

Development of an algorithm to identify urgent referrals for suspected cancer from the Danish Primary Care Referral Database

Berit Skjødeberg

Toftgaard^{1,2,3}

Louise Mahncke

Guldbrandt^{1,2,3}

Kaare Rud Flarup^{1,2}

Hanne Beyer¹

Flemming Bro^{1,3}

Peter Vedsted^{1,2,4}

¹Department of Public Health, Research Unit for General Practice,

²Department of Public Health, Research Centre for Cancer Diagnosis in Primary Care (CaP), Aarhus University, ³Department of Public Health, Section for General Medical Practice, ⁴Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

Background: Accurate identification of specific patient populations is a crucial tool in health care. A prerequisite for exploring the actions taken by general practitioners (GPs) on symptoms of cancer is being able to identify patients urgently referred for suspected cancer. Such system is not available in Denmark; however, all referrals are electronically stored. This study aimed to develop and test an algorithm based on referral text to identify urgent cancer referrals from general practice.

Methods: Two urgently referred reference populations were extracted from a research database and linked with the Primary Care Referral (PCR) database through the unique Danish civil registration number to identify the corresponding electronic referrals. The PCR database included GP referrals directed to private specialists and hospital departments, and these referrals were scrutinized. The most frequently used words were integrated in the first version of the algorithm, which was further refined by an iterative process involving two population samples from the PCR database. The performance was finally evaluated for two other PCR population samples against manual assessment as the gold standard for urgent cancer referral.

Results: The final algorithm had a sensitivity of 0.939 (95% confidence intervals [CI]: 0.905–0.963) and a specificity of 0.937 (95% CI: 0.925–0.963) compared to the gold standard. The positive and negative predictive values were 69.8% (95% CI: 65.0–74.3) and 99.0% (95% CI: 98.4–99.4), respectively. When applying the algorithm on referrals for a population without earlier cancer diagnoses, the positive predictive value increased to 83.6% (95% CI: 78.7–87.7) and the specificity to 97.3% (95% CI: 96.4–98.0).

Conclusion: The final algorithm identified 94% of the patients urgently referred for suspected cancer; less than 3% of the patients were incorrectly identified. It is now possible to identify patients urgently referred on cancer suspicion from general practice by applying an algorithm for populations in the PCR database.

Keywords: cancer, referral, algorithm, general practice, early diagnosis, Denmark

Introduction

Cancer is a major healthcare burden in Denmark as in most other countries.¹ Many resources are thus allocated to research and quality improvement of the entire cancer care pathway. Urgent referral for cancer investigation was introduced in Denmark in 2008 to reduce the waiting time from first symptom presentation to treatment initiation, particularly the interval from referral to treatment.² Nevertheless, Denmark still lags behind most of the European countries in cancer survival.³ Research has revealed that high use of urgent referrals among general practitioners (GPs) and short

Correspondence: Berit Skjødeberg Toftgaard
Research Unit for General Practice, Aarhus University, Bartolins Allé 2, DK-8000 Aarhus C, Denmark
Tel +45 87168151
Email berit.toftgaard@feap.dk

time to diagnosis may improve the prognosis.^{4–8} However, selecting patients for urgent cancer referral is a complex process; each GP is involved in the diagnosis of only 8–10 new cancer patients per year, but patients with potential cancer symptoms consult GPs on a daily basis.^{9,10} A Scottish study found a 6-fold variation in the use of urgent referrals among GPs, after excluding the highest and lowest 10% of the referring practices.¹¹ Thus, identifying the patients who have been urgently referred from general practice is crucial for achieving better insight into the use of urgent referrals for suspected cancer among GPs and the potential implications for cancer survival rates. Such new insight is of particular interest in healthcare systems, where GPs act as gatekeepers and initiators of the diagnostic pathway for 85% of all cancer patients.^{12–14}

Urgent cancer referrals from general practice are currently not registered systematically in Denmark (like in the UK).¹⁵ Unpublished studies have found that the ad hoc registration at Danish hospitals cannot validly identify all urgent referrals from GPs and cannot appropriately distinguish between urgent referrals from GPs and those from other healthcare professionals. Hospital consultants can redirect an urgent referral to a different route or convert an ordinary referral requested by a GP to an urgent referral. A sustainable method to identify urgent referrals from general practice is warranted: a way forward could be to exploit the systematically collected routine data from the existing electronic referral system, which is used for all referrals to private specialists and hospitals. These data have not been previously used for research and monitoring purposes.

The aim of this study was to develop and test a text-based algorithm for searching all referrals to identify urgent cancer referrals from general practice.

Methods

Setting and referral procedure

The study was conducted in one of the five Danish regions; the Central Denmark Region with 1.3 million inhabitants, 417 GP clinics with 831 GPs and ~8,000 new cancer patients annually. The Danish healthcare system provides free (tax-funded) access to general practice and hospital care, and most Danish citizens (98%) are listed with a particular general practice. The GP acts as gatekeeper to the rest of the healthcare system by referring patients to hospitals or private specialists when relevant.¹⁴

An online guideline for urgent cancer referral describes the symptoms for several specific types of cancer; this also includes clinical and para-clinical findings that could release

an urgent cancer referral from the GP. The guideline requests the referring GP to include the term “cancer” in the diagnostic text of the referral.¹⁶

All Danish GPs use computerized administrative systems with electronic patient records. These systems communicate electronically with the rest of the healthcare system, including all referrals from general practices and private specialists. A copy of each referral is stored in the Central Referral Database.

Data

The Primary Care Referral (PCR) Database

The PCR database was established in 2014 by extracting referrals made by GPs from the Central Referral Database. Patient names, addresses, and phone numbers were deleted to ensure that all referrals included only one unique identifier: the civil registration (CPR) number.¹⁷

The PCR database included information about referral date, referring GP, receiving department, text fields for diagnostic, patient history and clinical information, patient's age and gender, and patient's CPR number in encrypted form. The PCR database did not include referrals for radiology departments.

Populations derived from the PCR database

The PCR database had not been previously used for research purposes. The authors gradually increased their knowledge about the database during the study, which allowed them to further restrict the selection criteria to achieve the most optimal target population (Table 1). The sampling period for the populations was arbitrary; although no periods were overlapping. The sample size of each population was selected to ensure small confidence intervals (CIs) on the estimates.

The PCR populations were used in the last two steps of the incremental development and in the subsequent evaluation of the algorithm. In order to be able to measure the performance of each stepped algorithm, a gold standard for urgent cancer referrals was developed. Each sampled referral was assessed by a researcher (BST) and categorized as “urgent cancer referral” or “not urgent cancer referral.” In case of any doubt, another researcher (LMG) conducted an independent assessment. In case of disagreement (<2% of all cases), referral texts were discussed until consensus had been reached on categorization. Referrals categorized as “not urgent cancer referral” were further subcategorized to obtain data on the reasons for false-positive findings. The assessors were blinded to the results of the stepped algorithms.

Table I Characteristics of the four PCR-P

Populations	Inclusion criteria			Age of referred patients	Males	Urgently referred patients
	Requesters	Receivers	Patients	Median/years (IQI)	% (95% CI)	% (95% CI)
PCR-P I N=1,000	GPs in DK	Hospital (DK)	All age	52.0 (31.9–67.9)	40.3 (37.3–43.3)	9.1 (7.3–10.9)
PCR-P II N=2,000	GPs in CDR	Hospital (CDR) exc. psychiatric Private specialists ^a	≥ 40 years	62.8 (51.4–72.5)	43.5 (41.3–45.7)	10.7 (9.3–12.0)
PCR-P III N=2,172	GPs in CDR Exc. out-of- hours GPs	Hospital (CDR) exc. psychiatric Private specialists ^a	≥ 40 years	60.7 (50.0–71.1)	42.7 (40.6–44.8)	13.5 (12.1–14.9)
PCR-P IV N=1,949	GPs in CDR Exc. out-of- hours GPs	Hospital (CDR) exc. psychiatric Private specialists ^a	≥40 years No prior cancer	59.2 (49.3–70.0)	42.5 (40.3–44.8)	12.8 (11.4–14.4)

Notes: ^aPrivate specialists include the fields of otolaryngology, dermatology, gynecology, surgery, plastic surgery, or neurology in the Central Denmark Region.

Abbreviations: GP, general practitioner; DK, Denmark; CDR, Central Denmark Region; CRR-P, Cancer Referral Reference population; PCR-P, Primary Care Referral population; CI, confidence interval; IQI, interquartile intervals; exc., excluding.

Primary Care Referral Population I (PCR-P I)

One thousand unique consecutive patients were selected from the PCR database from 4 March 2013 to 14 March 2013. The referral requests that have been included should have been submitted by a Danish general practice and directed to a hospital department. No other criteria were used in the selection. This population was labelled the PCR-P I.

Primary Care Referral Population II (PCR-P II)

Two thousand unique consecutive patients aged ≥40 years were selected from the PCR database from 1 February 2013 to 8 February 2013. The referral requests that have been included should have been submitted by a general practice from the Central Denmark Region and directed to a medical unit located in the Central Denmark Region. Such medical unit could be a somatic hospital department or one of the following (privately practicing) medical specialties, which also form part of the national urgent cancer pathways: otorhinolaryngology, dermatology, gynecology, surgery, plastic surgery, or neurology. This population was labelled the PCR-P II.

Primary Care Referral Population III (PCR-P III)

The third population included 2,500 unique consecutive patients with the same inclusion criteria for receivers as for the PCR-P II (described earlier). This population was selected from the PCR database from 9 February 2013 to 18 February 2013; 328 patients referred from GPs in out-of-hours services were excluded, which resulted in a total of 2,172 patients (all referred from GPs during daytime). This population was labelled the PCR-P III.

Primary Care Referral Population IV (PCR-P IV)

A total of 1,949 patients were selected on the basis of the same inclusion criteria as applied to the PCR-P III, except that patients with prior cancer diagnosis identified in the Danish Cancer Registry (ICD 10, chapter II, C0–C9 except from C44) were excluded.¹⁸ This population was labelled the PCR-P IV.

The Cancer Referral Reference (CRR) database

Data from a parallel research study, which aimed to evaluate the impact of a continuing medical education in early cancer diagnosis, formed the basis of the CRR database. This database consisted of records on patients known to have been urgently referred for suspected cancer. The data were collected during an 8-month period from 1 September 2012 to 30 April 2013 by requesting all GPs in the Central Denmark Region to complete a one-page form with patient information each time they urgently referred a patient for cancer investigation.¹⁹ The form comprised the patient's CPR number, the referral date, and the type of urgent cancer pathway.

Populations derived from the CRR database

Two populations were extracted from the CRR database; these were labelled the Cancer Referral Reference Population I (CRR-P I) and the Cancer Referral Reference Population II (CRR-P II). From 1 to 20 September 2012, 127 consecutive patients were identified for the CRR-P I. From 1 November 2012 to 4 January 2013, 496 consecutive patients were identified for the CRR-P II.

Patients referred to the fast-track pathway for suspected breast cancer were excluded as their referrals were directed to radiology departments and thus would not be stored in the PCR database. For each urgently referred patient, GP referrals in the PCR database were identified within a period of ± 2 weeks surrounding the referral date registered by the GP. Patients could be included only once (the first referral). The CRR populations were used in the first two steps of the development of the algorithm.

Algorithm development and stepwise refinement

To identify text fields and words used by GPs which could identify urgent referrals, populations known to be urgently referred (CRR-P I and CRR-P II) were used. To develop an algorithm that would best balance sensitivity (proportion of true positives correctly identified as such) and specificity (proportion of true negatives correctly identified as such), unselected populations (PCR-P I and PCR-P II) were used.²⁰

In order to avoid that the performed refinements and measured performance of the stepped algorithms would be tailored to a specific population, a new population was used for each successive step of the development. It was considered important that the populations (PCR-P III and PCR-P IV) were “unknown” in the developing steps of the algorithm in order to validly measure the performance of the final algorithm.

Step 1: Identifying search fields and most frequent words in urgent cancer referrals

Two researchers (BST and LMG) independently reviewed the 127 CRR-P I referrals and identified the text fields that were most frequently used by GPs while filling out an urgent cancer referral. Five fields were identified, namely diagnosis, supplementary diagnoses, patient history (anamnesis), supplementary information and supplementary clinical information ([Supplementary material](#)). Each referral was then thoroughly examined to identify the most commonly used words linked to cancer suspicion. The selected words were “*pakke*” (Danish term for cancer pathway), “cancer,” “malignant,” general medical terms such as “tumor” and “lump,” and cancer-specific terms such as “leukemia” and “lymphoma.” In order to embrace typical misspellings, spelling variations, and use of abbreviations, letter strings of the identifying words were made as short as possible. A step-1 algorithm was developed; this included the most frequently used strings for cancer, cancer pathways, and cancer types.

Step 2: Refining the step-1 algorithm to enhance the sensitivity

The step-1 algorithm was applied to the CRR-P II. Strings were then identified and included until the step-2 algorithm appeared saturated and obtained a sensitivity >95%. Among the added strings were symptoms such as “hemoptysis,” “hematuria,” “rectal bleeding,” and procedures like “colonoscopy” and “gastroscopy”.

Step 3: Refining the step 2-algorithm by differentiated search

The step-2 algorithm was applied to the PCR-P I. A frequent reason for false-positive (incorrectly identified) findings was the search fields “anamnesis,” “additional information,” or “additional diagnoses.” Therefore, specifically problematic combinations of words and fields were identified in an iterative process by reviewing text, omitting fields in a new algorithm, and performing repeated calculations of specificity and sensitivity. The increase in specificity was achieved under strict supervision to ensure the lowest possible drop in sensitivity.

Step 4: Refining the step-3 algorithm by excluding misleading strings

The step-3 algorithm was applied to the PCR-P II to further reduce the false-positive rates. Combinations of words that were especially problematic were identified. For example, “*pakke*,” which was one of the strongest identification words, was also used in combination with “heart” and “blood test.” Therefore, the algorithm was set to exclude the identified case if “*pakke*” appeared in the same text field as “heart,” “angina,” or “rheumatic.” These combinations of identifying words were gradually incorporated to maximize the specificity while maintaining a sensitivity >93%.

Step 5: Assessing the performance of the final algorithm

The final algorithm was applied to the PCR-P III and the PCR-P IV to calculate the performance. The two populations comprised patients aged ≥ 40 years who had been referred from general practice in the daytime; one included patients with prior cancer diagnosis (PCR-P III), whereas the other (PCR-P IV) did not.

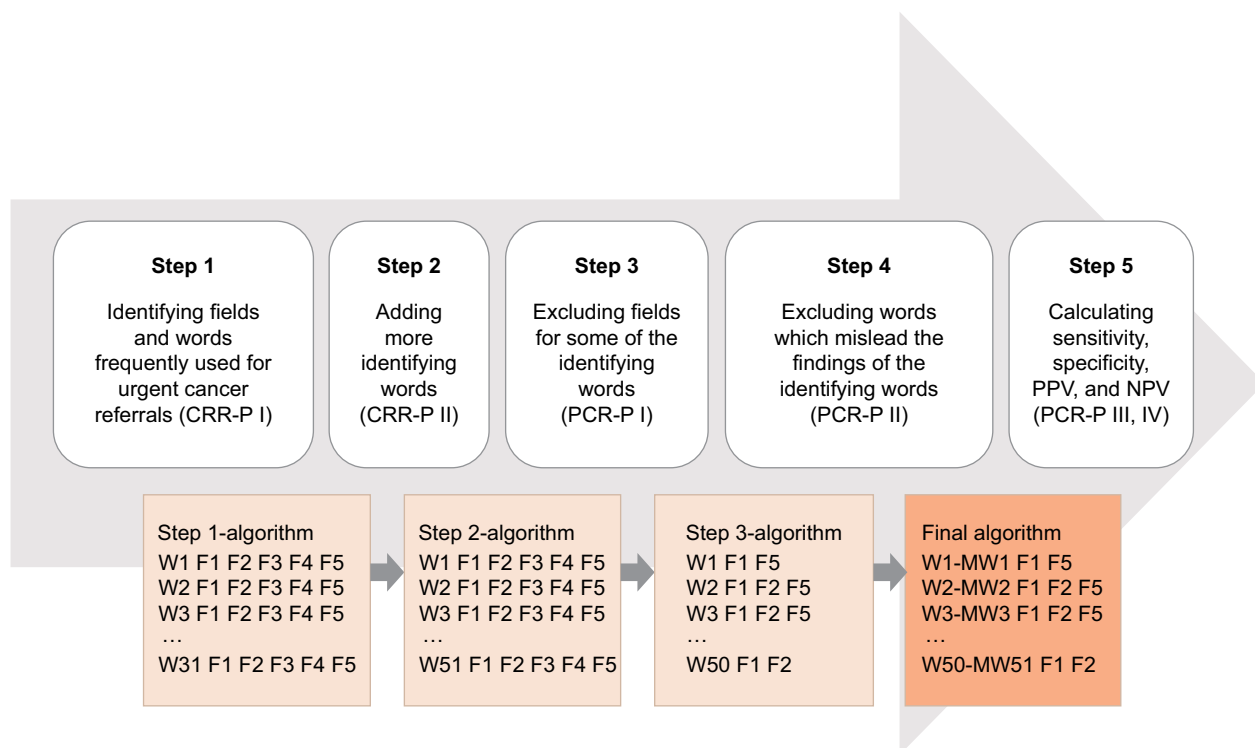
In order to test the consistency of the performance at each step, the stepped algorithms were applied to each successive population (Table 2). The five steps in the development of the algorithm are shown in Figure 1.

Table 2 The performance of the stepped algorithms on each population

	Step-1 algorithm	Step-2 algorithm	Step-3 algorithm	Final algorithm
CRR-P I^a				
Sensitivity % (95% CI)	89.8 (84.4–95.1)	96.9 (93.8–99.9)	95.3 (91.5–99.0)	95.3 (91.5–99.0)
CRR-P II^a				
Sensitivity % (95% CI)	85.7 (82.6–88.8)	94.8 (92.8–96.7)	94.6 (92.6–96.6)	93.8 (91.6–95.9)
PCR-P I^b				
Sensitivity % (95% CI)	63.7 (53.0–73.6)	97.8 (92.3–99.7)	94.5 (87.6–98.2)	94.5 (87.6–98.2)
Specificity % (95% CI)	90.8 (88.7–92.6)	83.2 (80.6–85.5)	92.2 (90.2–93.8)	95.7 (94.2–96.9)
PPV % (95% CI)	40.8 (32.7–49.4)	36.8 (30.7–43.2)	54.8 (46.6–62.7)	68.8 (59.9–76.8)
NPV % (95% CI)	96.2 (94.6–97.3)	99.7 (99.1–100.0)	99.4 (98.6–99.8)	99.4 (98.7–99.8)
PCR-P II^b				
Sensitivity % (95% CI)	78.9 (72.8–84.2)	97.2 (94.0–99.0)	94.8 (90.9–97.4)	93.0 (88.7–96.0)
Specificity % (95% CI)	84.9 (83.2–86.6)	76.6 (74.6–78.6)	90.4 (88.9–91.7)	95.2 (94.2–96.2)
PPV % (95% CI)	38.4 (33.9–43.2)	33.1 (29.4–37.0)	54.0 (48.8–59.1)	70.0 (64.3–75.2)
NPV % (95% CI)	97.1 (96.2–97.9)	99.6 (99.1–99.8)	99.3 (98.8–99.7)	99.1 (98.6–99.5)
PCR-P III^b				
Sensitivity % (95% CI)	76.1 (70.8–80.9)	96.2 (93.4–98.1)	93.9 (90.5–96.3)	93.9 (90.5–96.3)
Specificity % (95% CI)	81.6 (79.8–83.3)	73.9 (71.8–75.8)	88.3 (86.8–89.7)	93.7 (92.5–96.3)
PPV % (95% CI)	39.2 (35.2–43.3)	36.5 (33.1–40.9)	55.6 (51.1–60.0)	69.8 (65.0–74.3)
NPV % (95% CI)	95.6 (94.5–96.6)	99.2 (98.6–99.6)	98.9 (98.3–99.4)	99.0 (98.4–99.4)
PCR-P IV^b				
Sensitivity % (95% CI)	N/A	N/A	N/A	93.6 (89.8–96.3)
Specificity % (95% CI)	N/A	N/A	N/A	97.3 (96.4–98.0)
PPV % (95% CI)	N/A	N/A	N/A	83.6 (78.7–87.7)
NPV % (95% CI)	N/A	N/A	N/A	99.0 (98.4–99.5)

Notes: ^aThe performance of the algorithms on the Cancer Referral Reference populations (known to be urgent cancer referrals) was assessed by sensitivity. ^bThe performance of the algorithms on the Primary Care Referral populations was assessed by sensitivity, specificity, and positive and negative predictive values; all with 95% CIs.

Abbreviations: GP, general practitioner; CI, confidence interval, CRR-P, Cancer Referral Reference population; PCR-P, Primary Care Referral Population; PPV, positive predictive value; NPV, negative predictive value; N/A, not applicable.

**Figure 1** The stepwise development of the algorithm.

Abbreviations: CRR-P, Cancer Referral Reference population; PCR-P, Primary Care Referral population; PPV, positive predictive value; NPV, negative predictive value; W, word; F, Fields; MW, misleading word.

Statistics

The populations were described in terms of inclusion criteria for requester, receiver and patient, number of patients, patient age (median and interquartile intervals), proportion of males, and urgent cancer referrals for all referred patients according to manual assessment with a CI of 95%.

As the stepped algorithms provided dichotomous outcomes in terms of urgent cancer referral (“yes” or “no”), simple performance statistics was applied to calculate the sensitivity, specificity, positive predictive value (PPV; proportion of test positives that are truly positive) and negative predictive value (NPV; proportion of test negatives that are truly negative) with 95% CI for each of the stepped algorithms to allow for comparisons.²⁰ Coding and analyses were performed using Stata version 13.0 software (StataCorp LP, College Station, TX, USA).

Ethical approval

The study was approved by the Danish Data Protection Agency (file no. 2009-41-3471). The establishment of the PCR database was approved by Danish Regions. The collection of data for the CRR database was approved by the Danish Health Authority (file no. 3-3013-149/1/HKR) and the Danish Data Protection Agency, the Central Denmark Region (file no. 1-16-02-262-12).

Results

Identification strings and search fields are provided for each of the four-stepped algorithms. The final algorithm consisted of 50 identifying words; 35 were combined with the exclusion of misleading words.

Gradual improvement of the first and second algorithms increased the sensitivity and the NPV, whereas the specificity

and the PPV decreased. Specificity and PPV were both improved from the third to the final version of the algorithm (Table 2).

The final algorithm, which was applied to the PCR-P III, correctly identified 275 referrals as urgent cancer referrals and 1,760 as “not urgent cancer referrals.” However, it misclassified 119 referrals as “urgent cancer referrals” and 18 as “not urgent cancer referrals” (Table 2).

Three out of ten urgent cancer referrals included the word “cancer”; one of the words (defined as “cancer wording”) “*kræft*” (Danish term for cancer), “*cancer*,” “*pakke*,” or “*c.*” (abbreviation for cancer) was included in half of the referrals (Table 3). The most frequent identifying words that resulted in a false-positive finding were “*kræft*,” “*cancer*,” and “*malign*” (Table 3). These words could not be removed without decreasing the sensitivity of the algorithm. False-positive findings were mostly related to previously diagnosed cancer disease, earlier cancer investigation, hereditary cancer predisposition, or fear of cancer (Table 4).

Exclusion of patients with prior cancer in the PCR-P IV implied that the PPV of the final algorithm increased to 83.6% (95% CI: 78.7–87.7), the specificity increased to 0.973 (95% CI: 0.964–0.980), whereas the sensitivity remained unchanged at 0.936 (95% CI: 0.898–0.963) (Table 2).

Discussion

Main findings

This study developed an algorithm capable of identifying urgent cancer referrals among all electronic referrals from Danish general practice to hospitals and private specialists. When applied for patients ≥ 40 years of age who were not diagnosed with cancer earlier, the algorithm was capable of identifying 93.6% of all urgent cancer referrals; <3%

Table 3 True- and false-positive findings of important identifying words from the final algorithm

Important identifying words	Positive findings of algorithm applied to PCR-P III (n=394)		Positive findings of algorithm applied to PCR-P IV (n=280)	
	True positives (%) n=275 (69.8)	False positives (%) n=119 (30.2)	True positives (%) n=234 (83.6)	False positives (%) n=46 (16.4)
<i>Pakke</i> ^a	82 (29.8)	4 (3.4)	74 (31.6)	3 (6.5)
<i>Kræft</i> ^b	59 (21.5)	22 (18.5)	53 (22.7)	9 (19.6)
Cancer	93 (33.8)	44 (37.0)	73 (31.2)	8 (17.4)
Malign	24 (8.7)	17 (14.3)	19 (8.1)	7 (15.2)
<i>c.</i> ^c	14 (5.1)	2 (1.7)	11 (4.7)	0 (0.0)
<i>Pakke</i> or <i>kræft</i> or cancer or <i>c.</i> ^d	143 (52.0)	64 (53.8)	126 (53.9)	24 (52.2)

Notes: ^aDanish term for cancer pathway. ^bDanish term for cancer. ^c*c.* includes *c.*, *c.pulm.*, *c.panc.*, *c.cerv.*, *c.uteri.*, *c.mamma.*, *c.cere.*, and *c.test.* ^dThe last row gives the combination of words that measure the appearance of “cancer wording.”

Abbreviation: PCR-P, Primary Care Referral Population.

Table 4 Subgroup categories of false-positive findings of the final algorithm

	Main category N (%)	Subgroup categories, n (%)					
	Not urgent cancer referral	Earlier diagnosed cancer	Earlier suspicion for cancer rejected	Urgent referral for heart disease	Nonmelanoma skin cancer	Hereditary disposure or fear for cancer	Others
False positives	119 (100.0)	75 (63.0)	15 (12.6)	0 (0.0)	2 (1.7)	11 (9.2)	16 (13.4)

Notes: The final algorithm was applied to the Primary Care Referral population III. A false-positive finding was manually assessed as “not urgent cancer referral” and incorrectly identified by the algorithm. “Not urgent cancer referrals” were manually divided into subgroup categories.

were false positives, and a PPV of 84% was obtained. This provides an acceptably low misclassification rate for use in epidemiological research and quality development.

Strengths and weaknesses

The strength of the algorithm development was the stepwise approach, which ensured optimal balance between sensitivity and specificity. However, this balance depends on the specific purpose. A higher sensitivity could be obtained (eg, to ensure that all urgently referred patients are included), but this would require acceptance of lower specificity and PPV. In epidemiological research and quality development, fairly high specificity and PPV were generally required to enable restriction. It was realized that some urgently referred patients might have been missed in this attempt to ensure that all included patients had actually been urgently referred.

Urgent referrals for breast cancer could not be identified as these referrals were generally directed directly to a radiology department and not stored in the PCR database. Manual assessment, which established the gold standard, was made before applying the algorithm to the data, and the assessors were blinded to the results of the algorithm in order to ensure that misclassification was reduced to a minimum. The process was initiated by reviewing referrals that were known to be urgent cancer referrals (the CRR populations). This provided the insight that “cancer wording” is stated in only half of all referrals, and thus it was known that more cancer identification words than originally anticipated had to be included. The method based on actual referrals is considered more precise than identification of frequently used words through, for example, interviews or brainstorming sessions with GPs.

The measured performance of the final algorithm can be generalized to the rest of Denmark as there is no reason to believe that GPs in the Central Denmark Region communicate urgent referrals differently than other Danish GPs. Nevertheless, the GPs may improve their wording over time while writing an urgent referral letter and more often

remember to include “cancer” in the diagnostic text as they are informed to do so in the guideline.¹⁶ Such development would require regular adaptations of the algorithm based on measurements of performance.

One weakness of using a text-based algorithm to identify patients in the PCR database is the potential variation among GPs in the way they make referrals and communicate through referrals. The findings in this study suggest that it could be helpful to introduce a specific template for each urgent cancer pathway in the electronic referral system as this would require a simpler algorithm and allow distinction between different cancer pathways.

Implications and comparison with literature

Being able to identify the patients who have been urgently referred for suspected cancer from general practice opens up some new possibilities for register-based retrospective research. More accurate knowledge on cancer-related referral activity including referral, conversion, and detection rates in primary care may contribute to a better and more coherent healthcare organization.¹⁵

Text-based algorithms can be developed and applied to the PCR database for other research purposes, for example, to investigate diabetes or thyroid disorders. Other countries with an electronic referral system may use the explained stepped approach to develop a comparable algorithm. The essential studies on PPVs for symptoms of cancer and the QCancer web tool are based on word recognition from routinely collected patient data in a large number of GP clinics in England.^{21–29} However, to our knowledge, no other studies have reported the methodological development of an algorithm aiming to sample specific patient populations through pattern recognition of word strings and text fields in referral letters.

Conclusion

This study has successfully developed an algorithm that identifies 94% of all patients urgently referred for suspected

cancer. The PPVs were 69% for all adults aged ≥ 40 years and 84% when the cases with prior cancer were excluded.” The final algorithm helps to identify patients urgently referred for suspected cancer, and it can be used to monitor the cancer-related referral performance in general practice. The algorithm may constitute a valuable research tool in future studies as it could facilitate evaluations of the performance in general practice and the outcome for the patients. Similar algorithms may be developed for other types of disease.

Acknowledgments

The authors thank the GPs in the Central Denmark Region, Anne Gammelgaard at Quality and Informatics in the Central Denmark Region, and the four radiology departments in Aarhus, Holstebro, Randers, and Viborg for providing the information to build the CRR database. This work was supported by the Novo Nordisk Foundation, the Danish Cancer Society, and the Foundation for Primary Health Care Research (*Praksisforskningsfonden*) in the Central Denmark Region (grant no. 1-15-1-72-13-09).

Author contributions

KF and HB established the central referral database and the PCR database. BST and LG developed the algorithm supported by PV and FB. BST evaluated the algorithm and wrote the first draft of the article. PV, FB, LG, KF, and HB critically reviewed the article. All the authors read and commented on the final manuscript.

Disclosure

The authors report no conflicts of interest in this work.

References

- Engholm G, Ferlay J, Christensen N, et al. NORDCAN – a Nordic tool for cancer information, planning, quality control and research. *Acta Oncol.* 2010;49(5):725–736.
- Probst HB, Hussain ZB, Andersen O. Cancer patient pathways in Denmark as a joint effort between bureaucrats, health professionals and politicians – a national Danish project. *Health Policy.* 2012;105(1):65–70.
- Coleman MP, Forman D, Bryant H, et al. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *Lancet.* 2011;377(9760):127–138.
- Torrington ML, Frydenberg M, Hamilton W, Hansen RP, Lautrup MD, Vedsted P. Diagnostic interval and mortality in colorectal cancer: U-shaped association demonstrated for three different datasets. *J Clin Epidemiol.* 2012;65(6):669–678.
- Neal RD, Tharmanathan P, France B, et al. Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *Br J Cancer.* 2015;112(Suppl):S92–S107.
- Torrington ML, Frydenberg M, Hansen RP, Olesen F, Vedsted P. Evidence of increasing mortality with longer diagnostic intervals for five common cancers: a cohort study in primary care. *Eur J Cancer.* 2013;49(9):2187–2198.
- Richards MA, Smith P, Ramirez AJ, Fentiman IS, Rubens RD. The influence on survival of delay in the presentation and treatment of symptomatic breast cancer. *Br J Cancer.* 1999;79(5):858–864.
- Moller H, Gildea C, Meechan D, Rubin G, Round T, Vedsted P. Use of the English urgent referral pathway for suspected cancer and mortality in patients with cancer: cohort study. *BMJ.* 2015;351:h5102.
- Hamilton W. Cancer diagnosis in primary care. *Br J Gen Pract.* 2010;60(571):121–128.
- Nielsen TN, Hansen RP, Vedsted P. [Symptom presentation in cancer patients in general practice]. *Ugeskr Laeger.* 2010;172(41):2827–2831. Danish.
- Baughan P, Keatings J, O'Neill B. Urgent suspected cancer referrals from general practice: audit of compliance with guidelines and referral outcomes. *Br J Gen Pract.* 2011;61(592):700–706.
- Allgar VL, Neal RD. Delays in the diagnosis of six cancers: analysis of data from the national survey of NHS patients: cancer. *Br J Cancer.* 2005;92(11):1959–1970.
- Hansen RP, Vedsted P, Sokolowski I, Sondergaard J, Olesen F. Time intervals from first symptom to treatment of cancer: a cohort study of 2,212 newly diagnosed cancer patients. *BMC Health Serv Res.* 2011;11(1):284.
- Pedersen KM, Andersen JS, Sondergaard J. General practice and primary health care in Denmark. *J Am Board Fam Med.* 2012;25(Suppl 1):S34–S38.
- Meechan D, Gildea C, Hollingworth L, Richards MA, Riley D, Rubin G. Variation in use of the 2-week referral pathway for suspected cancer: a cross-sectional analysis. *Br J Gen Pract.* 2012;62(602):590–597.
- Praksisinformatoren RM. Kræftpakker, Region Midtjylland [Guidelines for fast track pathways in the Central Denmark Region]. Available from: <https://www.sundhed.dk/sundhedsfaglig/praksisinformatoren/almen-praksis/midtjylland/patientforloeb/kræftpakker/>. Accessed May 9, 2016.
- Pedersen CB. The Danish Civil Registration System. *Scand J Public Health.* 2011;39(7):22–25.
- Gjerstorff ML. The Danish Cancer Registry. *Scand J Public Health.* 2011;39(7):42–45.
- Toftegaard B, Bro F, Vedsted P. A geographical cluster randomised stepped wedge study of continuing medical education and cancer diagnosis in general practice. *Implement Sci.* 2014;9(1):159.
- Kirkwood BR, Sterne JAC, editors. Chapter 36: Measurement error: assessment and statistics. In: *Essential Medical Statistics*. 2nd ed. Blackwell Publishing; 2003:429–432.
- Hamilton W. The CAPER studies: five case-control studies aimed at identifying and quantifying the risk of cancer in symptomatic primary care patients. *Br J Cancer.* 2009;101(Suppl 2):S80–S86.
- Hamilton W, Peters TJ, Bankhead C, Sharp D. Risk of ovarian cancer in women with symptoms in primary care: population based case-control study. *BMJ.* 2009;339(959–535):b2998.
- Hamilton W, Kernick D. Clinical features of primary brain tumours: a case-control study using electronic primary care records. *Br J Gen Pract.* 2007;57(542):695–699.
- Hamilton W, Sharp DJ, Peters TJ, Round AP. Clinical features of prostate cancer before diagnosis: a population-based, case-control study. *Br J Gen Pract.* 2006;56(531):756–762.
- Hamilton W, Peters TJ, Round A, Sharp D. What are the clinical features of lung cancer before the diagnosis is made? A population based case-control study. *Thorax.* 2005;60(12):1059–1065.
- Hamilton W, Round A, Sharp D, Peters TJ. Clinical features of colorectal cancer before diagnosis: a population-based case-control study. *Br J Cancer.* 2005;93(4):399–405.
- Shephard EA, Stapley S, Neal RD, Rose P, Walter FM, Hamilton WT. Clinical features of bladder cancer in primary care. *Br J Gen Pract.* 2012;62(602):598–604.
- Hippisley-Cox J, Coupland C. Symptoms and risk factors to identify men with suspected cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract.* 2013;63(606):e1–e10.
- Hippisley-Cox J, Coupland C. Symptoms and risk factors to identify women with suspected cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract.* 2013;63(606):e11–e21.

Clinical Epidemiology**Publish your work in this journal**

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification,

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

systematic reviews, risk and safety of medical interventions, epidemiology and biostatistical methods, and evaluation of guidelines, translational medicine, health policies and economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Dovepress