

ORIGINAL ARTICLES

A diagnostic model for the detection of sensitization to wheat allergens was developed and validated in bakery workers

Eva Suarthana^{a,b,c,*}, Yvonne Vergouwe^b, Karel G. Moons^b, Jan de Monchy^d,
Diederick Grobbee^b, Dick Heederik^a, Evert Meijer^a

^aEnvironmental Epidemiology Division, Institute for Risk Assessment Sciences (IRAS), Utrecht University, 3508 TD, Utrecht, The Netherlands

^bJulius Center for Health Sciences and Primary Care, University Medical Center Utrecht, 3508 GA, Utrecht, The Netherlands

^cCommunity Medicine Department, Faculty of Medicine, University of Indonesia, JL. Pegangsaan Timur 16, Jakarta Pusat 10320, Indonesia

^dDepartment of Allergology, University Hospital Groningen, 9700 RB Groningen, The Netherlands

Accepted 12 October 2009

Abstract

Objectives: To develop and validate a prediction model to detect sensitization to wheat allergens in bakery workers.

Study Design and Setting: The prediction model was developed in 867 Dutch bakery workers (development set, prevalence of sensitization 13%) and included questionnaire items (candidate predictors). First, principal component analysis was used to reduce the number of candidate predictors. Then, multivariable logistic regression analysis was used to develop the model. Internal validation and extent of optimism was assessed with bootstrapping. External validation was studied in 390 independent Dutch bakery workers (validation set, prevalence of sensitization 20%).

Results: The prediction model contained the predictors nasojunctival symptoms, asthma symptoms, shortness of breath and wheeze, work-related upper and lower respiratory symptoms, and traditional bakery. The model showed good discrimination with an area under the receiver operating characteristic (ROC) curve area of 0.76 (and 0.75 after internal validation). Application of the model in the validation set gave a reasonable discrimination (ROC area = 0.69) and good calibration after a small adjustment of the model intercept.

Conclusion: A simple model with questionnaire items only can be used to stratify bakers according to their risk of sensitization to wheat allergens. Its use may increase the cost-effectiveness of (subsequent) medical surveillance. © 2010 Elsevier Inc. All rights reserved.

Keywords: Baker; Diagnostic model; Occupational allergic disease; Validity; Wheat; Sensitization; Wheat allergens

1. Introduction

In the past few decades, occupational allergic diseases (OADs) including occupational asthma (OA) have been shown to be a major respiratory problem. Many OA cases are related to exposure to high-molecular weight (HMW) allergens, including wheat flour and fungal α -amylase allergens in bakery and flour producing industries [1–3]. Although highly exposed individuals are more likely to have serious complaints and disability [4], workers with intermittent or occasional exposures may also be affected [5]. Recent studies suggest that there is no reason to assume that the burden of OAD is decreasing. So, prevention should focus on finding new cases, detecting subclinical illness, and early intervention in already asthmatic workers [1,3].

A genuine diagnosis of OAD, such as OA, can only be made in a particular person at a particular time in a clinical setting. Therefore, identification of OAD in bakers exposed to flour allergens should not focus on clinically established allergic disease but on highly associated preliminary symptoms and signs. Sensitization to allergens is an outcome strongly associated with OA and the most appropriate characteristic that can easily be investigated. Sensitized subjects, when continuously exposed to such allergens, are at high risk of developing occupational allergy outcomes [6]. A logical approach is therefore to first identify sensitized workers. After detecting workers with a high risk of wheat sensitization, additional clinical investigations may be restricted to this group, leaving a considerable number of workers with a low risk in which no or less far reaching medical investigations are needed [7].

For that reason, a diagnostic questionnaire model was developed to estimate the individual probability of the presence of sensitization to wheat allergens. This model enables objective and standardized quantification of the individual

* Corresponding author. Environmental Epidemiology Division, Institute for Risk Assessment Sciences (IRAS), Utrecht University, 3508 TD, Utrecht, The Netherlands. Tel.: +31-30-2539480; fax: +31-30-2539499.
E-mail address: eva.suarthana@gmail.com (E. Suarthana).

What is new?

A simple diagnostic questionnaire model for detection of sensitization to wheat allergens was developed and validated. This model will enable stratification of risk of sensitization to wheat allergens among bakers.

probability of wheat sensitization in a cost-effective manner by avoiding as much as possible the use of (invasive) advanced and costly reference tests. With the development of a diagnostic model for sensitization to wheat allergens, the authors aimed to easily detect individuals with a high probability of having OAD [8,9].

The accuracy of prediction rules is generally lower compared with the accuracy in the data on which the prediction rule is developed [10,11]. Therefore, the authors evaluated whether the developed diagnostic model performed accurately in new workers, and finally whether the performance of the diagnostic model could be improved (updated) before reliable implementation will be accepted in daily care use by occupational physicians in medical surveillance programs. The so obtained final diagnostic model was then simplified into an easy-to-use scoring rule to enhance its use in practice.

2. Methods

2.1. Populations

Data from a study performed in the framework of a National Occupational Respiratory Allergy Surveillance Program among bakery workers in The Netherlands were used to develop the diagnostic model (development set) [12]. Between June 2005 and June 2006, 341 and 28

randomly selected traditional and industrial bakeries, respectively were approached. From these companies, 64.5% (238 of 369) participated in the study, comprising 1,249 workers. Twelve bakeries no longer existed, 116 companies refused to cooperate, whereas three bakeries were withdrawn from the invitation list for different reasons. From these 1,249 workers, 186 did not want to participate, 173 were not available for reason of vacation, sick leave, or leaving their employer, resulting in 890 workers for serology testing and participation in responding to an extensive self-administered questionnaire. Twenty-three of these workers refused to give blood leaving 867 (80.6% = 867/[1249 – 173]) workers for further analysis (Fig. 1).

A second data set from a cross-sectional study in four sectors of baking and flour producing industries was used to validate and update the diagnostic model (validation set). This study included 153 workers from traditional bakeries, 85 from industrial bakeries, 83 from the milling industry, and 69 from the baking product manufacturers [13]. The authors considered the latter three groups as workers from nontraditional bakeries. Informed consent was obtained from all participating workers in both studies.

2.2. Reference standard

In both studies, sensitization to wheat allergens was chosen as the outcome. Specific IgE antibodies were measured with a modified enzyme immunoassay (EIA) [14]. Sensitization to wheat allergens was considered present if the optical density (OD) of 492 exceeded the OD + 0.10 of the reagent blank (no serum control). In the validation set, specific IgE antibodies were measured with a commercial immunoassay (Pharmacia CAP system; Pharmacia Diagnostics, Uppsala, Sweden). A concentration of 0.7 kU/L or greater (class II) was considered positive. The EIA for wheat allergens has

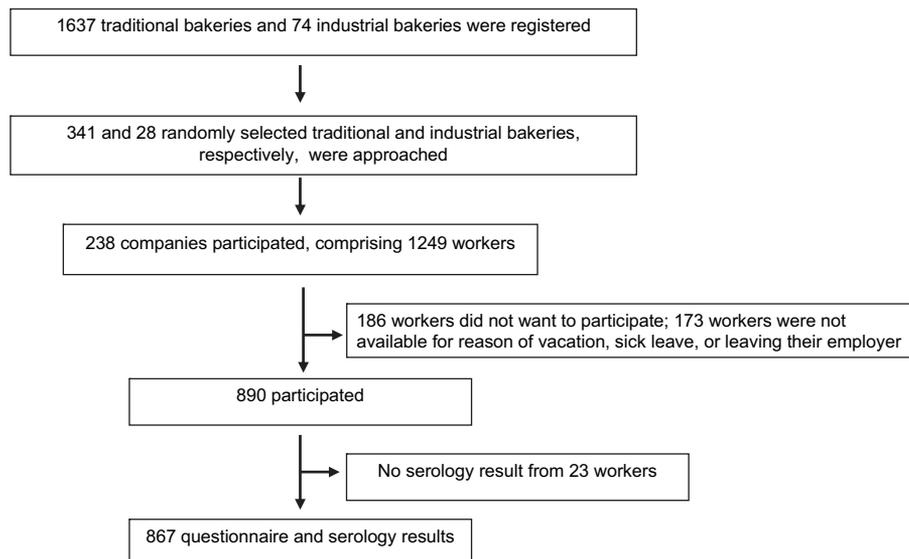


Fig. 1. Recruitment of participants in the development set.

been compared earlier to the CAP assay. The overall agreement was satisfactory and good at higher titers [14].

2.3. Potential predictors

Diagnostic information was collected with a questionnaire. In both studies, workers were asked to complete a self-administered respiratory questionnaire, which contained items from internationally accepted questionnaires [15,16]. The questionnaire comprised items on employment data (job, tasks), history of lower and upper respiratory symptoms, allergic symptoms because of common allergens, symptoms suggesting bronchial hyperresponsiveness (BHR), work-related upper and lower respiratory symptoms, skin symptoms, absenteeism, medication use, changes in tasks or jobs, and smoking habits.

Given the 108 sensitized workers (number of events) in the development set with 75 questionnaire items on symptoms as candidate predictors, variable reduction was required. When too many candidate predictors are considered in the development process of a prediction model, the model shows too extreme predictions in new patients. A rule of thumb is to have at least 10 events per candidate predictor [17]. The number of candidate predictors was reduced with principal component analysis (PCA) that identified clusters of correlated questionnaire items [18]. Clusters with eigenvalues greater than one (i.e., explaining more variance than a single predictor) were selected and considered as candidate predictors. Each selected cluster contained questionnaire items with factor loadings higher than 0.6.

The PCA resulted in 14 clusters with eigenvalues greater than one. The following clusters were identified and named (post hoc) as adequately as possible: “asthma symptoms in the last 12 months,” “shortness of breath and wheeze,” “chronic cough and phlegm in the last 12 months,” “upper respiratory symptoms,” “allergic symptoms in contact with house dust mites and/or plants,” “allergic symptoms in contact with pets,” “allergic symptoms because of certain food,” “symptoms suggestive of BHR induced by smoke,” “symptoms suggestive of BHR induced by change of temperature,” “skin symptoms,” “upper respiratory symptoms during work,” “lower respiratory symptoms during work,” “upper respiratory symptoms after work,” and “lower respiratory symptoms after work.” For details regarding included questionnaire items in each cluster, see Appendix. A cluster was regarded positive if at least one of the included questionnaire items was present.

To further reduce the number of clusters, the authors combined clusters with a Pearson correlation coefficient of 0.7 or higher. As a consequence, the clusters “upper respiratory symptoms during work” and “upper respiratory symptoms after work” were combined and renamed as “work-related upper respiratory symptoms.” Similarly, “lower respiratory symptoms during work” and “lower respiratory symptoms after work” were combined and renamed as “work-related lower respiratory symptoms.”

As a result, 13 candidate predictors were defined: 12 clusters of questionnaire items and type of bakery (industrial or traditional). Age, sex, and smoking habits were not considered as candidate predictors, based on the literature [2].

2.4. Data analysis

2.4.1. Model development

Multivariable logistic regression analysis starting with all 13 above-defined candidate predictors (12 clusters plus type of bakery) was done to develop a diagnostic model for sensitization to wheat allergens using the development set. Backward stepwise selection was applied using the likelihood ratio test with a *P*-value corresponding to Akaike’s information criterion (*P*-value < 0.157) to select the most important predictors [19,20].

The model’s ability to discriminate sensitized from non-sensitized workers (discrimination) was assessed with the receiver operating characteristic (ROC) area. The ROC area can range from 0.5 (no discrimination) to 1.0 (perfect discrimination) [21]. Internal validity and adjustment for overfitting of the model was executed with a bootstrapping procedure (S-Plus 6 for windows; Insightful Corporation, Seattle, WA) [22]. One hundred bootstrap samples were drawn from the development set. In each bootstrap sample, the entire modeling (predictor selection) process was repeated. The bootstrap procedure yields a ROC area corrected for optimism and a shrinkage factor to adjust the model for overfitting. The regression coefficients of the predictors in the model were multiplied by the shrinkage factor to prevent too extreme predictions for new workers [10,22,23].

2.4.2. Methods to validate and update the diagnostic model

The accuracy of internally validated and adjusted model was tested on the data of the validation set. The regression formula from the developed model was applied to all bakery workers of the validation set. The agreement between the predicted probabilities and the observed frequencies for sensitization (calibration) was evaluated graphically by plotting the predicted probabilities (x-axis) by the observed frequencies (y-axis) of the outcome [24]. The association between predicted probabilities and observed frequencies can be described by a line with an intercept and a slope. An intercept of zero and a slope of one indicate perfect calibration. The unreliability statistic *U* was used to perform one test that the intercept is not statistically significantly different from zero and the slope not different from one (Chi-square distribution with two degrees of freedom) [25]. The discrimination was assessed with the ROC area. If the model did not perform well in the validation set, the authors updated the model to improve the performance using methods that were previously described, starting with adjustment of only the intercept [10,11].

2.4.3. Model application

To facilitate the application of the model in practice, the final model was converted to an easy-to-use scoring rule. The regression coefficients of the predictors were divided by the smallest one and rounded to the nearest half integer. For each worker, the sum score was calculated. The sum scores were related to the corresponding probabilities of being sensitized to wheat allergens. To enhance decision making, workers were stratified into three different (low, intermediate, and high) probability groups. Different additional medical investigations may be applied to each probability group.

2.4.4. Missing data

Deleting subjects with a missing value leads to a loss of statistical power and can lead to biased results. Therefore, imputing missing values is generally preferred to complete case analysis [26,27]. Missing data were single imputed with values obtained from regression equations using SPSS version 15.0 (Statistical Package for Social Sciences, Chicago, IL, USA). The development set contained 286 (32.1%) subjects with one or more missing questionnaire items, of which

135 had one missing item. The validation set contained 144 (36.9%) subjects with at least one missing questionnaire items, of which 100 had one missing item.

3. Results

Table 1 shows the distribution of the workers' characteristics in the development and validation sets. Most workers in the development set were traditional bakers (74%) compared with 39% in the validation set. The prevalence of wheat sensitization was 13% and 20% in the development and validation sets, respectively. Table 2 shows the univariable association between the candidate predictors and sensitization in the development set. Nasoconjunctival symptoms, asthma symptoms, shortness of breath and wheeze, work-related respiratory symptoms, both upper and lower, were strongly associated with wheat sensitization.

3.1. Model development

Table 3 shows the final, internally validated, and adjusted model obtained from the development set. The

Table 1
Distribution of workers' characteristics in the development and validation sets

Workers Characteristics	Development set (<i>n</i> = 867)	Validation set (<i>n</i> = 390)
Age (yr) ^a	40 (17–79)	37 (17–63)
Male (gender)	820 (95)	380 (97)
Ever smoker	509 (59)	220 (56)
Traditional bakers	645 (74)	153 (39)
Nasoconjunctival symptoms in the last 12 mo ^b	560 (65)	274 (70)
Asthma symptoms in the last 12 mo ^c	32 (4)	36 (9)
Shortness of breath and wheeze ^d	261 (30)	136 (35)
Chronic cough and phlegm in the last 12 mo ^e	138 (16)	Not available
Allergic symptoms in contact with HDM and/or plants ^f	226 (26)	93 (24)
Allergic symptoms in contact with pets ^g	95 (11)	38 (10)
Allergic symptoms after ingestion of particular food ^h	103 (12)	9 (2)
Symptoms suggestive of BHR induced by dust and/or smoke ⁱ	253 (29)	104 (27)
Symptoms suggestive of BHR induced by change in temperature ^j	145 (17)	66 (17)
Skin symptoms ^k	357 (41)	171 (44)
Work-related upper respiratory symptoms ^l	200 (23)	159 (41)
Work-related lower respiratory symptoms ^m	104 (12)	88 (23)
IgE sensitization to wheat allergens	108 (13)	76 (20)

Data presented as *n* (%) unless otherwise stated.

Abbreviations: HDM, house dust mite; BHR, bronchial hyperresponsiveness.

^a Median (range).

^b Present if experience allergic symptoms, including hay fever, blocked or runny nose, and sneezing awakened because of nasal symptoms and/or itchy or watery or red eyes in the last 12 mo.

^c Present if experience at least one of the following in the last 12 mo: asthma attack, awakened by an asthma attack, asthma attack induced by exercise.

^d Present if experience at least one of the following: respiratory problem, shortness of breath while walking at normal pace with someone of same age or wheezing.

^e Present if experience chronic (3 mo) cough or cough with phlegm in the last 12 mo.

^f Present if experience upper and/or lower respiratory and/or skin symptoms in contact with HDM or plants.

^g Present if experience upper and/or lower respiratory and/or skin symptoms in contact with pets.

^h Present if experience upper and/or lower respiratory and/or skin symptoms after ingestion of certain food.

ⁱ Present if experience respiratory problem because of dust, smoke, or tobacco.

^j Present if experience respiratory problem because of change in temperature.

^k Present if experience dry skin, itchy skin, and/or eczema in the last 12 mo.

^l Present if experience itchy or blocked nose and/or itchy eyes during or after work, which improve when away from work.

^m Present if experience asthma, wheezing, shortness of breath, and/or chest tightness during or after work, which improve when away from work.

Table 2

Univariable association between the candidate predictors and sensitization to wheat allergens in the development set

Workers Characteristics	Nonsensitized (<i>n</i> = 759)	Sensitized (<i>n</i> = 108)	OR	95% CI
Traditional bakers	556 (73)	89 (82)	1.7	1.0, 2.8
Nasoconjunctival symptoms in the last 12 mo	465 (61)	95 (88)	4.6	2.5, 8.4
Asthma symptoms in the last 12 mo	18 (2)	14 (13)	6.1	3.0, 12.7
Shortness of breath and wheeze	193 (25)	68 (63)	5.0	3.3, 7.6
Chronic cough and phlegm in the last 12 mo	114 (15)	24 (22)	1.6	1.0, 2.7
Allergic symptoms in contact with HDM and/or plants	177 (23)	49 (45)	2.7	1.8, 4.1
Allergic symptoms in contact with pets	74 (10)	21 (19)	2.2	1.3, 3.8
Allergic symptoms after ingestion of particular food	81 (11)	22 (20)	2.1	1.3, 3.6
Symptoms suggestive of BHR induced by dust and/or smoke	206 (27)	47 (44)	2.1	1.4, 3.1
Symptoms suggestive of BHR induced by change in temperature	112 (15)	33 (31)	2.5	1.6, 4.0
Skin symptoms	306 (40)	51 (47)	1.3	0.9, 2.0
Work-related upper respiratory symptoms	144 (19)	56 (52)	4.6	3.0, 7.0
Work-related lower respiratory symptoms	64 (8)	40 (37)	6.4	4.0, 10.2

Data presented as *n* (%) unless otherwise stated.

Abbreviations: OR, odds ratio; CI, confidence interval; HDM, house dust mite; BHR, bronchial hyperresponsiveness.

shrinkage factor for adjusting the regression coefficients obtained from the bootstrapping procedure was 0.89. The model showed good discrimination, with a ROC area of 0.75 after correction based on bootstrapping (95% confidence interval (CI): 0.71, 0.81).

3.2. Model validation and updating

Without adjustment, the discrimination of the developed model in the validation set was reasonable (ROC area of 0.69, 95% CI: 0.62, 0.75) but the calibration was poor (*U* Chi-square = 19.9, *P*-value < 0.001). The calibration plot shows that the predicted probabilities were systematically too low (Fig. 2a). Reestimation of the intercept (from −3.66 to −3.05) of the model improved the calibration (*U* Chi-square = 2.2, *P*-value = 0.335, see Fig. 2b). The difference of intercept between the development and validation sets (0.61) was explained by the different outcome prevalence (13% and 20%). The difference in intercept was 1.8 ($\exp^{0.61}$) on the odds scale, which was very similar to the differences in prevalence on the odds scale ([0.20/

0.80]/[0.13/0.87] = 1.7). These findings suggest that the differences in prevalence could not be explained by differences in predictor values. Hence the final model consisted of the six predictors with two possible intercepts.

3.3. Model application

To enhance its application in practice, the regression coefficients of the predictors were simplified into easy-to-use numbers (Fig. 3). The corresponding predicted probability of wheat sensitization for each sum score is shown for a prevalence of 13% and for a prevalence of 20% (Fig. 3, lower part). As an example, a traditional baker with no symptom except hay fever had a sum score of two (1 + 1 + 0 + 0 + 0 + 0), which corresponds to a predicted probability of wheat sensitization of 9% or 15%, depending on the population to which the baker belongs.

The workers were finally categorized into three groups: low probability (sum score 0–1), intermediate probability (sum score 2–3), and high probability (sum score ≥4). A clear association could be observed between these

Table 3

Multivariable associations between the predictors and sensitization to wheat allergens in the development set

Intercept and predictors	β^a	OR	95% CI
Intercept	−3.66		
Traditional baker	0.67	2.2	1.2, 3.9
Nasoconjunctival symptoms in the last 12 mo	0.72	2.3	1.2, 4.5
Asthma symptoms in the last 12 mo	0.63	2.0	0.9, 4.4
Shortness of breath and wheeze	0.61	2.3	1.3, 3.8
Work-related upper respiratory symptoms	0.47	1.7	0.9, 3.1
Work-related lower respiratory symptoms	0.61	2.2	1.1, 4.4
ROC area (95% CI)	0.75 (0.71–0.81)		

The predicted probability of wheat sensitization can be calculated using the following formula: $P(\text{sensitization}) = 1/(1 + \exp(-(-3.66 + \text{traditional baker} \times 0.67 + \text{nasoconjunctival symptoms in the last 12 mo} \times 0.72 + \text{asthma symptoms in the last 12 mo} \times 0.63 + \text{shortness of breath and wheeze} \times 0.61 + \text{work-related upper respiratory symptoms} \times 0.47 + \text{work-related lower respiratory symptoms} \times 0.61)))$. Predictor value is one when present and zero when absent.

Abbreviations: CI, confidence interval; OR, odds ratio; HDM, house dust mite; BHR, bronchial hyperresponsiveness; ROC, receiver operating characteristic area.

^a Regression coefficient multiplied with a shrinkage factor (obtained from the bootstrapping procedure) of 0.89.

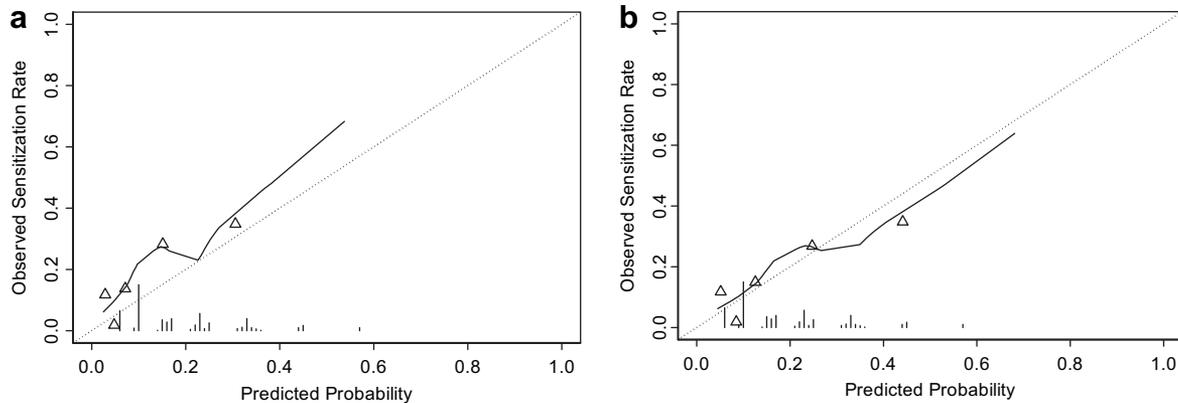


Fig. 2. Calibration plots of the wheat questionnaire model in the validation set when the original model was applied without adjustment (a) and when the intercept of the model was reestimated (b). The solid line is a smoothed curve that represents a nonparametric estimate of the relation between the predicted probability and the observed sensitization rate. Ideally, this line fits the dotted line that represents perfect calibration. Triangles indicate the observed sensitization rate per equal-size-quintiles of the predicted probability. Distribution of the predicted probabilities is indicated with vertical lines at the bottom.

probability groups and the prevalence of sensitization and symptoms in the development set (Table 4). Only 2–6% of the workers in the low probability group had allergic symptoms, 4% had a doctor's visit for respiratory complaints, 4% used medication to improve their respiratory complaints, 2% was absent from work because of allergic symptoms, and 0.3% changed jobs because of allergic symptoms. These figures were much lower compared with workers in the intermediate or high probability group. The sensitivity and specificity were 0.88 and 0.43, respectively, when the cut point for a positive test was between the low and intermediate probability groups, and were 0.46 and 0.89, respectively, when the cut point was between the intermediate and high probability groups.

4. Discussion

A simple diagnostic model for the detection of sensitization to wheat allergens was developed and validated in bakery workers. The model included six items: nasojunctival symptoms in the last 12 months, asthma problems in the last 12 months, shortness of breath and wheeze, work-related upper and lower respiratory symptoms, and type of bakery (traditional yes or no). Exposure to wheat allergens was presented by a proxy variable, that is, type of bakery. The authors intentionally did not include raw exposure variables to have a simple model consisting of questionnaire items that are available in daily occupational health practice.

The diagnostic model showed adequate discrimination in the validation set, but the predictions were in general too low. Simple reestimation of the intercept improved the model's calibration. In its form as an easy-to-use scoring rule, the model enables manual calculation of the individual probability of being sensitized to wheat allergens. Our findings support the idea that simple questionnaire items can be used to identify bakery workers at high risk for sensitization to wheat allergens.

We started by reducing 75 questionnaire items on symptoms to 12 candidate predictors. The multivariable modeling was based on 108 cases and 13 candidate predictors (ratio 8:1). A bootstrapping technique was used to check whether the model produced optimistic (i.e., too high or too low) estimates. Finally, an external data set was used to test our model in a new population of bakers. It is generally acknowledged that differences between the population where a model is derived and where a model is applied may influence its performance and therefore, external validation is recommended to determine whether a model is applicable in new populations [10,23]. The calibration plot in Fig. 2a indeed demonstrates that when the original diagnostic model was applied in the validation set, it produced systematically too low predicted probabilities—despite adjustment for overfitting using bootstrapping techniques [19,28]. Reestimation of the intercept alone already improved the calibration of the model in the validation set. This improvement suggests that different intercepts may be required to obtain correct absolute predicted risks for the two populations. For future applications in other baker populations, the same predictors could likely be applied. As for the choice of the intercept, it is more reasonable to use the intercept from the derivation set (−3.66) because this set consisted of workers from randomly selected bakeries with a prevalence rate of wheat sensitization in the range of what is commonly found [4,29].

Simple reestimation of the intercept can only improve the calibration of the model. By default the model discrimination will not change because the relative ranking of the sensitized workers vs. nonsensitized is not altered by this update method. Therefore, in an extra analysis a more advanced update method [8,9] was performed to evaluate if the model's discrimination could be improved. A total reestimation of the intercept and all regression coefficients in the validation set yielded a ROC area of 0.71, which was comparable to the ROC area from simple adjustment of the intercept (0.69).

Variable in the model	Value	Score
Type of bakery	If traditional baker	1
Nasoconjunctival symptoms in the last 12 months	If in the last 12 months experience at least one of the following: ●allergic symptoms including hay-fever ●blocked/runny nose or sneezing ●awakened due to nasal symptoms ●itchy/ watery/ red eyes	1
Asthma symptoms in the last 12 months	If in the last 12 months experience at least one of the following: ●asthma attack ●awakened due asthma attack ●asthma attack induced by exercise	1
Shortness of breath and wheeze	If experience at least one of the following: ●respiratory problem ●shortness of breath while walking at normal pace with someone of same age ●wheezing in the last 12 months	1
Work-related upper respiratory symptoms	If experience at least one of the following symptom during or after work: ●itchy nose ●blocked nose ●itchy eyes All symptoms should improve when away from work	1
Work-related lower respiratory symptoms	If experience at least one of the following symptom during or after work: ●asthma ●wheezing ●shortness of breath ●chest tightness All symptoms should improve when away from work	1
	Sum score	...

Sum score	0	1	2	3	4	5	6
Development set (prevalence=13%)							
N	75	262	257	141	70	49	13
Observed wheat sensitization (%)	1	5	10	14	26	47	69
Mean predicted probability (%)	3	5	9	15	24	38	54
Validation set (prevalence=20%)							
N	46	112	74	79	53	19	7
Observed wheat sensitization (%)	11	6	19	29	21	63	57
Mean predicted probability (%)	5	9	15	23	37	53	68

Fig. 3. Scoring rule for sensitization to wheat allergen and the corresponding predicted probability.

IgE sensitization to wheat allergens is closely associated with the development of allergic respiratory diseases among bakery workers [30]. The diagnosis of any allergic disease (especially bronchial asthma) can only be made at an individual level in a clinical setting. However, at the population level, precursors of the disease of interest can be used for the detection of workers with an elevated risk of sensitization. Subsequent diagnostic investigations can be limited to these high-risk workers, leaving a considerable amount of bakers in which no further medical investigations are needed. Therefore, instead of using a clinical

definition, such as OA, the authors used class II positive IgE serology to wheat allergens as the reference standard.

As an illustration of how the model can formalize decision making and improve the efficiency of surveillance, the workers were stratified into three probability groups. In workers with a high probability of wheat sensitization, advanced medical investigations are needed and may be best performed in an occupational respiratory health clinic. Workers with an intermediate probability can be evaluated further by occupational physicians. No further medical investigations are advised for workers with a low probability;

Table 4

General and health characteristics across low, intermediate, and high probability groups in the development set

Workers Characteristics	Low probability (sum score 0–1) (<i>n</i> = 337)	Intermediate probability (sum score 2–3) (<i>n</i> = 398)	High probability (sum score ≥ 4) (<i>n</i> = 132)
Age (yr) ^a	43 (18–79)	37 (17–79)	37 (19–63)
Allergic symptoms in contact with HDM and/or plants	21 (6)	123 (31)	82 (62)
Allergic symptoms in contact with pets	7 (2)	44 (11)	44 (33)
Allergic symptoms after ingestion of particular food	20 (6)	48 (12)	35 (27)
Symptoms suggestive of BHR induced by dust and/or smoke	49 (15)	122 (31)	82 (62)
Symptoms suggestive of BHR induced by change in temperature	11 (3)	65 (16)	69 (52)
Use of medication to improve respiratory complaints in the last 12 mo (e.g., inhalants)	12 (4)	90 (23)	72 (55)
Doctor visit for respiratory complaints	12 (4)	79 (20)	79 (60)
Absenteeism because of allergic symptoms in the last 12 mo	6 (2)	24 (6)	19 (14)
Change of function or task because of respiratory symptoms	1 (0.3)	0 (0)	7 (5)
Mean predicted probability of wheat sensitization, % ^b	4 (0.1)	11 (2)	32 (0.8)
IgE sensitization to wheat allergens present ^c	13 (4)	45 (11)	50 (38)
IgE sensitization to wheat allergens absent	324 (96)	353 (89)	82 (62)
Sensitivity (95% CI)		0.88 (0.82–0.94)	0.46 (0.37–0.56)
Specificity (95% CI)		0.43 (0.40–0.47)	0.89 (0.87–0.91)

Data presented as *n* (%) unless otherwise stated.

Abbreviations: HDM, house dust mite; BHR, bronchial hyperresponsiveness; CI, confidence interval.

^a Median (range).

^b Mean (standard error).

^c The percentages equal to positive predictive values.

they will be enrolled in the next round of detection program. If a sum score of four or higher was proposed as a high threshold, 132 (15%) workers would be advised for additional clinical investigations, of which 50 (38%) of them were found to be sensitized to wheat allergens. Nonetheless, this group showed very high proportions of allergic and BHR symptoms, on average two and 10 times higher than what was found in the intermediate and low probability group, respectively. This group also showed higher rates of medication use, doctor visit, absenteeism, and change in function because of respiratory or allergic symptoms than the other groups. In the intermediate group, these figures were lower than the high probability group but definitely higher than the low probability group. Thirteen (4%) sensitized workers will be falsely classified as not being sensitized in the low probability group, but serological tests will be correctly withheld in 324 (37% of the total population) workers. None of these 13 false-negative cases used medication; they also had no doctor visit, absenteeism, or change in function because of respiratory or allergic symptoms (results not shown). These findings suggested that the choice of the score thresholds was clinically sound.

For decision-making purposes, the choice of the threshold value is an important issue. This value should be based on the balance between the acceptable proportion of missed cases and the reduction of unnecessary tests. In general, a higher threshold leads to fewer subjects in the high probability group; the specificity is higher but at the cost of lower sensitivity. The context where the model will be applied also determines the choice of the threshold value. In case of (occupational) allergic disease, false-positive cases will undergo a simple serological test to confirm the presence of work-related sensitization and investigations will

end when they eventually have a negative serology result. The false-negative cases might have a more serious consequence because further exposure might lead to the development of OADs. Nevertheless, our findings suggest that they belong to a group of sensitized workers with less serious health problems that did not influence their work capability at that moment.

Recent study shows that continued exposure to the same occupational HMW allergens increased the risk of developing OAD [6]. Unfortunately, continued exposure to the same occupational sensitizers is unavoidable for some occupations (e.g., veterinarian, traditional baker). However, many exposed workers are employed by companies that have the means to reduce exposure of the identified sensitized person, that is, by conducting job rotation, enforcing the use of personal protection device, and create a more hygienic working environment (e.g., good ventilation system).

In conclusion, an easy-to-use and accurate diagnostic model for sensitization to wheat allergens was developed. The model uses simple questionnaire items, which are commonly available in occupational health practice. By applying this model and stratifying the workers based on their risk of being sensitized to wheat allergens, the efficiency of a medical surveillance program in the baking industry may be increased. External validation and updating in a sample of independent bakery workers yielded a model that is ready to be used in a wider population of bakery workers.

Acknowledgments

This study was supported by Productschap GZP, The Hague, and Aladdin program for occupational health research, The Netherlands Ministry Social Affairs and Employment.

Appendix

Supplementary information

Supplementary data associated with this article can be found, in the online version, at [10.1016/j.jclinepi.2009.10.008](https://doi.org/10.1016/j.jclinepi.2009.10.008).

References

- [1] Brant A. Baker's asthma. *Curr Opin Allergy Clin Immunol* 2007;7:152–5.
- [2] Balmes J, Becklake M, Blanc P, Henneberger P, Kreiss K, Mapp C, et al. Environmental and Occupational Health Assembly, American Thoracic Society. American Thoracic Society statement: occupational contribution to the burden of airway disease. *Am J Respir Crit Care Med* 2003;167:787–97.
- [3] Nicholson PJ, Cullinan P, Taylor AJ, Burge PS, Boyle C. Evidence based guidelines for the prevention, identification, and management of occupational asthma. *Occup Environ Med* 2005;62:290–9.
- [4] Houba R, Heederik D, Doekes G. Wheat sensitization and work-related symptoms in the baking industry are preventable. An epidemiologic study. *Am J Respir Crit Care Med* 1998;158(5 Pt 1):1499–503.
- [5] Gautrin D, Ghezzi H, Infante-Rivard C, Malo JL. Incidence and determinants of IgE-mediated sensitization in apprentices. A prospective study. *Am J Respir Crit Care Med* 2000;162(4 Pt 1):1222–8.
- [6] Gautrin D, Ghezzi H, Infante-Rivard C, Magnan M, J L'Archevêque RT, Suarathana E, et al. Long-term outcome of a prospective cohort of apprentices exposed to high-molecular-weight agents. *Am J Respir Crit Care Med* 2008;177:871–9.
- [7] Suarathana E, Meijer E, Grobbee DE, Heederik D. Predicting occupational diseases. *Occup Environ Med* 2009;66:713–4.
- [8] Meijer E, Grobbee DE, Heederik D. A strategy for health surveillance in laboratory animal workers exposed to high molecular weight allergens. *Occup Environ Med* 2004;61:831–7.
- [9] Suarathana E, Vergouwe Y, Nieuwenhuijsen M, Heederik D, Grobbee DE, Meijer E. Diagnostic model for sensitization in workers exposed to occupational high molecular weight allergens. *Am J Ind Med* 2005;48:168–74.
- [10] Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med* 2004;23:2567–86.
- [11] Janssen K, Moons K, Kalkman C, Grobbee D, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol* 2008;61:76–86.
- [12] Jacobs JH, Meijster T, Meijer E, Suarathana E, Heederik D. Wheat allergen exposure and the prevalence of work-related sensitization and allergy in bakery workers. *Allergy* 2008;63:1597–604.
- [13] Oostenbrink JHTJ, Tempels Z, Heide S, Steketee HA, Kerkhof M, Monchy JGR. [Prevalence of occupation related symptoms in bakeries, flour industries, and bakery ingredient factories.] Groningen, The Netherlands: Academic Medical Center; 2002.
- [14] Doekes G, Douwes J, Wouters I, de Wind S, Houba R, Hollander A. Enzyme immunoassays for total and allergen specific IgE in population studies. *Occup Environ Med* 1996;53:63–70.
- [15] Burney PG, Laitinen LA, Perdrizet S, Huckauf H, Tattersfield AE, Chinn S, et al. Validity and repeatability of the IUATLD (1984) Bronchial Symptoms Questionnaire: an international comparison. *Eur Respir J* 1989;2:940–5.
- [16] van der Lende R, Orie NG. The MRC-ECCS questionnaire on respiratory symptoms (use in epidemiology). *Scand J Respir Dis* 1972;53:218–26.
- [17] Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373–9.
- [18] Bollen K. Structural equations with latent variables. New York, NY: John Wiley & Sons; 1989.
- [19] Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.
- [20] Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. *Med Decis Making* 2001;21:45–56.
- [21] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29–36.
- [22] Van Houwelingen JC, Le Cessie S. Predictive value of statistical models. *Stat Med* 1990;9:1303–25.
- [23] Toll D, Janssen K, Vergouwe Y, Moons K. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol* 2008;61:1085–94.
- [24] Lee CJ, Miller LT. The process of evidence-based clinical decision making in occupational therapy. *Am J Occup Ther* 2003;57:473–7.
- [25] Harrell FE Jr, Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Stat Med* 1984;3:143–52.
- [26] Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006;59:1087–91.
- [27] van der Heijden GJ, Donders AR, Stijnen T, Moons KGM. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol* 2006;59:1102–9.
- [28] Bleeker SE, Moll HA, Steyerberg EW, Donders ART, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 2003;56:826–32.
- [29] Brant A, Nightingale S, Berriman J, Sharp C, Welch J, Newman Taylor AJ, et al. Supermarket baker's asthma: how accurate is routine health surveillance? *Occup Environ Med* 2005;62:395–9.
- [30] Chan-Yeung M, Malo J. Natural history of occupational asthma. In: Bernstein IL, Chan-Yeung M, Malo J, Bernstein DI, editors. *Asthma in the workplace*. 2nd ed. New York, NY: Marcel Dekker; 1999. p. 129–44.