ORIGINAL ARTICLE

Relationship between chemical structure and the occupational asthma hazard of low molecular weight organic compounds

J Jarvis, M J Seed, R A Elton, L Sawyer, R M Agius

Occup Environ Med 2005;**62**:243–250. doi: 10.1136/oem.2004.016402

See end of article for authors' affiliations

Correspondence to: Prof. R Agius, Centre for Occupational and Environmental Health, The University of Manchester, Oxford Road, Manchester M13 9PL, UK; Raymond.agius@manchester.ac.uk

Accepted 6 January 2005

Aims: To investigate quantitatively, relationships between chemical structure and reported occupational asthma hazard for low molecular weight (LMW) organic compounds; to develop and validate a model linking asthma hazard with chemical substructure; and to generate mechanistic hypotheses that might explain the relationships.

Methods: A learning dataset used 78 LMW chemical asthmagens reported in the literature before 1995, and 301 control compounds with recognised occupational exposures and hazards other than respiratory sensitisation. The chemical structures of the asthmagens and control compounds were characterised by the presence of chemical substructure fragments. Odds ratios were calculated for these fragments to determine which were associated with a likelihood of being reported as an occupational asthmagen. Logistic regression modelling was used to identify the independent contribution of these substructures. A post-1995 set of 21 asthmagens and 77 controls were selected to externally validate the model.

Results: Nitrogen or oxygen containing functional groups such as isocyanate, amine, acid anhydride, and carbonyl were associated with an occupational asthma hazard, particularly when the functional group was present twice or more in the same molecule. A logistic regression model using only statistically significant independent variables for occupational asthma hazard correctly assigned 90% of the model development set. The external validation showed a sensitivity of 86% and specificity of 99%.

Conclusions: Although a wide variety of chemical structures are associated with occupational asthma, bifunctional reactivity is strongly associated with occupational asthma hazard across a range of chemical substructures. This suggests that chemical cross-linking is an important molecular mechanism leading to the development of occupational asthma. The logistic regression model is freely available on the internet and may offer a useful but inexpensive adjunct to the prediction of occupational asthma hazard.

Occupational asthma (OA) can present in a variety of clinical and pathophysiological forms. A recent classification by Vandenplas and Malo[1] defines four main types in order to distinguish irritant and other mechanisms from what most would regard as "true" OA; immunological OA characterised by a latency period necessary to acquire immunologically induced sensitisation. Within this type of OA approximately 40% of cases are due to low molecular weight (LMW) chemicals.[2] A literature search by van Kampen and colleagues[3] for all substances reported to have caused occupational asthma in humans identified approximately 250 substances, of which approximately 90 were LMW organic chemicals. The mechanism by which LMW chemicals cause asthma due to sensitisation tends to differ from that of high molecular weight (HMW) substances in that the latter group is often associated with pre-existing atopy and detectable serum IgE antibodies to the allergen.[4] Specific IgE antibodies (to chemical-protein conjugates) are commonly found in cases due to some LMW agents (for example, acid anhydrides), but rarely in others (for example, isocyanates).[2] The chemical mechanism by which LMW chemicals cause asthma remains generally poorly defined.

There are approximately 100 000 individual chemical entities on the European Inventory of Existing Commercial Substances. For many of these substances the available toxicological data is very sparse; this is particularly true for asthmagenicity.[5] Despite recent experiments with cytokine profiling in mice and IgE production in rats,[6] no single experimental animal test has been well validated for predicting asthma hazard potential of individual chemicals. As a result the asthmagenic potential of a chemical usually only comes to light when human cases occur.

Observations of possible links between a chemical's structure and its asthmagenic potential have been made.[7–12] (Quantitative) structure-activity relationships ((Q)SARs) have been developed in the pharmaceutical industry for prediction of adverse effects of drugs such as skin sensitisation, mutagenicity, carcinogenicity and teratogenicity.[13 14] The potential for predicting human health effects of chemicals from their chemical structure has been recognised by the European Centre for Ecotoxicology and Toxicology of Chemicals.[15] However, this report has explicitly acknowledged that there is no well validated (Q)SAR in respect of respiratory sensitisation. We set out to develop and validate such a model and to present it for peer scrutiny on the world wide web.

The model described in this paper and its validation results from an assessment of substructures in reported asthmagens compared to control chemicals. Logistic regression analysis was used to generate a computer software program that would give a quantitative estimate of a chemical's asthmagenic potential (hazard index) from its chemical structure.

Abbreviations: HMW, high molecular weight; LMW, low molecular weight; OA, occupational asthma; QSAR, quantitative structure-activity relationship; SAR, structure-activity relationship

Main messages

- Human data (clinical and epidemiological) show that there are significant differences in chemical structure between organic low molecular weight occupational asthmagens and control chemicals also used in the workplace and which are not known to be asthmagenic.
- The occupational asthma hazard appears to be associated with specific reactive chemical groups and especially if two or more of them are present in the molecule.

Policy implications

- If further corroborated, logistic regression models can be used to help predict the occupational asthma hazard of novel chemical entities.
- An inexpensive and rapid web based tool which is relatively user friendly can undertake the predictive assessments.

METHODS

Data and selection criteria

Two separate searches were carried out by different observers to avoid bias in the first instance to develop the model and later to validate it independently. The model development set consisted of LMW asthmagens reported to cause OA in the literature before 1 January 1995, while the validation asthmagens were selected from the literature after 1 January 1995. Control chemicals for model development were selected from the 1994 edition of the UK Health & Safety Executive's document *Occupational exposure limits* (EH40).[16] Controls for validation were selected from the 2002 editions of both EH40[17] and the American Conference of Governmental Industrial Hygienists document, *Threshold limit values and biological exposure indices*.[18]

Pre-1995 search for model development set of asthmagenic and control chemicals

A search of the medical literature up to the end of December 1994 was carried out (JJ) using the Medline database to identify publications relating to occupational asthma. Since the use of the Medline ''keyword'' system is not adequately sensitive, merely specific, several textword based searches were performed in which words potentially synonymous with the general topics of occupational disease, sensitisation, and respiratory disease were used. Approximately 2000 references resulted. The title and abstract were reviewed in order to exclude papers that did not describe occupational asthma ascribed to LMW organic chemicals. The remaining references were appraised, and chemicals were accepted as asthmagens if a physician had, in a peer reviewed report, clearly diagnosed occupational asthma arising following a latent period of exposure. Moreover, good evidence for the chemical identity (formula and structure) of the substance responsible for the occupational asthma was required. Chemicals of molecular weight greater than 1000 were excluded. For the purposes of this study ''organic'' was defined as ''carbon containing''. Organometallic complexes were excluded but organic salts formed with metal cations were accepted. A systematic protocol was applied to the appraisal and abstraction of information, and a database was set up. Since the degree of certainty of the occupational asthma hazard of particular chemicals may have been variable, data on the number of reported cases in relation to each chemical, and the extent of the confirmatory investigation (for example, whether challenge testing confirmed the diagnosis) was recorded prospectively. Other data obtained included clinical type of response, evidence of IgE involvement, latency, and persistence of disease. The control chemicals for the model development set were selected from tables 1 and 2 of the Health & Safety Executive's EH40/94 *Occupational exposure limits 1994* document which detail occupational exposure indices. All the listed organic compounds fulfilling the chemical criteria of composition and molecular weight as for the ''active'' chemicals and for which a structure could be identified were used (excepting those which were already included in the study as active). If an exposure limit had been set for each structural isomer of a compound, then all those isomers were included. The resulting 78 known sensitisers and 301 controls were stored electronically in a topological chemical database (IsisBase 1.2.1, MDL)[19] along with the tabulated data. The model validation sets are described later.

Chemical substructure fragment selection

A substructure fragment was defined as an atom or group of atoms connected by a defined bond type. The fragments were selected from the represented atom types and known chemical groups, and tended to be fairly small and fairly common. Fragments were not necessarily exclusive, such that the C = O carbonyl-like fragment for example would also be found within the carboxylic acid fragment. Two methods of fragment frequency generation were used. The more general substructure fragment technique allowed the rapid evaluation of the occurrence frequency of over 120 substructure fragments in the dataset of structures. These substructure fragments were explicit and specific with respect to which atom and bond patterns allowed for a match. From this, a second method utilising a refined set of more complex fragments was created in which some atoms were ''wild cards''—that is, they could match any one of several atoms. Since the chemistry of fragments adjacent to aromatic rings is different to that of fragments attached to alkyl groups, due to the stabilising effect of delocalised electrons, a distinction was made between fragments adjacent to aliphatic groups and those adjacent to aromatic rings. A computer program was written that interpreted standard chemical connection tables (MDL *.mol or *.sdf file formats)[19] and tabulated the number of times each fragment occurred within each of the 379 compounds. This automation avoided observer error or bias in the manner in which fragment distribution in the asthmagens and in the control sets was determined.

Statistical methods

Two approaches were used for structure-activity relationship studies. Using the 379 compounds in the database, the odds ratio for a LMW compound having been reported as a cause of occupational asthma based on specific components of its chemical substructure was calculated. This approach was analogous to a case-control study method, in which the ''cases'' were reported causative agents of occupational asthma (asthmagens), whereas the ''controls'' were the matched hazardous occupational agents which were not reported causes of occupational asthma, while the ''exposures'' were the selected chemical sub-structure fragments.[8] Therefore, for each substructure fragment, the resultant ''hazard odds ratios'' was a ratio of the two odds: the odds of an asthmagen containing the substructure fragment,

Table 1 Chemical substructure fragments with hazard odds ratios

Chemical substructure fragment	Occurrence rate (per molecule)	Hazard odds ratio	95% CI	Chemical substructure fragment	Occurrence rate (per molecule)	Hazard odds ratio	95% CI
Aliphatic amine	**Any**	**6.0**	**3.2 to 11**	β lactam ring	**Any**	**∞**	**6.3 to ∞**
	1	**18**	**4.3 to 110**		**1**	**∞**	**6.3 to ∞**
	2	**3.0**	**1.4 to 6.4**	Benzyl	**Any**	**2.4**	**1.4 to 4.2**
	3	**37**	**4.2 to 1700**		**1**	**2.2**	**1.2 to 3.9**
	4	**∞**	**2.4 to ∞**		2	3.1	0.9 to 9.7
Aliphatic carboxylic acid	**Any**	**5.6**	**1.9 to 16**		3	5.6	0.4 to 79
	1	**5.8**	**1.9 to 18**		4	∞	0.14 to ∞
	2	4.5	0.06 to 350	Ethene derivative	**Any**	**3.2**	**1.8 to 5.8**
Aliphatic chlorine	**Any**	**0.28**	**0.07 to 0.81**		**1**	**2.8**	**1.1 to 7**
	1	0.22	0.01 to 1.5		2	3.3	0.8 to 11
	2	0.5	0.05 to 2.3		**3**	**2.7**	**1.3 to 5.4**
	3	0	0 to 1.4		**4**	**5.6**	**1.3 to 23**
Aliphatic nitrogen	**Any**	**6.3**	**3.5 to 11**	Imine derivative	**Any**	**8.2**	**3.5 to 20**
	1	**4.0**	**2.0 to 8.1**		1	2.5	0.5 to 9.7
	2	**9.0**	**3.4 to 24**		**2**	**13**	**3.1 to 80**
	3	**11**	**1.2 to 140**		**3**	**∞**	**4.4 to ∞**
	4	**∞**	**1.4 to ∞**		4	∞	0.13 to ∞
Aliphatic oxygen	**Any**	**2.6**	**1.4 to 5.0**	Carbonyl	**Any**	**3.2**	**1.8 to 5.5**
	1	1.1	0.43 to 2.9		1	1.2	0.54 to 2.4
	2	**2.3**	**1.1 to 5.1**		**2**	**7.3**	**3.3 to 16**
	3	**5.4**	**1.8 to 16**		**3**	**∞**	**10 to ∞**
	4	2.6	0.90 to 7.4		4	∞	0.2 to ∞
Aliphatic sulphur	**Any**	**3.7**	**1.7 to 8.1**	Carboxylic acid	**Any**	**5.5**	**2.3 to 13**
	1	**5.4**	**2.2 to 13**		**1**	**5.6**	**2.2 to 14**
	2	1.6	0.03 to 20		2	4.7	0.06 to 370
	3	0	0 to 25	Isocyanate	**Any**	**∞**	**6.3 to ∞**
	4	0	0 to 25		1		
Any amine	**Any**	**5.2**	**2.9 to 9.3**		**2**	**∞**	**3.9 to ∞**
	1	**2.6**	**1.2 to 5.2**		3	∞	0.81 to ∞
	2	**16**	**4.9 to 61**	Any nitrogen	**Any**	**5.6**	**3.1 to 10**
	3	**23**	**4.2 to 230**		**1**	**2.4**	**1.1 to 5.1**
	4	**∞**	**1.2 to ∞**		**2**	**6.7**	**3.0 to 15**
Any anhydride	**Any**	**31**	**3.8 to 1400**		**3**	**15**	**3.8 to 60**
	1	**26**	**3.1 to 1200**		**4**	**23**	**3.4 to 240**
	2	∞	0.11 to ∞	Any oxygen	**Any**	**2.8**	**1.4 to 5.5**
Aromatic to carbonyl	**Any**	**3.7**	**1.8 to 7.3**		1	1.2	0.44 to 3.2
	1	2.5	0.97 to 6.2		**2**	**2.3**	**1.0 to 5.4**
	2	**4.4**	**1.5 to 13**		**3**	**4.9**	**1.7 to 14**
	3	∞	0.9 to ∞		4	2	0.57 to 6.4
	4	∞	0.12 to ∞	Any sulphur	**Any**	**4.3**	**2.1 to 8.8**
Aromatic to nitrogen	**Any**	**2.8**	**1.5 to 5.4**		**1**	**4.3**	**1.8 to 10**
	1	1.5	0.55 to 3.9		2	3.3	0.27 to 29
	2	**3.3**	**1.2 to 9.1**		3	4.9	0.35 to 69
	3	**9.8**	**1.4 to 110**	Any double bond	**Any**	**3.8**	**2.0 to 7.2**
	4	0	0 to 190		1	1.3	0.53 to 3.1
Aromatic to oxygen	**Any**	**2.1**	**1.0 to 4.8**		**2**	**3.1**	**1.3 to 7.2**
	1	2	0.8 to 4.6		**3**	**10**	**3.6 to 31**
	2	2.8	0.69 to 10		**4**	**14**	**4.6 to 46**
	3	0	0 to 24	Ethanolamine backbone	**Any**	**20**	**6.1 to 84**
	4	4.5	0.06 to 350		**1**	**25**	**6.7 to 140**
Aromatic to sulphur	**Any**	**13**	**3.2 to 78**		2	0	0 to 200
	1	**7.5**	**1.4 to 49**		3	∞	0.13 to ∞
	2	∞	0.11 to ∞	Acrylate derivative	**Any**	**13**	**2.3 to 140**
	3	∞	0.82 to ∞		**1**	**13**	**2.3 to 140**

Statistically significant hazard odds ratios given in bold.
Odds ratios given to 2 significant figures.

compared to the odds of a control chemical containing the same substructure fragment. The calculation of these ''hazard odds ratios'' was carried out using the Epi-Info statistical computer program package available from the Centers for Disease Control, Atlanta, Georgia, USA. The exact method was used when low expected occurrence frequencies caused the Cornfield confidence limits to be unreliable. Odds ratios for reporting of the particular structural fragments as asthmagens were calculated with upper and lower 95% confidence intervals (CIs), and therefore where the value of 1.00 fell within this interval this odds ratio was deemed non-significant.

In the second approach, logistic regression analysis was performed using the Statistical Package for Social Scientists (SPSS, from SPSS Inc.) package. A model was obtained using a backward stepwise method. The significance of the likelihood ratio was used to determine which independent variables remained in the model. Variables (chemical substructure fragments) were entered and removed from the model on the basis of an entry probability of 0.05 and a removal probability of 0.10. Other model methods and parameters were tested but are not described.

The logistic regression model was then used as a tool to predict a hazard index, with predictions falling in the range of 0 to 1. Compounds were classified as active (that is, predicted as occupational asthma hazards) in the model testing if the predicted hazard index was greater than 0.5. Values of hazard index less than 0.5 were classified as inactive.

A kappa value (measure of model reliability) for the model agreement with the observed data was calculated,[20] as a means of internal validation of the model.

Post-1995 chemicals for external validation

Besides the internal validation of the model, external validation was conducted using a new three part dataset of chemicals, none of which contributed to the original model and which were identified and handled by a different observer (MS) from the original set. This consisted of:

- Chemicals (LMW) identified as new occupational asthmagens through a literature search covering the period January 1995 to December 2003, using the same criteria as for the model development, and which yielded 21 new asthmagenic chemicals.
- Agents reported to the SWORD scheme[21] as possible causes of occupational asthma, but which did not already appear as asthmagens in either the original or validation datasets. These consisted of 16 LMW chemicals which were presumed to be less certain as causes of OA than the substances which fulfilled the criteria of asthmagens since they had not been reported following a peer review process.
- New controls consisted of 77 LMW chemicals which did not appear in any previous category. These fulfilled the same chemical criteria as the other compounds. The 30 compounds in tables 1 and 2 of the 2002 edition of EH40 which were in neither of the learning dataset groups nor the validation asthmagen or SWORD groups were selected. A further 47 chemicals also fulfilling the criteria for atomic content and molecular weight range were selected at random from the 2002 edition of ACGIH to make the total number of 77.

The chemicals in these three categories were presented single blind in random order for the purposes of validating the logistic regression model.

RESULTS

General data

CAS Registry Numbers were found for all but five of the asthmagens used in model development (Appendix), although a structure was identified. A large proportion of the 78 asthmagens (33) were identified as asthma hazards on the basis of single case reports, with over half the compounds (48) having three or fewer cases recorded. Twenty two compounds had accounted for between 3 and 33 reported cases each, and only eight compounds had resulted in more than 33 reported cases. Evidence of specific IgE antibodies to the hapten-protein conjugate was found for 20 of the 78 (26%) asthmagens.

Structure-activity relationship studies: odds ratios

The 78 asthmagens were compared to the 301 control compounds. Significantly increased ''hazard odds ratios'' were noted both for aliphatic and aromatic occurrences of the following hetero-atom atom types (table 1): nitrogen, oxygen, sulphur and the acid anhydride, isocyanate, amine, imine, and carbonyl groups. Significantly increased ''hazard odds ratios'' were also noted for (poly) unsaturated aliphatic fragments. Furthermore, the presence of two or more imine or carbonyl groups was significant while a single occurrence was not significantly associated with reported hazard. The ''hazard odds ratios'' increased significantly with substructure fragment occurrence frequency within each molecule for fragments as diverse as carbonyl and amine.

Logistic regression model

Logistic regression models were created from the fragment data using various probability statistics for entry and removal of variables from the model. A final model based on 0.05 entry and 0.10 removal probabilities correctly classified 90% of the 379 compounds. With a prediction threshold set at a hazard index of 0.5 (see fig 1), sensitivity of the final model was 69% (that is, 54 of 78 asthmagens correctly classified). The specificity was 93% (279 of 301 controls correctly classified). The kappa value for the model was 0.63.

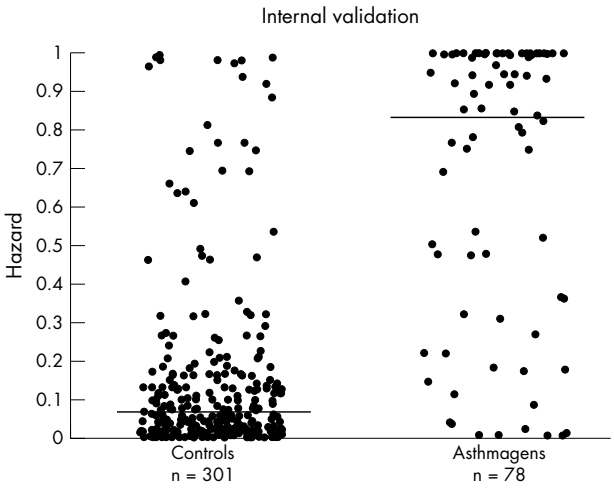

Figure 1 Internal validation showing hazard indices for the pre-1995 model development control and asthmagenic chemicals (median hazard index for each group indicated by horizontal line).

The model is freely accessible through the following web page:

- http://www.coeh.man.ac.uk/research/asthma/

The 16 asthmagenic compounds that were most seriously misclassified on internal validation (that is, with predicted hazard index less than 0.25) were: fenthion (0.01), acetic acid (0.15), hexachlorophene (0.01), ethylene oxide (0.01), chloroxylenol (0.02), furfuryl alcohol (0.01), styrene (0.04), paraphenylene diamine (0.22), hydroquinone (0.01), chloramine-T (0.22), tetrachloroisophthalonitrile (0.11), plicatic acid (0.01), 3-carene (0.18), phenylglycine acid chloride (0.18), dobutamine (0.17), Black GR Reactive Dye (BK-5) (0.04).

The 15 control compounds that were seriously misclassified (that is, with predicted hazard index greater than 0.75) were *o*-acetylsalicylic acid (0.81), strychnine (0.75), methacrylic acid (0.75), warfarin (0.77), diethyl phthalate (0.97), diisobutylphthalate (0.99), dibutyl phthalate (0.99), benzyl butyl phthalate (0.98), dibenzoyl peroxide (0.92), acetic anhydride (0.89), 2,2′-iminodiethanol (diethanolamine) (0.98), hexahydro-1,3,5-trinitro-1,3,5-triazine (0.77), dimethyl phthalate (0.94), diallyl phthalate (0.97), benomyl (0.98).

External validation

Figure 2 shows the application of the logistic regression model to the external validation set consisting of chemical compounds, which did not contribute to the derivation of the original model. The new asthmagenic chemicals whose reports as asthma causes were peer reviewed and appraised, achieved high predicted hazard indices (median 0.98). The new controls showed a low hazard index (median 0.03), whereas the chemicals reported through SWORD (and which had not also been reported in the peer reviewed literature) occupied an intermediate position (median 0.45), with a wide scatter. Using a threshold hazard index of 0.5 for prediction of asthmagenicity in the control and asthmagen groups correctly identified 18 of 21 asthmagens (sensitivity 86%) and 76 of 77 controls (specificity 99%).

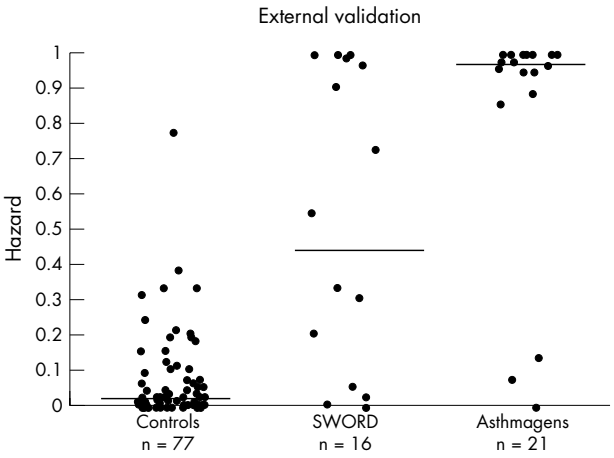

Figure 2 External validation showing hazard indices for post-1995 sets of chemicals; controls, asthmagens, and possible asthmagens reported to SWORD (median values indicated by horizontal line).

DISCUSSION

The logistic regression model that has been developed in this study is a quantitative structure-activity relationship (QSAR) for prediction of the likelihood (hazard index between 0 and 1) that a chemical has asthmagenic potential. Internal validation using the pre-1995 reported asthmagens and control chemicals and external validation using post-1995 reported asthmagens and newer controls have been performed. If a threshold hazard index of 0.5 is used, internal validation gave a sensitivity of 69% and specificity of 93% whilst external validation had a sensitivity of 86% and specificity of 99%.

Alongside clinical, pathophysiological, immunological, and epidemiological data a QSAR has two purposes. Firstly, at a practical level the capacity to make cost effective predictions about the asthma hazard that may be associated with novel chemical entities, would be a useful tool in helping to determine policy on issues such as degree of control of exposure and prospective health surveillance. Secondly, an understanding of structure-activity relationships may permit mechanistic conclusions at a molecular level regarding the initial chemical reactions between environmental chemicals and the human host, which lead to the development of occupational asthma. This may also give insights into the molecular mechanisms of non-atopic asthma, which is of increasing prevalence in the general population.

Prediction of occupational asthma hazard of novel chemical entities

(Quantitative) structure-activity relationships ((Q)SARS) have been developed for prediction of several human health endpoints as well as physicochemical properties and environmental endpoints of chemicals. The European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) published a report[15] summarising the commercially available (Q)SARS. This was in response to proposals for a new European Union Chemicals Policy, which would require extensive toxicological data to be generated on thousands of chemicals. (Q)SARS may provide a quicker and cheaper alternative to current toxicity testing and also have the potential to reduce the need for animal testing. The ECETOC report summarises validation data of (Q)SARS for the following human health endpoints: skin and eye irritation, skin sensitisation, acute and chronic oral toxicity, teratogenicity, mutagenicity, and carcinogenicity.

Early animal models of asthmagenicity included inhalational challenges in guinea pigs.[22] These were costly and required considerable technical expertise. While cytokine profiles and total IgE measurements have shown promise,[6] animal tests, besides being expensive, require wider validation for the prediction of asthmagenicity in humans.

Graham and colleagues[11] developed a structure-activity model for respiratory sensitisation using the CASE/MultiCASE systems and the resulting predictive model achieved sensitivities and specificities of 95%. However, the control chemicals were selected from known dermatological nonsensitisers on the assumption that skin nonsensitisers are also respiratory nonsensitisers. This is not necessarily true.[23] A more recent model by Karol and colleagues[12] was developed using more appropriate criteria for selection of negative controls. While this model provided mechanistic insights into the physicochemical properties associated with asthmagenicity, internal validation of its predictive power gave a sensitivity of 85% and specificity of only 74%.

The number of asthmagens used by Graham *et al* was only 40; the selection criterion was a positive challenge test with a fall in FEV_1 of 20% or more. The model development set in our study included all chemicals (n = 78) reported as asthmagens in peer reviewed literature, whether or not a formal challenge test had been performed (Appendix). In the Graham study, 37% of predictions were accompanied by a warning of the presence of structural fragments not present in the learning dataset, which might influence activity. The greater number of active chemicals used in the learning dataset for our study reduced the likelihood of the presence of such unaccounted fragments.

The identification and selection of the active and control compounds for all studies may affect the validity of the results. Our selection of controls was based on agents that are classified as hazardous and to which there is significant occupational exposure to warrant the setting of exposure limits. It was deemed more appropriate to use these controls than a random selection of chemicals (to which humans might not have been significantly exposed). The requirement that an active compound be a documented cause of human occupational respiratory sensitisation rather than merely suspected from animal experiments, was fundamental to the study reported here, so as to focus exclusively on hazards proven to be specific to man.

The QSAR methods used to correlate chemical function with chemical structure have varied, although all are based on the assumption that chemical structure is the primary determinant of reactive properties and hence biological activity. An important problem in structure based studies relates to the adequate representation of chemical structures and a number of methods can be used. Empirical formulae indicate the proportions of elements present; chemical formulae indicate actual numbers of each element present; and structural formulae allow a *topological* representation of the elements present. To be of use in all but the simplest QSAR models, the representation of chemicals must indicate at least the topological relationship between the constituent atoms—that is, which atoms are bonded to which. More sophisticated databases may also indicate the *topographical* relationship between atoms—that is, their relative positions in three dimensional space. The present study used a *topological* database and substructure fragments to investigate QSARs. It is possible that a future topographical approach might improve the prediction model so far described.

A number of problems must be overcome in relating chemical structure to biological activity. Traditional QSAR studies require relatively similar structures and an activity, which can be readily described by a continuous variable. Respiratory sensitisers do not readily comply with either of

these criteria. At present, the existing cases are too poorly characterised to provide a suitable measure of potency (although our unpublished observations suggest that the more potent chemicals tend to have a higher hazard index). The predictors used are solely derived from chemical structure and so this essentially employs a Free-Wilson approach to QSAR.[24] This was considered appropriate given the relative chemical diversity of sensitising compounds and the intent to undertake the study with no a priori aetiological assumptions.

Simple ranking of the odds ratios for reported asthma hazard associated with specific substructures is not an adequately sophisticated output, because of the internal confounding that may result from the complex interactions between the substructures—some of which (such as non-carbon, non-hydrogen ''hetero''-atoms) are subsumed within other reactive groups. Therefore more sophisticated logistic regression modelling was warranted. The use of the stepwise logistic regression facility within the SPSS program was considered appropriate since this work falls into both the exploratory and the predictive categories described by Menard.[24] The significance of the likelihood ratio was used in preference to the Wald statistic because some of the fragments, for example the isocyanate fragment, were rare but occurred exclusively in active compounds. That is, the presence of the fragment guarantees activity but its absence does not preclude activity.

However, the results of the regression model suggest that it may be of use in the initial screening of compounds for occupational asthma hazard. The kappa value of 0.63 for the model is satisfactory, but equally important is the distribution of predictions across the range of values from zero to one. In fig 1 showing the results of internal validation, analysis of the misclassified compounds highlights some points about the model. If we assume the model is a good predictor of hazard, the misclassified compounds may simply represent data outliers. By addressing reported asthma causing compounds to which the model assigned a low hazard index such as styrene, ethylene oxide, and chloroxylenol, it is possible that a mechanism not adequately catered for by the model may be responsible. Correspondingly, for the agents not described as occupational asthmagens, such as methacrylic acid, where the model has assigned a high hazard index, it might be correctly identifying compounds with true sensitising hazard but which have not yet been reported as such. Thus the ''control'' compound methacrylic acid is predicted as a hazard (0.75), but it should be noted that methyl methacrylate is an accepted respiratory sensitiser and acrylic acid is a suspected one.[25] Another curious ''false positive'' was strychnine (hazard index 0.75). It can be postulated that strychnine may be a potential sensitiser, but because of its deserved reputation as a poison, exposure to it may have been more tightly controlled than compounds of equivalent sensitisation but less notorious toxicity. At the time of model development, diethanolamine met the criteria for a control chemical but had a hazard index of 0.98. In the post-1995 literature search for validation asthmagens, a case report was found of asthma caused by diethanolamine.[26] This illustrates the possible explanation of ''false positives''—that is, they have asthmagenic potential but a case is yet to occur.

Figure 2 shows the results of external validation of the model—that is, using cases published post-1995 and more up to date controls. The control and asthmagen columns illustrate the value of the model in correctly predicting whether or not a chemical has the potential to cause occupational asthma. Moreover, speculative explanations can be offered for the 1 of 77 controls and for one of the 3 of 21 asthmagens that are seemingly misclassified.

The control with hazard index greater than 0.5 is propranolol, a β-adrenergic blocking drug which is known to provoke bronchoconstriction.[27] The three validation asthmagens with a hazard index less than 0.5 are fluazinam (0), morphine (0.08), and 1,2-benzisothiazolin-3-one (0.14). The mechanism by which morphine has allergic effects is of interest because it causes direct degranulation of mast cells, and skin prick tests cause a weal and flare reaction in control subjects, which is indistinguishable from opiate sensitive subjects.[28] This exceptional mechanism may be responsible for its asthmagenicity and explain why its structure-activity relationship does not follow this model. There is no obvious explanation for the misclassification of fluazinam and 1,2-benzisothiazolin-3-one.

The middle column in fig 2 comprises the hazard indices of the 16 novel asthmagenic chemicals reported to the SWORD scheme from 1995 that were also absent from any other group. These data had a wider scatter and a median value (0.45) between that of the control group (0.03) and the asthmagen group (0.98). This might be expected since some of the reported cases, particularly if they are single cases, possibly have incorrect causal attribution (for example, if the reporting physician attributes the asthmagenicity to a solvent or vehicle comprising the bulk of the agent rather than to the active agent which is present in a low concentration and perhaps not chemically defined for reasons of trade secrecy).

It is accepted that models such as this require further corroboration and testing. However, based on the indices of sensitivity and of specificity discussed above, it was considered appropriate to make our model available on the world wide web for other researchers and for practitioners to test it and to provide us with comments. The version is amenable to improvement by iteration, and in due course might prove to be a valuable adjunct to the a priori hazard assessment of potential LMW occupational asthmagens.

Mechanistic conclusions

As regards the interpretation of the findings, one of the interesting outcomes of this study is that a higher frequency of occurrence of certain substructure fragments is associated with a larger odds ratio for the reported occupational asthma hazard (table 1). Thus the odds ratios for reported asthma hazard for single occurrence of chemical substructures such as imine fragments, aliphatic or aromatic amines, or carbonyls were not significant at the 95% level. In contrast, for many reactive fragment groups, two or more fragment occurrences result in statistically significant odds ratios. This supports earlier hypotheses that chemicals which are bi- or poly-functional (that is, have cross-linking potential) are more likely to be asthmagens.[7 8] This hypothesis requires further study, for example, by investigating workers exposed to monofunctional LMW chemicals such as monoisocyanates. The literature search identified no reports of monoisocyanates causing asthma. Moreover a study of workers at a chemical facility could not detect any deterioration in lung function from long term exposure to methyl isocyanate.[29]

It could be argued that among the few fragments for which a single functional group occurrence is significantly associated with reported occupational asthma hazard is the (cyclic) acid anhydride fragment. Although acid anhydrides contain two carbonyl groups and in this respect may be considered bifunctional, these two groups are not independent and it has been suggested that they link to free amino groups in lysine residues to form hapten-protein adducts.[30 31]

A mechanistic hypothesis based on the need for chemical bifunctionality is not inconsistent with the immunobiochemistry of antibody-allergen recognition. Immunologically recognisable epitopes could be exposed as a result of covalent cross-linking of otherwise non-antigenic ''self'' proteins. Whether this mechanism is important through

Table A1 Model development asthmagens (literature pre-1995)

CAS no.	Chemical name	Hazard index	CAS no.	Chemical name	Hazard index
50-00-0	**Formaldehyde**	**1**	**141-43-5**	**Ethanolamine**	**0.64**
52-67-5	**Penicillamine**	**0.97**	148-01-6	2-methyl-3,5-dinitrobenzamide	0.32
54-85-3	**Isoniazid**	**0.79**	**304-20-1**	**Hydralazine**	**0.78**
55-38-9	Fenthion	0.01	305-80-6	Pauli's Reagent (4-diazobenzenesulphonic acid)	0.31
55-56-1	**Chlorhexidine**	**1**	**317-34-0**	**Aminophylline**	**0.99**
60-54-8	**Tetracycline**	**1**	**514-10-3**	**Abietic acid**	**0.94**
64-19-7	Acetic acid	0.15	**551-16-6**	**6-aminopenicillanic acid**	**1**
69-52-3	**Ampicillin**	**1**	**552-30-7**	**Trimellitic anhydride**	**1**
69-57-8	**Benzyl penicillin**	**1**	**555-30-6**	**Methyl DOPA**	**0.94**
70-30-4	Hexachlorophene	0.01	**561-27-3**	**Diacetyl morphine**	**0.5**
72-14-0	Sulfathiazole	0.37	**584-84-9**	**Toluene diisocyanate**	**1**
75-21-8	Ethylene oxide	0.01	**822-06-0**	**Hexamethylene diisocyanate**	**1**
80-62-6	Methylmethacrylate	0.48	**957-68-6**	**7-aminocephalosporanic acid**	**1**
85-42-7	**Hexahydrophthalic anhydride**	**0.95**	1897-45-6	Tetrachloroisophthalonitrile	0.11
85-44-9	**Phthalic anhydride**	**1**	**2746-19-2**	**Himic anhydride**	**0.95**
88-04-0	Chloroxylenol	0.02	**2939-80-2**	**Captafol**	**0.52**
89-32-7	**Pyromellitic dianhydride**	**1**	**3173-72-6**	**1,5-naphthalene diisocyanate**	**1**
98-00-0	Furfuryl alcohol	0.01	**3779-63-3**	**1,3,5-tris-(6-isocyanato-hexyl)-<1,3,5>triazinane-2,4,6-trione**	**1**
100-37-8	**2-Diethylethanolamine**	**0.86**	**4035-89-6**	**Biuret of hexamethylene diisocyanate**	**1**
100-42-5	Styrene	0.04	**4098-71-9**	**Isophorone diisocyanate**	**1**
100-97-0	**Hexamethylene tetramine**	**0.9**	**7085-85-0**	**Ethyl cyanoacrylate**	**0.75**
101-68-8	**Diphenylmethane di-isocyanate**	**1**	7531-92-2	Plicatic acid	0.01
106-50-3	Paraphenylene diamine	0.22	**8025-81-8**	**Spiramycin 1**	**1**
107-15-3	Ethylenediamine	0.27	13466-78-9	3-carene	0.18
108-01-0	**Dimethylethanolamine**	**0.77**	**15686-71-2**	**Cephalexin**	**1**
108-31-6	**Maleic anhydride**	**0.85**	**28983-56-4**	**Methyl blue**	**0.92**
109-02-4	**N-methylmorpholine**	**0.54**	**31330-63-9**	**Tetrazene**	**1**
109-55-7	3-dimethylaminopropylamine	0.47	**36519-31-0**	**Salbutamol**	**0.84**
110-85-0	Piperazine	0.36	39878-87-0	Phenylglycine acid chloride	0.18
111-30-8	**Glutaraldehyde**	**0.82**	49745-95-1	Dobutamine HCl	0.17
111-41-1	**Aminoethylethanolamine**	**0.92**	**51481-61-9**	**Cimetidine**	**1**
112-24-3	**Trientine**	**0.93**	**61336-70-7**	**Amoxicillin**	**1**
117-08-8	**Tetrachlorophthalic anhydride**	**1**	123354-92-7	Sodium iso-nonanoyloxybenzene sulphonate	0.48
117-81-7	**Dioctylphthalate**	**1**		Black GR Reactive Dye (BK-5)	0.04
121-25-5	**Amprolium hydrochloride**	**0.95**		**Orange-GR reactive dye**	**0.99**
123-31-9	Hydroquinone	0.01		**Red-BBN reactive dye**	**1**
123-77-3	**Azodicarbonamide**	**0.92**		**Glycyl compound**	**0.8**
124-04-9	**Adipic acid**	**0.75**		**MM22383**	**1**
127-65-1	Chloramine-T	0.22			
137-05-3	**Methyl-2-cyanoacrylate**	**0.69**			

intra- or inter-molecular cross-linking requires further investigation. Among the compounds for which there is cross-linking potential are the polyamines and the polyisocyanates, both of which may cause sensitisation without any observed IgE response despite a clinical picture, which tends to be consistent with an allergic response. The strong possibility remains that some important chemically bifunctional asthmagens may sensitise via a non-classical IgE mechanism, perhaps by causing membrane damage or cross-linking cell surface receptors. This may result in asthma directly and/or in an increased propensity to develop asthma perhaps because of persistent epithelial damage. These hypotheses warrant further investigations.

At present the data are not adequate to permit robust comparisons in chemical substructure distribution between those asthmagens that are apparently mediated by an IgE mechanism and those that are not. However, the model development data is consistent with published observations that specific IgE is often detected in cases due to acid anhydrides and reactive dyes but rarely when the cause is a diisocyanate or bifunctional aldehyde, such as glutaraldehyde. It is hoped that further formal scientific investigation will corroborate the findings and interpretations discussed here and additionally that practitioners can participate in the iteration of the web based program and comment on its usefulness.

ACKNOWLEDGEMENTS

James Jarvis was supported by a Colt Foundation Fellowship supervised by Raymond Agius and Lindsay Sawyer. The Surveillance of Work-related and Occupational Respiratory Disease (SWORD) project is funded by the Health & Safety Executive. The views expressed here are those of the authors and not of the funding bodies.

Authors' affiliations

J Jarvis, University Computing Services, The University of Edinburgh, UK
M J Seed, R Agius, Centre for Occupational & Environmental Health, The University of Manchester, UK
R Elton, The University of Edinburgh, Public Health Sciences, Edinburgh, UK
L Sawyer, School of Biological Sciences, The University of Edinburgh, Edinburgh, UK

Competing interests: the authors do not have any competing interests to declare

Ethics approval: the only human data used in this study was that already published in peer reviewed literature or reported to the SWORD scheme (and which has been approved by the relevant Multi-Centre Research Ethics Committee)

APPENDIX

Model development asthmagens are listed in table A1 and external validation asthmagens in table A2.

Table A2 External validation asthmagens (literature post-1995)

CAS no.	Chemical name	Hazard index
52-26-6	Morphine hydrochloride	0.08
111-42-2	**Diethanolamine**	**0.98**
115-27-5	**Chlorendic anhydride**	**0.86**
485-47-2	**Ninhydrin**	**1**
860-22-0	**Indigotine**	**0.89**
2451-62-9	**Triglycidyl isocyanurate**	**1**
2634-33-5	1,2-benzisothiazolin-3-one	0.14
7696-12-0	**Tetramethrin**	**0.97**
8001-54-5	**Benzalkonium chloride**	**0.98**
19438-64-3	**Methyltetrahydrophthalic anhydride**	**0.95**
38661-72-2	**1,3-Bis(isocyanatomethyl)cyclohexane**	**1**
52185-43-0	**Piperidinyl chlorotriazine derivative**	**1**
56796-20-4	**Cefmetazole**	**1**
59703-84-3	**Piperacillin**	**1**
64265-57-2	**polyfunctional aziridine**	**1**
64401-02-1	**Diacrylate**	**0.98**
65271-80-9	**Mitoxanthrone**	**0.95**
66592-87-8	**Cefadroxil**	**1**
72558-82-8	**Ceftazidime**	**1**
79622-59-6	Fluazinam	0
82547-81-7	**Cefteram pivoxil**	**1**
125700-67-6	**2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate**	**0.96**

REFERENCES

1 **Vandenplas O**, Malo JL. Definitions and types of work-related asthma: a nosological approach. *Eur Respir J* 2003;**21**:706–12.
2 **Bernstein DI**. Occupational asthma caused by exposure to low-molecular-weight chemicals. *Immunol Allergy Clin North Am* 2003;**23**:221–34, vi.
3 **van Kampen V**, Merget R, Baur X. Occupational airway sensitisers: an overview on the respective literature. *Am J Ind Med* 2000;**38**:164–218.
4 **Wild L**, Lopez M. Occupational asthma caused by high-molecular-weight substances. *Immunol Allergy Clin North Am* 2003;**23**:235–50.
5 **Fairhurst S**. Hazard and risk assessment of industrial chemicals in the occupational context in Europe: some current issues. *Food Chem Toxicol* 2003;**41**:1453–62.
6 **Dearman RJ**, Skinner RA, Humphreys NE, *et al*. Methods for the identification of chemical respiratory allergens in rodents: comparisons of cytokine profiling with induced changes in serum IgE. *J Applied Toxicol* 2003;**23**:199–207.
7 **Agius RM**, Nee J, McGovern G, *et al*. Structure activity hypotheses in occupational asthma caused by low molecular weight substances. *Ann Occup Hyg* 1991;**35**:129–37.
8 **Agius RM**, Elton RA, Sawyer L, *et al*. Occupational asthma and the chemical properties of low molecular weight organic substances. *Occup Med* 1994;**44**:34–6.
9 **Agius RM**. Why are some low-molecular-weight agents asthmagenic? *Occup Med (Rev)* 2000;**15**:369–83.
10 **Karol MH**, Graham C, Gealy R, *et al*. Structure-activity relationships and computer-assisted analysis of respiratory sensitisation potential. *Toxicol Lett* 1996;**86**:187–91.
11 **Graham C**, Rosenkranz HS, Karol MH. Structure-activity model of chemicals that cause human respiratory sensitisation. *Regul Toxicol Pharmacol* 1997;**26**:296–306.
12 **Karol MH**, Macina OT, Cunningham A. Cell & molecular biology of chemical allergy. *Ann Allergy Asthma Immunol* 2001;**87**(suppl):28–32.
13 **Barratt MD**, Rodford RA. The computational prediction of toxicity. *Curr Opin Chem Biol* 2001;**5**:383–8.
14 **Pearl GM**, Livingston-Carr S, Durham S. Integration of computational analysis as a sentinel tool in toxicological assessments. *Curr Top Med Chem* 2001;**1**:247–55.
15 **ECETOC**. European Centre for Ecotoxicology of Chemicals. Technical Report No.89. *(Q)SARS: evaluation of the commercially available software for human health and environmental endpoints with respect to chemical management applications*. Brussels, September 2003. ISSN-0773-8072-89.
16 **Health & Safety Executive**. EH40/94. *Occupational exposure limits 1994*, HSE, HMSO, UK. ISBN 0-7176-0722-4.
17 **Health & Safety Executive**. EH40/2002. *Occupational exposure limits 2002*. HSE, HMSO, UK. ISBN 0-7176-2194-4.
18 **American Conference of Governmental Industrial Hygienists**. *Threshold limit values and biological exposure indices*. ACGIH, 2002.
19 **Dalby A**, Nourse JG, Hounshell WD, *et al*. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. *J Chem Inf Comput Sci* 1992;**32**:244–55.
20 **Armitage P**, Berry G. *Statistical methods in medical research,* 3rd edn. Oxford: Blackwell Scientific Publications, 1994:443.
21 **Meyer JD**, Holt DL, Cherry NM, *et al*. SWORD '98: surveillance of work-related and occupational respiratory disease in the UK. *Occup Med (Oxf)* 1999;**49**:485–9.
22 **Sarlo K**, Ritz HL. Predictive assessment of respiratory sensitising potential in guinea pigs. In: Kimber I, Dearman RJ, eds. *Toxicology of chemical respiratory hypersensitivity*. London: Taylor and Francis, 1997:107–20.
23 **Kimber I**, Pichowski JS, Betts CJ, *et al*. Alternative approaches to the identification and characterisation of chemical allergens. *Toxicology in Vitro* 2001;**15**:307–12.
24 **Menard S**. Stepwise logistic regression. In: Lewis-Beck MS, ed. *Applied logistic regression*. Thousand Oaks: Sage University Press, 1995:54–7.
25 **Savonius B**, Keskinen H, Tuppurainen M, *et al*. Occupational respiratory disease caused by acrylates. *Clin Exp Allergy* 1993;**23**:416–24.
26 **Piipari R**, Tuppurainen M, Tuomi T, *et al*. Diethanolamine-induced occupational asthma, a case report. *Clin Exp Allergy* 1998;**28**:358.
27 **Terpstra GK**, Raaijmakers JAM, Wassink GA. Propranolol-induced bronchoconstriction: a non-specific side-effect of B-adrenergic blocking therapy. *Eur J Pharmacol* 1981;**73**:107–8.
28 **Nasser SMS**, Ewan PW. Opiate-sensitivity: clinical characteristics and the role of skin prick testing. *Clin Exp Allergy* 2001;**31**:1014–22.
29 **Avashia B**, Battigelli MC, Morgan WKC, *et al*. Effects of prolonged low exposure to methyl isocyanate. *J Occup Environ Med* 1996;**38**:625–30.
30 **Zeiss CR**. Advances in acid anhydride induced occupational asthma. *Curr Opin Allergy Clin Immunol* 2002;**2**:89–92.
31 **Venables KM**. Low molecular weight chemicals, hypersensitivity, and direct toxicity: the acid anhydrides. *Br J Ind Med* 1989;**46**:222–32.

Relationship between chemical structure and the occupational asthma hazard of low molecular weight organic compounds

J Jarvis, M J Seed, R Elton, et al.

Occup Environ Med 2005 62: 243-250
doi: 10.1136/oem.2004.016402