Development and validation of a clinical prediction rule for acute appendicitis in children in primary care

Blok, Guus; Burger, Huib; van der Lei, Johan; Berger, Marjolein; Holtman, Gea

Link to publication in University of Groningen/UMCG research database

Development and validation of a clinical prediction rule for acute appendicitis in children in primary care

Guus Blok, Huib Burger, Johan van der Lei, Marjolein Berger & Gea Holtman

View supplementary material

Published online: 14 Aug 2023.

Submit your article to this journal

Article views: 118

View related articles

View Crossmark data

RESEARCH ARTICLE

Development and validation of a clinical prediction rule for acute appendicitis in children in primary care

Guus Blok[a] , Huib Burger[a] , Johan van der Lei[b] , Marjolein Berger[a] and Gea Holtman[a]

[a]Department of General Practice and Elderly Care Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; [b]Department of Medical Informatics, Erasmus Medical Center, Rotterdam, The Netherlands

KEY MESSAGES

- For GPs, recognising appendicitis in children with acute abdominal pain is challenging and clinical prediction rules are unavailable.
- We developed and validated a clinical prediction rule to stratify these children's risk of acute appendicitis.
- For the three risk groups, we suggest parental observation, further assessment and direct referral.

ABSTRACT

Background: Recognising acute appendicitis in children presenting with acute abdominal pain in primary care is challenging. General practitioners (GPs) may benefit from a clinical prediction rule.

Objectives: To develop and validate a clinical prediction rule for acute appendicitis in children presenting with acute abdominal pain in primary care.

Methods: In a historical cohort study data was retrieved from GP electronic health records included in the Integrated Primary Care Information database. We assigned children aged 4–18 years presenting with acute abdominal pain ($\leq$ 7 days) to development (2010–2012) and validation (2013–2016) cohorts, using acute appendicitis within six weeks as the outcome. Multiple logistic regression was used to develop a prediction model based on predictors with $>$ 50% data availability derived from existing rules for secondary care. We performed internal and external temporal validation and derived a point score to stratify risk of appendicitis into three groups, i.e. low-risk, medium-risk and high-risk.

Results: The development and validation cohorts included 2,041 and 3,650 children, of whom 95 (4.6%) and 195 (5.3%) had acute appendicitis. The model included male sex, pain duration ($<$24, 24–48, $>$ 48 h), nausea/vomiting, elevated temperature ($\geq$ 37.3 °C), abnormal bowel sounds, right lower quadrant tenderness, and peritoneal irritation. Internal and temporal validation showed good discrimination (C-statistics: 0.93 and 0.90, respectively) and excellent calibration. In the three groups, the risks of acute appendicitis were 0.5%, 7.5%, and 41%,

Conclusion: Combined with further testing in the medium-risk group, the prediction rule could improve clinical decision making and outcomes.

Introduction

Acute appendicitis in children has a population incidence of approximately 0.12% per year, with 20%–74% of cases developing potentially life-threatening complications [1]. The general practitioner (GP) must not only recognise cases and refer them to secondary care promptly but also avoid unnecessary referrals for children with less harmful conditions, such as gastroenteritis or constipation. This balance is especially challenging for the GP because individual clinical features of appendicitis have low predictive value [2–4].

Although guidelines exist to help with decisions about referral [2, 3], these specify about 20–40 clinical features related to acute appendicitis and a myriad of

other abdominal diseases to be considered. Moreover, we have reported that GPs record only a few clinical features after assessing children with acute abdominal pain and that these vary widely [4]. GPs instead use heuristic decision-making [5], relying on their clinical judgement rather than formal algorithms or complete investigations of all features. However, prediction rules can support clinical assessment by using optimally weighed predictive variables for risk stratification [6]. Although several clinical prediction rules have been developed and validated for diagnosing appendicitis in children in secondary care, discrimination and calibration in primary care may differ. This is caused by selective referral of patients to secondary care with more severe disease and a higher probability of disease [7–9]. In this study, we first develop and validate a clinical prediction rule for children with acute abdominal pain in primary care, using registration data [10, 11]. Second, we classify risk groups based on preset diagnostic accuracy criteria and report the observed probabilities of acute appendicitis for each group.

Methods

Study design

We performed a historical cohort study of children presenting with acute abdominal pain to primary care in the Netherlands between November 2010 and November 2016 [4]. The cohort was split into development and validation cohorts based on data obtained before and after November 2013, respectively. Reporting follows the TRIPOD statement [12].

Data source

We used medical records from the longitudinal International Primary Care Information (IPCI) database managed by Erasmus Medical Centre, Rotterdam, which contains complete pseudonymised data for 1.5 million patients from approximately 600 Dutch GP practices [13]. All Dutch residents must enlist with a GP, who must code (or record as free text) all clinical data in electronic patient records. Medical specialists must also report all patient contacts to GPs, irrespective of the referral source. The IPCI database contains data from six GP software platforms, and for this study, we used the three platforms with the most complete specialist reports, accounting for 32.9% of the total (368 practices). The database complies with European Union guidelines on using medical data for research.

Study population

GP consultation records for children aged 4–18 years with 'abdominal pain' noted in the free text and an International Classification of Primary Care (ICPC) code for the gastrointestinal tract (D.xx) were identified automatically. Index consultations were those for children who presented to their GP for the first time with acute abdominal pain ($\leq$ 7 days) during the study period, excluding cases with a history of prior appendicitis or appendectomy.

Outcome

Presence of appendicitis was based on imaging, intraoperative findings and/or histopathology as reported by the secondary care specialist. Absence of appendicitis was either based on the GP's medical records within six weeks from the index consultation or on the specialist reports, if available. Four medical students extracted outcome data from free text.

Candidate predictors

Thirteen candidate predictors were derived from existing prediction rules: male sex, right lower quadrant (RLQ) pain, RLQ pain migration, pain intensity, pain duration, increasing pain, nausea and/or vomiting, anorexia, RLQ tenderness, peritoneal irritation (rebound tenderness, guarding, abdominal wall rigidity, and/or pain at jarring movements), abnormal bowel sounds (absent, tinkling or high-pitched), elevated temperature ($\geq$ 37.3 °C by recall or measure), and difficulty walking [11]. All predictors were binary, except pain duration ($<$ 24, 24–48 h, $>$ 48 h). Using standardised criteria, the coders extracted the candidate predictors from the free text, as assessed during the index consultation [4, 14]. Demographic data was retrieved automatically from the patients' records. Candidate predictors were only included in the model when recorded in $>$ 50% of consultations, whether positive or negative [6].

Missing data

We considered items missing at random (MAR) if the observed values of other predictors or outcomes could predict missingness [15]. For data presumed MAR or missing completely at random, we used multiple imputation to replace missing values. Based on all known predictors, demographic data, referral data, and outcomes, 20 datasets were constructed with the missing data imputed. Acute appendicitis and

perforation of the appendix were included as predictive variables in the multiple imputation. We used predictive mean matching for imputation because the incomplete variables were not normally distributed. As a sensitivity analysis, we performed a complete case analysis in the development cohort [15].

Statistical analysis

Model development and validation.
In the development cohort, we fitted a logistic regression model with acute appendicitis as the dependent variable and all candidate predictors as independent variables. Backward selection was used to identify predictors meeting the Akaike information criterion ($p < 0.157$) in over 50% of the 20 imputed datasets. For internal validation, we performed bootstrapping with 250 iterations to obtain an overfitting-adjusted concordance index (C-statistic) and calculate a shrinkage factor for uniform application to the beta coefficients to obtain an adjusted slope. Final model calibration was achieved by adjusting the intercept [16]. Model performance was assessed by using the C-statistic pooled across imputed data sets for the discrimination and calibration curves constructed separately for each imputed set.

External temporal validation of the final model involved using data from the validation cohort to calculate the C-statistic and calibration curves. The model was accepted without updating if the results were satisfactory.

Subgroup analysis by age.
Because the differential diagnosis of acute abdominal pain differs between pre- and post-pubertal children [11], we conducted two posthoc analyses in the development cohort. First, by adding age group to the other candidate predictors, we assessed whether age group (4–10 and 11–18 years) independently predicted acute appendicitis. Second, we assessed the interaction between the linear predictor of the final model and age group. To preserve any potential interaction with age, missing predictor variables were imputed separately for each age group [17].

Score development.
We constructed a clinical practice score based on the final model as the sum of points for all predictors. This involved assigning points to each predictor equal to its beta coefficient divided by the absolute lowest beta coefficient and rounded this to the nearest integer [18]. We plotted the predicted probability of acute appendicitis for each point score.

Performance.
Model performance was further assessed in the validation cohort by calculating sensitivity, specificity, positive predictive value, and negative predictive value with binomial proportion 95% confidence intervals (95%CIs) at all possible cut-off values. We report observed numbers of children with and without acute appendicitis at different cut-off values in the validation cohort. Parameter estimates were pooled using Rubin's rules, which were also used to obtain 95%CIs [19].

Additionally, we constructed a percentile curve showing the proportion of children below each point score in the validation cohort. Benchmarking was done by calculating the test characteristics and the probability of appendicitis based on the GP's decision to refer during the index consultation in the validation cohort. Finally, we built an online calculator for estimating the theoretical probability of acute appendicitis [18].

Risk groups.
Aiming for arbitrarily defined sensitivity of 95% and specificity of 95%, we defined two cut-offs for the score. The lower cut-off corresponded with 95% model sensitivity and the upper cut-off with 95% model specificity. The two cut-offs were used to assign children to the low-risk, medium-risk or high risks groups. Referral rates by the GP within the risk groups were assessed. Analyses were performed in IBM SPSS version 26.0 (IBM Corp, Armonk, NY, USA) and R version 4.0.4.

Results

Participants

During the study period, 15,607 children presented to their GP and 5,691 met the inclusion criteria. Of these, we included 2,041 in the development cohort (referred: 84 of 96 with acute appendicitis and 196 without appendicitis) and 3,650 in the validation cohort (referred: 152 of 195 with acute appendicitis and 366 without appendicitis). Table 1 shows the broadly comparable characteristics between the cohorts.

Included candidate predictors

Table 1 shows that seven candidate predictors met the criterium of > 50% available values. These were used for analysis: male sex, pain duration, nausea/vomiting, elevated temperature, abnormal bowel sounds, RLQ tenderness, and peritoneal irritation.

Table 1. Availability and values of predictors and outcomes in the development and validation cohorts.

	Development cohort		Validation cohort	
	Number Available (%)	Number positive (%)	Number Available (%)	Number positive (%)
Demographic characteristics				
Age, median (Q1, Q3) (years)	2041 (100)	10 (7, 14)	3650 (100)	10 (7, 14)
Male, n (%)	2041 (100)	951 (47)	3650 (100)	1667 (46)
Symptoms				
Pain duration				
<24 hrs, n (%)	1492 (73)	476 (32)	2813 (77)	985 (35)
24–48 hrs, n (%)	1492 (73)	182 (12)	2813 (77)	290 (10)
≥48 hrs, n (%)	1492 (73)	834 (56)	2813 (77)	1538 (55)
Pain severity, n (%)	233 (11)			
Mild-to-moderate, n (%)		36 (15)	681 (19)	197 (29)
Severe, n (%)		197 (85)	681 (19)	484 (71)
Migration of pain, n (%)	65 (3)	58 (89)	126 (3)	117 (93)
Increased pain, n (%)	198 (10)	150 (76)	459 (13)	370 (81)
Pain RLQ, n (%)	546 (27)	320 (59)	1103 (30)	664 (60)
Anorexia, n (%)	515 (25)	298 (58)	1001 (27)	591 (59)
Nausea/Vomiting, n (%)	1176 (58)	799 (68)	2245 (62)	1525 (68)
Signs				
Abnormal Bowel Sounds, n (%)	1160 (57)	154 (13)	2289 (63)	306 (13)
Peritoneal Irritation, n (%)	1436 (70)	294 (21)	2695 (74)	993 (21)
Elevated temperature (≥ 37.3 °C), n (%)	1140 (56)	370 (33)	2319 (64)	781 (34)
Tenderness RLQ, n (%)	1115 (55)	560 (50)	2037 (56)	993 (49)
Difficulty walking, n (%)	132 (6)	73 (55)	283 (8)	174 (61)
Appendicitis, n (%)	2041 (100)	96 (5)	3650 (100)	195 (5)

Number available = number of children with the characteristic recorded. Number positive = number of children with the characteristic (with percentage relative to Number available). Median age is reported.
Abbreviations: Q1, first quartile; Q3, third quartile; RLQ, right lower quadrant.
Only predictors with >50% data availability were considered as candidate predictors for the model.

Missing data

The missing data analysis in Supplementary Table S1 shows that children with peritoneal irritation (odds ratio, 0.62; 95%CI, 0.45–0.84) and acute appendicitis (odds ratio, 0.57; 95%CI, 0.34–0.96) were less likely to have missing values of the other predictors. Thus, we assumed the data were at least partially MAR.

Development of the prediction model

All seven candidate predictors were independently associated with an increased risk of acute appendicitis and were included in the initial model (Table 2), which had an excellent discriminatory power (pooled C-statistic, 0.93; 95%CI, 0.92–0.95). The complete case sensitivity analysis produced a C-statistic of 0.96 (95%CI, 0.92–1.0; Supplementary Table S2) and the calibration curves showed good calibration for all 20 datasets and complete case analysis (Supplementary Figure S1).

Subgroup analysis by age

In the development cohort, backward selection led to age group (4–10 versus 11–18 years old) being deselected in all imputed datasets. No interaction effect existed between the linear predictor and age group (p for interaction: 0.15).

Internal and external validation

Table 2 shows the final model after adjusting the initial beta coefficients with a shrinkage factor of 0.95. The C-statistic was 0.93 (95%CI, 0.92–0.95) and the calibration curves showed good calibration. In the external temporal validation, the C-statistic decreased to 0.90 (95%CI, 0.89–0.92) (Supplementary Figure S2) and the calibration curves again showed good calibration (Figure 1). Therefore, the final model did not require updating and was used to build the online calculator available at https://jscalc.io/calc/YKeGYmYoeBSgK7FC for demonstration purposes.

Performance

Point scores ranged from 0 to 18, corresponding with model-based predicted probabilities of acute appendicitis from 0.05% to 85% (Figure 2). In the validation cohort, 29.3% of children referred and 1.4% of children not referred had appendicitis. Agreement between the classifications based on the point score and the GP's decision to refer was maximal at a cut-off value ≥11 points (Table 3), with 24.8% above and 1.5% below this cut-off having appendicitis.
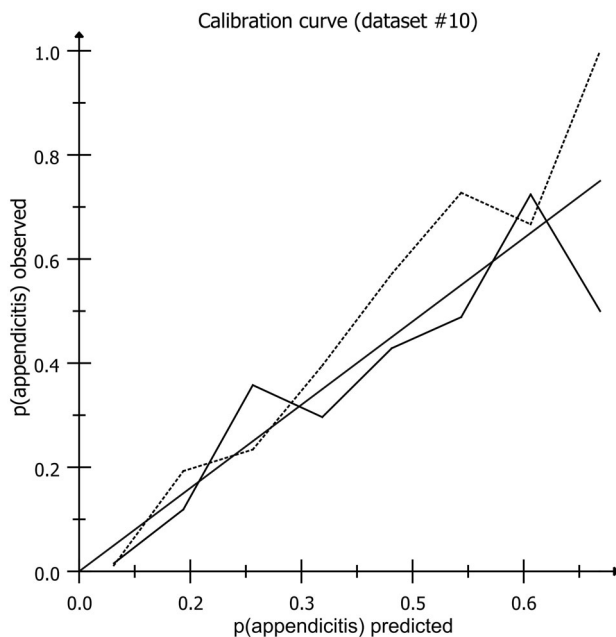
Risk groups

Using two cut-offs corresponding to a 95% sensitivity and 95% specificity, we categorised the point scores

Table 2. Adjusted model after internal validation, with point scores for the prediction rule.
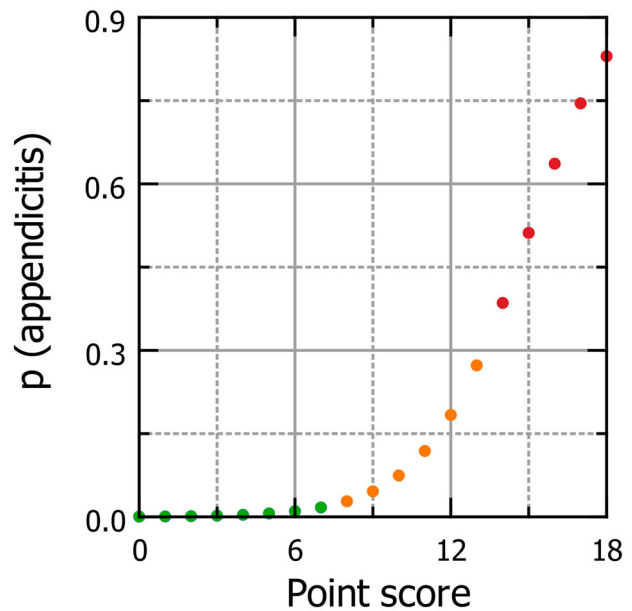
Predictor	Adjusted Beta coefficient	Point score
Demographic characteristics		
Male sex	1.27	2
Symptoms		
Nausea vomiting	1.17	2
Pain duration (<24 hrs)	0.82	1
Pain duration (24–48 hrs)	1.19	2
Signs		
Elevated temperature ($\geq 37.3\,°C$)	0.88	1
Abnormal bowel sounds	0.61	1
RLQ tenderness	2.76	5
Peritoneal irritation (abdominal rigidity, guarding, rebound tenderness and/or pain at jarring motions)	2.27	4
Intercept	−8.96	

The development cohort was used to develop and adjust the model. Beta coefficients of predictors were adjusted by bootstrapping (shrinkage factor: 0.95), and the intercept was adjusted using an offset procedure. Points were assigned by dividing al Beta coefficients by the smallest Beta coefficient and rounding to the nearest integer. The total score is defined as the sum of the points assigned to each feature. Abbreviations: RLQ, right lower quadrant.

Figure 2. Predicted probability of acute appendicitis by point score the predicted probability of acute appendicitis in the low-risk (green), medium-risk (orange) and high-risk (red) groups, based on the final model, is juxtaposed against the point score. For example, in a boy (2 points) with fever (1 point), vomiting (2 points), pain duration between 24 and 48 h (2 points), absent bowel sounds (1 point), and peritoneal irritation (4 points), but without right lower quadrant tenderness, the point score is 11. As the graph shows, this corresponds to a predicted probability of acute appendicitis of 0.12.

Figure 1. Calibration plot showing the agreement between expected and observed probabilities in the development cohort (dotted line) and the validation cohort (solid line). the best model's results would align closely with the diagonal line X = Y, were the observed numbers are equal to the predicted numbers. The data are based on predictions by the final model applied in the development and validation cohorts in two randomly chosen imputed datasets (#10 of each cohort).

and defined low risk as ≤ 7, medium risk as 8–13, and high risk as ≥ 14 with observed probabilities of appendicitis of 0.5% (95%CI, 0.2–0.9), 7.5% (95%CI, 6.2–9.0), and 41.0% (95%CI, 34.0–48.3), respectively. Of the 3,650 children in the validation cohort, 2,053 (56%) had low risk, 1,402 (38%) had medium risk, and 195

(5%) had high risk (Supplementary Figure S3); and, of the 195 children with acute appendicitis, 10 (5%, as pre-defined) had low risk, 105 (54%) had medium risk, and 80 (41%) had high risk. In the low, medium and high-risk groups, 2.9% (95% CI: 2.3% − 3.5%), 27.1% (24.9% − 29.4%) and 71.5% (65.9% − 77.2%) of children were referred by the GP during the index consultation, respectively.

Discussion

Main findings

Our clinical prediction rule uses seven easily assessed predictors of appendicitis in children with acute abdominal pain: male sex, pain duration, nausea/vomiting, elevated temperature, abnormal bowel sounds, RLQ tenderness, and peritoneal irritation. The rule can stratify children presenting with acute abdominal pain in primary care into three risk groups. Among the children, 56% were classified as low risk, 38% as medium risk, and 5% as high risk, with corresponding probabilities of appendicitis estimated at 0.5%, 7.5%, and 41.0%, respectively. Calibration and discrimination

Table 3. Diagnostic performance and observed diagnoses at each point score cut-off in the validation cohort.

| | Diagnostic Performance | | | | Observed numbers | | | |
| | | | | | ≥ Point score cut-off | | < Point score cut-off | |
Point Score	Sensitivity	Specificity	PPV	NPV	Appendicitis (TP)	No appendicitis (FP)	No appendicitis (TN)	Appendicitis (FN)
Low risk								
1	1.00 (1.00–1.00)	0.04 (0.04–0.05)	0.06 (0.05–0.06)	1.00 (1.00–1.00)	195	3307	148	0
2	1.00 (1.00–1.00)	0.09 (0.08–0.10)	0.06 (0.05–0.06)	1.00 (1.00–1.00)	195	3154	301	0
3	0.99 (0.98–1.00)	0.21 (0.19–0.22)	0.07 (0.06–0.07)	1.00 (1.00–1.00)	194	2739	716	1
4	0.98 (0.96–0.99)	0.32 (0.31–0.34)	0.08 (0.07–0.08)	1.00 (0.99–1.00)	191	2338	1117	4
5	0.97 (0.96–0.99)	0.42 (0.41–0.44)	0.09 (0.08–0.10)	1.00 (0.99–1.00)	190	1988	1467	5
6	0.96 (0.94–0.98)	0.52 (0.51–0.54)	0.10 (0.09–0.11)	1.00 (0.99–1.00)	188	1647	1808	7
7	0.95 (0.93–0.97)	0.59 (0.58–0.61)	0.12 (0.10–0.13)	1.00 (0.99–1.00)	185	1412	2043	10
Medium risk								
8	0.93 (0.91–0.96)	0.68 (0.66–0.69)	0.14 (0.12–0.16)	0.99 (0.99–1.00)	182	1120	2335	13
9	0.89 (0.85–0.92)	0.76 (0.74–0.77)	0.17 (0.15–0.19)	0.99 (0.99–0.99)	173	847	2609	22
10	0.84 (0.79–0.88)	0.82 (0.81–0.83)	0.21 (0.18–0.23)	0.99 (0.99–0.99)	163	621	2833	32
11	0.77 (0.72–0.81)	0.87 (0.86–0.88)	0.25 (0.22–0.28)	0.99 (0.98–0.99)	150	456	2998	46
12	0.70 (0.65–0.75)	0.91 (0.90–0.91)	0.30 (0.26–0.33)	0.98 (0.98–0.99)	136	326	3129	59
13	0.63 (0.57–0.68)	0.94 (0.93–0.95)	0.37 (0.32–0.41)	0.98 (0.97–0.98)	122	213	3243	73
High risk								
14	0.41 (0.35–0.47)	0.97 (0.96–0.97)	0.41 (0.35–0.47)	0.97 (0.96–0.97)	80	115	3340	115
15	0.26 (0.21–0.31)	0.99 (0.98–0.99)	0.55 (0.46–0.64)	0.96 (0.95–0.96)	50	42	3413	145
16	0.10 (0.06–0.13)	1.00 (1.00–1.00)	0.67 (0.52–0.83)	0.95 (0.95–0.96)	19	9	3446	176
GP	0.76 (0.70–0.82)	0.90 (0.89–0.91)	0.29 (0.26–0.34)	0.99 (0.98–0.99)	149	357	3098	46

Data show the diagnostic performance of the model and observed numbers for each point score, with the GP's decision to refer during index consultations in the validation cohort (3650 children: 195 with acute appendicitis). The numbers for the estimates were averaged over 20 imputed datasets and rounded to the nearest integer. Abbreviations: FN, false negative; FP, false positive; NPV, negative predictive value; PPV, positive predictive value; TN, true negative; TP, true positive.

of the model was good, with a C-statistic of 0.90 at external validation.

Strengths and limitations

This study benefitted from a large cohort with an adequate number of patients with acute appendicitis (96 of 2041 in the development cohort and 195 of 3,650 in the validation cohort), allowing for over 10 events per variable [8]. In fact, the number of events exceeded the minimum of 100 and approached the more ideal 200 recommended for external validation in simulation studies [20]. Although new methods have become available since designing the study and collecting the data [21], we decided against applying elaborate posthoc sample size calculations. The successful temporal external validation, with good calibration and discrimination, represents a strength of this study, although the validation was narrowed by using a similar cohort, limiting generalisability. External validation in different datasets, such as prospective studies or studies conducted in other countries, could yield less favourable results and future studies are required to achieve broad validation and generalisability [6].

Using routine healthcare data resulted in many unavailable data on predictors, with only 7 of the 13 candidate predictors being sufficiently available [6]. Predictors were selected as candidate predictors only when > 50% of their values were available, based on

judgement as strict criteria are lacking [6]. However, we assume each GP will record the most important predictors to their decision [4]. Information bias and problems in selecting a representative cohort may also limit research using registration data [22]. To reduce these biases, we used a combination of free text, diagnostic codes, and structured extraction forms. Additionally, defining the absence of acute appendicitis based on an observation period may have introduced differential verification bias. However, this approach ensured that no children with clinically manifest appendicitis will have gone unnoticed, which makes differential verification bias less likely [24]. Although information and differential verification biases can lead to test characteristics being overestimated [22], the study predictors were assessed according to daily practice, which is recognised to result in realistic accuracy estimates [23].

Comparison with existing literature

Several clinical prediction rules exist for estimating the probability of appendicitis in secondary care, but none are validated in primary care. Moreover, we are aware of no other research into developing or validating a clinical prediction rule for acute appendicitis in primary care [11]. Our study is per existing recommendations stating that the setting affects how prediction models perform; existing secondary care prediction models need to be

validated and modified for primary care [9, 18]. Unfortunately, the limited availability of predictor values in our cohort and the inclusion of white blood counts in most existing models meant that we could not directly compare our rule with these models. Although age is reported to affect clinical presentation, it did not influence the interpretation or selection of predictors in this study [11].

Clinical prediction rules for acute appendicitis in children have been studied extensively in secondary care, producing C-statistics of 0.61–0.85, sensitivities of 0.72–1.00, and specificities of 0.34–0.98 [10, 11, 24]. External temporal validation of the model in this study produced a C-statistic of 0.90 (95%CI, 0.89–0.92). A decrease in the discriminative value of diagnostic tests in secondary care when patients with positive test results are likely to be referred could account for this difference [25]. In addition, the baseline prevalence (prior) was lower in primary care compared to secondary care (5% vs 27%) [11]. Furthermore, low and high cut-offs have been presented for different prediction rules for appendicitis, with sensitivity at low cut-offs ranging from 88.1 to 97.2 and specificity at high cut-offs ranging from 70.1 to 98.9 as compared to the pre-defined value of 95% for both sensitivity and specificity in our study [26].

Implications for research and practice

GPs miss approximately 19% of children with appendicitis and refer 69% to other, mainly self-limiting, conditions [4]. The developed rule could quickly evaluate symptoms and signs to support the structured evaluation of children with acute abdominal pain. It could serve as a memory aid for the GP to ensure that the most predictive clinical features of acute appendicitis are evaluated, making care more homogeneous.

Risk stratification could offer a more efficient approach to usual care by reducing the number of referrals without missing more cases of acute appendicitis. This may include using C-reactive protein testing, which has most impact on the decision process in those at medium risk [14]. CRP point-of-care testing has become increasingly available in primary care in many European countries and its introduction in the UK has been advocated [27]. We did not consider CRP for the prediction model because a CRP-test is only appropriate in children with a medium risk of appendicitis [14].

Thus, when using the prediction rule in practice, we suggest parental observation for children at low risk (0.5%), further testing and/or reassessment by the GP for those at medium risk (7.5%), and direct referral to secondary care for those at high risk (41.0%) or for those at medium risk when the suspicion is high. This diagnostic strategy may potentially decrease the referrals of children without acute appendicitis without delaying a diagnosis of appendicitis compared with current GPs' practice. However, it is recommended that GPs should combine the probabilistic data obtained from our tool with their clinical judgement in every case. Further research could investigate factors influencing the decision to refer, given the 2.9% referral rate for low-risk children and the 28.5% non-referral rate for high-risk children. Given that other diagnoses are usually considered when assessing the need for referral, clinical impact studies must now evaluate the impact of our proposed diagnostic strategy on clinical decision-making and patient outcomes in primary care [18].

Conclusion

Using registration data in a historical cohort study, we developed a clinical prediction rule for acute appendicitis in children with acute abdominal pain in primary care. The underlying model's performance was good at internal and temporal validation. We suggest a diagnostic strategy with risk stratification based on this clinical prediction rule. Clinical-impact studies are warranted to evaluate its effect on clinical decision-making and patient outcomes in primary care.

Disclosure statement

The authors alone are responsible for the content and writing of the paper.

Ethical approval

The Medical Ethics Review Committee (METc) of the University Medical Centre Groningen approved this study (METc2017/541) and concluded that this study did not fall under the scope of the Dutch *Medical Research Involving Human Subjects* Act (WMO).

ORCID

Guus Blok http://orcid.org/0000-0003-1654-6844
Huib Burger http://orcid.org/0000-0001-6484-0349

Johan van der Lei 🔘 http://orcid.org/0000-0002-9203-3410
Marjolein Berger 🔘 http://orcid.org/0000-0001-6774-5961
Gea Holtman 🔘 http://orcid.org/0000-0001-6579-767X

References

[1] Kumar JS, Jaya Kumar S. Trends in incidence of acute appendicitis in children. Acad J Ped Neonatol. 2017; 3(5):1–5. doi:10.19080/AJPN.2017.03.555682.

[2] Eizenga WH, Gieteling MJ, Berger MY, et al. Summary of the NHG guideline "abdominal pain in children", the 100th NHG guideline. Ned Tijdschr Geneeskd. 2013; 157(15):A6191. http://www.ncbi.nlm.nih.gov/pubmed/23575296

[3] National Institute for Health and Care Excellence. Appendicitis. Clinical Knowledge Summary. Published 2015. [cited 2019 November 20]. Available form: https://cks.nice.org.uk/appendicitis#!topicSummary.

[4] Blok CGH, Veenstra LMM, van der Lei J, et al. Appendicitis in children with acute abdominal pain in primary care, a retrospective cohort study. Fam Pract. 2021;38(6):758–765. doi:10.1093/fampra/cmab039.

[5] Marewski JN, Gigerenzer G. Heuristic decision making in medicine. Dialog Clin Neurosci. 2012;14(1):77–89. doi:10.31887/dcns.2012.14.1/jmarewski.

[6] Steyerberg E. Clinical prediction models. New York (NY): Springer-Verlag; 2009. doi:10.1007/978-0-387-77244-8.

[7] Craig S, Dalton S. Diagnosing appendicitis: what works, what does not and where to go from here? J Paediatr Child Health. 2016;52(2):168–173. doi:10.1111/jpc.12998.

[8] Riley RD, Ensor J, Snell KIE, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016;353(mar31_1):i3140. doi:10.1136/bmj.i3140.

[9] Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51–58. doi:10.7326/M18-1376/SUPPL_FILE/M18-1376_SUPPLEMENT.PDF.

[10] Kulik DM, Uleryk EM, Maguire JL. Does this child have appendicitis? A systematic review of clinical prediction rules for children with acute abdominal pain. J Clin Epidemiol. 2013;66(1):95–104. doi:10.1016/j.jclinepi.2012.09.004.

[11] Nepogodiev D, Wilkin RJ, Bradshaw CJ, et al. Appendicitis risk prediction models in children presenting with right iliac fossa pain (RIFT study): a prospective, multicentre validation study. Lancet Child Adolesc Heal. 2020;4(4):271–280. doi:10.1016/S2352-4642(20)30006-7.

[12] Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMC Med. 2015;13:1. doi:10.1186/s12916-014-0241-z.

[13] de Ridder MAJJ, de Wilde M, de Ben C, et al. Data resource profile: the integrated primary care information (IPCI) database, The Netherlands. Int J Epidemiol. 2022;51(6):e314–e323. doi:10.1093/IJE/DYAC026.

[14] Blok GCGH, Nikkels ED, van der LJ, et al. Added value of CRP above clinical features for appendicitis in children. Eur J Gen Pract. 2022;28(1):1–26. doi:10.1080/13814788.2022.2067142.

[15] Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. Stat Methods Med Res. 2014;23(5):440–459. doi:10.1177/0962280213476378.

[16] Heymans MW, Anema JR, van Buuren S, et al. Return to work in a cohort of low back pain patients: development and validation of a clinical prediction rule. J Occup Rehabil. 2009;19(2):155–165. doi:10.1007/s10926-009-9166-3.

[17] Tilling K, Williamson EJ, Spratt M, et al. Appropriate inclusion of interactions was needed to avoid bias in multiple imputation. J Clin Epidemiol. 2016;80:107–115. doi:10.1016/j.jclinepi.2016.07.004.

[18] Cowley LE, Farewell DM, Maguire S, et al. Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature. Diagn Progn Res. 2019;3(1):1–23. doi:10.1186/s41512-019-0060-y.

[19] Marshall A, Altman DG, Holder RL, et al. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Med Res Methodol. 2009;9:57. doi:10.1186/1471-2288-9-57.

[20] Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Stat Med. 2016;35(2):214–226. doi:10.1002/SIM.6787.

[21] van Smeden M, Moons KGM, de Groot JAH, et al. Sample size for binary logistic prediction models: beyond events per variable criteria. Stat Methods Med Res. 2019;28(8):2455–2474. doi:10.1177/0962280218784726.

[22] Oostenbrink R, Moons KGM, Bleeker SE, et al. Diagnostic research on routine care data: prospects and problems. J Clin Epidemiol. 2003;56(6):501–506. doi:10.1016/S0895-4356(03)00080-5.

[23] Grobbee DE, Arno WH. Clinical Epidemiology Principles, Methods and Applications for Clinical Research. 2015.

[24] van Amstel P, Gorter RR, van der Lee JH, et al. Ruling out appendicitis in children: can We use clinical prediction rules? J Gastrointest Surg. 2019;23(10):2027–2048. doi:10.1007/s11605-018-3997-1.

[25] Sackett DL, Haynes RB. Evidence base of clinical diagnosis: the architecture of diagnostic research. BMJ. 2002;324(7336):539–541. doi:10.1136/bmj.324.7336.539.

[26] Gudjonsdottir J, Marklund E, Hagander L, et al. Clinical prediction scores for pediatric appendicitis. Eur J Pediatr Surg. 2021;31(3):252–260. doi:10.1055/s-0040-1710534.

[27] John AS, Hopstaken R, Tirimacco R, et al. Analysis implementing point-of-care CRP testing for better diagnosis of acute respiratory infections. Br J Gen Pract. 2022; 72(715):87–88. doi:10.3399/bjgp22X718517.