

ORIGINAL ARTICLE

# A principal factor analysis to characterize agricultural exposures among Nebraska veterans

Lisa Weissenburger-Moser<sup>1</sup>, Jane Meza<sup>2</sup>, Fang Yu<sup>2</sup>, Oyewale Shiyabola<sup>3</sup>, Debra J. Romberger<sup>4</sup> and Tricia D. LeVan<sup>1,4</sup>

Agricultural workers are at an increased risk of developing chronic respiratory disorders. Accurate estimation of long-term agricultural exposures based on questionnaires has been used to improve the validity of epidemiologic investigations and subsequent evaluation of the association between agricultural exposures and chronic diseases. Our aim was to use principal factor analysis (PFA) to distill exposure data into essential variables characterizing long-term agricultural exposures. This is a cross-sectional study of veterans between the ages of 40 and 80 years and who worked on a farm for  $\geq 2$  years. Participant characteristics were: 98.1% were white males with a mean age  $65 \pm 8$  (SD) years and 39.8% had chronic obstructive pulmonary disease. The final model included four factors and explained 16.6% of the variance in the exposure data. Factor 1 was a heterogeneous factor; however, Factor 2 was exclusively composed of exposure to livestock such as hogs, dairy and poultry. Factor 3 included exposures from jobs on or off the farm such as wood dust, mineral dust, asbestos and spray paint. Crop exposure loaded exclusively in Factor 4 and included lifetime hours of exposure and maximum number of acres farmed in the participants' lifetime. The factors in the final model were interpretable and consistent with farming practices.

*Journal of Exposure Science and Environmental Epidemiology* (2017) **27**, 214–220; doi:10.1038/jes.2016.20; published online 6 April 2016

**Keywords:** agriculture; factor analysis; occupational; principal factor analysis

## INTRODUCTION

Agricultural workers are at an increased risk of developing chronic respiratory disorders including chronic bronchitis, occupational asthma and obstructive lung disease, and these diseases are likely caused by multiple agricultural exposures.<sup>1</sup> Epidemiological studies can provide evidence of an exposure–response relationship, an important factor for the suggestion of a causal association.<sup>2</sup> However, causal inference from epidemiological studies of chronic disease in agricultural populations is often limited because of a lack of long-term exposure measurements.<sup>3</sup> The types of methods used to assess agricultural exposures have included direct measurement of personal exposure,<sup>4,5</sup> biomarkers of exposure<sup>6,7</sup> and self-report questionnaires.<sup>8,9</sup> Although the direct measurement method is often a precise approach, it may not be relevant for studies of disease with long latency periods such as obstructive lung diseases. Accurate estimation of long-term agricultural exposures based on questionnaire data has been used to improve the validity of epidemiologic investigations and subsequent evaluation of the association between agricultural exposures and chronic lung diseases.<sup>8</sup> The questionnaire is usually designed to ask a large set of questions about agricultural tasks and exposures with the purpose of obtaining enough information for chronic exposure assessment. However, oftentimes the designed questions are not direct indicators of the true exposure. Sorting out useful information from the large amount of questionnaire data is challenging, yet essential in obtaining

objective, unbiased and interpretable exposure assessments in an epidemiological study.

Here we use a statistical method, principal factor analysis (PFA), to summarize a large amount of important agricultural exposure variables from questionnaires designed to assess the relationship between agricultural exposures and respiratory disease. PFA is a statistical method that has been proposed to characterize heterogeneous exposures when exposure monitoring is unavailable and short-term exposure measurements are inadequate.<sup>10,11</sup> To our knowledge, there has been no assessment of agricultural exposures, such as animals, crops and farm tasks based on factor analysis.

Our overall objective was to identify a set of essential agriculturally related exposures that should be considered when assessing respiratory outcomes. Using data from a cross-sectional study of veterans who worked on a farm or in production agriculture as an adult for  $\geq 2$  years, we applied the method of factor analysis to two questionnaires. Questionnaire 1 (Q1) assessed agricultural exposures in 263 individuals, and Questionnaire 2 (Q2, extended version) evaluated exposure in another 418 individuals. We first compared the pattern of clustered agricultural exposures of Questionnaire 1 with Questionnaire 2. Second, we ascertained whether utilization of dichotomous (yes/no) vs intensity exposure variables (years) yielded similar factor loading models. Finally, we evaluated whether there was greater variation explained using agricultural intensity exposure variables coded as total lifetime hours compared with exposure intensity variables coded as total lifetime years.

<sup>1</sup>Department of Epidemiology, University of Nebraska Medical Center, Omaha, Nebraska, USA; <sup>2</sup>Department of Biostatistics, University of Nebraska Medical Center, Omaha, Nebraska, USA; <sup>3</sup>Department of Population Health, University of Wisconsin-Madison, Madison, Wisconsin, USA and <sup>4</sup>Department of Internal Medicine, University of Nebraska Medical Center and Veterans Affairs Nebraska Western Iowa Healthcare System, Omaha, Nebraska, USA. Correspondence: Dr. Tricia D. LeVan, Department of Internal Medicine, 985910 Nebraska Medical Center, Omaha, NE 68198-5910, USA.

Tel.: +1 402 559 3985. Fax: +1 402 559 4878.

E-mail: tlevan@unmc.edu

Received 2 September 2015; accepted 5 February 2016; published online 6 April 2016

## METHODS

### Study Population

We used agricultural exposure data from a cross-sectional study designed to assess the relationship between agricultural exposures and chronic respiratory disease in veterans utilizing the VA Nebraska Western Iowa Health Care System. Potential study participants were approached in the primary care outpatient clinics if they had worked on a farm as an adult for  $\geq 2$  years. Eligibility criteria for the study included individuals between the ages of 40 and 80 years. Individuals who had been diagnosed by a physician with asthma, lung cancer or interstitial lung disease such as pulmonary fibrosis, sarcoidosis and hypersensitivity pneumonitis were excluded from the study. Recruitment into the study began in March 2008 and continued through December 2013, with a total of 681 participants. Demographic information, smoking status and agricultural-related exposures were obtained at the time of enrollment. COPD was defined as post-bronchodilator  $FEV_1/FVC < 0.70$  by the Global Initiative for Chronic Obstructive Lung Disease (GOLD) classification criteria.<sup>12</sup> The study was approved by the VA Nebraska Western Iowa Healthcare Systems institutional review board and all participants signed a written informed consent document.

### Exposure Questionnaires

Agricultural exposures were assessed using Q1 from March 2008 to July 2010. Q2 was developed to obtain more detailed agricultural exposure data and was utilized from August 2010 to December 2013. All participants answered either Q1 or Q2.

**Questionnaire 1.** Q1 was a telephone questionnaire conducted by the Nebraska Department of Health and Human Services. Participants were contacted at their preferred phone number within 30 days of enrollment. For Q1, participants were asked to provide total years of working or living on a farm, as well as intensity of farm work (weeks per year, hours per week) during their 20s, 40s and 60s. Of the population, 24% was under the age of 60 years and did not have agricultural exposure data during this time period (60s), and thus the PFA for Q1 only examined intensity of farm work in participants' 20s and 40s. Information (yes/no) on their farm and off the farm exposures (farm tasks, livestock, crops and "other exposures", i.e., wood dust, grain dust, silica/mineral dust, asbestos, smoke other than cigarette, chemical solvents, spray paint, welding fumes) and whether they worked on a farm (yes/no) during their 20s and 40s were obtained. Farm tasks were assessed by asking "What were the tasks you performed on the farm?" and included spread manure, grind animal feed, handle silage, grind hay, till soil, drive combines, drive diesel tractors and repair engines. Total years worked or lived on the farm were calculated by taking the age last lived or worked on a farm minus the age first lived or worked on a farm and subtracting any time between these two points when the participant did not live or work on a farm. The variables for weeks per year ( $\leq 4$ , 5–20, 21–40,  $> 41$ ) and hours per week ( $< 20$ , 20–40, 41–60,  $> 60$ ) working on the farm during the participants' 20s and 40s were collected as categorical variables.

**Questionnaire 2.** Q2 was administered in person by the study coordinator at the time of enrollment. In contrast to Q1, Q2 assessed lifetime exposures (birth to 80 years) and more detailed information about intensity of farm work (hours per week, weeks per year, total years), farm tasks (ever/never), livestock (total years, maximum number of livestock), crops (total years, maximum number of acres) and "other exposures" on and off the farm (hours per week, weeks per year, total years). A composite intensity exposure variable, total lifetime hours, was calculated as (total years  $\times$  total hours/week  $\times$  total weeks/year). Additional exposure variables were collected in Q2 such as worked with diesel-powered farm equipment (maximum days per year, total years) and worked with gas-powered farm equipment (maximum days per year, total years). In order to compare Q1 and Q2, we recoded Q2 to represent exposures during the participants' 20s and 40s similar to Q1 (yes/no), except Q2 data for farm tasks were utilized as ever/never. The intensity of farm work variables were collected as continuous variables in Q2 then coded as categorical variables ( $\leq 4$ , 5–20, 21–40,  $> 41$  weeks per year and  $< 20$ , 20–40, 41–60,  $> 60$  hours per week) during the participants' 20s and 40s for Q1 and Q2 comparisons.

### Principal Factor Analysis

Analyses were conducted using SAS/STAT software for Windows version 9.2 (SAS Institute, Cary, NC, USA). We first standardized all time-related

exposure variables that were continuous and described as total years or total hours to zero mean and unit variance, so that all these variables would enter PFA under similar scales. PFA was conducted using SAS PROC FACTOR using a polychoric correlation covariance matrix; a method for estimating correlations among theorized normally distributed continuous latent variables from observed ordinal variables.<sup>13,14</sup> With this method, factors, that are independent of each other, were extracted in descending order of importance with respect to the proportion of the variance accounted for by each factor.<sup>15</sup> For example, the first factor was derived from a weighted linear combination of agricultural variables that accounted for the largest total variation in the data. The second factor derived contained another linear combination of agricultural variables and accounted for variance not accounted for by the first factor.

The number of factors in the model was determined based on the following criteria: at least two variables with a loading score of  $\geq 0.5$  in a factor; factors must have an eigenvalue  $> 1.0$ ; and each factor must account for at least 1% of the total variance. For every variable in each factor, a factor loading score was calculated that represents the correlations between each of the variables included in each factor, similar to Pearson's correlation coefficients.<sup>16,17</sup> Generally, a factor loading score of 0.30 to 0.40 is considered meaningful,<sup>15–18</sup> however, we used a factor loading score of  $\geq 0.5$  to identify the most highly correlated variables in each factor. In addition, the eigenvalue for each factor was calculated and an eigenvalue  $> 1$  indicated that the factor explained more of the variance than could be accounted for by any one variable.<sup>15,19</sup> We used a promax (oblique) rotated factor pattern because we assumed that the factors were correlated.<sup>10</sup> We determined the number of factors using the scree test plot. The scree test plots the factors on the x axis and the corresponding eigenvalues on the y axis.<sup>20</sup> The test drops factors after the break of inflexion. This test is reliable when the sample size is at least 200.<sup>15</sup> The scree test plot was first viewed to determine the number of factors to include and PROC FACTOR was again conducted where the number of factors were specified.

In total, four models were run. Model 1 used data from Q1 and was compared with Q2. Models 1 and 2 differed only by the way farm task questions were asked; that is, for Q1, farm tasks were asked as "yes/no" during the participants' 20s and 40s and, for Q2, farm tasks were asked as "ever/never" during their lifetime. Model 2 was then compared with Model 3 to ascertain whether utilization of dichotomous vs intensity exposure variables (years), respectively, yielded similar factor loading models. Finally, Model 3 was compared with Model 4 to determine whether agricultural exposure variables coded as total lifetime years compared with total lifetime hours, respectively, generated a greater percentage of variation explained.

## RESULTS

A total of 263 eligible subjects were enrolled using Q1 and 418 participants enrolled using Q2, all with the exposure questionnaire completed. The two populations were primarily white males with  $\sim 55\%$  of the participants having greater than a high school education (Table 1). The prevalence of COPD in this population was 39.8%. Of note, participants enrolled using the Q2 were older ( $P = 0.007$ ,  $Q1 = 63.5$  years  $\pm 8.1$  SD vs  $Q2 = 65.3$  years  $\pm 8.7$  SD), worked on a farm for longer ( $P = 0.001$ ,  $Q1 = 24.6$  years  $\pm 19.6$  SD vs  $Q2 = 29.6$  years  $\pm 18.5$  SD) and were more likely to be exposed to hogs in open pen, beef cattle, dairy cattle, poultry and crops than those enrolled with Q1.

### Questionnaire 1

For development of Model 1, agricultural exposure data were obtained from Q1. Q1 collected mostly dichotomous exposure data (yes/no) during the participants' 20s and 40s, except duration (years lived/worked) and intensity of farm work (weeks per year and hours per week) were obtained as continuous variables. The factors for Model 1 yielded eigenvalues  $> 1$  and explained 24.4% of the variance in the exposure data (Table 2). Factor 1 explained 7.3% of the variance in the observed data, Factor 2 explained 7.0%, and Factors 3 and 4 explained 7.0% and 3.1%, respectively. The proportion of variance explained by each of the remaining

**Table 1.** Study population characteristics.

Characteristic	Questionnaire 1, n = 263	Questionnaire 2, n = 418
<b>Sex*</b>		
Male	261 (99)	406 (97)
Female	2 (1)	12 (3)
<b>Age (years)*</b>		
≤ 49	14 (5)	20 (5)
50–59	59 (22)	71 (17)
60–69	126 (48)	191 (46)
70–80	64 (24)	136 (32)
<b>Race*<sup>a</sup></b>		
White	259 (98)	391 (95)
Other	4 (2)	20 (5)
<b>Education**</b>		
≤ High school	99 (39)	180 (44)
> High school	131 (51)	230 (56)
Refused	27 (10)	0 (0)
<b>Smoking status<sup>a</sup></b>		
Current	53 (21)	86 (21)
Former	147 (59)	240 (58)
Never	49 (20)	87 (21)
Refused	1 (< 1)	0 (0)
Worked on farm (years) <sup>b***</sup>	23.9 ± 19.7	29.4 ± 18.6
<b>Agricultural exposures (yes/no)</b>		
Hogs, confinement	58 (26)	84 (20)
Hogs, open pen**	115 (51)	296 (71)
Beef cattle**	136 (60)	309 (74)
Dairy cattle**	67 (30)	243 (58)
Poultry**	85 (38)	286 (69)
Crops**	155 (68)	399 (96)
<b>COPD<sup>a</sup></b>		
Yes	102 (42)	151 (37)
No	141 (58)	257 (63)

Data are presented as n (%). \*P < 0.05, \*\*\*P < 0.001. <sup>a</sup>Numbers do not add up to 100% because of missing values. <sup>b</sup>Data presented as mean ± SD.

factors was 6.2% and these factors were not included in the final model because of our *a priori* inclusion criteria. Variables loading high on Factor 1 (i.e., factor loading scores ≥ 0.50) were exclusively “other exposures” from a job on or off the farm during the participants’ 20s or 40s, including wood dust, grain dust, rock dust, asbestos, smoke other than cigarette, chemical solvents, spray paint and welding fumes. Loading high on Factor 2 were live/work on farm (weeks per year, hours per week) during their 20s, farm tasks such as spread manure, handle silage and grind hay during their 20s or 40s and exposure to many types of livestock. Variables substantial to Factor 3 were total lifetime years lived or worked on the farm as well as worked on the farm during their 40s (weeks/year and hours/week). Farm tasks performed during their 20s or 40s, such as grinding animal feed, driving combines, driving diesel tractors, along with exposure to pesticides, loaded high in Factor 3. Factor 4 included two variables: exposure to hogs in closed lots and crops.

#### Questionnaire 2

Because there were two questionnaires, two phases of population recruitment and more detailed exposure information collected in Q2 compared with Q1, we wanted to determine whether the factor models obtained by each questionnaire were qualitatively

**Table 2.** Principal factor analysis results for Questionnaire 1 (Model 1; n = 263).<sup>a</sup>

Exposures	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
<b>Live/work on farm (20s, 40s)</b>				
Lived on farm, years (lifetime)	–28	1	<b>82</b>	6
Worked on farm, years (lifetime)	3	9	<b>61</b>	8
Worked on farm, weeks/year (20s)	–1	<b>63</b>	27	7
Worked on farm, weeks/year (40s)	10	–8	<b>109</b>	–32
Worked on farm, hours/week (20s)	5	<b>55</b>	19	12
Worked on farm, hours/week (40s)	1	–14	<b>110</b>	–16
<b>Farm tasks, yes/no (20s, 40s)<sup>b</sup></b>				
Spread manure	17	<b>59</b>	36	–7
Grind animal feed	7	39	<b>52</b>	17
Handle silage	21	<b>50</b>	19	12
Grind hay	10	<b>67</b>	30	7
Till soil	20	34	48	19
Drive combines	9	33	<b>67</b>	0
Drive diesel tractors	34	7	<b>50</b>	23
Repair engines	38	28	25	–4
<b>Livestock, yes/no (20s, 40s)<sup>b</sup></b>				
Hogs in closed lots	14	–8	10	<b>85</b>
Hogs in open lots	–12	<b>62</b>	24	29
Beef cattle	20	<b>55</b>	27	12
Dairy cattle	–18	<b>120</b>	–22	–9
Poultry	8	<b>69</b>	1	5
Other livestock	31	<b>83</b>	–16	33
<b>Crops, yes/no (20s, 40s)<sup>b</sup></b>				
Corn, soybeans, hay, grain sorghum, wheat, oats	–5	0	–25	<b>115</b>
<b>Other exposures, yes/no (20s, 40s)<sup>b</sup></b>				
Wood dust	<b>89</b>	7	–4	–25
Grain dust	<b>53</b>	32	22	6
Silica/sand/rock/mineral dust	<b>84</b>	22	–30	3
Asbestos	<b>76</b>	–20	10	9
Smoke other than cigarette	<b>90</b>	13	–18	3
Chemical solvents	<b>82</b>	2	10	17
Spray paint	<b>80</b>	–12	14	7
Welding fumes	<b>74</b>	2	23	2
Pesticide	28	11	<b>56</b>	4
Eigenvalue	19.0	2.1	1.8	1.4

Bold values represent factor loading score of +0.50 or higher. <sup>a</sup>For ease of presentation, all values were multiplied by 100 and rounded to the nearest integer. <sup>b</sup>If they answered yes in their 20s and/or 40s.

comparable when using similar exposure variables. Data for Q2 were recoded to represent exposures (lived/worked on a farm and variables for exposure to livestock, crops and “other exposures”) during the participants’ 20s and 40s. Data for farm tasks were utilized as lifetime exposure (ever, never). In Model 2, four factors were retained in the model and explained 14.5% of the total variance in the observed data (Table 3). The remaining factors accounted for 5.3% of the variance. Variables loading high on Factor 1 were heterogeneous and included worked on a farm during the participants’ 20s (weeks per year and hours per week) and exposure to hogs in open lots, beef cattle, dairy cattle, poultry,

crops and grain dust in their 20s or 40s. Factor 1 explained 4.3% of the variance in the observed data. Factor 2 explained 3.9% of the variance in the observed data and was a homogeneous factor comprising many farming tasks performed in their lifetime such as spread manure, grind animal feed, handle silage, grind hay, till soil and drive combines and diesel tractors. Variables included in Factor 3 were years lived and worked on the farm (weeks per year and hours per week) in the participants' 40s. Factor 3 explained 3.6% of the variance in the observed data. Factor 4 explained 2.7% of the variance and included exposure to wood dust, rock dust, asbestos, chemical solvents and spray paint during their 20s or 40s with asbestos, smoke, chemical solvents and welding fumes near the cutoff loading score of 0.5.

Q2 collected detailed exposure data over the participant's lifetime. We wanted to ascertain whether utilization of these intensity exposure variables (years) yielded more homogeneous factors compared with using dichotomous (yes/no) exposure variables. In Model 3, we incorporated lifetime agricultural exposures (continuous variables) and compared the factors and factor loading scores with Model 2, where dichotomous exposure variables (20s and 40s) were utilized. For Model 3, three factors explained 10.5% of the total variance in the observed data (Table 4). The proportion of variance explained by the remaining factors was 5.6%. Factor 1 was a heterogeneous factor explaining 4.7% of the variance and included years lived and worked on the farm, years worked with beef cattle, crops, grain dust and pesticide. Factor 2 in Model 3 loaded similar variables as Factor 2 in Model 2 and explained 3.5% of the variance, that is, farming tasks such as spread manure, grind animal feed, handle silage, grind hay, till soil and drive combines. Factor 3 explained 2.3% of the variance and included the lifetime exposure (years) to wood dust, rock dust, asbestos, chemical solvents and spray paint.

We developed Model 4 to assess whether more detailed lifetime intensity variables resulted in unique principal factors and exposure patterns that captured a greater variation than Model 3. Model 4 employed total lifetime hours for worked on farm, worked with livestock, exposure to crops and "other exposures" (Table 5). Additional variables utilized in Model 4 were the summation of maximum number of livestock, maximum number of acres of crops and diesel/gas exposure. Model 4 included four factors and explained 16.6% of the variance. The remaining factors accounted for 11.5% of the variance. Factor 1 explained 7.8% of the total variance and included years lived on the farm, total hours worked on the farm, total years worked with diesel power, total days/year worked with gas-powered equipment and farm tasks performed over a lifetime, such as till soil, drive combines and drive diesel tractors. Total years worked with beef cattle, total years worked with crops, total number of acres of crops and total hours exposed to grain dust, pesticides and diesel fuel were also included in Factor 1. Factor 2 included total years exposed to hogs in open lots, total years of exposure and number of dairy cattle and poultry. Factor 2 explained 3.6% and Factor 3 explained 2.7% of the total variance. Factor 3 included lifetime total hours exposed to rock dust and spray paint. Factor 4 included total years and acres of other crops and explained 2.5% of the total variance.

In order to reduce bias, a sensitivity analysis was performed for Model 4 by stratifying by COPD status (Supplementary Table S1). Similar clustering patterns were found for Factors 1 and 2 in the total population and those with COPD and those without COPD, and were identical when the factor loading score was relaxed to 0.4. Factor 3 in Model 4 for the total population loaded similar variables to those with COPD, whereas Factor 4 contained variables from both COPD and no COPD. In addition, age and smoking status were tested in all models; however, these variables had a loading score < 0.5, and thus were not included in the final models.

**Table 3.** Principal factor analysis results for Questionnaire 2 (Model 2; n = 418).<sup>a</sup>

Exposures	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
<i>Live/work on farm (20s, 40s)</i>				
Lived on farm, years (lifetime)	8	28	<b>52</b>	-23
Worked on farm, years (lifetime)	9	18	<b>65</b>	-17
Worked on farm weeks/year (20s)	<b>66</b>	0	-13	-13
Worked on farm weeks/year (40s)	-11	-16	<b>105</b>	8
Worked on farm hours/week (20s)	<b>50</b>	-7	-13	-8
Worked on farm hours/week (40s)	9	-10	<b>92</b>	4
<i>Farm tasks, ever/never (lifetime)</i>				
Spread manure	1	<b>62</b>	-1	7
Grind animal feed	3	<b>77</b>	6	-10
Handle silage	0	<b>73</b>	-12	25
Grind hay	-3	<b>52</b>	-13	7
Till soil	3	<b>70</b>	17	-13
Drive combines	23	<b>75</b>	5	0
Drive diesel tractors	16	<b>51</b>	29	34
Repair engines	-9	20	17	35
<i>Livestock, yes/no (20s, 40s)<sup>b</sup></i>				
Hogs, closed lots	24	7	5	4
Hogs, open lots	<b>58</b>	27	8	-18
Beef cattle	<b>56</b>	7	18	2
Dairy cattle	<b>65</b>	17	-15	2
Poultry	<b>79</b>	2	-3	10
Other livestock	12	5	7	33
<i>Crops, yes, no (20s, 40s)<sup>b</sup></i>				
Corn, soybeans, hay, grain sorghum, wheat, or oats	<b>76</b>	-5	30	4
<i>Other exposures, yes, no (20s, 40s)<sup>b</sup></i>				
Wood dust	10	-1	-6	<b>63</b>
Grain dust	<b>74</b>	-8	21	17
Silica/sand/rock/mineral dust	16	-9	-15	<b>62</b>
Asbestos	-20	14	0	<b>51</b>
Smoke, not cigarette	-12	-6	17	45
Chemical solvents	-12	-2	6	<b>52</b>
Spray paint	-9	20	-17	<b>59</b>
Welding fumes	20	-3	-1	44
Pesticides	35	8	36	7
Eigenvalue	7.3	2.9	2.3	2.0

Bold values represent factor loading score of +0.50 or higher. <sup>a</sup>For ease of presentation, all values were multiplied by 100 and rounded to the nearest integer. <sup>b</sup>If they answered yes in their 20s and/or 40s.

## DISCUSSION

The ultimate goal of the veteran cohort is to describe long-term agricultural exposures and their relation to respiratory outcomes. Existing studies have shown the harmful effects of the farming environment on COPD, asthma and other airway diseases.<sup>1</sup> Specifically, exposures such as animals, hay and grains are known to have an adverse effect on respiratory health,<sup>21</sup> as well as agricultural pesticides.<sup>22</sup> Long-term work in large animal-feeding operations, particularly swine confinement facilities and cattle feedlots,<sup>23</sup> also contribute to chronic respiratory disease with dairy farming associated specifically with COPD.<sup>24</sup>

In this exploratory statistical analysis, we utilized principal factor analysis to examine the correlation among a large number of

**Table 4.** Principal factor analysis results using Questionnaire 2 (Model 3; n = 418).<sup>a</sup>

Exposures	FACTOR 1	FACTOR 2	FACTOR 3
<i>Live/work on farm (lifetime)</i>			
Lived on farm, years	<b>68</b>	11	10
<i>Farm tasks, ever/never (lifetime)</i>			
Worked on farm, years	<b>94</b>	-9	0
Spread manure	-3	<b>65</b>	5
Grind animal feed	15	<b>72</b>	-9
Handle silage	-16	<b>77</b>	14
Grind hay	-15	<b>58</b>	1
Till soil	17	<b>70</b>	-15
Drive combines	18	<b>72</b>	-7
Drive diesel tractors	23	46	24
Repair engines	20	13	40
<i>Livestock, total years (lifetime)</i>			
Hogs in closed lots	5	24	2
Hogs in open lots	49	19	-13
Beef cattle	<b>61</b>	7	1
Dairy cattle	23	19	2
Poultry	17	0	6
Other livestock	10	5	17
<i>Crops, total years (lifetime)</i>			
Corn, soybeans, hay, grain sorghum, wheat, oats	<b>96</b>	-5	0
<i>Other exposures, total years (lifetime)</i>			
Wood dust	-4	-5	<b>56</b>
Grain dust	<b>84</b>	-9	11
Silica/sand/rock/mineral dust	-15	0	<b>62</b>
Asbestos	6	-1	<b>57</b>
Smoke other than cigarette	8	-3	44
Chemical solvents	8	11	<b>50</b>
Spray paint	-16	15	<b>53</b>
Welding fumes	22	8	44
Pesticide	<b>68</b>	8	7
Eigenvalue	6.8	2.7	1.4

Bold values represent factor loading score of +0.50 or higher. <sup>a</sup>For ease of presentation, all values were multiplied by 100 and rounded to the nearest integer.

exposure variables as well as to reduce the number of variables into domains of agricultural exposure patterns without loss of a significant amount information. Model 1 utilized Q1 that collected dichotomous (yes/no) exposure data during the participants' 20s and 40s. Models 2–4 utilized variables collected from Q2 that quantitated lifetime agricultural exposures as total years, weeks per year and hours per week. Overall, we found that duration and intensity of farm work, farm tasks, livestock exposure, crop exposure and "other exposures" were independent entities and their clustering within a model was modified by the intensity units of exposure (dichotomous vs continuous).

There were four principal factors derived for Model 1 using Q1. Factor 1 had a homogeneous cluster composed of variables in the "other exposures" category and represented job exposures on or off the farm such as wood dust, grain dust, rock dust, asbestos, smoke other than cigarette, chemical solvents, spray paint and welding fumes. These exposures are often categorized as vapor, dust and smoke, and have been associated with occupational respiratory disease such as asthma and COPD.<sup>25</sup> Factor 2 was heterogeneous yet interpretable and included variables such as duration of farm work during the participants' 20s, select farm tasks and livestock exposures. Of note, the farm tasks in this factor were related to animal husbandry such as spread manure and exposure to dairy cattle. Individuals who farmed during their 20s were more likely to

have exposure to animals than those who farmed during their 40s. In contrast, individuals that farmed during their 40s were more likely to perform less strenuous tasks such as drive combines and diesel tractors and this pattern was observed in Factor 3. There are many reasons why younger farmers have different exposures than older farmers. Open cabbled tractors, although rare today, were the norm for older farmers, and therefore they were more exposed to pesticides and dust.<sup>26</sup> We see this in Model 1 where working on the farm in the participants' 40s clustered with driving of combines and diesel tractors as well as pesticides. There was a clear separation between all factors such that each variable loaded significantly on only one factor. The variables with loading scores of  $\pm 0.50$  or higher within a factor were correlated most likely because of the fact that many of the variables within a cluster, such as farm tasks, are done collectively when working in agriculture.

Model 2 was derived using Q2 variables that were recoded to replicate exposure variables similar to Q1. The variables were dichotomous for exposure during the participants' 20s and 40s. As in Model 1, there were four factors and each variable loaded significantly on only one factor. The first factor included working on a farm during their 20s and this was correlated to animal exposures such as beef and dairy cattle and hogs in open lots (marginal correlation). This pattern was also observed in Model 1. In addition to animal exposures during the participants' 20s, crops and grain dust were included in Factor 1 and are consistent with livestock production practices. Factor 2 aligned with many of the farming tasks, whereas Factor 3 consisted of lifetime years lived and worked on a farm along with intensity of farm work during the participants' 40s. This clustering of lifetime years and intensity of farm work was similar to that observed in Model 1. Factor 4 included variables from "other exposures" and this same pattern was seen for Model 1. Overall, the factors in Model 1 and 2 were similar with clustering of lifetime years worked/lived on a farm, intensity of farm work, livestock exposure and "other exposures". The major difference between the two models was that farm tasks loaded heavily in Model 2 compared with Model 1. In Q2, these farm task questions were asked as "ever/never" during their lifetime, whereas in Q1 these questions were asked with "yes/no" answers for their 20s and 40s. These observations suggest that collecting information on farm tasks is important in accounting for the variability in agricultural exposures because of their heavy loading in the model and that the "ever/never" during a person's lifetime would be more all-inclusive. Furthermore, the dissimilarities of factors in Models 1 and 2 may be because of the different age structures of these two populations. The population from Q2 had a greater proportion of people >70 years old than the population from Q1. Of note, the percentage of variation explained is a measurement of fit. Q2 had a lower percentage of variation explained compared with Q1, and this may be because of greater variability as it was used on a larger population with more workers (86% vs 59%) working on a farm for >10 years.

For Model 3, we used lifetime exposures with intensity units as total years, except for farm tasks as ever/never. The principal factors for Model 3 had three distinct patterns. The first factor contained heterogeneous exposure variables including live/work on the farm, livestock, crops and "other exposures". Farming tasks clustered and loaded heavily in Factor 2 as with "other exposure" variables in Factor 3. In addition, these domains were predominant in Model 2. Even though the percentage of variance explained in Model 3 was less than that in Model 2, there was utilization of more complete exposure variables (lifetime) in Model 3 compared with Model 2 (20s and 40s).

As a final Model, we included all of the collected exposure and intensity variables as total lifetime hours, maximum lifetime number of livestock or acres of crops or ever/never farm tasks. We observed four distinct factors in Model 4. Factor 1 was a heterogeneous factor that included exposures related to crops and livestock, whereas the main Factor 2 domain was livestock.

**Table 5.** Principal factor analysis results using Questionnaire 2 (Model 4;  $n = 418$ ).<sup>a</sup>

Exposures	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
<i>Live/work on farm (lifetime)</i>				
Lived on farm, years	<b>64</b>	15	-25	3
Worked on farm, total hours <sup>b</sup>	<b>88</b>	3	-21	10
<i>Farm tasks, never/ever (lifetime)</i>				
Spread manure	20	48	25	-20
Grind animal feed	36	36	8	-42
Handle silage	15	33	42	-29
Grind hay	7	27	20	-26
Till soil	<b>53</b>	22	10	-7
Drive combines	<b>50</b>	14	11	-39
Drive diesel tractors	<b>64</b>	-21	37	-10
Repair engines	27	-12	38	-3
<i>Livestock, total hours<sup>b</sup> and max no.<sup>c</sup> (lifetime)</i>				
Hogs in closed lots, total years	12	17	5	-8
Hogs in open lots, total years	36	<b>51</b>	-31	-15
Max no. of hogs in open lots	30	38	-19	-34
Beef cattle, total years	<b>58</b>	23	-13	7
Max no. of beef cattle	43	7	6	5
Dairy cattle, total years	4	<b>66</b>	-2	6
Max no. of dairy cattle	-4	<b>58</b>	6	0
Poultry, total years	-8	<b>78</b>	-5	20
Max no. of poultry	-26	<b>72</b>	-6	14
Other livestock, total years	17	30	31	40
Max no. of other livestock	10	29	36	35
<i>Crops, total hours<sup>b</sup> and max no.<sup>c</sup> (lifetime)</i>				
Corn, soybeans, hay, grain sorghum, wheat, oats, total years	<b>85</b>	2	-22	6
Max no. of acres, corn, soybeans, hay, grain sorghum, wheat, oats	<b>71</b>	-25	6	-4
Other crops, years	11	17	6	<b>80</b>
No. of acres, other crops	15	12	9	<b>79</b>
<i>Other exposures, total hours<sup>b</sup> (lifetime)</i>				
Wood dust	-14	12	49	8
Grain dust	<b>79</b>	-1	-4	12
Silica/sand/rock/mineral dust	-24	14	<b>54</b>	-7
Asbestos	-3	-8	49	0
Smoke other than cigarette	8	11	37	31
Chemical solvents	-2	-4	39	9
Spray paint	-6	-10	<b>57</b>	1
Welding fumes	14	0	40	-6
Pesticide	<b>70</b>	3	1	12
<i>Diesel/gas (lifetime)</i>				
Worked diesel power, years	<b>93</b>	-23	10	13
Worked gas powered equipment, days/year	<b>72</b>	0	-3	-4
Diesel fuel/fumes/exhaust	<b>75</b>	-7	16	6
Eigenvalue	9.3	2.9	2.4	2.2

<sup>a</sup>For ease of presentation, all values were multiplied by 100 and rounded to the nearest integer. Bold values represent factor loading score of +0.50 or higher.

<sup>b</sup>Total hours in lifetime = (total years) × (total hours/week) × (total weeks/year). <sup>c</sup>Maximum no. = average number of livestock or average number of acres of crops in lifetime.

Factor 3 included variables from “other exposures” and Factor 4 was solely “other crops”. Model 4 captured a higher percentage of variance, suggesting that detailed intensity variables for agricultural exposure are advantageous in capturing a greater percentage of variance than dichotomous (yes/no) or even the variables coded as total years. We observed that diesel/gas exposure variables were important to include in Model 4 as it loaded high in Factor 1. Model 4 included additional crop variables that were asked in Q2 and resulted in a distinct factor pattern of crops (Factor 4). This was not found in previous models.

Many studies have found the utility of factor analysis. The Agricultural Health Study utilized factor analysis to identify clusters of pesticide exposures that relate to prostate cancer.<sup>27</sup> Another study clustered respiratory phenotypes of COPD to explain the

heterogeneity of COPD.<sup>28</sup> PFA is not only used to assess the effect of occupational exposures on respiratory diseases, but is also used to evaluate the reproducibility and validity of questionnaires<sup>29</sup> as Hammond *et al.*<sup>29</sup> tested the validity and reliability of the English Evaluation of Daily Activity Questionnaire.

In this study, factor analysis was used to extract the useful information from a complex data set to interpret the agricultural exposure data. Studies have found the importance of including the use of solvents, paint, exposure to welding fumes<sup>30</sup> and pesticide use<sup>17</sup> when investigating exposure–respiratory disease associations. We found these exposure variables to be also important in our analysis in describing long-term agricultural exposures.

This study has some important strengths. First, the exposure data were comprehensive, including hours per week, weeks per

years and total years, and were collected by trained study personnel. Second, the agricultural population is large and all have worked in Nebraska or Iowa, and thus have similar exposures. Finally, the statistical methods used allow unbiased analyses that are not based on any *a priori* assumptions. This study does have limitations. Recall bias is probable as participants were asked to retrospectively recall their lifetime farming exposures. This could have resulted in overestimation or underestimation of the exposure that could ultimately impact factor weighting and subsequent regression analysis. There is a potential for interviewer bias as there were two methods to obtain exposure information: telephone interviews for Q1 and in-person interviews for Q2. It would be difficult to determine whether this would be an over- or under-reporting of exposures. In addition, there is the issue of generalizability of these results. The population included veterans with agricultural exposure utilizing the VA Nebraska Western Iowa Health Care System. They were primarily white males with a mean age of 64 years; therefore, their agricultural exposures may be different from younger workers because of technological advances in farming. In addition, direct measurement of agricultural exposures was not performed.

In summary, we found that PFA was an effective statistical method for characterizing exposure patterns in our population of agricultural workers. We have identified clusters in a large data set that describes the heterogeneity of exposures including duration and intensity of farm work, farm tasks, livestock exposure, crop exposure and "other exposures". We examined four models and found that Model 4, with the most detailed exposure information, captured the highest percentage of variance compared with the other models. The resulting factor patterns were clearly interpretable and logical in terms of farming practice. From this study, we also determined that the most important exposure variables to be asked in questionnaires when evaluating agricultural exposures and respiratory diseases are years worked on a farm, farm tasks and exposure to livestock, crops and "other exposures" as these consistently loaded high across the four models. The next step is to further explore these patterns in Model 4 to examine the relationship between agricultural exposures and respiratory diseases such as COPD in this population.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

This work was supported by VA Merit Awards for TDL and DJR. We acknowledge the hard work of the study coordinators Robin Zotti-Pierce, Eric Chickris and Kelsey Palm, as well as the participation of veterans.

## REFERENCES

- 1 Poole JA, Romberger DJ. Immunological and inflammatory responses to organic dust in agriculture. *Curr Opin Allergy Clin Immunol* 2012; **12**: 126–132.
- 2 Thomas KW, Dosemeci M, Coble JB, Hoppin JA, Sheldon LS, Chapa G *et al*. Assessment of a pesticide exposure intensity algorithm in the agricultural health study. *J Expo Sci Environ Epidemiol* 2010; **20**: 559–569.
- 3 Pope CA 3rd, Burnett RT, Thurston GD, Thun MJ, Calle EE, Krewski D *et al*. Cardiovascular mortality and long-term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease. *Circulation* 2004; **109**: 71–77.
- 4 Basinas I, Sigsgaard T, Erlandsen M, Andersen NT, Takai H, Heederik D *et al*. Exposure-affecting factors of dairy farmers' exposure to inhalable dust and endotoxin. *Ann Occup Hyg* 2014; **58**: 707–723.
- 5 Moran RE, Bennett DH, Garcia J, Schenker MB. Occupational exposure to particulate matter from three agricultural crops in California. *Int J Hyg Environ Health* 2014; **217**: 226–230.

- 6 Ljubicic A, Varnai VM, Vucemilo M, Matkovic K, Milic D, Macan J. Exhaled breath condensate pH and FeNO as biomarkers of acute and chronic exposure to hazards at swine farms. *J Occup Environ Med* 2014; **56**: 946–952.
- 7 Sutuluk Z, Kecec Z, Daglioglu N, Hant I. Association of chronic pesticide exposure with serum cholinesterase levels and pulmonary functions. *Arch Environ Occup Health* 2011; **66**: 95–99.
- 8 Wurtz ET, Schlunssen V, Malling TH, Hansen JG, Omland O. Occupational chronic obstructive pulmonary disease in a Danish population-based study. *COPD* 2014; **12**: 435–443.
- 9 Valcin M, Henneberger PK, Kullman GJ, Umbach DM, London SJ, Alavanja MC *et al*. Chronic bronchitis among nonsmoking farm women in the agricultural health study. *J Occup Environ Med* 2007; **49**: 574–583.
- 10 Gorsuch RL. *Factor Analysis*. 2nd edn. Lawrence Erlbaum Associates: Hillsdale, NJ, 1983.
- 11 Suhr D. Principal component analysis vs. exploratory factor analysis. 2009.
- 12 Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for Diagnosis, Management and Prevention of COPD 2014. Available from: [http://www.goldcopd.org/uploads/users/files/GOLD\\_Report\\_2014\\_Jan23.pdf](http://www.goldcopd.org/uploads/users/files/GOLD_Report_2014_Jan23.pdf). Accessed 20 November 2015.
- 13 Olsson U. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika* 1979; **44**: 443–460.
- 14 Baglin J. Improving your exploratory factor analysis for ordinal data: a demonstration using FACTOR. *PARE* 2014; **19**.
- 15 Yong AG, Pearce S. A beginner's guide to factor analysis: focusing on exploratory factor analysis. *Tutor Quant Methods Psychol* 2013; **9**: 79–94.
- 16 Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. *Psychol Assessment* 1995; **7**: 286–299.
- 17 Samanic C, Hoppin JA, Lubin JH, Blair A, Alavanja MC. Factor analysis of pesticide use patterns among pesticide applicators in the Agricultural Health Study. *J Expo Anal Environ Epidemiol* 2005; **15**: 225–233.
- 18 Tabachnick BG, Fidell LS. *Using Multivariate Statistics*, 5th edn. Allyn & Bacon: Boston, 2007.
- 19 Kaiser HF. The application of electronic computers to factor analysis. *Educ Psychol Meas* 1960; **20**: 141–151.
- 20 Colgan P. The scientific use of factor analysis in behavioral and life sciences Raymond B Cattell. Plenum Press, New York and London, 1978. xxii + 618 pp., US \$ 32.50, ISBN 0-306-30939-4. *Behav Processes* 1981; **6**: 385–386.
- 21 Dosman J, Husman K, Saiki C, Schenker MB. Respiratory health hazards in agriculture. *Am J Respir Crit Care Med* 1998; **158**(Pt 2): 51–576.
- 22 Hoppin JA. Pesticides and respiratory health: where do we go from here? *Occup Environ Med* 2014; **71**: 80–2013-101876.
- 23 Wells AD, Poole JA, Romberger DJ. Influence of farming exposure on the development of asthma and asthma-like symptoms. *Int Immunopharmacol* 2014; **23**: 356–363.
- 24 Marescaux A, Degano B, Soumagne T, Thaon I, Laplante JJ, Dalphin JC. Impact of farm modernity on the prevalence of chronic obstructive pulmonary disease in dairy farmers. *Occup Environ Med* 2015; **73**: 127–133.
- 25 LeVan TD, Koh WP, Lee HP, Koh D, Yu MC, London SJ. Vapor, dust, and smoke exposure in relation to adult-onset asthma and chronic respiratory symptoms: the Singapore Chinese Health Study. *Am J Epidemiol* 2006; **163**: 1118–1128.
- 26 Voaklander D, Day L, Dosman J, Hagel L, Pickett W. Older farmers and machinery exposure-cause for concern? *Am J Ind Med* 2012; **55**: 1044–1050.
- 27 Alavanja MC, Samanic C, Dosemeci M, Lubin J, Tarone R, Lynch CF *et al*. Use of agricultural pesticides and prostate cancer risk in the Agricultural Health Study cohort. *Am J Epidemiol* 2003; **157**: 800–814.
- 28 Postma DS, Anzueto AR, Jenkins C, Make BJ, Similowski T, Ostlund O *et al*. Factor analysis in predominantly severe COPD: identification of disease heterogeneity by easily measurable characteristics. *Respir Med* 2013; **107**: 1939–1947.
- 29 Hammond A, Tennant A, Tyson SF, Nordenskiold U, Hawkins R, Prior Y. The reliability and validity of the English version of the Evaluation of Daily Activity Questionnaire for people with rheumatoid arthritis. *Rheumatology (Oxford)* 2015; **54**: 1605–1615.
- 30 Coble J, Hoppin JA, Engel L, Elci OC, Dosemeci M, Lynch CF *et al*. Prevalence of exposure to solvents, metals, grain dust, and other hazards among farmers in the Agricultural Health Study. *J Expo Anal Environ Epidemiol* 2002; **12**: 418–426.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Supplementary Information accompanies the paper on the Journal of Exposure Science and Environmental Epidemiology website (<http://www.nature.com/jes>)