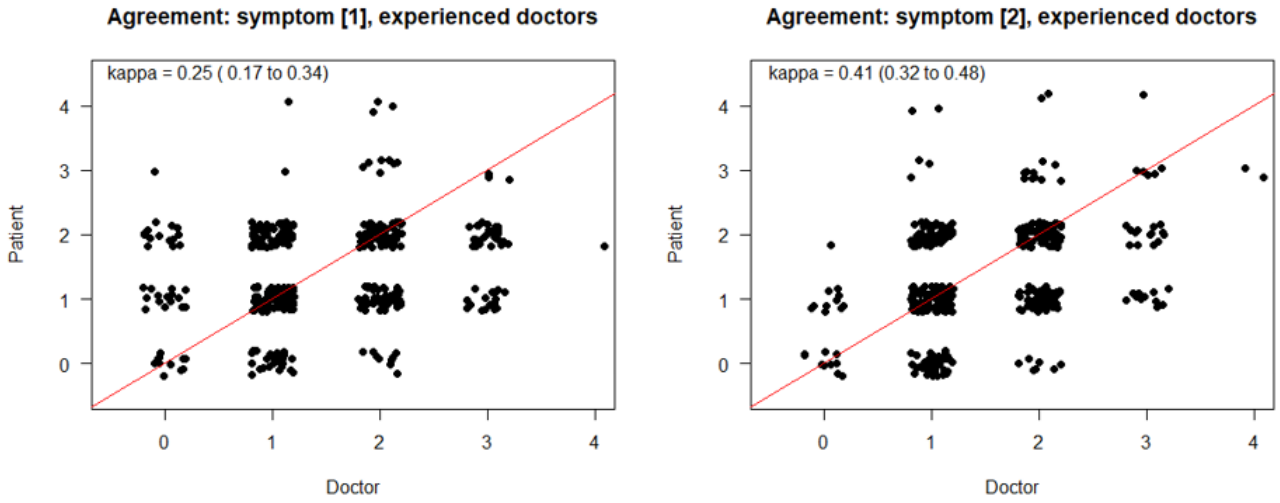
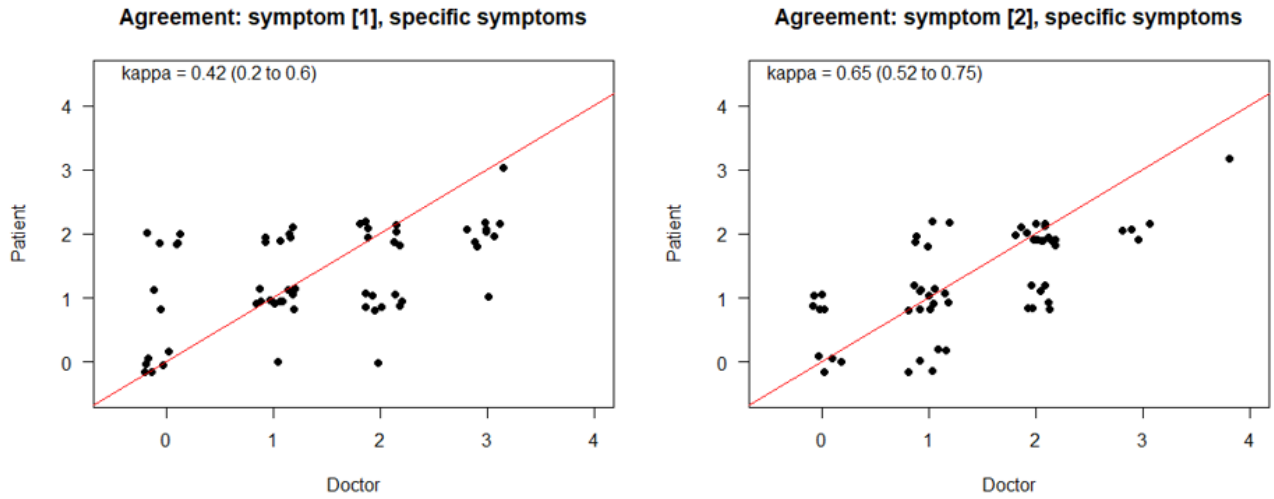


Figure 5. Estimations of agreement levels of experienced physicians specifically for the most frequent symptoms.

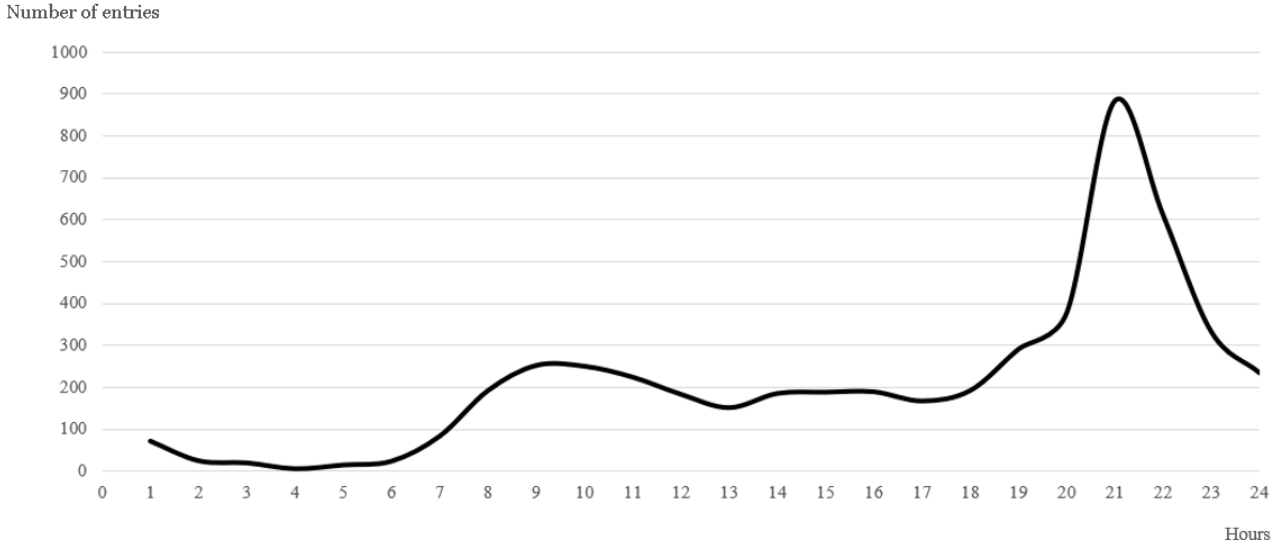


Well-being and Symptoms

Regarding well-being, 4762 data entries were derived from 210 evaluable patients during the observation period. Patients reported their well-being almost every single day and in a classical circadian rhythm (Figure 6). Because well-being was reported independently of the underlying diagnosis or symptoms, we assumed that this indicated a pattern of app use. Users preferred to use the app in the morning and also used it during the evening hours. Therefore, a circadian pattern of symptom reporting seemed to be favored. The degree to which the app's functions (eg, occasional push notifications, design features, tips for self-care, or effects of collaborative review and shared reporting) affected data entries remains unclear, as this evaluation was not addressed.

Overall, 210 patients generated a large absolute number of 42,142 electronically reported symptoms and side effects, suggesting easy handling of the app for an effective symptom history insight. Given the observational period of 84 days, this resulted in an average number of 2 to 3 entries per patient and day. The most commonly reported symptom was fatigue, which was indicated significantly more often in the breast cancer and lymphoma groups (data not shown) compared to other cancer entities. Due to the heterogeneity of drugs and limited information on dosage, a potential association of symptoms with the respective cancer type, medication, or regimen, could not be performed sufficiently. However, more than 32.59% (13,734/42,142) of all data entries affected usual activities of daily living and symptoms such as pain/discomfort (8370/42,142, 19.86%), self-care (3475/42,142, 8.24%), anxiety/depression (1458/42,142, 3.45%), and mobility (431/42,142, 1.02%), all of which potentially represent components of the 5-level EQ-5D questionnaire.

Figure 6. Circadian distribution of well-being entries (y-axis) over 24 hours (x-axis).



Unplanned Consultations and Serious Adverse Events

Although fewer than 18.2% (33/181) of the participants with solid cancer (breast, colon, lung, prostate) required unplanned consultations or emergency services due to treatment-related side effects and toxicities, more than twice this proportion (4/9, 44%) was recorded in patients with lymphoma, mostly attributed due to fatigue and fever. An association with a possible benefit from app use cannot be made, as data from a matched analysis (age, cancer type, therapy) of patients from 2 larger participating cancer centers indicated only a nonsignificant decline in these events (data not shown). Importantly, no serious adverse events related to the use of the app were recorded during the entire study period.

Usability and Usefulness of the App

Questionnaires from 171 patients included were available for the rating of the app at the time of this survey. Six patients died due to cancer progression during the study, from whom surveys were not available for analysis. A utility analysis could not be conducted on 16 patients, as they were not correctly included into the study, withdrew informed consent, had technical problems, or lacked a sufficient number of data entries. The results are displayed in Multimedia Appendix 2.

Discussion

The systematic electronic recording of PRO by smartphone has not yet been extensively explored in cancer treatment. Previous studies indicate that the range of measures used and symptoms captured seem to vary greatly across studies, and that, regardless of the concordance metric employed, the reported agreement between clinician-based CTCAE and PRO seems to be moderate, at best [17]. In one study that retrospectively applied CTCAE patient language adaptations, including the Symptom Tracking and Reporting system, to assess specific symptoms, extracted clinician- and patient-reported adverse event ratings were considered poor to moderate, at best, when the applied rating sources for each of the adverse events were compared [18]. In an attempt to improve these differences, we explored integrating ePRO and clinician reporting with a standardized, shared review process, according to CTCAE criteria with adapted patient-oriented language by testing the level of agreement between the patients' and physicians' judgment on the severity of patient symptoms with 3 weekly reviews of randomly selected symptoms at any severity grade.

Overall, we found fair agreement for long-lasting symptoms, whereas for the more recent symptoms (defined as those recorded 1-7 days earlier), the degree of agreement in symptom reporting between the patient and physician was moderate and comparable to results from a study in early breast cancer [19]. However, the congruence between patients and physicians gained substantial reliability when analyses on levels of agreement in subgroups of the specific symptoms (ie, pain, fever, diarrhea, obstipation, nausea, vomiting, and stomatitis) were performed and also in an identical manner to that in the most frequently occurring symptoms, including fatigue, hot flashes, sleep disorder, headache, and taste disorder. Together, data entries from these symptoms covered about 50% of all recorded symptom-related entries during this study. As patients obviously recalled recent symptoms more clearly, the high trustworthiness of symptom rating could be sufficiently proven by 1 review in this context. Congruency of rating seemed to be independent of the reviewers' experience, and no outlier result in congruency of symptom reporting could be demonstrated for any specific patient cohort, indicating the potentially broad acceptance and use of such an approach. Additionally, no differences in symptom congruency were noted with respect to light or more severe symptom grading.

Compliance for the use of the consilium care app was high as evidenced by the high number of 2 to 3 data entries per patient and by the response from questionnaires, and was found to be comparable with results from other studies that used more standardized questionnaires for different devices [20]. In a recent study, patients were invited to complete the European Organization for Research and Treatment of Cancer (EORTC) Core Quality of Life questionnaire (QLQ-C30) and cancer site–specific modules before each visit on tablets or computers in the hospital or at home. An adequate compliance (at least 66% of health-related quality of life assessments were

completed) was demonstrated for the cohort of breast cancer (96%), colorectal cancer (98%), and lung cancer (91%) [21], which we consider comparable to the results of our study.

In one study that administered weekly PRO from the National Cancer Institute's PRO-CTCAE item library (for symptoms such as pain, nausea, and diarrhea) via mail or telephone and assessed them by using a 5-point ordinal verbal descriptor scale and via PRO questions about physical performance (ECOG) and financial toxicities (The Comprehensive Score for Financial Toxicity [COST-FACIT] questionnaire), it was found that most patients agreed that weekly reporting was a favored frequency for ePRO questionnaire administration in the context of advanced and metastatic cancer treatment [22]. However, during a more complex or intensive treatment phase, a more frequent (even daily) assessment of more than 8 symptoms might be well regarded and positively associated with an increased use of educational materials about home symptom management. In another trial, almost 40% of patients (particularly older patients and those living in rural areas) chose to use an automated telephone interface rather than a web interface or preferred personal contact in the case of severe symptoms affecting cognitive or sexual dysfunction [23]. Web-based, guided, self-help interventions can provide clinically meaningful improvements in quality of life; however, producing a meaningful effect might require punctual psychological interventions [24]. Although no such findings were apparent in our study, following advice and using tips for timely self-care and compliance remains challenging for patients and caregivers. The consilium app contains 20 tips for the most common symptoms. In personal communication with patients, it was suggested that this opportunity of self-help intervention should be linked to the appropriate symptom or grade, as patients perceived this to be a component of personalized medicine [25,26].

There were potential limitations to this study. The frequency of the completed symptom reviews varied between the 3 German-speaking countries conducting the trial, most patients were suffering from breast cancer, and the study was not randomized, which precluded analysis in regard to the effects of empowered self-care and the potential impact on unplanned consultations. Statistical limitations evolved from the data set when there were multiple observations per patient; thus, observations could not be considered independent. Furthermore, there were limitations to the interpretation of Cohen κ values. In this study, we used magnitude guidelines proposed by Landis and Koch [14] to describe levels of interrater reliability; however, other guidelines exist, such as those of Fleiss [27]. Because of the ongoing debate about the correct description of κ values, the interpretation we employed can still be subject to scrutiny. Importantly, due to the lack of appropriate accrual in the context of the COVID-19 pandemic, the trial was ended prematurely.

In summary, we demonstrated that a shared monitoring and review process to assess symptoms between patients and physicians has the potential to improve the quality of future patient self-reporting. Our study indicated that the integration of ePRO into oncological clinical research and continuous clinical practice should leverage monitoring of side effects and symptom management [28,29] using the rapidly developing digital mobile and sensor technologies, which can provide more objective measures and facilitate the active and passive collection of detailed, personalized data.

Acknowledgments

The authors would like to thank all patients who participated in this study. Furthermore, we would like to thank the Swiss Tumor Institute, Zürich, and the Hirslanden Forschungsstiftung, Zürich, for their financial support.

Authors' Contributions

AT, UH, and CJ were responsible for the integrity of the entire study. AT, TW, and UH were responsible for the study concept and design. AT, TW, and NL conducted the literature research. AT, CT, AR, TW, AJ, CE, and NL conducted the trial. UH, NL, and AT performed the data analysis. UH and NL performed statistical analyses. NL, AT, UH, and CJ prepared the manuscript. NL, AT, UH, and CJ edited the manuscript. NL and AT contributed equally as the main authors. All authors approved the final manuscript.

Conflicts of Interest

AT is the founder and chief medical officer of Mobile Health AG, a startup company that operates the consilium care smartphone app. He also owns stock in the company. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1

Overview flow chart of the patient enrollment.
[PNG File , 37 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Results of the questionnaire about the usability and usefulness of the app.
[PNG File , 57 KB-Multimedia Appendix 2]

References

1. Egbring M, Far E, Roos M, Dietrich M, Brauchbar M, Kullak-Ublick GA, et al. A mobile app to stabilize daily functional activity of breast cancer patients in collaboration with the physician: a randomized controlled clinical trial. J Med Internet Res 2016 Sep 06;18(9):e238 [FREE Full text] [doi: 10.2196/jmir.6414] [Medline: 27601354]
2. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
3. Basch E. High compliance rates with patient-reported outcomes in oncology trials submitted to the US Food and Drug Administration. J Natl Cancer Inst 2019 May 01;111(5):437-439. [doi: 10.1093/jnci/djy183] [Medline: 30561701]
4. Snyder C, Smith K, Holzner B, Rivera YM, Bantug E, Brundage M, PRO Data Presentation Delphi Panel. Making a picture worth a thousand numbers: recommendations for graphically displaying patient-reported outcomes data. Qual Life Res 2019 Feb;28(2):345-356 [FREE Full text] [doi: 10.1007/s11136-018-2020-3] [Medline: 30306533]
5. Basch E, Dueck AC. Patient-reported outcome measurement in drug discovery: a tool to improve accuracy and completeness of efficacy and safety data. Expert Opin Drug Discov 2016 Aug;11(8):753-758. [doi: 10.1080/17460441.2016.1193148] [Medline: 27310432]
6. Iivanainen S, Alanko T, Peltola K, Konkola T, Ekström J, Virtanen H, et al. ePROs in the follow-up of cancer patients treated with immune checkpoint inhibitors: a retrospective study. J Cancer Res Clin Oncol 2019 Mar;145(3):765-774 [FREE Full text] [doi: 10.1007/s00432-018-02835-6] [Medline: 30666409]
7. Trojan A, Huber U, Brauchbar M, Petrausch U. Consilium smartphone app for real-world electronically captured patient-reported outcome monitoring in cancer patients undergoing anti-PD-L1-directed treatment. Case Rep Oncol 2020;13(2):491-496 [FREE Full text] [doi: 10.1159/000507345] [Medline: 32518544]
8. Schmalz O, Jacob C, Ammann J, Liss B, Iivanainen S, Kammermann M, et al. Digital monitoring and management of patients with advanced or metastatic non-small cell lung cancer treated with cancer immunotherapy and its impact on quality of clinical care: interview and survey study among health care professionals and patients. J Med Internet Res 2020 Dec 21;22(12):e18655 [FREE Full text] [doi: 10.2196/18655] [Medline: 33346738]
9. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. JAMA 2017 Jul 11;318(2):197-198 [FREE Full text] [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
10. Lizée T, Basch E, Trémolières P, Voog E, Domont J, Peyraga G, et al. Cost-effectiveness of web-based patient-reported outcome surveillance in patients with lung cancer. J Thorac Oncol 2019 Jun;14(6):1012-1020. [doi: 10.1016/j.jtho.2019.02.005] [Medline: 30776447]
11. Licqurish SM, Cook OY, Pattuwage LP, Saunders C, Jefford M, Koczwara B, et al. Tools to facilitate communication during physician-patient consultations in cancer care: an overview of systematic reviews. CA Cancer J Clin 2019 Nov;69(6):497-520 [FREE Full text] [doi: 10.3322/caac.21573] [Medline: 31339560]
12. Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan A, King MT, the SPIRIT-PRO Group, et al. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the spirit-pro extension. JAMA 2018 Feb 06;319(5):483-494. [doi: 10.1001/jama.2017.21903] [Medline: 29411037]
13. Trojan A, Bättig B, Mannhart M, Seifert B, Brauchbar MN, Egbring M. Effect of collaborative review of electronic patient-reported outcomes for shared reporting in breast cancer patients: descriptive comparative study. JMIR Cancer 2021 Mar 17;7(1):e26950 [FREE Full text] [doi: 10.2196/26950] [Medline: 33729162]
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
15. The R Project for Statistical Computing. R Foundation for Statistical Computing. URL: https://www.R-project.org/ [accessed 2021-03-29]
16. Picecchi D, Bertram K, Brücher D, Bauer M. Towards novel reimbursement models for expensive advanced therapy medicinal products (ATMPs). Swiss Med Wkly 2020 Dec 14;150:w20355 [FREE Full text] [doi: 10.4414/smw.2020.20355] [Medline: 33378542]
17. Atkinson TM, Ryan SJ, Bennett AV, Stover AM, Saracino RM, Rogak LJ, et al. The association between clinician-based common terminology criteria for adverse events (CTCAE) and patient-reported outcomes (PRO): a systematic review. Support Care Cancer 2016 Aug;24(8):3669-3676 [FREE Full text] [doi: 10.1007/s00520-016-3297-9] [Medline: 27260018]
18. Atkinson TM, Rogak LJ, Heon N, Ryan SJ, Shaw M, Stark LP, et al. Exploring differences in adverse symptom event grading thresholds between clinicians and patients in the clinical trial setting. J Cancer Res Clin Oncol 2017 Apr;143(4):735-743 [FREE Full text] [doi: 10.1007/s00432-016-2335-9] [Medline: 28093637]
19. Nyrop KA, Deal AM, Reeve BB, Basch E, Chen YT, Park JH, et al. Congruence of patient- and clinician-reported toxicity in women receiving chemotherapy for early breast cancer. Cancer 2020 Jul 01;126(13):3084-3093 [FREE Full text] [doi: 10.1002/cncr.32898] [Medline: 32315091]
20. Howell D, Rosberger Z, Mayer C, Faria R, Hamel M, Snider A, iPEHOC Collaborative Team. Personalized symptom management: a quality improvement collaborative for implementation of patient reported outcomes (PROs) in 'real-world' oncology multisite practices. J Patient Rep Outcomes 2020 Jun 17;4(1):47 [FREE Full text] [doi: 10.1186/s41687-020-00212-x] [Medline: 32556794]

21. Mouillet G, Falcoz A, Fritzsch J, Almotlak H, Jacoulet P, Pivot X, et al. Feasibility of health-related quality of life (HRQoL) assessment for cancer patients using electronic patient-reported outcome (ePRO) in daily clinical practice. Qual Life Res 2021 Jan 02. [doi: 10.1007/s11136-020-02721-0] [Medline: 33387291]
22. Basch E, Stover AM, Schrag D, Chung A, Jansen J, Henson S, et al. Clinical utility and user perceptions of a digital system for electronic patient-reported symptom monitoring during routine cancer care: findings from the PRO-TECT trial. JCO Clin Cancer Inform 2020 Oct;4:947-957 [FREE Full text] [doi: 10.1200/CCI.20.00081] [Medline: 33112661]
23. Gressel GM, Dioun SM, Richley M, Lounsbury DW, Rapkin BD, Isani S, et al. Utilizing the Patient Reported Outcomes Measurement Information System (PROMIS®) to increase referral to ancillary support services for severely symptomatic patients with gynecologic cancer. Gynecol Oncol 2019 Mar;152(3):509-513. [doi: 10.1016/j.ygyno.2018.10.042] [Medline: 30876496]
24. Urech C, Grossert A, Alder J, Scherer S, Handschin B, Kasenda B, et al. Web-based stress management for newly diagnosed patients with cancer (STREAM): a randomized, wait-list controlled intervention study. J Clin Oncol 2018 Mar 10;36(8):780-788 [FREE Full text] [doi: 10.1200/JCO.2017.74.8491] [Medline: 29369731]
25. Pircher M, Winder T, Trojan A. Response to vemurafenib in metastatic triple-negative breast cancer harbouring a BRAF V600E mutation: a case report and electronically captured patient-reported outcome. Case Rep Oncol 2021;14(1):616-621 [FREE Full text] [doi: 10.1159/000513905] [Medline: 33976643]
26. Richardson D, Zhan L, Mahtani R, McRoy L, Mitra D, Reynolds M, et al. A prospective observational study of patient-reported functioning and quality of life in advanced and metastatic breast cancer utilizing a novel mobile application. Breast Cancer Res Treat 2021 May;187(1):113-124 [FREE Full text] [doi: 10.1007/s10549-020-06082-7] [Medline: 33428072]
27. Fleiss J. Statistical Methods for Rates and Proportions (Wiley Series in Probability and Statistics). New York, USA: Wiley-Interscience; 1981.
28. Thong MSY, Chan RJ, van den Hurk C, Fessele K, Tan W, Poprawski D, MASCC EMA working group. Going beyond (electronic) patient-reported outcomes: harnessing the benefits of smart technology and ecological momentary assessment in cancer survivorship research. Support Care Cancer 2021 Jan;29(1):7-10 [FREE Full text] [doi: 10.1007/s00520-020-05648-x] [Medline: 32844316]
29. Riis CL, Stie M, Bechmann T, Jensen PT, Coulter A, Möller S, et al. ePRO-based individual follow-up care for women treated for early breast cancer: impact on service use and workflows. J Cancer Surviv 2021 Jan 07. [doi: 10.1007/s11764-020-00942-3] [Medline: 33415653]

Abbreviations

COST-FACIT: The Comprehensive Score for Financial Toxicity
CTCAE: Common Terminology Criteria of Adverse Events
ECOG: Eastern Cooperative Oncology Group
ePRO: electronic patient-reported outcomes
EORTC: European Organization for Research and Treatment of Cancer
PRO: patient-reported outcomes

Edited by G Eysenbach; submitted 05.04.21; peer-reviewed by B Koczwara, C Jacob, P Innominato; comments to author 26.04.21; revised version received 13.05.21; accepted 03.06.21; published 05.08.21

Please cite as:
Trojan A, Leuthold N, Thomssen C, Rody A, Winder T, Jakob A, Egger C, Held U, Jackisch C
The Effect of Collaborative Reviews of Electronic Patient-Reported Outcomes on the Congruence of Patient- and Clinician-Reported Toxicity in Cancer Patients Receiving Systemic Therapy: Prospective, Multicenter, Observational Clinical Trial
J Med Internet Res 2021;23(8):e29271
URL: https://www.jmir.org/2021/8/e29271
doi: 10.2196/29271
PMID:

Original Paper

Evaluating Epidemiological Risk by Using Open Contact Tracing Data: Correlational Study

Stefano Piotto[1,2], PhD; Luigi Di Biasi[3], MSc; Francesco Marrafino[1], PhD; Simona Concilio[1,2], PhD

[1]Department of Pharmacy, University of Salerno, Fisciano, Italy

[2]Bionam Research Center for Biomaterials, University of Salerno, Fisciano, Italy

[3]Department of Computer Sciences, University of Salerno, Fisciano, Italy

Corresponding Author:
Stefano Piotto, PhD
Department of Pharmacy
University of Salerno
Via Giovanni Paolo II, 132
Fisciano, 84084
Italy
Phone: 39 3204230068
Email: piotto@unisa.it

Abstract

Background: During the 2020s, there has been extensive debate about the possibility of using contact tracing (CT) to contain the SARS-CoV-2 pandemic, and concerns have been raised about data security and privacy. Little has been said about the effectiveness of CT. In this paper, we present a real data analysis of a CT experiment that was conducted in Italy for 8 months and involved more than 100,000 CT app users.

Objective: We aimed to discuss the technical and health aspects of using a centralized approach. We also aimed to show the correlation between the acquired contact data and the number of SARS-CoV-2–positive cases. Finally, we aimed to analyze CT data to define population behaviors and show the potential applications of real CT data.

Methods: We collected, analyzed, and evaluated CT data on the duration, persistence, and frequency of contacts over several months of observation. A statistical test was conducted to determine whether there was a correlation between indices of behavior that were calculated from the data and the number of new SARS-CoV-2 infections in the population (new SARS-CoV-2–positive cases).

Results: We found evidence of a correlation between a weighted measure of contacts and the number of new SARS-CoV-2–positive cases (Pearson coefficient=0.86), thereby paving the road to better and more accurate data analyses and spread predictions.

Conclusions: Our data have been used to determine the most relevant epidemiological parameters and can be used to develop an agent-based system for simulating the effects of restrictions and vaccinations. Further, we demonstrated our system's ability to identify the physical locations where the probability of infection is the highest. All the data we collected are available to the scientific community for further analysis.

(J Med Internet Res 2021;23(8):e28947) doi: 10.2196/28947

KEYWORDS

SARS-CoV-2; COVID-19; contact tracing; Bluetooth Low Energy; transmission dynamics; infection spread; mobile apps; mHealth; digital apps; mobile phone

Introduction

In China, during December 2019, SARS-CoV-2 was identified as a novel beta coronavirus. At the time of writing this paper (December 2020), SARS-CoV-2 has caused almost 60 million confirmed human infections worldwide and more than 1 million deaths since its discovery [1,2]. The disease caused by SARS-CoV-2 is called COVID-19, and the disease was declared a global pandemic on March 11, 2020 [3]. Containment measures are the first and most crucial step for rapidly halting an outbreak that could otherwise become an epidemic or even turn into a pandemic, such as the COVID-19 outbreak [4].

Notable examples of disease epidemics with a high occurrence of superspreading events (SSEs) are the SARS-CoV (severe acute respiratory syndrome coronavirus; 2002-2003) and MERS-CoV (Middle East respiratory syndrome coronavirus; since 2013) epidemics [5-9]. The basic reproduction number (R_0) is a key measure of transmissibility. It is defined as the number of infected contacts that 1 infected individual generates on average during their infectious period. An R_0 value of >1 means that a virus will continue its propagation among susceptible hosts. In contrast, an R_0 of <1 means that it is certain that epidemic spread will stop [10,11]. The SARS-CoV and MERS-CoV have an R_0 of around 3 [12]. For SARS-CoV-2, the estimated R_0 ranges between 2 and 3 [9,13]. However, it is unknown as to what extent SSEs are involved in the spread of SARS-CoV-2 infection.

Lockdown was the most widespread pandemic containment response, and it was introduced at different levels by most affected countries. As already predicted by mathematical models [14] and proven by trends that were updated at the time of writing this paper, the contagion's spread resumed rapidly when lockdown countermeasures were lifted. Rapid and automatic contact tracing (CT) is an essential intervention for contagion containment [15-19]; however, user localization poses a privacy risk and reduces compliance rates [20]. According to the World Health Organization, CT involves the following three steps: the identification of a contact (identifying those that a confirmed positive patient had contact with based on the transmission modalities of the pathogen of interest), the listing of contacts (keeping a record of individuals who possibly had contact with infected patients and informing these individuals), and contact follow-up [21]. CT has a dual purpose—treating people who have possibly been exposed to infectious diseases and stopping the transmission chain to contain an epidemic. Due to the prevalence of smartphones, CT has the potential to become a powerful intervention; the vast majority of smartphone users carry their smartphone devices with them throughout the day, and smartphones can generate detailed GPS location information. However, due to the availability of users' location data, there is growing concern about the infringement of an individual's right to privacy. An alternative is using other contact monitoring technologies that are based on proximity assessments rather than those based on location information [22]. It is important to note that this study does not constitute an endorsement or rejection of CT based on potential data security risks or privacy limitations. This study intends to assess whether and to what extent the acquisition of contact data helps with assessing the spread of SARS-CoV-2.

Technologies such as Bluetooth Low Energy allow for the evaluation of the distance between users without locating them and thus help with addressing the privacy issue. The number of CT apps that have been introduced since the beginning of the SARS-CoV-2 pandemic is considerable [23,24] and reflects governments' interest in automating the tracing of people who have had recent contact with individuals who tested positive for COVID-19. An app that uses a centralized approach was developed by the academic spin-off company of the University of Salerno—SoftMining (SM). The app [25] was supported by government agencies such as the Campania Region and was validated by more than 120,000 users; the app had peaks of more than 15,000 active daily users.

CT is a fundamental intervention for acquiring population data, which show how different population groups can behave differently. Such behaviors result in different risks of infection among group members. In Multimedia Appendix 1, we describe how CT data were acquired via the Bluetooth Low Energy technology of the SM-COVID-19 app and how data were clustered to obtain different mobility and behavior groups. In this paper, we discuss how we used Italian National Institute of Health data on contagion trends in Italy [26] to estimate a more precise number of SARS-CoV-2–positive cases that was less influenced by the number of tests performed on the population. In addition, we show the link between the acquired CT data and the number of new SARS-CoV-2–positive cases. This allowed us to define an epidemiological risk function that was based on the number of, frequency of, and distance between contacts. The risk function expresses the probability that an individual will become ill as a function of their age within a given period of time. This study aims to evaluate whether the use of CT can support the containment of an epidemic. The data acquired from CT were analyzed and correlated with data on the progression of SARS-CoV-2 infection.

This study was not conducted for commercial purposes; it was conducted for the purposes of academic research and aims to make CT data available to the scientific community for future research.

Methods

CT Data Acquisition

During the CT phase, the SM-COVID-19 app analyzed the environment and, at regular intervals, sent data on the duration of a contact and the instantaneous and average distances (over the time) of a contact to the server. App users could voluntarily decide to share location data as well. If they did, the server also received latitude, longitude, precision, and smartphone provider data. We provide the full description of the data acquisition procedures in Multimedia Appendices 1 and 2. The developed technologies allowed for high precision in distance calculations (less than 0.5 m under optimal conditions and after device calibration) and were implemented via the SM-COVID-19 app, which is available on Android and iOS smartphones (via TestFlight; Apple Inc). Daily data were anonymized and saved for further use, as described in Multimedia Appendix 3, in accordance with the General Data Protection Regulation. Anonymity was also guaranteed when the GPS localization function was enabled, as data were stored randomly in the database; the database did not present an individual user's location in a precise way. The app only used random 128-bit proximity IDs, and only the user's device kept track of the device IDs. The app's functions were conducted and maintained with a back-end server, on which arbitrary identifiers were stored. Users could not be identified directly with app data, as only the app's random identifiers were stored on the server.

Social Mobility Analysis

The data set obtained from the SM-COVID-19 app in the period of April to November 2020 was analyzed. The data set's structure is described in Multimedia Appendix 2. Reported data from August 1 to August 30, 2020, were obtained to analyze mobility data from a period when no lockdown measures were in place. Such data are useful for tracking movements in real situations. We removed users with less than 15 days of activity from our analysis to exclude users who may have deactivated the app. The cleaned data set was clustered. Before the clustering process, the t-distributed stochastic neighbor embedding machine learning algorithm was applied to the data set to reduce its dimensionality to 2. The clustering was carried out by using the Ward linkage method. This method allows the user to select the number of clusters arbitrarily. We analyzed the distribution of data for different numbers of clusters (2-10 clusters); the optimal distribution was obtained with 5 clusters. The average number of daily contacts and the SDs for the clusters are reported in Table 1. SDs were high, since every cluster had many users with 0-contact days among those with low- and high-contact days. As shown in Table 1, the population was divided into clusters of approximately the same size. However, cluster 5 was larger and included users who had a larger number of contacts. This cluster accounted for the population with the highest number of contacts and included users with the highest number of contacts and the highest mobility.

Table 1. The cluster data of active users for the period of August 1 to August 30, 2020.

Cluster number	Number of daily contacts based on Bluetooth Low Energy technology, mean (SD)[a]	Percentage of active users
1	23.40 (38.55)	14
2	12.05 (22.62)	19
3	41.95 (75.79)	17
4	69.91 (103.76)	20
5	121.48 (145.05)	30

[a]The average number of daily contacts for each cluster and SDs were calculated based on all cluster data (ie, from days 1 to 30).

Data Availability

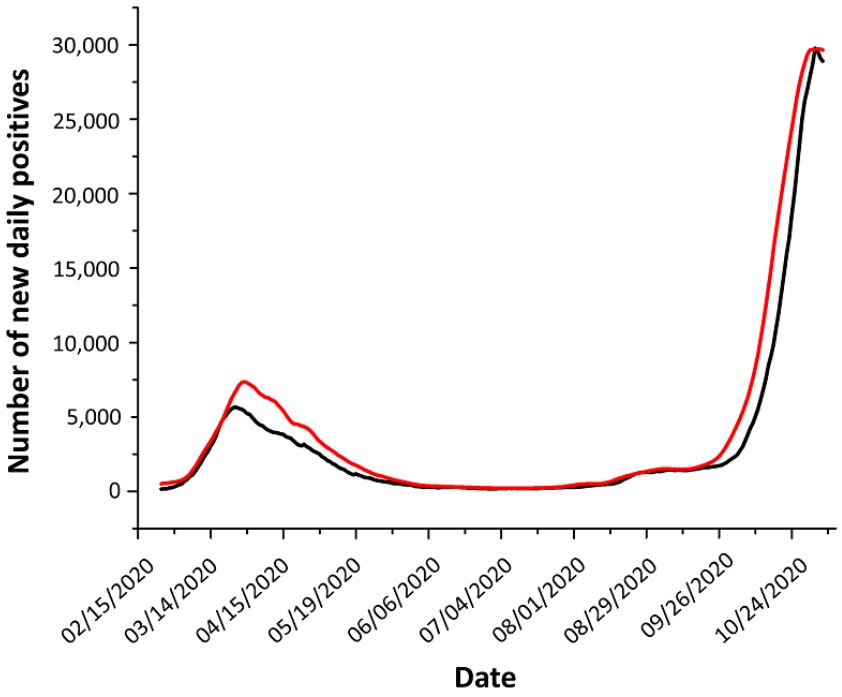
All data can be made available upon request from the authors or the SM-COVID-19 team [27].

Results

Statistical Analysis and Estimates of the Real Number of SARS-CoV-2–Positive Cases

For our statistical analysis, we relied on official data on the daily SARS-CoV-2–related trends in Italy, which were released by the Italian National Institute of Health and aggregated by the Department of Civil Protection of the Presidency of the Council of Ministers [17]. We estimated the possible number of real infections that may have occurred during the epidemic in Italy. We obtained the daily number of newly performed tests based on the total number of tests performed. This was calculated by using equation 1 in Multimedia Appendix 4. The method for estimating the number of new daily SARS-CoV-2–positive cases is detailed in Multimedia Appendix 4. We performed data smoothing via sliding-window averaging to reduce each day's variability, which was the result of the cumulative regional data's intrinsic variability. The SARS-CoV-2–related trends over a given period were roughly linear; there were no sudden peaks. Additionally, the averaging process performed allowed us to smoothen the curves, which were in line with these trends. Equation 2 in Multimedia Appendix 4 was used to define the ratio between the number of daily tests and the number of daily reported SARS-CoV-2–positive cases. The estimated number of new SARS-CoV-2–positive cases ($EP_{[k]}$) for each day was calculated with equation 3 in Multimedia Appendix 4. With our method, we estimated a correction for the number of real SARS-CoV-2–positive cases that occurred during the pandemic period. We observed that around 224,000 cases were not diagnosed, and of these cases, nearly around 81,000 were missed in the period of March to May 2020. The difference between the official number of cases and the estimated number of new SARS-CoV-2–positive cases is shown in Figure 1.

Figure 1. Comparison of the official number of daily new SARS-CoV-2–positive cases reported by the ISS (black line) and the estimated number of daily new SARS-CoV-2–positive cases (red line). The difference was higher during the initial phases of the pandemic. ISS: Istituto Superiore di Sanità (Italian Superior Institute of Health).



Correlation Between CT Data and Contagion Trends

The correct number of daily SARS-CoV-2–positive cases was calculated to perform correlation analyses with the data obtained from CT. The data distributed by the ISS, due to how the data were structured, showed considerable fluctuations based on the number of tests performed. It was also possible to observe a weekly trend in the number of SARS-CoV-2–positive cases recorded due to the reduced number of tests performed during weekends. Such data therefore presented fluctuations that could alter the analysis. Data smoothing via sliding-window averaging also provided an additional element for alleviating the issue with fluctuations.

We then examined whether the contact index (CI) and the alpha index (α) correlated with the number of daily new SARS-CoV-2–positive cases. These two parameters are indices of effective contacts and account for the distance between two users who come into contact with each other and the contact's duration. These parameters and the related equations are described in detail in Multimedia Appendix 5 [4,28,29]. These parameters were necessary, since not all of the contacts recorded by the app involved people who could effectively transmit the virus. CI_k is a value that indicates a user's risk of infection on day k based on the number of effective contacts that the user had on the same day. CI_k was calculated with equation 4 in Multimedia Appendix 5 [4,28,29]. α_k is a risk index, and it is based on data from the previous k−14 days (excluding day k). α_k reflects a user's behavior. The optimization of these parameters will be the subject of future studies.

The SM-COVID-19 data set lists the CI and α values for each day and every user. Therefore, to evaluate daily trends, we calculated the total CI and α values for each day (k) by summing each individual users' values. As such, it was possible to evaluate the trends for CI and α values and exclude users who deactivated the app for a given period. The values were smoothed by using a sliding window of 7 days. In Figure 2, we show the temporal evolution of CI values over 160 days. For visualization, in Figure 2, we report the logarithm of the number of new SARS-CoV-2–positive cases. There is an evident, rough correlation between the CI and the number of new SARS-CoV-2–positive cases. For each CI_k and α_k value, we calculated the Pearson correlation coefficient based on the estimated number of SARS-CoV-2–positive cases to assess how the number of contacts varied before and after a confirmation of COVID-19 positivity. It was very interesting to note that the correlation coefficient for CI_k reached its maximum at k+7 days. The high correlations observed in the subsequent days correlated with SARS-CoV-2 incubation times, and COVID-19 positivity occurred in the days following an effective contact. The α_k value reached its maximum at k+5 days. The differences between the α and CI values' correlation coefficients (ie, their correlation with the number of new SARS-CoV-2–positive cases) were attributable to the different calculation methods that were used for the two parameters, as the α value accounts for the risk of infection in the 14 days before day k. The correlation between CI values and the number of new SARS-CoV-2–positive cases is shown in Figure 3. We reported the correlation data that corresponded to the period of June to October 2020 because of the high availability of more consistent CT data. This correlation was also monitored for the previous studied period (March to May 2020) to confirm that the obtained values were not the result of artifacts or autocorrelations.

Figure 2. Temporal evolution of the CI values (black line) and the logarithm of the number of new SARS-CoV-2–positive cases (red line) during a 160-day period. CI: contact index.

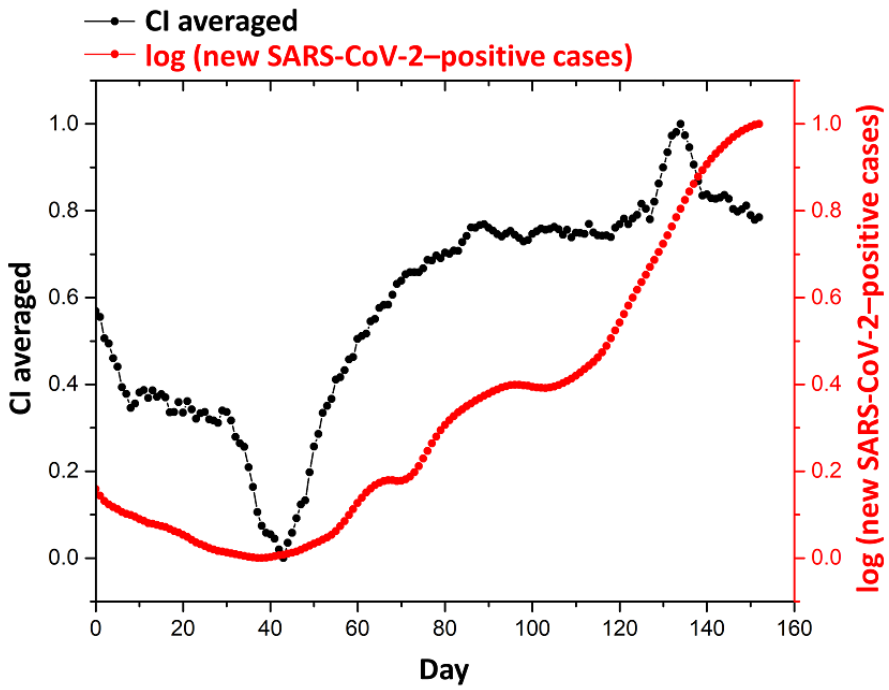
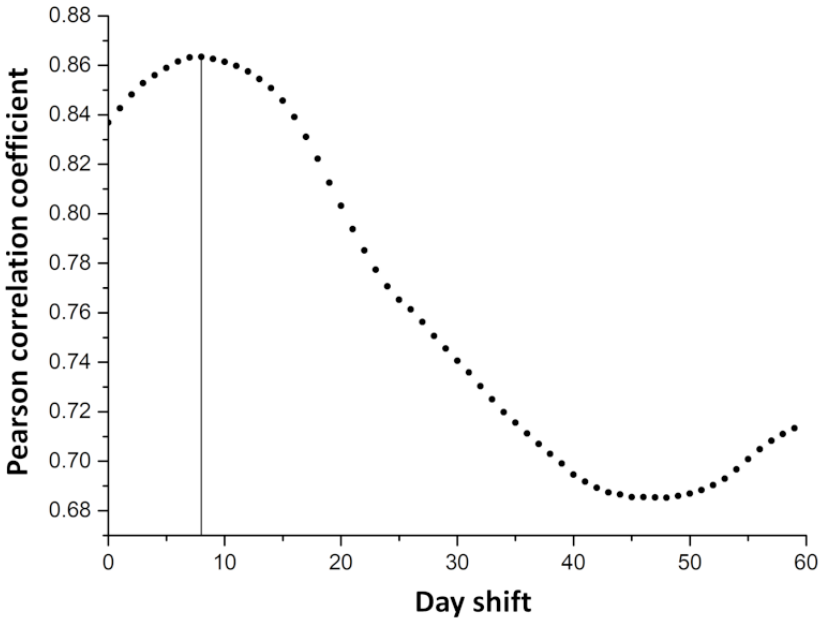


Figure 3. Pearson correlation analysis between CIk and the number of new SARS-CoV-2–positive cases with a time shift of 0 days to 60 days for the period of June 1 to October 31, 2020. The highest correlation value was observed at k+7 days. CI: contact index.



Discussion

The analysis of the collected data allowed us to determine the aspects of CT that are essential for the evaluation of the progression of the SARS-CoV-2 pandemic. These essential aspects were identified via the estimation of the real number of new SARS-CoV-2–positive cases and the correlation of the number and frequency of contacts with the probability of infection.

Estimation of the Total Number of People Who Tested Positive for SARS-CoV-2

At the beginning of the pandemic in Italy, during the period of March to May 2020, the substantial underestimation of the total number of people who tested positive for SARS-CoV-2 in Italy

was a likely scenario. This was undoubtedly due to the reduced number of tests that were performed during the first phase of the SARS-CoV-2 pandemic and the lack of an adequate response for tracing infections. One method for estimating a realistic number of SARS-CoV-2–positive cases is to use the ratio between the number of tests carried out and the number of SARS-CoV-2–positive cases detected every day. We chose this ratio because as the number of tests carried out increases, this number eventually plateaus. These data are collected throughout the country and are therefore subject to regional and local variability. It has been assumed that the ratio between the number of positive cases and the number of tests performed varies slowly over time in the absence of hospitalization problems. This ratio has been used to estimate the actual number of SARS-CoV-2–positive cases, which is always greater than or equal to the official number of cases. As shown in Figure 1, the difference between the official number of daily new SARS-CoV-2–positive cases and the estimated number of cases was higher during the initial phases of the pandemic (ie, during the period of March to May 2020). During this period, according to our analysis, at least 81,000 patients with SARS-CoV-2 infection were not diagnosed with COVID-19. As already mentioned, calculating the real number of new SARS-CoV-2–positive cases was necessary because the data provided by the Istituto Superiore di Sanità (Italian Superior Institute of Health) varied according to the number of tests performed each day. In the initial stages of the pandemic, the number of tests was remarkably low due to the lack of adequate diagnostic tools.

Ethical and Practical Issues of CT Apps

CT apps have generated much discussion, particularly discussions regarding privacy and such apps' susceptibility to attacks. Considerations of data security and possible privacy violations are certainly essential elements and have resulted in the creation of numerous solutions that have been adopted at the national level. This paper does not aim not to take a position on the security and privacy of CT apps, although the developers of SM-COVID-19 have considered these aspects. Rather, we are concerned with assessing whether CT apps, that is, those that can be developed based on currently available technology, can impact communities' health. Several apps have been adopted at a national level by multiple countries. However, during our research, we did not find any information on the availability of data collected by these apps. CT data provide useful information on various aspects of the SARS-CoV-2 pandemic (eg, the pandemic course) and the behavior and mobility of app users, thereby allowing researchers to map the frequency of contacts and identify high-risk areas. Our CT data set allowed us to analyze data and identify different classes of behavior among the population.

The SM-COVID-19 app uses a centralized model [23,24]. However, despite using a centralized model, users' privacy is completely protected via anonymization, as per the General Data Protection Regulation. The advantage of using a centralized model is that data stored on the server can be anonymized via aggregation and used by public authorities as a source of important aggregate information about the number of contacts in the population, the app's effectiveness in tracing and alerting contacts, and the aggregate number of people who could potentially develop symptoms. Unlike a decentralized model, a centralized model provides access to CT data, thereby making these data available for analysis and the improvement of epidemiological models. As already stated by Ferretti et al [19], the control of the SARS-CoV-2 epidemic via manual CT is impossible, as CT introduces a time lag resulting from the need to notify individuals about having contact with infected individuals. Such lag exacerbates the spread the infection, which is already remarkable given the infectivity of SARS-CoV-2 and the high percentage of transmission by presymptomatic individuals. The use of this app model, in which individuals are immediately notified about having contact with people who tested positive for SARS-CoV-2, would be sufficient for stopping the epidemic if the app is used by an adequate number of people [30] and would provide valuable data for creating accurate and valid predictive and epidemiological models. The choice of using a centralized model allows for the reconstruction of the chains of contagion transmission and the rapid propagation of risk indices (calculated with mathematical models)—operations that are difficult to implement when tracing data are only kept on devices.

By using data from August 2020, during which no lockdown measures or restrictions on mobility were in place and only partial restrictions were placed on gatherings, it was possible to identify 5 different behavior classes (or mobility classes). Table 1 shows the data from the clustering process. The five groups had approximately the same population size except for cluster 5, which had the largest number of people and included individuals with the highest mobility. The high amount of deviation in cluster 5 shows how users in this class alternated between experiencing days with 0 contacts (ie, no mobility; eg, days when they could be working from home) and experiencing days with a very high number of contacts (eg, due to a commute or due to work involving contact with the public). From these clusters, it is impossible to define the reasons behind a given number of contacts, but this is irrelevant as long as similar behaviors are present among the users belonging to a certain cluster. However, this clustering process provided interesting insights; it showed that there are classes of people with very low mobility (eg, older people) and classes of people with high mobility who experience a high number of contacts (eg, working in a hospital, supermarket, etc). This information can be even more useful when using a localized approach, such as using GPS data, as such data would help with providing more appropriate definitions for categories. The contacts registered by the app allowed us to trace the frequency of contacts and the trend in the number of contacts for a given period, a single user, a cluster, or the whole data set.

Correlation Between CT and the Total Number of New SARS-CoV-2–Positive Cases

CT data correlated with the growth in the number of new SARS-CoV-2–positive cases, and the highest correlation was observed 5 to 7 days after day k. This observation is in line with the hypothesis that an increase in the number of contacts is linked to an increased risk of infection. The most interesting element of the correlation is the time gap. The differences in the correlation values were probably related to the incubation

period of SARS-CoV-2. Consequently, a contact that occurs on day k will not result in COVID-19 positivity on day k but on day k+n. This time gap is in line with the estimated incubation time for SARS-CoV-2 [4,28], and our analysis shows the effectiveness of using CT data to predict the number of new SARS-CoV-2–positive cases. This high correlation means that CT data can be used to develop new and more accurate epidemiological models and predictive tools.

Although a distributed approach that involves the use of a central advertising server makes it possible to alert individuals in direct contacts (the first contact between a newly infected individual and another person) about an eventual infection, flooding operations are necessary on CT networks to warn individuals about contacts of level 2 or higher. The decentralized model provides only 1 degree of separation from a CT app user who tested positive for COVID-19 (user A). To obtain data on a longer chain of contacts, which would have a decreasing risk gradient, it would be necessary for user B (a user in user A's contact chain) to publish their identifier so that user C (a user who had contact with user B but not with user A) is alerted. This could prove particularly dangerous when an asymptomatic or low-symptomatic individual who has not been tested for SARS-CoV-2 infection could infect another person and even cause another person's death. [31] In such a situation, decentralized CT would fail. On the other hand, the centralized model allows for the instant tracing of all contacts, regardless of the degree of separation. This would result in the more effective containment of the contagion, since all individuals in a contact chain that are deemed to be at risk for infection would be notified immediately about the danger. In this model, voluntary data input by individuals involved in first-degree contacts for informing those involved in second-degree contacts would not be required whenever the former was notified about having contact with a person who tested positive for COVID-19. Similar conclusions were reached by Aleta et al [30], who proved the effectiveness of using an automatic and extensive CT system to contain the spread of SARS-CoV-2 when lockdown measures are lifted. The work of Aleta et al [30] confirmed the usefulness of CT data collected from the population and provided an excellent basis for improving predictions and reducing the social and economic impact of SARS-CoV-2 prior to the effective vaccination of the entire population. At the time of writing this paper, we did not find any other available data sets with real CT data.

Geolocalization

CT data can be beneficial for evaluating SARS-CoV-2 propagation data. The data set that was made available by the app is particularly interesting because, due to its structure, it can be used as the basis for tracing SSEs. SSEs are generally defined as outbreaks in which a small number of individuals infect a large number of secondary individuals (ie, well-above the expected average number of individuals) [32]. The CT data that allowed us to define behavioral clusters for the population can also help with determining the SARS-CoV-2 pandemic's potential for generating SSEs. Although lower than those of the SARS-CoV and MERS-CoV pandemics, the SARS-CoV-2 pandemic's potential for generating SSEs is significant. In the absence of interventions such as social distancing, this potential would be even more significant. When developing disease control measures, people should focus on the rapid CT and quarantining of infected individuals and policies for physical distancing or targeted shutdowns to prevent the occurrence of SSEs. Having the ability to predict a pandemic's potential for generating SSEs would be vital in preventing outbreaks, and it would considerably reduce a contagion's overall R_0 value. The use of GPS data that are made anonymous with an appropriate protocol would enable researchers to use a rapid localized approach to significantly reducing the risk of contagion spread in certain areas and act in a targeted and localized manner. This type of information can prove very useful for planning the possible containment of a contagion in defined areas. The tests we performed that used GPS data showed the potential of this approach. For these tests, CT data that were acquired during the lockdown period (April 14 to May 3, 2020) from SM-COVID-19 users who had explicitly activated GPS tracing and whose GPS coordinates included the Campania Region were used (Figure 4). The simulations showed that a higher number of alerts were generated in locations that corresponded to the outbreaks that occurred during the lockdown (Figure 5). This type of voluntarily provided information can be a handy tool for confining and preventing contagion spread.

Figure 4. A map showing contact tracing app users' GPS locations on September 10, 2020. These data were used for the tests.

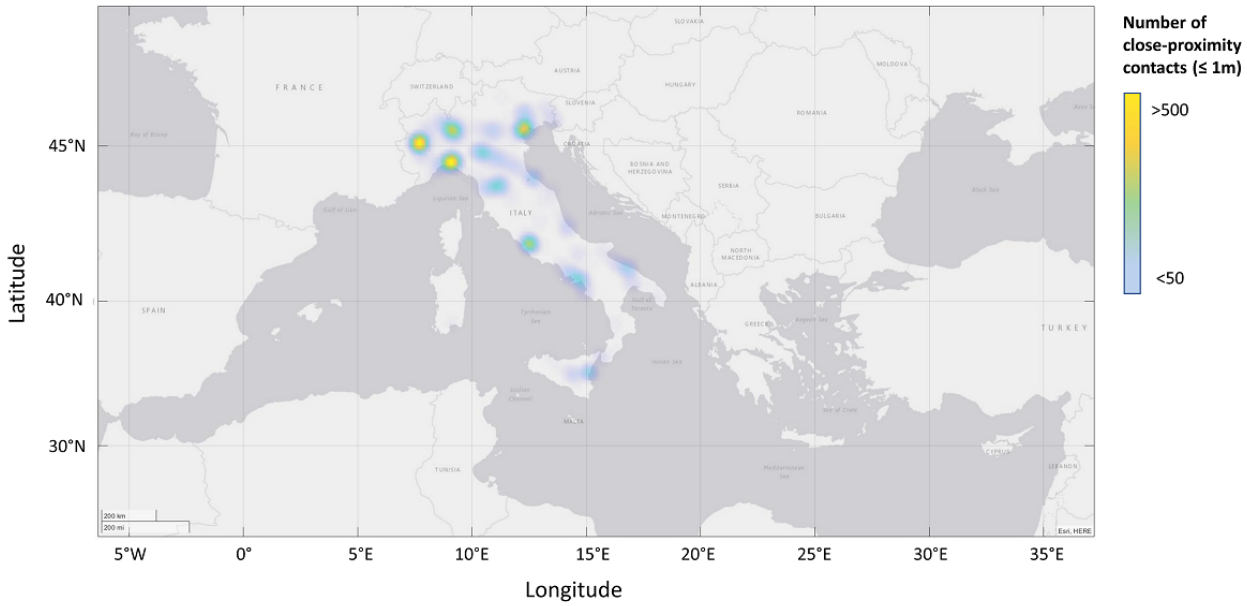
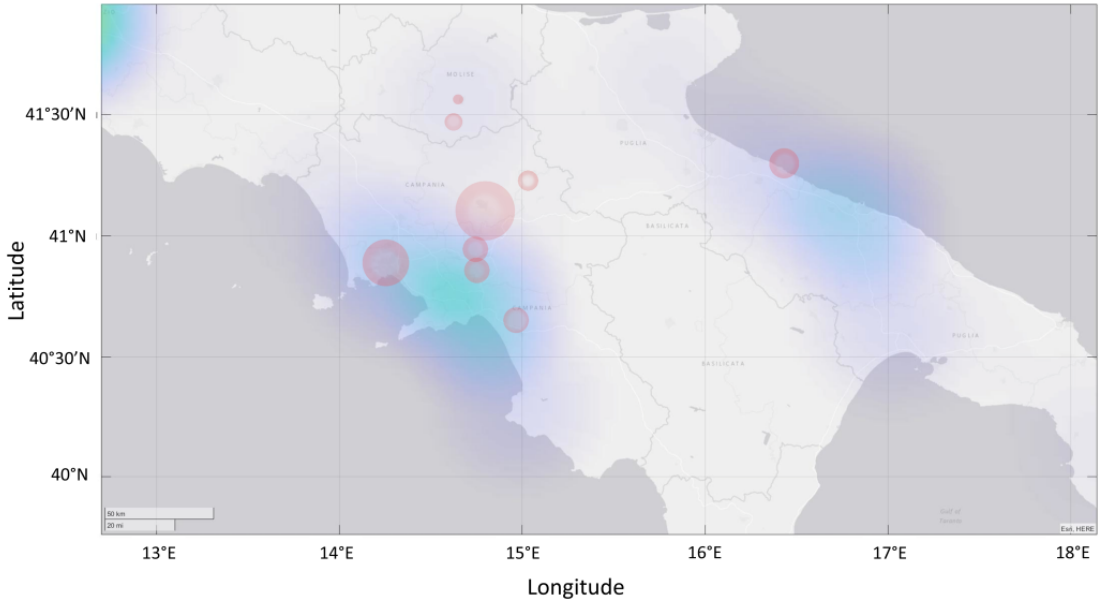


Figure 5. GPS test results. Green areas indicate locations that had a low risk of SARS-CoV-2 infection. Red areas indicate locations that had a high risk of infection. The red areas correspond to locations where SARS-CoV-2 outbreaks happened during the lockdown period.



Conclusions

The high correlation between CT data and the number of recorded SARS-COV-2–positive cases (with a delay of 5-7 days) was remarkable. The number of registered contacts and the number of new SARS-COV-2–positive cases showed the same weekly trend fluctuations, which not only depended on the number of tests but also on the different mobility abilities of people. Moreover, there was a time lag between the two factors, and this was the result of the incubation time of SARS-CoV-2. This time lag can be used to estimate the real incubation time of SARS-CoV-2. Further, this correlation can be extremely useful for defining and predicting infection trends and can be used to improve predictive models that only use health authorities' data. Regardless of the effectiveness of CT, the collected data provided a powerful tool for improving predictive and epidemiological models and could be integrated

into different types of analyses to improve the accuracy and efficiency of predictions based on real data.

This study lays the foundation for our upcoming papers. In future papers, we will show how CT data were implemented in a CT simulator to turn it into a real data-based contagion spread simulator, which provided us with data on the mobility of the different clusters that were defined in this study. The agents' mobility data will be used to determine the risk of infection, identify epidemiological parameters, and simulate the spread of SARS-CoV-2 in different contexts. The SM-COVID-19 data set is open and free for use by the scientific community. This paper does not represent a policy pronouncement, as this would not be a scientific objective. We believe that our study may prompt informed discussions of the possible risks and likely benefits of our approach to using CT data. For these reasons, all collected data are available for further analysis.

Acknowledgments

The authors thank the Campania Region for supporting our scientific research and the development of the app. We express our gratitude to Cesare Pianese and Luca Canepa for the numerous opportunities to discuss ethical and privacy issues. We also thank the entire SM-COVID-19 team [33] for supporting and distributing the data used in this work.

Conflicts of Interest

SP and LDB are members of the academic spin-off company SM, and they were involved in the development of the SM-COVID-19 app.

Multimedia Appendix 1

Acquisition of SM-COVID-19 app data on contacts.
[DOCX File , 41 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Open data format.
[DOCX File , 15 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Dumping and anonymization.
[DOCX File , 15 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Statistical analysis and estimates of the real number of SARS-CoV-2–positive cases.
[DOCX File , 14 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Contact index and alpha values.
[DOCX File , 20 KB-Multimedia Appendix 5]

References

1. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2021-07-15]
2. Weekly epidemiological update - 5 October 2020. World Health Organization. 2020 Oct 05. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20201005-weekly-epi-update-8.pdf [accessed 2021-07-15]
3. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2021-07-15]
4. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann Intern Med 2020 May 05;172(9):577-582 [FREE Full text] [doi: 10.7326/M20-0504] [Medline: 32150748]
5. Cho SY, Kang JM, Ha YE, Park GE, Lee JY, Ko JH, et al. MERS-CoV outbreak following a single patient exposure in an emergency room in South Korea: an epidemiological outbreak study. Lancet 2016 Sep 03;388(10048):994-1001 [FREE Full text] [doi: 10.1016/S0140-6736(16)30623-7] [Medline: 27402381]

6. Kim KH, Tandi TE, Choi JW, Moon JM, Kim MS. Middle East respiratory syndrome coronavirus (MERS-CoV) outbreak in South Korea, 2015: epidemiology, characteristics and public health implications. J Hosp Infect 2017 Feb;95(2):207-213 [FREE Full text] [doi: 10.1016/j.jhin.2016.10.008] [Medline: 28153558]
7. Shen Z, Ning F, Zhou W, He X, Lin C, Chin DP, et al. Superspreading SARS events, Beijing, 2003. Emerg Infect Dis 2004 Feb;10(2):256-260 [FREE Full text] [doi: 10.3201/eid1002.030732] [Medline: 15030693]
8. Wang SX, Li YM, Sun BC, Zhang SW, Zhao WH, Wei MT, et al. The SARS outbreak in a general hospital in Tianjin, China -- the case of super-spreader. Epidemiol Infect 2006 Aug;134(4):786-791. [doi: 10.1017/S095026880500556X] [Medline: 16371174]
9. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. Int J Infect Dis 2020 Mar;92:214-217 [FREE Full text] [doi: 10.1016/j.ijid.2020.01.050] [Medline: 32007643]
10. Domingo E. Chapter 7 - Long-term virus evolution in nature. In: Virus as Populations (Second Edition). Cambridge, Massachusetts, United States: Academic Press; 2020:225-261.
11. Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, WHO Rapid Pandemic Assessment Collaboration. Pandemic potential of a strain of influenza A (H1N1): early findings. Science 2009 Jun 19;324(5934):1557-1561 [FREE Full text] [doi: 10.1126/science.1176062] [Medline: 19433588]
12. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol 2004 Sep 15;160(6):509-516 [FREE Full text] [doi: 10.1093/aje/kwh255] [Medline: 15353409]
13. Zhang S, Diao M, Yu W, Pei L, Lin Z, Chen D. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. Int J Infect Dis 2020 Apr;93:201-204 [FREE Full text] [doi: 10.1016/j.ijid.2020.02.033] [Medline: 32097725]
14. Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College COVID-19 Response Team. 2020 Mar 16. URL: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf [accessed 2021-07-15]
15. Fraser C, Riley S, Anderson RM, Ferguson NM. Factors that make an infectious disease outbreak controllable. Proc Natl Acad Sci U S A 2004 Apr 20;101(16):6146-6151 [FREE Full text] [doi: 10.1073/pnas.0307506101] [Medline: 15071187]
16. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Glob Health 2020 Apr;8(4):e488-e496 [FREE Full text] [doi: 10.1016/S2214-109X(20)30074-7] [Medline: 32119825]
17. Klinkenberg D, Fraser C, Heesterbeek H. The effectiveness of contact tracing in emerging epidemics. PLoS One 2006 Dec 20;1(1):e12. [doi: 10.1371/journal.pone.0000012] [Medline: 17183638]
18. Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert JHHM, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. Lancet Public Health 2020 Aug;5(8):e452-e459 [FREE Full text] [doi: 10.1016/S2468-2667(20)30157-2] [Medline: 32682487]
19. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491):eabb6936 [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
20. Bengio Y, Janda R, Yu YW, Ippolito D, Jarvie M, Pilat D, et al. The need for privacy with public digital contact tracing during the COVID-19 pandemic. Lancet Digit Health 2020 Jul;2(7):e342-e344 [FREE Full text] [doi: 10.1016/S2589-7500(20)30133-3] [Medline: 32835192]
21. Dar AB, Lone AH, Zahoor S, Khan AA, Naaz R. Applicability of mobile contact tracing in fighting pandemic (COVID-19): Issues, challenges and solutions. Comput Sci Rev 2020 Nov;38:100307 [FREE Full text] [doi: 10.1016/j.cosrev.2020.100307] [Medline: 32989380]
22. Chan J, Foster D, Gollakota S, Horvitz E, Jaeger J, Kakade S, et al. PACT: Privacy Sensitive Protocols and Mechanisms for Mobile Contact Tracing. arXiv. 2021 Preprint posted online on May 7, 2020. [FREE Full text]
23. Ahmed N, Michelin RA, Xue W, Ruj S, Malaney R, Kanhere SS, et al. A survey of COVID-19 contact tracing apps. IEEE Access 2020 Jul 20;8:134577-134601 [FREE Full text] [doi: 10.1109/access.2020.3010226]
24. Braithwaite I, Callender T, Bullock M, Aldridge RW. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. Lancet Digit Health 2020 Nov;2(11):e607-e621 [FREE Full text] [doi: 10.1016/S2589-7500(20)30184-9] [Medline: 32839755]
25. SM-Covid-19 App — Tracing app against Covid-19. SM-COVID-19. URL: https://www.smcovid19.org [accessed 2021-07-15]
26. COVID-19 Italia - Monitoraggio situzaione. GitHub. URL: https://github.com/pcm-dpc/COVID-19/ [accessed 2021-07-15]
27. SM-COVID-19: Contact Tracing Open Data. SM-COVID-19. URL: https://smp.softmining.it/smp-smc19od.aspx [accessed 2021-07-15]
28. Yang L, Dai J, Zhao J, Wang Y, Deng P, Wang J. Estimation of incubation period and serial interval of COVID-19: analysis of 178 cases and 131 transmission chains in Hubei province, China. Epidemiol Infect 2020 Jun 19;148:e117 [FREE Full text] [doi: 10.1017/S0950268820001338] [Medline: 32594928]

29. Xiao Y, Torok ME. Taking the right measures to control COVID-19. Lancet Infect Dis 2020 May;20(5):523-524 [FREE Full text] [doi: 10.1016/S1473-3099(20)30152-3] [Medline: 32145766]
30. Aleta A, Martín-Corral D, Piontti APY, Ajelli M, Litvinova M, Chinazzi M, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. Nat Hum Behav 2020 Sep;4(9):964-971. [doi: 10.1038/s41562-020-0931-9] [Medline: 32759985]
31. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 2020 May 01;368(6490):489-493 [FREE Full text] [doi: 10.1126/science.abb3221] [Medline: 32179701]
32. Galvani AP, May RM. Epidemiology: dimensions of superspreading. Nature 2005 Nov 17;438(7066):293-295 [FREE Full text] [doi: 10.1038/438293a] [Medline: 16292292]
33. Team — SM-Covid-19 App. SM-COVID-19. URL: https://www.smcovid19.org/team/ [accessed 2021-07-20]

Abbreviations

CI: contact index
CT: contact tracing
MERS-COV: Middle East respiratory syndrome coronavirus
SARS-CoV: severe acute respiratory syndrome coronavirus
SM: SoftMining
SSE: superspreading event

Edited by C Basch; submitted 19.03.21; peer-reviewed by F Tonelli, A Lahiri; comments to author 29.04.21; revised version received 10.05.21; accepted 15.05.21; published 02.08.21

Please cite as:
Piotto S, Di Biasi L, Marrafino F, Concilio S
Evaluating Epidemiological Risk by Using Open Contact Tracing Data: Correlational Study
J Med Internet Res 2021;23(8):e28947
URL: https://www.jmir.org/2021/8/e28947
doi: 10.2196/28947
PMID: 34227997