Hypersensitivity Pneumonitis in Workers Exposed to Metalworking Fluids

Christopher M. Barber, MD,1 Clare M. Burton, MD,1 David J. Hendrick, MD,2 C Anthony C. Pickering, MD,3 Alastair S. Robertson, MD,4 Wendy Robertson, PhD,5 and P Sherwood Burge, MD6

Background This study used data from a large UK outbreak investigation, to develop and validate a new case definition for hypersensitivity pneumonitis due to metalworking fluid exposure (MWF-HP).

Methods The clinical data from all workers with suspected MWF-HP were reviewed by an experienced panel of clinicians. A new MWF-HP Score was then developed to match the “gold standard” clinical opinion as closely as possible, using standard diagnostic criteria that were relatively weighted by their positive predictive value.

Results The new case definition was reproducible, and agreed with expert panel opinion in 30/37 cases. This level of agreement was greater than with any of the three previously utilized case definitions (agreement in 16–24 cases). Where it was possible to calculate, the MWF-HP Score also performed well when applied to 50 unrelated MWF-HP cases.

Conclusions The MWF-HP Score offers a new case definition for use in future outbreaks.

KEY WORDS: hypersensitivity pneumonitis; diagnostic criteria; metalworking fluid

INTRODUCTION

Hypersensitivity pneumonitis (HP) is an allergic lung disease, most commonly caused by inhalation of proteins or chemicals at home or in the workplace. The presenting symptoms of HP can be non-specific, and vary with acute, sub-acute or chronic forms of disease.

A new cause of HP was first recognized in the United States in 1990s [Bernstein et al., 1995; Kreiss and Cox-Ganser, 1997], and despite a number of detailed workplace investigations, cases continue to be reported due to exposure to metalworking fluid mists (MWFs) [Rosenman, 2009]. Modern MWFs are complex mixtures of oil and water that are used as coolants and lubricants in a wide-range of industries.

Although the exact cause of the outbreaks remains elusive [Burton et al., 2012], it is thought that the re-circulation of the MWFs leading to microbiological contamination is most likely to be responsible [Barber et al., 2014].
et al., 2011, 2013]. The outbreaks of MWF-HP that have been reported from the US [Mirer, 2010], UK [Robertson et al., 2007], and France [Tillie-Leblond et al., 2011], have usually involved workplaces with several hundred exposed workers. Investigating the cause of symptoms among such a large workforce is logistically challenging, and a range of case definitions have therefore previously been developed and utilized [Zacharisen et al., 1998; Fox et al., 1999; Hodgson et al., 2001; Dangman et al., 2002; Weiss et al., 2002; Gupta and Rosenman, 2006; Robertson et al., 2007]. Little comparative data exists however [Dangman et al., 2002], and no single case definition has emerged as an accepted standard to allow comparisons between future outbreaks [Barber et al., 2012].

The aim of this study was to use existing data from a large UK investigation of an outbreak [Robertson et al., 2007] to develop a simple and inclusive case definition for MWF-HP, and attempt to validate it by applying it to cases from unrelated outbreaks.

MATERIALS AND METHODS

In the absence of any validated method of diagnosing MWF-HP, this study utilized the clinical opinion of an expert panel of occupational respiratory disease consultants as the diagnostic “gold standard.”

Case Selection

In 2003–2004, a large outbreak of respiratory ill health was investigated in three phases, at Powertrain, an automobile engine manufacturing plant in Birmingham UK. In total 37,808 workers were suspected of having MWF-HP and/or humidifier fever, with 19 workers going on to meet a case definition for HP [Robertson et al., 2007]. The main study population for this project comprised all of the 37 workers recorded as having suspected HP and/or humidifier fever), following assessment by the clinical team at Birmingham Chest Clinic during Phase 3 of the outbreak investigation (Fig. 1).

Selection of Expert Panel

The panel comprised five consultants selected from the UK Group of Occupational Respiratory Disease Specialists (GORDS). This included two members (PSB/ASR) who had been involved both in the outbreak investigation, and clinical follow-up of affected workers at Birmingham Chest Clinic. The third member (CACP) had assisted in the outbreak investigation, but not the clinical follow-up of affected workers. The fourth and fifth members (DJH/CMB) had not participated in the investigation or follow-up of workers.

Expert Panel Meeting

The panel reviewed hospital records and clinical data documented during the outbreak investigation and subsequent hospital follow-up for each worker. This included predicted lung function measures previously calculated utilizing the European Steel and Coal Community reference equations [Cotes et al., 1993; Quanjer et al., 1993]. Following discussion of each case, individual panel member independently recorded their clinical opinion as to whether there was sufficient clinical evidence to classify the patient as a definite case, a possible case, or definitely not a case of MWF-HP. The panel was not provided with any information during the meeting relating to whether each case had previously met the case definition of HP used at the time of the outbreak investigation. A definite clinical case of MWF-HP required at least four of the panel members to agree. Similarly MWF-HP was excluded where at least four panel members agreed it was definitely not a case. Any other combination of opinions resulted in a possible case of MWF-HP. An Expert Panel Percentage (EPP range 0%–100%) was calculated for each case, representing how likely it was that this represented MWF-HP. This was calculated based on the total of five scores, where each definite opinion equaled 20%, each possible opinion equaled 10%, and each definitely not opinion equaled 0%.

Reproducibility

In order to assess the reproducibility of the panel opinion, 10 randomly selected cases were re-presented for review, with the panel being blinded to their previous opinion. Results from the first and second opinion were compared. Reproducibility of the EPP was calculated using the concordance correlation coefficient.

MWF-HP Case Definition Development

Based on the opinion of the panel, the 37 workers with suspected MWF-HP during the outbreak investigation, were divided into two groups based on whether they had been classified as a definite clinical case, or not. Demographic and clinical data were compared between these groups. Continuous data were compared using Student t-tests, and categorical data were compared using Fisher’s Exact tests.

Utilizing the clinical differences found between these two groups of workers, a new diagnostic case definition (the MWF-HP Score) was developed. The final model was selected to fit as closely as possible with the panel opinion, whilst remaining clinically valid and inclusive. To facilitate this, each diagnostic element was weighted based on its positive predictive value for MWF-HP as defined by the panel opinion. The new scoring system was then applied to the 37
FIGURE 1. Outbreak investigation and worker selection criteria for expert panel review.

Phase 1. 11 point screening questionnaire
N=808/836

Phase 2. Detailed questionnaire 108 questions, spirometry and blood sampling.
N=510

Phase 2. Clinical impression of possible EAA and/or humidifier fever.
N= 101

Phase 3. Clinical review and testing at Birmingham Chest Clinic
N= 78

Clinical suspicion of EAA and/or humidifier fever
N=37

Expert panel review
N=14 definite cases of EAA
workers, and compared with the EPP. Suitable cut-offs for definite, possible and definitely not MWF-HP were chosen, to best match the opinion of the panel.

Validation of MWF-HP Score

The performance of the MWF-HP Score and three other previously published case definitions were compared by applying each of them to the 37 workers reviewed [Fox et al., 1999; Dangman et al., 2002; Robertson et al., 2007]. Performance was assessed by comparing the proportion of workers correctly identified as definite, possible and definitely not MWF-HP, against the panel opinion. In an attempt to externally validate the MWF-HP Score, it was also applied to 50 previously published US case reports of workers developing HP due to MWF exposure [Bernstein et al., 1995; Trout et al., 1996; Kreiss and Cox-Ganser, 1997; Zacharisen et al., 1998; Fox et al., 1999; Hodgson et al., 2001; Dangman et al., 2002; Weiss et al., 2002; Trout et al., 2003; Gupta and Rosenman, 2006]. These cases had previously been identified during a separate systematic literature review of all previously published MWF ill health outbreaks [Burton et al., 2012]. MWF-HP Scores, or the range of possible scores, were calculated, depending on the level of detail presented in the case reports.

Ethics Approval

The data utilized in developing the MWF-HP Score was anonymised, and had been collected as part of a health investigation carried out by the Health and Safety Executive. The Birmingham Ethics Committee approved the study to proceed, and written informed consent from individual workers was not required.

RESULTS

Participants

The panel classified 14 workers as definite cases of MWF-HP (all had an EPP of 100%), 12 as possible cases (EPP 20–80%), and 11 as definitely not cases of MWF-HP (EPP 0–10%). The reproducibility of the Expert Panel Percentage was good, with a concordance correlation coefficient of 0.98, with a mean difference of 5% (range 0–20%). More detail is provided in the online Supplemental Material.

Demographics of Workers With and Without MWF-HP

No statistically significant demographic differences were found between the 14 workers with definite MWF-HP, the other 23 workers reviewed by the panel, and the remaining workforce (who had taken part in the original investigation due to reporting symptoms, but had no clinical suspicion of MWF-HP). Smoking prevalence was lower amongst those with definite MWF-HP, but did not reach statistical significance (Table I).

Significant clinical differences between workers with and without definite MWF-HP are shown in Table II, and where available, comparative figures for the remaining workforce are also shown. CT findings compatible with HP included centrilobular nodularity, ground glass, mosaic change, or established interstitial fibrosis (including usual interstitial pneumonitis pattern). Biopsy findings compatible with HP included granulomas, interstitial expansion, lymphocytic alveolitis, or established usual interstitial pneumonitis [Girard et al., 2009]. Further details of all of the clinical data comparisons can be found in the online Supplemental Material.

Development of MWF-HP Score and Case Definition

The MWF-HP Score developed is shown in Table III, where the highest score from each of the five sections is applied, and then added together up to a maximum of 41 points. Individual scores calculated for each of the 37 workers with suspected MWF-HP in the outbreak are shown in the online Supplemental Material.
Comparison of Expert Panel Opinion and MWF-HP Score

There was a good level of correlation between the derived MWF-HP Score, and the Expert Panel Percentage, with a Pearson correlation coefficient of 0.85 (P < 0.01). By using suitable cut-offs for the MWF-HP Scores (definite case > 26, possible case 19–26, and definitely not a case < 19), it was possible to show agreement (shown in bold) with the panel opinion in 30/37 (81%) cases (Table IV). The level of agreement of three other MWF-HP case definitions with the Expert Panel opinion was 59% [10], 43% [12], and 65% [7].

External Validation

Figure 2 shows the MWF-HP Score, or range of possible scores, for 50 previously published cases of MWF-HP from a range of workplaces in the United States [Bernstein et al., 1995; Trout et al., 1996; Kreiss and Cox-Ganser, 1997; Zacharisen et al., 1998; Fox et al., 1999; Hodgson et al., 2001; Dangman et al., 2002; Weiss et al., 2002; Trout et al., 2003; Gupta and Rosenman, 2006]. In 22/50 of these cases, the MWF-HP Score or score range would have identified them as a definite HP case. In the remaining cases it was not possible to assess this for certain, due to the range of possible scores. Taking the mid-point of the score range as an approximation of the actual MWF-HP Score, a further 20 cases would also have been identified as definite MWF-HP. The calculated scores or range of possible scores for each case are shown in the online Supplemental Material.

DISCUSSION

Study Findings

This study developed a new case definition for assessing workers with suspected MWF-HP, based on the evidence-
The generation of this was novel, in that each diagnostic element was weighted based on its positive predictive value for HP in the outbreak. It has been designed to be simple to use, reproducible, inclusive of different types of clinical presentation, and to match expert opinion as closely as possible. Although difficult to truly validate, the MWF-HP Score performed well both in comparison to other published case definitions, and when applied to previously published MWF-HP cases from unrelated outbreaks.

**Study Limitations**

Due to the lack of a single diagnostic test for HP [Girard et al., 2009], we used the consensus clinical opinion of five UK occupational respiratory disease specialists as our gold standard. Their final opinion was only made after review of all available clinical information, and a face-to-face case discussion, aimed at replicating clinical practice. This opinion showed good reproducibility, with unanimous agreement among the panel for all of the definite cases of MWF-HP. The study aimed to be inclusive and the panel also reviewed cases of suspected humidifier fever, in order not to miss cases of acute HP, with similar constitutional symptoms. The main limitation of the study design was that the panel opinion could only be based on retrospective review of data documented during the outbreak investigation and subsequent outpatient visits, rather than data taken directly and contemporaneously from a clinic patient. In particular, there may have been logistical delays between the workplace investigation, and obtaining certain hospital investigations (such as CT scans), which may have affected their predictive value, if recovery had occurred in that time interval. It is also possible that in some cases MWF-HP could not be confirmed by the panel due to a lack of clinical information, for example, where patients declined certain tests, or did not attend follow-up appointments.

Another study limitation related to the assumptions that had to be made in applying other case definitions to the study population, where the exact wording of symptoms may have varied, and some of the diagnostic tests (e.g., gallium scan, alveolar-arterial gradient, and erythrocyte sedimentation rate) had not been performed [Fox et al., 1999; Dangman et al., 2002; Robertson et al., 2007]. This was particularly relevant to the hypersensitivity pneumonitis diagnostic index (HPDI) [Dangman et al., 2002], which is likely to have performed better if the same range of tests had been carried out in the Powertrain investigation [Robertson et al., 2007].

Table III. MWF-HP Score for Workers With Suspected MWF-HP

<table>
<thead>
<tr>
<th>Diagnostic element</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiratory symptoms</td>
<td></td>
</tr>
<tr>
<td>Work-related cough/wheeze/dyspnoea/cough</td>
<td>+4</td>
</tr>
<tr>
<td>at own pace on level ground</td>
<td>+6</td>
</tr>
<tr>
<td>Previous time off work with any chest illness</td>
<td>+7</td>
</tr>
<tr>
<td>Constitutional symptoms</td>
<td></td>
</tr>
<tr>
<td>Recurrent flu-like symptoms worse at the end of the working week</td>
<td>+5</td>
</tr>
<tr>
<td>Unexplained weight loss</td>
<td>+7</td>
</tr>
<tr>
<td>Physiology</td>
<td></td>
</tr>
<tr>
<td>FVC &lt; 80% predicted</td>
<td>+3</td>
</tr>
<tr>
<td>FVC &lt; 70% predicted and/or Tlco &lt; 80% predicted</td>
<td>+5</td>
</tr>
<tr>
<td>Tlco &lt; 60% predicted</td>
<td>+10</td>
</tr>
<tr>
<td>Radiology/clinical examination</td>
<td></td>
</tr>
<tr>
<td>Abnormal CXR (diffuse ground)</td>
<td>+6</td>
</tr>
<tr>
<td>Abnormal HRCT (ground glass, noddularity)</td>
<td>+7</td>
</tr>
<tr>
<td>Fine end-inspiratory crepitations on auscultation</td>
<td>+7</td>
</tr>
<tr>
<td>Evidence of inflammation</td>
<td></td>
</tr>
<tr>
<td>Neutrophilia &gt; 7 x 10^9/l or CRP &gt; 10 mg/l</td>
<td>+5</td>
</tr>
<tr>
<td>BAL lymphocytosis ≥ 20%</td>
<td>+8</td>
</tr>
<tr>
<td>Lung biopsy typical of HP</td>
<td>+10</td>
</tr>
<tr>
<td>(sub-acute HP or usual interstitial pneumonitis)</td>
<td></td>
</tr>
<tr>
<td>Total (max 41)</td>
<td></td>
</tr>
</tbody>
</table>

Table IV. Comparison Between MWF-HP Case Definitions, Versus Panel Opinion

<table>
<thead>
<tr>
<th>Expert panel opinion</th>
<th>Fox et al. [1999]</th>
<th>Dangman et al. [2002]</th>
<th>Robertson et al. [2007]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Definite</td>
<td>Possible</td>
<td>Not a case</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

Finally, the external validation of the MWF-HP Score was limited by the lack of detail found in some of the...
published case series from the United States. An exact score could only be calculated in 4 of the 50 cases, and for the remainder it was only possible to calculate a range of possible scores based on the information provided.

**Comparison With Other Work**

A number of general HP diagnostic criteria are widely used [Terho, 1986; Richerson et al., 1989; Cormier and Lacasse, 1996; Schuyler and Cormier, 1997], but little or no comparative data relating to their validity has been published [Girard et al., 2009]. The HP Study Group has however published a prediction rule, applicable to patients presenting to hospital with possible HP [Lacasse et al., 2003]. The six significant predictive factors for HP in this study were found to be; two different symptom patterns; weight loss; inspiratory crackles; positive precipitins; and exposure to a known cause. In addition, their study highlighted the importance of a reduced pulmonary gas transfer in HP [Girard et al., 2009]. The findings in our study relating to the high predictive value of unexplained weight loss, and of a reduced gas transfer in MWF-HP are therefore in keeping with the findings of this international multi-center cohort study. The HP prediction rule however was not developed for use in outbreaks, and is likely to be of limited usefulness where all workers are exposed to a potential cause. In addition, standardized serum IgG precipitin testing for MWF-HP is still in development [Tillie-Leblond et al., 2011].

A number of specific case definitions for MWF-HP have also been published, following recognition of outbreaks in exposed workers. The most widely used of these comprised seven diagnostic elements, resulting in definite (6–7 points), probable (5 points) and possible (4 points) cases of MWF-HP [Fox et al., 1999]. The authors compared 34 demographic risk factors between 18 cases of MWF-HP and 51 controls from the same workplace, but no clear risk factors were identified. The authors acknowledged that their conclusions may have been limited due to the epidemiological case definition not having been validated prior to usage. The findings in our study were however in keeping with this study, and unable to demonstrate clear demographic risk factors for MWF-HP.

A further case definition (the HPDI) was developed from outbreak data, by comparing data from 16 cases of biopsy proven MWF-HP, with that from the 14 workers in the outbreak thought least likely to have the disease [Dangman et al., 2002]. Perhaps not surprisingly, the workers with MWF-HP had more symptoms, higher inflammatory markers, lower gas transfers, higher alveolar-arterial oxygen gradients, lower vital capacities, more abnormal CT scans, and more abnormal gallium scans. The analysis in our study was broadly similar in principal, although we used the more inclusive panel clinical opinion (which included biopsy findings where available) as our gold standard for MWF-HP diagnosis. Our study comparison groups were also different, as we used a more clinically relevant scenario, looking for clinical differences between workers with definite MWF-HP, and workers without definite MWF-HP (many having occupational asthma or humidifier fever), within a group who had all previously been suspected of having HP during an outbreak investigation. As a result of this, our study found

![Figure 2: MWF-HP Scores (or score range) for 50 US workers with HP due to MWF.](image-url)
less marked differences in symptoms between the two groups of workers, with the only statistically significant differences being for MRC Grade 4 breathlessness, and unexplained weight loss. Other constitutional symptoms such as fever, shivering, tiredness, weakness, joint, or muscle pains were not predictive of MWF-HP within the group studied, as these symptoms were also common in the other workers.

Dangman et al. [2002] found elevated levels of erythrocyte sedimentation rate (ESR) in workers with MWF-HP, but not neutrophil count or C-reactive protein (CRP). ESR was not measured during the UK outbreak, but we did find significantly higher mean levels of total white cell and neutrophil counts in workers with MWF-HP. Although mean CRP was also higher in the workers with MWF-HP, it was only available in half of the workers studied, and the lack of a significant difference may reflect a type 2 error. The clinical findings in sub-acute HP may vary with exposure, and the most useful marker of inflammation in an outbreak is likely to be dependant on the logistics of where and when the blood sampling takes place (e.g., workplace versus outpatient clinic), based on ease of processing, and the half-life of the inflammatory marker selected.

The diagnostic criteria selected by Dangman et al. [2002] were designed to avoid the need for invasive biopsies, and were based on scoring points for; symptoms; crackles; abnormal pulmonary physiology; raised inflammatory markers; and abnormal radiology (on CXR, CT, or gallium scan). This formed the basis of the HPDI Score with 6–9 being a definite case, 4–5 a probable case, and 3 a possible case. Although similar in concept to the Fox criteria [Fox et al., 1999], Dangman et al. [2002] applied some weighting to two of the criteria, so that more abnormal results contributed twice as many points to the total. They also attempted to validate the HPDI by applying it to the other 31 workers from their outbreak, and demonstrating that their case definition agreed with that developed by Fox et al. [1999] in 55 of the 61 workers. The lack of an agreed “gold standard” method of diagnosis in HP however, makes any form of validation difficult.

The diagnostic criteria in our study were more heavily weighted, based on their positive predictive value for MWF-HP (range 30–100%), giving each a possible score of 3–10. The diagnostic elements selected were chosen to be as inclusive as possible, including both invasive and non-invasive test results, and aimed at identifying the full range of possible types of HP. As examples, exertional breathlessness was included to identify workers with chronic HP (without possible types of HP. As examples, exertional breathlessness was included to identify workers with chronic HP (without

### Study Implications

Investigating any ill health outbreak in large workplaces is logistically challenging, due to the unexpected nature of the outbreak, and the need to screen a large number of workers in a short period [Barber et al., 2012]. This study has developed and attempted to validate a new case definition for HP, and although it has been developed using data from workers exposed to MWFs, it may also be useful in other similar types of HP outbreaks [Ganier et al., 1980; Rose et al., 1998; Iossifova et al., 2011].

### ACKNOWLEDGMENTS

This study was funded by the Health and Safety Executive. WR is part funded by the National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Health Research and Care West Midlands (CLAHRC-WM) initiative.

### REFERENCES


**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of this article at the publisher’s web-site.