

# Predicting occupational diseases

Eva Suartha<sup>1,2,3</sup> Evert Meijer,<sup>1</sup> Diederick E Grobbee,<sup>2</sup> Dick Heederik<sup>1,2</sup>

Prediction research is relatively new in the occupational health field,<sup>1-4</sup> although it is well established in clinical medicine.<sup>5,6</sup> Prediction models are developed to estimate the individual probability of the presence (diagnostic model) or future occurrence (prognostic model) of an outcome (ie, disease). As an example from clinical practice, Wells and colleagues demonstrated that a diagnostic model (comprised of the patient's history and physical examination) in combination with impedance plethysmography can safely rule out the presence of deep vein thrombosis. This approach reduced patient health care costs by avoiding expensive venography.<sup>5</sup> Assessment of the 10-year risk of coronary heart disease (CHD) using the Framingham scores is a well known example of prognostic prediction.<sup>6</sup> Such prediction allows physicians to identify a subset of patients with a higher probability of CHD in whom preventive action should be more effective.

The development of prediction models makes it possible to identify a small number of predictors to provide the best possible knowledge base for diagnosis. These models enable risk groups to be easily identified by quantification of the individual probability of having an occupational disease. Recently, we developed and validated diagnostic models to predict IgE sensitisation to occupational high molecular weight allergens in laboratory animal workers and bakers, and silicosis in construction workers. The models resulted in user-friendly prediction rules. This approach allows sequential testing by more burdensome tests only in the high-risk group.<sup>1,3,4</sup>

For example, the traditional approach does not specify which workers exposed to crystalline silica should undergo

periodic health surveillance with extensive questionnaires, chest x rays and lung function tests.<sup>7</sup> On the basis of disease prevalence, a high number of negative x ray results can be expected. However, a diagnostic model based on questionnaire results and lung function tests accurately identified workers with a low probability of pneumoconiosis and ruled them out from further investigations and thus improved diagnostic efficiency over no model.<sup>3</sup>

## ISSUES IN MODEL DEVELOPMENT

The first consideration when developing prediction models is the choice of the relevant outcome and how to define it.<sup>8</sup> In occupational allergy, identification should focus not on clinically established allergic disease but on highly associated preliminary symptoms and signs. Sensitisation to occupational allergens is an outcome strongly associated with work-related allergy and is the most appropriate characteristic that can easily be investigated.<sup>9</sup> A logical approach is therefore to first identify sensitised workers, and then carry out sequential diagnostic investigations at the clinical level only in these workers. We therefore aimed to identify individuals with a high probability of sensitisation.<sup>1,4</sup>

The second consideration is the choice of predictors to be included in the model. In contrast to aetiological studies, prediction studies do not aim to identify causal associations.<sup>10</sup> Furthermore, a time sequence (a cause should precede an effect) is not required in diagnostic studies. These studies are inherently cross-sectional and most predictors are actually the consequence of disease presence. In the diagnostic model for pneumoconiosis, being a "current smoker" was selected as an important independent predictor with an odds ratio (OR) of 2.4. The OR means that, compared to non-smokers, current smokers have 2.4 times higher probability of having a positive chest x ray result, without any reference to causality between smoking and pneumoconiosis.<sup>3</sup>

In practice, diagnosis is rarely established by a single test result and many test results generate more or less identical

diagnostic information. Therefore, multivariable regression analysis is used to evaluate the independent diagnostic or prognostic value of each test.<sup>10</sup> Candidate predictors with a univariable p value (<0.5) are included in the multivariable regression analysis. The final multivariable model can usually be generated with a stepwise selection method using  $p < 0.157$  for inclusion. To adjust for over-optimism, bootstrap procedures are performed and shrinkage factors calculated.<sup>10</sup>

## ISSUES IN MODEL APPLICATION

In general, prediction models show a lower performance in populations compared to that from which the model was derived. Therefore, external validation of the model in a new, but related, population is necessary to ensure its generalisability. A model is also more generalisable when applied in the same domain.<sup>11</sup> For example, the Dutch diagnostic model for sensitisation to laboratory animal allergens was externally validated and updated to be generalisable in a Canadian animal health apprentice setting.<sup>12</sup> Once we have a valid model, we can transform it into an easy-to-use score chart or nomogram to facilitate its use in practice.

The next important issue is to determine probability thresholds for risk stratification. Overall, a higher threshold leads to a smaller high-risk group (lower referral number) and higher specificity (lower false positive rate), but at the cost of lower sensitivity (higher false negative rate). Therefore, the choice of threshold must be a balance between the proportion of missed cases and a reduction in unnecessary diagnostic tests. It will also depend on the severity of disease, disease prognosis when missed, or, alternatively, improved prognosis when detected.

Ethical and legal aspects should also be carefully considered when applying a prediction model in occupational practice. Generally the application of prediction models should be limited to health surveillance programs and to occupational physicians screening individuals for disease and subclinical illness for early intervention.

## THE FUTURE OF PREDICTION MODELLING IN OCCUPATIONAL HEALTH

The application of prediction models in occupational health care is a novel approach. Detection of disease and subclinical illness may decrease the burden of occupational disease. Nevertheless,

<sup>1</sup> IRAS (Institute for Risk Assessment Sciences), Division of Environmental Epidemiology, Utrecht University, Utrecht, The Netherlands; <sup>2</sup> Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands; <sup>3</sup> Community Medicine Department, Faculty of Medicine, University of Indonesia, Jakarta Pusat, Indonesia

**Correspondence to:** Dr Eva Suartha, IRAS (Institute for Risk Assessment Sciences), Environmental Epidemiology Division, Utrecht University, PO Box 80178, 3508 TD, Utrecht, The Netherlands; E.Suartha@uu.nl

investigators should carefully design the model and clearly state the context where and how it can be used. There is now an opportunity to develop prediction models for diverse occupational diseases, especially given the existing large epidemiological studies. Future research is also needed to more explicitly evaluate the cost-effectiveness of these models in occupational health practice.

**Competing interests:** None.

**Provenance and peer review:** Not commissioned; externally peer reviewed.

Accepted 20 June 2009

Published Online First 4 August 2009

*Occup Environ Med* 2009;**66**:713–714.  
doi:10.1136/oem.2008.045609

## REFERENCES

1. **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.
2. **Kuijpers T**, van der Windt DA, van der Heijden GJ, et al. A prediction rule for shoulder pain related sick leave: a prospective cohort study. *BMC Musculoskeletal Disord* 2006;**7**:97.
3. **Suarthana E**, Moons KG, Heederik D, et al. A simple diagnostic model for ruling out pneumoconiosis among construction workers. *Occup Environ Med* 2007;**64**:595–601.
4. **Suarthana E**, Vergouwe Y, Nieuwenhuijsen M, et al. Diagnostic model for sensitization in workers exposed to occupational high molecular weight allergens. *Am J Ind Med* 2005;**48**:168–74.
5. **Wells PS**, Anderson DR, Bormanis J, et al. Value of assessment of pretest probability of deep-vein thrombosis in clinical management. *Lancet* 1997;**350**:1795–8.
6. **Wilson PW**, D'Agostino RB, Levy D, et al. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;**97**:1837–47.
7. **American College of Occupational and Environmental Medicine**. *Policies and position statements: Medical surveillance of workers exposed to crystalline silica*. Elk Grove Village, IL: ACOEM, 2008. Available at <http://www.acoem.org/guidelines.aspx?id=746> (accessed 7 August 2009).
8. **Coggon DMC**, Palmer KT, Evanoff B. Assessing case definitions in the absence of a diagnostic gold standard. *Int J Epidemiol* 2005;**34**:949–52.
9. **Gautrin D**, Ghezzi H, Infante-Rivard C, et al. Long-term outcomes in a prospective cohort of apprentices exposed to high-molecular-weight agents. *Am J Respir Crit Care Med* 2008;**177**:871–9.
10. **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.
11. **Justice AC**, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999;**130**:515–24.
12. **Suarthana E**, Meijer E, Heederik D, et al. The Dutch diagnostic model for laboratory animal allergen sensitization was generalizable in Canadian apprentices. *J Clin Epidemiol* 2009;**62**:542–9.