Preventing Coal Mine Dust Lung Disease: Application of Bayesian Hierarchical Framework for Occupational Exposure Assessment in The South African Coal Mining Industry

Felix Made

Student number: 1239045

Supervisors:
Prof Derk Brouwer
Prof Jerome Lavoue
Prof Ngianga-Bakwin Kandala

This is an original published work submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy (Public Health)

Johannesburg, South Africa, 2023

Declaration

I, Felix Made, declare that this thesis is my own work. It is being submitted for the degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg, South Africa. The work has not been submitted previously for any degree or examination at any other, including the University of the Witwatersrand.

I have read the sections on referencing and plagiarism in the Wits Plagiarism Policy, and I have followed the required conventions in referencing the thoughts and ideas of others.

Signature:

Date: 3 October 2023

Place: Johannesburg, South Africa

Dedication

This PhD is dedicated to you, my mother; may your soul rest in eternal peace. You inspired me to keep trying even when things are hard and to never give up. I am forever grateful for your support and blessings even in death. I also want to thank my wife and children for giving me strength throughout my PhD journey.

Acknowledgements

My sincere gratitude goes to the supervisors, Prof. Derk Brouwer, Prof. Jerome Lavoué, and Prof. Ngianga-Bakwin Kandala, for their unwavering support, guidance, and mentorship from the beginning until the final submission of this thesis. I am forever grateful for your love, motivation, and support.

Funding

This work was partially supported financially by the Anglo-American Endowment Fund, and the National Health Laboratory Services. The financial supporters had no role in study design, data collection, data analysis, the decision to publish, or the preparation of the thesis or its content, whatsoever.

Presentations arising from this thesis

1. Made F, Kandala NB, Brouwer D. Application of Bayesian statistics in occupational hygiene Paper presented at: Conference of the Southern African Institute of Occupational Hygiene (SAIOH); 2019 September 5; Johannesburg, South Africa.
2. Made F, Brouwer D. Expert elicitation: application of the Bayesian framework in occupational exposure assessment in the South African coal mining industry. Presented at: Mine Ventilation Society of South Africa (MVSSA); 2022 May 27; Johannesburg, South Africa.
3. Made F. PhD results and policy proposal. Presented at: Meeting of the Collieries Dust Working Group. Minerals Council of South Africa (MCSA); 2022 November 18; Johannesburg, South Africa.

Publications arising from this thesis

1. Made F, Kandala NB, Brouwer D. Compliance Testing and Homogenous Exposure Group Assessment in the South African Coal Mining Industry. *Ann. Work Expo. Health*. 2021; 65(8):955-65.
2. Made F, Kandala NB, Brouwer D. Bayesian hierarchical modelling of historical data of the South African coal mining industry for compliance testing. *Int. J. Environ. Res. Public Health*. 2022; 19(8):4442. doi: 10.3390/ijerph19084442.
3. Made F, Kandala NB, Brouwer D. Bayesian Hierarchical Framework from Expert Elicitation in the South African Coal Mining Industry for Compliance Testing. *Int. J. Environ. Res. Public Health*. 2023 ; 20(3):2496. doi: 10.3390/ijerph20032496.

Preface

After completing my master’s degree, I was employed by Sydney Brenner Institute for Molecule Biosciences (SBIMB) and then by the National Institute for Occupational Health (NIOH). During my time at the NIOH, we did a statistical analysis on compensation of deceased miners for a mining company for workers that died after exposure to dust in the Gold and Coal mines. From this analysis on compensation of deceased miners, I developed an interest in exposure science and advanced biostatistics. I started to wonder what more can be done to reduce overexposure to coal dust. With the NIOH department, we consulted Prof Derk Brouwer who then gave me some documents to read alongside further literature search, to understand the current strategy of establishing HEGs and compliance testing in South Africa. Immediately we were able to identify gaps in the Department of mineral resources and energy (DMRE) guidelines where the constitution of homogenous exposure groups (HEG) tends to be done by common area and return aired intake where highly exposed workers or job titles are not considered as a single HEG. The DMRE's current guidelines and searches on international best practices form the backbone of this PhD study.

In the study, chapter 1 is the background which covered the coal industry, coal dust mine lung disease (CMDLD) occupational exposure categories, grouping of workers in DMRE alongside international practices, and the assessment of homogeneity strategy both in the DMRE and elsewhere. In the literature review in chapter 2, we presented the prevalence of CMDLD, the application of Bayesian statistics in industrial hygiene, expert elicitation and its various protocols and applications. This is followed by the method section in chapter 3, where data collection and dust sampling methods were described, Bayesian statistical modelling was used to draw inferences from the data and sensitivity analysis was applied for the various parameter space. Chapter 4 is the results section. Chapter 5 covered the discussion section, strengths and limitations of the study, PhD study’s new knowledge, future recommendations and policy impact resulting from the PhD findings. Finally, in chapter 6, the key conclusions from the study were drawn.

Abstract

Background: The world's largest energy source is coal with nearly 36% of all the fuel used to produce power. South Africa is the world's top exporter and the seventh-largest producer of coal. In the upcoming years, it is expected that South Africa's coal production output rate will rise. Coal mine dust lung disease (CMDLD) is an irreversible lung disease caused by the production of coal, the emission of dust, and prolonged exposure to the dust. When conducting safety evaluation, exposure is typically reported as an eight-hour time-weighted average dust concentration (TWA8h). In occupational exposure contexts, occupational exposure limits (OEL) are often used as a threshold where workers can be exposed repeatedly without adverse health effects. The workers are usually grouped into homogenous exposure groups (HEGs) or similar exposure groups (SEGs). In South Africa, a HEG is a group of coal miners who have had similar levels and patterns of exposure to respirable crystalline silica (RCS) dust in the workplace. Several statistical analysis methods for compliance testing and homogeneity assessment have been put into use internationally as well as in South Africa. The international consensus on occupational exposure analysis is based on guidelines from the American Industrial Hygiene Association (AIHA), the Committee of European Normalisation (CEN), and BOHS British and Dutch Occupational Hygiene Societies' guidelines (BOHS). These statistical approaches are based on Bayesian or frequentist statistics and consider the 90th percentile (P90) and 95th percentile (P95), with- and between-worker variances, and the lognormal distribution of the data. The current existing practices in South Africa could result in poor or incorrect risk and exposure control decision-making.

Study Aims: The study aimed to improve the identification of coal dust overexposure by introducing new methods for compliance (reduced dust exposure) and homogeneity (similar dust exposure level) assessment in the South African coal mining industry.

Study Objectives: The objectives of this study were:

1. To compare compliance of coal dust exposure by HEGs using DMRE-CoP approach and other global consensus methods.
2. To investigate and compare the within-group exposure variation between HEGs and job titles.
3. To determine the posterior probabilities of locating the exposure level in each of the OEL exposure categories by using the Bayesian framework with previous information from historical data and compare the findings and the DMRE-CoP approach.
4. To investigate the difference in posterior probabilities of the P95 exposure being found in OEL exposure category between previous information acquired from the experts and the current information from the data using Bayesian analysis. .

Methods: The TWA(8h) respirable coal dust concentrations were obtained in a cross-sectional study with all participants being male underground coal mine workers. The occupational hygiene division of the mining company collected the data between 2009 and 2018. The data were collected according to the South African National Accreditation System (SANAS) standards. From the data, 28 HEGs with a total of 728 participants were included in this study.

In objectives 1 and 2, all 728 participants from the 28 HEGs were included in the analysis. For exposure compliance, the DMRE-CoP accepts 10% exceedance of exposure above the OEL (P90 exposure values from HEGs should be below the OEL). The 10% exceedance was compared to the acceptability criterion from international consensus which uses 5% exceedance above the OEL (P95 exposure is below the OEL) of the lognormal exposure data. For exposure data to be regarded as homogenous, the DMRE-CoP requires that the arithmetic mean (AM) and P90 must fall into the same DMRE-CoP OEL exposure category. The DMRE-CoP on assessment of homogeneity was also compared with the international approaches which include the Rappaport ratio (R-ratio) and the global geometric standard deviation (GSD). A GSD greater than 3 and an R-ratio greater than 2 would both indicate non-homogeneity of the exposure data of a HEG. The GSD and DMRE-CoP criteria were used to assess the homogeneity of job titles exposure within a HEG.

In objective 3 a total of nine HEGs which have 243 participants, were included in the analysis. To investigate compliance, a Bayesian model was fitted with a Markov chain Monte Carlo (MCMC) simulation. A normal likelihood function with the GM and GSD from lognormal data was defined. The likelihood function was updated using informative prior derived as the GM and GSD with restricted bounds (parameter space) from the HEGs' historical data. The posterior probabilities of the P95 being located in each DMRE exposure band were produced and compared with the non-informative results and the DMRE approach DMRE-CoP using a point estimate inform of the 90 percentiles.

In objective 4, a total of 10 job titles were analysed and selected. The selection of the job titles was based on if they have previous year's data so it can be used to develop prior information in the Bayesian model. The same job titles were found across different HEGs, so to ensure the mean is not different across HEGs, the median difference of a job title exposure distribution across HEGs was statistically compared using the Kruskal-Wallis test, a non-parametric alternative to analysis of variance (ANOVA). Job titles with statistically non-significant exposure differences were included in the analysis.
Expert judgements about the probability of the P95 located in each of the DMRE exposure bands were elicited. The IDEA (Investigate", "Discuss", "Estimate" and "Aggregate) expert elicitation procedure was used to collect expert judgements. The SHELF tool was then used to produce the lognormal distribution of the expert judgements as GM and GSD to be used as informative prior. A similar Bayesian analysis approach as in objective 3 was used to produce the probability of the P95 falling in

each of the DMRE exposure bands. The possible misclassification of exposure arising from the use of bounds in the parameter space was tested in a sensitivity analysis.

Results: There were 21 HEGs out of 28 in objectives 1 and 2 that were non-compliant with the OEL across all methods. According to the DMRE-CoP approach, compliance to the OEL, or exposure that is below the OEL, was observed for 7 HEGs. The DMRE-CoP and CEN both had1 HEG with exposures below the OEL. While the DMRE-CoP showed 6 homogeneous HEGs, however, based on the GSDs 11 HEGs were homogeneous. The GSD and the DMRE-CoP agreed on homogeneity in exposures of 4 (14%) HEGs. It was discovered that by grouping according to job titles, most of the job titles within non-homogenous HEGs were homogenous. Five job titles had AMs above their parent HEG.

For objective 3, the application of the DMRE-CoP (P90) revealed that the exposure of one HEG is below the OEL, indicating compliance. However, no HEG has exposures below the OEL, according to the Bayesian framework. The posterior GSD of the Bayesian analysis from non-informative prior indicated a higher variability of exposure than the informative prior distribution from historical data. Results with a non-informative prior had slightly lower values of the P95 and wider 95% credible intervals (CrI) than those with an informative prior. All the posterior P95 findings from both non-informative and informative prior distribution were classified in exposure control category 4 (i.e., poorly controlled since exceeding the OEL), with posterior probabilities in the informative approach slightly higher than in the non-informative approach.

Job titles were selected as an alternative group to assess compliance in objective 4. The posterior GSD indicated lower variability of exposure from expert prior distribution than historical data prior distribution. The posterior P95 exposure was very likely (at least 98% probability) to be found in exposure control category 4 when using prior distribution from expert elicitation compared to the other Bayesian analysis approaches. The probabilities of the P95 from experts' judgements and historical data were similar. The non-informative prior generally showed a higher probability of finding the posterior P95 in lower exposure control categories than both experts and historical data prior distribution. The use of different parameter values to specify the bounds showed comparable results while the use of no parameter space at all put the posterior P95 in exposure category 4 with 100% probability.

Conclusions: In comparison to other approaches, the DMRE-CoP tend to show that exposures are compliant more often. Overall, all methods show that the majority of HEGs were non-compliant. The HEGs that suggest non-homogeneity revealed that the constituent job titles were homogenous. Application of the GSD criterion indicated that HEGs are more likely to be considered as homogeneous than when using the DMRE-CoP approach. When using the GSD and the DMRE-CoP guidelines, alternative grouping by specific job titles showed a greater agreement of homogeneity. The use of job

titles showed that using HEGs following the DMRE-CoP current guidelines might not show high-exposure job titles and would overestimate compliance. Additionally, since job titles within a HEG may be homogeneous or have a different exposure to the parent HEG, exposure variability is not properly recorded when using HEGs. In compliance assessment, it is important to use the P95 of the lognormal distribution rather than the DMRE-CoP approach that use the empirical P90. Our findings suggest that the subgrouping of exposure according to job titles within a HEG should be used in the retrospective assessment of exposure variability, and compliance with the OEL.

Our results imply that the use of a Bayesian framework with informative prior from either historical or expert elicitation may confidently aid concise decision making on coal dust exposure risk. Contrary to informative prior distribution derived from historical data or expert elicitation, Bayesian analysis using the non-informative uniform prior distribution places HEGs in lower exposure categories. Results from noninformative prior distributions typically show high levels of uncertainty and variability, so a decision on dust control would be reached with less confidence. The Bayesian framework should be used in the assessment of coal mining dust exposure along with prior knowledge from historical data or professional judgment, according to this study. For exposure, findings are to be reported with high confidence and for sound decisions to be reached about risk mitigation, an exposure risk assessment should be considered while using historical data to update the current data. The study also promotes the use of experts in situations where it is necessary to combine current data with historical data, but the historical data is unavailable or inapplicable.

Keywords: occupational exposure limit; Bayesian framework; non-informative prior; informative prior; 90th percentile; 95th percentile; South African Coal Mining Industry.

Table of Contents

List of Tables

List of Figures

Abbreviations and Acronyms

HEG	Homogenous Exposure Group
SEG	Similar Exposure Group
TWA8h	Eight-hour Time Weight Average Dust Concentration
BHF	Bayesian Hierarchical Framework
BDA	Bayesian Decision Analysis
DMRE	Department of Mineral Resources and Energy
CMDLD	Coal Mine Dust Lung Disease
CWP	Coal Worker's Pneumoconiosis
OEL	Occupational Exposure Limit
MCMC	Monte Carlo Markov Chain
MHSC	Mine Health and Safety Council
MHSA	Mine Health Safety Act
SANAS	South African National Accreditation System
AIHA	American Industrial Hygiene Association
CEN	Committee of European Normalisation
CoP	Code of Practice
P90	90th percentile
P95	95th percentile
AM	Arithmetic Mean
GM	Geometric mean
GSD	Geometric standard deviation
BOHS	British and Dutch Occupational Hygiene Societies' guidelines
SAMOHP	The South African Mines Occupational Hygiene Programme
SHELF	Sheffield Elicitation Framework

CHAPTER 1: BACKGROUND

In this chapter, the background on the coal industry globally and in South Africa, coal mine dust lung disease (CMDLD) and the main context of this PhD study were presented. The background also discussed the occupational exposure limit (OEL) and control categories, the constitution or grouping of homogenous exposure groups (HEG) using guidelines from the Department of Mineral Resources and Energy Code of Practice (DMRE-CoP). Information on compliance testing and homogeneity of exposure assessment in the DMRE-CoP and international consensus were also introduced. From the international consensus, the general concepts of Bayesian framework for analysis of overexposures in industrial hygiene was explored. Finally, the problem statement, study rationale, aims and objectives of this PhD thesis were presented.

1.1. Coal

Coal refers to flammable, carbon-like, sedimentary rocks containing carbon and hydrogen.[1, 2] Figure 1 depicts the coal feature.[3] Calorific value, ash, moisture, and sulphur are the four main characteristics that are found in coal. Bituminous coal makes up the majority, with only a trace percentage of anthracite and silica.[4] Anthracite, bituminous, black lignite, and lignite peat are the four subcategories of coal.[5] Coal have a higher combustibility because of free radicals.[6] Coal dust is produced during coal mining operations. Dust is defined as tiny "dry particles emitted into the air by natural forces, such as wind, and by mechanical operations, such as grinding, crushing, and drilling".[7] In the coal mine, these mechanical tasks are often performed.

Figure 1. An open-cast coal mine. Source: https://aidc.org.za/8160-2/

1.1.1. Global Coal Industry

A quarter of the world's energy comes from coal. Countries such as China, the USA, and India together account for 70% of global coal energy consumption.[8] Global coal consumption is predicted to increase by 2040 at a rate of 1.5% annually.[9] China produces the most coal globally, followed by India, the USA, and Indonesia.[8,9] Some countries have taken steps to decrease coal production but it is expected that annual growth in coal production will average 2.3%, reaching an estimated 8.8 billion metric tons by 2025.[10] In year 2019, about 36% of the total fuel used to produce electricity around the world came from coal.[11]

1.1.2. South African Coal Industry

In year 2022, the seventh largest producer of coal in the world is South Africa. The South African coal industry had 92,230 employees in year 2019, up from 86,647 in 2018, or 19% of all mining industry jobs.[12] The South African annual coal production is estimated to be 258.9 million tons.[12] The majority of the coal deposits in South Africa have shallow, thick seams, making mining them simpler and less expensive. Given the current rate of production, there should be enough coal for another 50 years.[12, 13] In South Africa, coal mining takes different forms. These activities primarily employ the longwall excavation and pillar method, and they are both opencast (49%) and underground (51%).[13] Figure 3 shows an example of underground coal mining in South Africa.[14] In other words, it is either underground mining or surface mining, with depths typically not exceeding 300 meters.[4] South Africa's coal mine fields are primarily found in the provinces of Mpumalanga, Limpopo, and KwaZulu-Natal.[15] There are currently 19 coalfields in the country. Many of these coalfields are in the two provinces of Limpopo and Mpumalanga, with a small number also in KwaZulu-Natal and the Free State.

The coal mining sector supplies more than 70% of the nation's energy needs.[15] South Africa still lacks an alternative to coal to produce energy, despite other countries like India and the US developing alternative sources of energy. As a result, there will still be a long-term reliance on coal. The South African chemical and steel industries both rely heavily on coal.

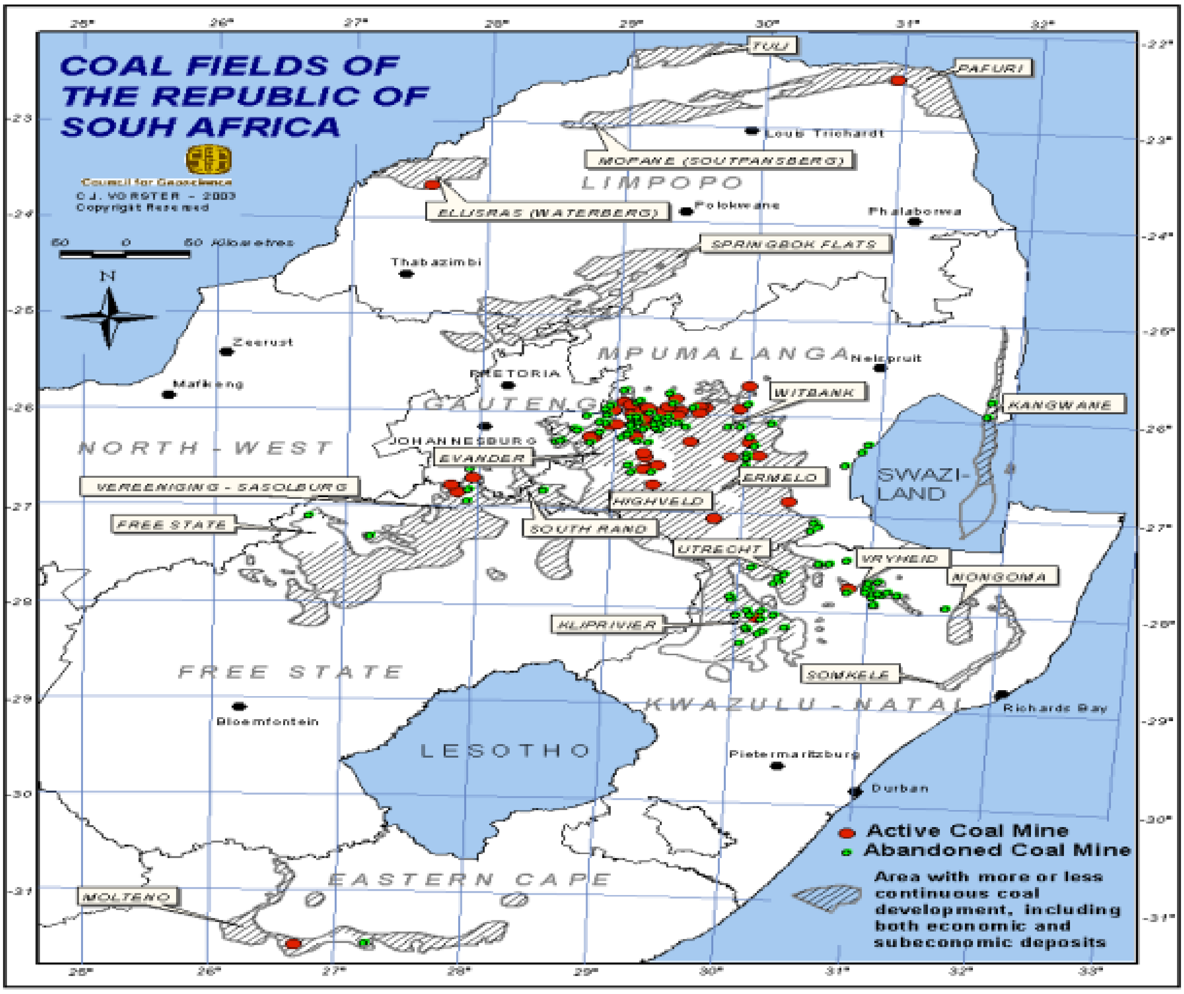

Figure 2. A South African map showing the distribution of coal fields. Source: http://www.southafricanmi.com/south-africa-mining-production-17jul2018.html.

Figure 3. South African underground coal mining.
Source: https://www.southafricanmi.com/south-africa-mining-production-17jul2018.html

1.1.3. Coal Mine Dust Lung Disease

According to the MHSA, a person involved in risk assessment must be familiar with the epidemiology of occupational lung diseases among mine workers.[4] This is important so that regular checks, treatment, and compensation for those who died or are unable to work can be provided. The Occupational Diseases in Mines and Workers Act (ODMWA) and the Compensation for Occupational Injuries and Diseases Act (COIDA) are the two acts that are used by the MHSA to cover compensation for occupational diseases.[4] Inhaling respirable coal dust increases the risk of getting coal mine dust lung disease (CMDLD), also known as coal worker's pneumoconiosis (CWP).[16,17] The CWP poses a serious threat to coalmine workers. Inhaling respirable dust, which is typically less than 10 µm in size D50 = 4 µm and invisible in ambient light, damages the lungs. The accumulation of dust particles in the lung can result in tissue inflammation and scarring.[18] The pathogenesis gradually manifests until it becomes irreversible when the lung is damaged. A well-known occupational health risk, respirable crystalline silica (RCS), is found as quartz in underground coal mines and can lead to silicosis, progressive massive fibrosis, and other silica dust-related diseases.[19, 20]

According to the International Labor Office Classification of Pneumoconiosis, CWP is defined as small opacities seen on chest radiographic film either as a profusion of 1/0 or a large opacity.[21] Soutar et al, 1993[22] reported that CWP poses a serious threat to the lives of coal mine workers in underdeveloped countries. The risk of developing CWP due to excessive coal dust exposure over the course of a worker's career depends on various work activities, as shown by Attfield & Seixas[23]. In the coal mining industry, work activities produce dust; these are attributable to poor organization or a lack of adequate exposure control for coal and other dust. Inadequate methods for removing coal, poor transportation networks, and coal spills, are a few of the activities.[24] The other mine conditions may include the generation of quartz dust, from the rock strata with thin seams; other coal seam-generated products; and the use of rock dusting products, e.g., limestone, to reduce explosions.[25, 26] All of these products are found in coal dust, which is produced during coal mining activities.

1.1.4. Occupational Exposure Control Categories

The Mine Health Safety Act (MHSA) 29 of year 1996 and the Occupational Health and Safety Act are both enforced by the South African Department of Minerals and Energy and Labour. These rules are in place for workplace health and safety in factories, offices, and mines.[4] In risk assessment, the DMRE-CoP requires that mining companies conduct regular risk assessments to identify potential hazards and assess the risk associated with these hazards. The DMRE-CoP also requires that mining companies provide appropriate PPE to workers to protect them from potential. To know that the DMRE-CoP measures are working, the occupational exposure limit (OEL) was imposed on all mine workers. The OEL refers to the amount of exposure to any potentially hazardous substance that should not have any negative effects on a worker's health in their working lifetime. In South Africa, the OEL for coal dust is 2 mg/m^3. The exposure of an individual to dust is defined as an eight-hour time-weighted average (TWA8h). The exposure of groups can also be categorized into four exposure categories, as a fraction of the OEL.[4,27] Table 1 lists the categories for exposure control groups and the sampling method for homogenous exposure groups (HEGs) in the DMRE-CoP.[27,28]

Table 1. The exposure classification categories and the sampling strategy in the DMRE-CoP[27, 29]

Category	Description	Statistical Illustration	Exposure Profile	Minimum Frequency
1	Exposures less than 10% of the OEL 10% of the time	P90 < 0.1% OEL	Very highly controlled	No sampling plan for this category. Measurement results that are below 10% of the OEL will be reported under this category
2	Exposures exceed 10% of the OEL and less than 50% of the OEL 10% of the time	P90 ≥ 0.1 OEL < 0.5 OEL	Highly controlled	Sample 5% of workers within a HEG on an annual basis with a minimum of 5 samples per HEG, whichever is greater.
3	Exposures exceed 50% of the OEL and less than OEL 10% of the time	P90 ≥ 0.5 OEL < OEL	Adequately controlled	Sample 5% of workers within a HEG on a 6-monthly basis with a minimum of 5 samples per HEG, whichever is greater
4	Exposures exceed the OEL 10% of the time	P90 ≥ OEL	Poorly controlled	Sample 5% of workers within a HEG on a 3-monthly basis with a minimum of 5 samples per HEG, whichever is greater

1.1.5. Grouping of Workers

In grouping of mine workers for surveillance or checking of similar exposure profiles, the constitution of HEGs is regarded as a key factor. The grouping of workers is important because only a small sample of workers can be used to find the exposures of the remaining workers since they will be assumed to be similar exposure.

In the DMRE-CoP, the HEGs are defined as a group of workers whose exposures to harmful agents have been proven to be statistically similar enough in such a way that by monitoring a small number of workers, the exposure profile of the remaining workers can be defined.[27] The exposures within a HEG should be sufficiently comparable that assessing the exposure yields information that can be used to forecast exposures for all other workers belonging to that HEG. A sample of HEG and its individual job titles are shown in Table 2 and the coal dust data capture sheet in Appendix A. Section 12 (3) of the MHSA requires ongoing evaluation of the medical surveillance database and occupational hygiene. In the DMRE- CoP, the HEGs in the underground area are created using a step-by-step procedure where the first step is to divide the mine into small areas with common air and return intake. In step two, the area is then divided into activity areas.[27,29] The exposures of job titles are not considered at this point.

Table 2. Example of subdivision of HEGs in DMRE-CoP

HEGs	Specific job titles found in a HEG
HEG A	Pump attendant
	Roofbolt operator
	Shuttle car operator
HEG B	Pump attendant
	Roofbolt operator
	Safety officer

Other countries employ the approach of similar exposure groups (SEGs) to appropriately define workers with similar exposure profiles and sample a group of workers for occupational safety and health assessments. SEGs are defined by job titles or tasks, task-specific exposure profiles, workforce experience, exposure duration and location, and equipment used.[30, 31] One major benefit of SEG is that it avoids the need to individually assess each worker by finding the extent of exposure risk to which each worker is likely to be exposed. According to Resources Safety and Health, 2022[32], a representative sample is likely to provide an estimate for the remaining workers in a SEG.

The American Industrial Hygiene Association (AHIA) guidelines have criteria for establishing SEGs.[33] These criteria include the observational approach which is based on processes, jobs, tasks and the environmental agents presented in Table 3. Secondly on sampling approach include monitoring to redefine groups. In the observational approach, workers will be assigned based on activities they perform, and personal judgement based on their expected similarity in the exposures of interest. Workers are divided into subgroups based on the hygienist's qualitative judgement of the work area. A basic definition of environmental agents is considered, for example, chemicals, by-products, products, biological agents etc. These environmental agents are combined with task duration and time on the production before linking to a SEG. So generally, the work is divided into processes, then processes are subdivided into jobs or specific tasks, and workers are identified and grouped into a SEG. Details on the observational process of establishing SEG are available in the AIHA Strat Monography 4th edition.[33]

Table 3. Description of observational approach in AHIA.

Process	Job	Task	Environmental Agent
Coil coating	Coal coat operator	Coal feed	Noise
Coil coating	Coal coat operator	Discharge	2-butoxyethanol
Coil coating	Coal coat operator	QC	MIBK
Coil coating	Coal coat operator	Clean up	MIBK
Coil coating	Coal coat operator	Clean	Cyclohexanone

The second approach used by the AHIA is by sampling where the environmental agents and workers that can potentially be exposed to the agents are identified, and then the monitoring data are collected and used to group workers into SEG through some statistical techniques.[34,35] Details on establishing SEG based on sampling are further discussed in section 1.1.8 on the international consensus on homogeneity.

The British and Dutch occupational hygiene societies (BOHS-NVvA)[36] guidelines and European standards (CEN)[37] on establishing SEG are like that of the AHIA. Grouping workers into SEG will require three factors to consider, namely, the materials, the processes and the way of doing the work activity. Workers performing different tasks in different work shifts must be allocated to more than one SEG. The BOHS-NVvA also emphasises that if workers are doing the same task but an individual worker is doing it in different ways or methods, they should be grouped in different SEG. The guidelines also require that when measurement samples are available, the homogeneity of the SEG can be statistically assessed, details about this are described in section 1.5.

1.1.6. Compliance Testing and Homogeneity in the DMRE-COP

The Mine Health and Safety Act (MHSA) of year 1996 required mines to implement health and safety management procedures, including workplace exposure risk assessment and medical surveillance, to make ensure the health and safety of the workers are considered.[29] It requires that 5% or a sample size of five workers are taken from each HEG whichever is higher, for exposure assessment. The DMRE-CoP guidance states that for a HEG to be certified as homogenous, the arithmetic mean (AM) and the 90th percentile (P90) of the exposure data should fall in the same exposure category that is associated with the OEL.[29] The DMRE-CoP approach and other approaches are summarized in Table 3. The current DMRE-CoP approach for compliance testing assumes that exposure samples have a normal distribution and that the P90 of the exposure should be below the OEL. According to this standard, a 10% overexposure above the OEL is considered acceptable.

1.1.7. International Consensus on Compliance Assessment

Exposure assessments are important to understand health risks and to document exposure patterns. The BOHS-NVvA guidelines test compliance with an 8-hour OEL airborne agent or aerosols. The guidelines assume that occupational exposure data follows lognormal distribution. The lognormal distribution refers to the continuous distribution of a variable that is skewed to the right or has a long tail (see Figure 4). The lognormal distribution is widely used in occupational exposure assessments and in non-occupational settings. In lognormal distribution, the logarithm of a random variable is normally distributed.[38, 39] Therefore, if the exposure has a lognormal distribution, then $\mathrm{Y} = \log(\mathrm{x})$ will be said to have a normal distribution. Data that have lognormal distribution take only non-negative values.

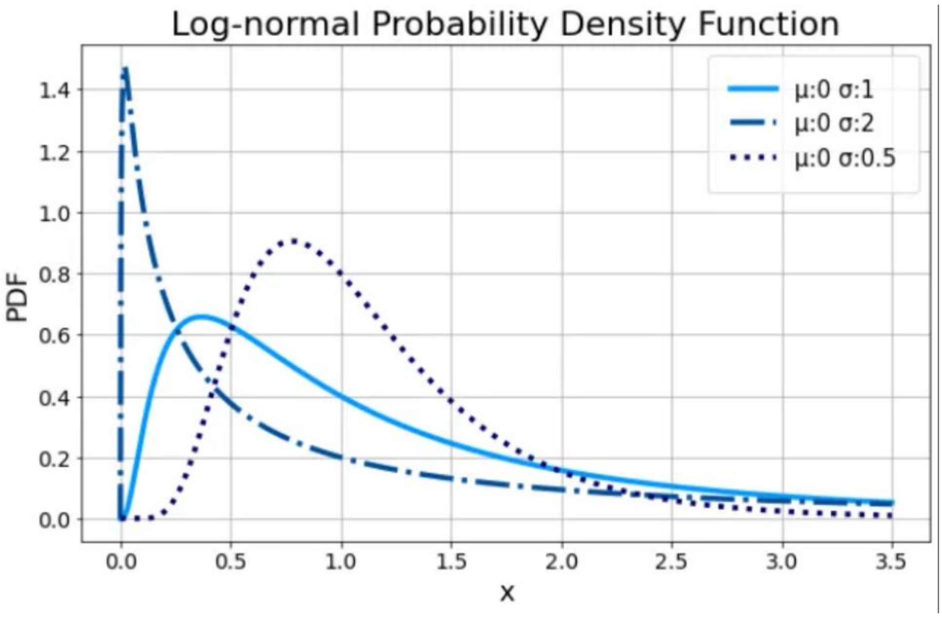

Figure 4. The example of the lognormal distribution. Source: https://towardsdatascience.com/log-normal-distribution-a-simple-explanation-7605864fb67c

The probability density function is defined by the mean μ and the variance σ where exposure x >0. In occupational hygiene exposure assessment, exposures are assumed to be always greater than 0. In other words, there is always some quantity of exposure even if it is below the limit of detection or quantification. The probability density function is given by the equation:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^2}$$

Where μ is the geometric mean (GM) which is the mean of the log-transformed data x, and σ is the geometric standard deviation (GSD), which is the standard deviation of the log-transformed data. The distribution is skewed due to the variability of exposure data. Based on the common variability of industrial hygiene data, the lognormal distribution assumption has gained widespread acceptance.

According to AHIA[33] and BOHS-NVvA[36] guidelines, each SEG must have at least six samples that can be used to decide whether less than 5% of exposures exceed the OEL. The SEG upper 99, 95 or 90th percentiles of exposure are selected to compare to the OEL to confirm a minimal exposure risk. Since there can be a high degree of uncertainty in exposure assessment, comparing the percentiles to the OEL should be accompanied by a degree of certainty. Uncertainty is the confidence level of making sure that the exposure level is indeed above or below the OEL. For example, whether there is a 95% certainty (confidence) that the 95th percentile (P95) value is below the OEL. The P95 of the exposure data must be below the OEL to certify that the exposure is under control or acceptable. According to the AHIA, BOHS-NVvA guidelines and the European Committee for Standardization (CEN). The P95 of the exposure below the OEL means that a 5% exceedance of the OEL is acceptable.[36, 37] Figure 5 below shows occupational exposure assessment procedures employed by the BOHS-NVvA guidelines. In the first step, SEG is established (details in section 1.3 above). In the second step, a total of 3 samples are randomly selected from the SEG for compliance testing. If the exposures are above the OEL, another 6 samples are taken from the same workers for further compliance testing. A total of 9 or more samples are recommended to establish uncertainty of exposure against the OEL. The uncertainty is based on 70% confidence that there is less than a 5% probability of the exposure being greater than OEL.[37] Furthermore, individual compliance can be conducted to see if there is less than 20% probability that 5% of the exposure is above the OEL. Exposures above the OEL should warrant further control measures. Uncertainty is discussed in addition to this acceptability criterion of 5% exceedances of the OEL. The BOHS-NVvA guidelines and CEN recommends that the probability of the exposure being above the OEL should be less than 30%.[36, 27] The probability should be at most 30%, that 5% of the exposures are above the OEL.[36, 37]Additionally, according to the AIHA, there should be less than 5% probability, that the actual P95 of the exposure is above the OEL.[22] The AHIA compliance testing is also similar to that of the BOHS-NVvA, which starts with data collection, data analysis, and data interpretation.

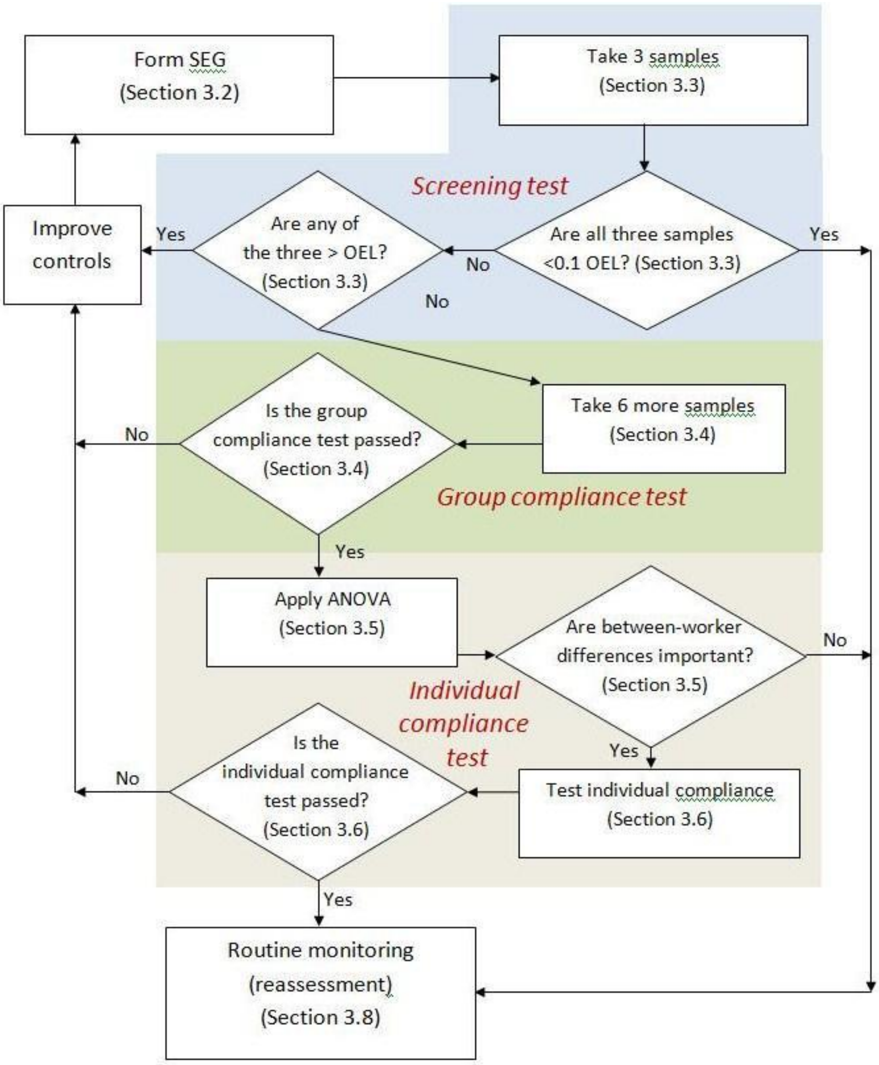

Figure 5. The British and Dutch guidelines occupational exposure assessment strategy. Source: https://www.arbeidshygiene.nl/-uploads/files/insite/2011-12-bohsnvva-sampling-strategy-guidance.pdf.

1.1.8. International Consensus on Homogeneity Assessment

The DMRE-Cop approach of homogeneity assessment is different from that of the AIHA, CEN, and BOHS-NVvA recommendations. The approach used in these guidelines to prospectively define SEG or HEGs (in the South African context) is predicated on the idea that grouping workers who perform related tasks would lead to similar exposure profiles. There are several methods for assessing the homogeneity of HEG or SEG. The analysis of variance (ANOVA)[40] is used by the AIHA and BOHS-NVvA to assess homogeneity.[41, 42]

The ANOVA is used to compare the exposure means of more than two groups of the population. For example, the mean of exposures per SEG, year or work shift. The ANOVA analysis is commonly used in log-transformed exposure data to compare their mean distribution. As stated earlier lognormal data produce the GM and GSD of the data, the GM is the parameter of interest compared between the groups. The analysis consists of the between-worker variation (differences of exposure between two groups) and the within-worker variation (differences of exposure arising from repeated measurements within one group).[37, 41] If the exposure profiles between-worker and within-worker are similar, the probability of observing a difference will be more than the p-value = 0.05.[40] If the p-value is less than 0.05, then the GM of exposure between the groups are different. The between-worker GSD ($GSD_{b)}$ and within-worker GSD (GSD_w) can also be calculated, where GSD_b is the overall variability of exposure and GSD_w is the individual exposure variability.[41,42]

Table 4 shows the various methods used in this study for assessing homogeneity and compliance testing.[29,36,37] The total variation of exposure in SEG is composed of individual exposure variation. Therefore, if the between-worker variation is greater or equal to 20% of the total variation, individual worker compliance is encouraged. The individual compliance test of exposure against the OEL will be passed if the between-worker variability is less than 20% of the total variance. Therefore, within-worker variation of more than 0.2 (20%) would show that the group is not homogeneous, per the BOHS guideline. The Rappaport ratio (R ratio), which displays the ratio of the highest exposed worker to the lowest exposed worker, is another indicator of homogeneity. It has been proposed that a ratio of two or less defines a homogeneous exposure distribution. The R-ratio is also defined as the "ratio of the 97.5th to the 2.5th percentiles of the distribution of individual AM." Another indicator of homogeneity is the global geometric standard deviation (GSD). A HEG or SEG exposure may not be homogeneous if the global GSD is greater than three.[37]

Table 4. The methods used to assess compliance and homogeneity.

		Methods based on various guidelines		
Compliance testing		**DMRE CoP**	**CEN/BOHS-NvVA**	**AIHA**
	For data to be compliant	The P90 of the data should be below the OEL	The P95 of the data should be below the OEL	The P95 of the data should be below the OEL
	Assumption of the underlying data	Normal distribution	Lognormal distribution	Lognormal distribution
	Degree of certainty	None	70%	95%
		Guidelines used by various methods		
Assessment of Homogeneity		**DMRE CoP**	**Global GSD**	**R-ratio**
	For data to be regarded as homogenous	The AM and P90 values must fall within the same OEL exposure band for homogeneity	A GSD of 3 or less was used to confirm homogeneity	The R-ratio point estimate of 95% and 90% of the workers, where a ratio of 2 or less was considered homogenous

1.2. Study rationale

The coal mine dust lung disease, an irreversible lung disease, is more likely to affect coal mine workers as they get exposed to coal dust. The risk of getting CMDLD is increased by exposures above the OEL. There are no established standard therapy modalities or illness management procedures for CMDLD, and the condition cannot be cured. Due to various methods of defining the constitution of HEGs, which can be overly heterogeneous, it is now more challenging to distinguish between under and overexposed workers. In addition, the DMRE approach uses frequentist statistics for each new set of monitoring data separately, without likelihood estimation to derive the posterior distribution.

In the coal mining industry, there has not been a full discussion of the need for an evidence-based assessment of exposure levels and the creation of frameworks to improve the way decisions are made about occupational hygiene exposure. Exposure levels must be considered to find workers who are most at risk and improve prevention plans that are already in place. The DMRE-CoP methods do not consider the uncertainty surrounding the P90 and simply assert that the P90 of the present exposure data should be below the OEL to certify compliance with the exposure. In the Bayesian framework, uncertainty about a parameter is shown by the credible interval. For instance, the credible interval refers to the probability that an estimate will be found inside a specific range. Bayesian statistics naturally allows a prior distribution derived from the common population distribution.

Several methods can improve the constitution of HEGs and compliance testing. These methods also consider variability and distribution of exposure data. In this PhD, the various methods that are used in other countries are compared with the current DMRE-CoP to develop, to improve procedures for constituting HEGs, and recommend compliance assessment tools for the coal mining industry.

1.2.1. Study question, aims and objectives

The research question: Can refined grouping and compliance testing using the BHF improve the identification of highly exposed workers compared to the current DMRE-CoP?

The study aimed to improve the identification of coal dust overexposure by introducing new methods for compliance (reduced dust exposure) and homogeneity (similar dust exposure level) assessment in the South African coal mining industry.

The specific objectives of the study were the following:

1. To compare compliance of coal dust exposure by HEGs using DMRE-CoP approach and other global consensus methods.
2. To investigate and compare the within-group exposure variation between HEGs and job titles.
3. To determine the posterior probabilities of locating the exposure level in each of the OEL exposure categories by using the Bayesian framework with previous information from historical data and compare the findings and the DMRE-CoP approach.
4. To investigate the difference in posterior probabilities of the P95 exposure being found in OEL exposure category between previous information acquired from the experts and the current information from the data using Bayesian analysis.

CHAPTER 2: LITERATURE REVIEW

2.1. Prevalence of Coal Mine Lung Disease in the World

The National Institute for Occupational Safety and Health (NOISH) of the Centre for Disease Control (CDC) reports that from year 1970 to 2016 in the United States, coal worker's pneumoconiosis (CWP) affected the lives of 75,178 miners, killing 4,118.[43] In countries with weak control procedures for exposure to coal dust, the death rate from Coal Mine Dust Lung Disease (CMDLD) among coal mine workers has increased. The onset of CMDLD was linked to greater death rates, according to a Polish study, with a standardized mortality ratio (SMR) of 383.[44] Similar studies from China and the Netherlands discovered that there was a higher risk of death and an SMR of 4,523 when 89% and 89% of the workers were diagnosed with CMDLD, respectively.[45, 46] The amount of time spent at coal mining sites and the work activities may have cumulatively harmful effects that result in developing CMDLD.[47] According to empirical data by Torres et al, 2014[48] , those who have worked in the coal industry for at least 25 years are four times more likely to get a radiological diagnosis of CMDLD.

2.1.1. Prevalence and Incidence of Coal Mine Lung Disease in South Africa

The epidemiology of CWP has received little attention. In recent years, there has been a sharp decline in CWP certification.[49] According to the information that is currently available, there were 25 cases of CWP per 1,000 coal mine workers in South Africa.[49] According to a recent radiological report in the province of KwaZulu-Natal, 15% of 187 workers with more than 5 years of experience had CWP.[50] Among young workers, CWP was discovered to be positively associated with number of years of work. An X-ray report showed that the prevalence of CWP varied from 1.8 to 4.2% in another study. The incidence of CWP increased as exposure to coal dust increased, which increased with years of employment. According to autopsy data, CWP was 7.3% compared to 10.8% for silicosis, both of which show an association with increased exposure years. The average number of years of work of employment for coal miners in South Africa is about 17 years.[51] Given the number of years of exposure, there may be a significant risk of developing CMDLD among work. From a study done in KwaZulu-Natal, the number of years of work was positively associated with the prevalence of CMDLD.[52] In another study, cumulative dust exposure increased the probability of developing CMLD. The data that is currently available from the Mine Health and Safety Council (MHSC) revealed that between 100 and 190 people per 1,000 workers in South Africa were estimated to have CMDLD in year 2005.[49]

2.1.2. Application of Bayesian Framework in Occupational Hygiene

Bayesian statistics is a type of statistical inference based on Bayes' Theorem. Bayes' Theorem is the conditional probability of an event given the data and the prior information.[53, 54] Bayesian statistical analysis methods have become very popular in scientific research and have been widely used in industrial hygiene for exposure assessment.[55, 56] Bayesian statistics have been used to determine the likely AHIA exposure control category from a SEG by Hewett et al, 2006[57]. The AHIA proposed that the Bayesian approach in which a judgment can be used to estimate the probability that the true exposure will be found within one of the exposure control categories is important in occupational risk assessment. To obtain the posterior probabilities in each of the exposure categories, the current monitoring data or measurements are combined with prior information from historical data or expert judgements in a Bayesian framework. Using an informative prior distribution derived from either historical data or expert judgements of exposure through elicitation can reduce the uncertainty of an estimate. Further details on expert elicitation are covered in section 2.1.3. While approaches in the DMRE guidelines and other Bayesian approaches use point estimates like mean, median, and P90, the Bayesian method available in the AHIA guideline estimates the probabilities of the P95 of an exposure falling in each AHIA exposure control category. This method is called the Bayesian Decision Analysis (BDA). Figure 6 below illustrates the decision charts from the BDA. The prior decision chart has the previous information from expert judgements or historical data which indicate that the P95 from a SEG falls or is rated in one of the AHIA exposure controls categories. When expert judgements or historical data are unavailable, a non-informative prior can be developed to give an equal probability of the exposure being located in each exposure control category. The likelihood functions based on monitoring data are calculated as described in the BDA. Finally, the posterior is the probability of the P95 being found in each exposure category after combining the prior and the likelihood function in a Bayesian framework.[57]

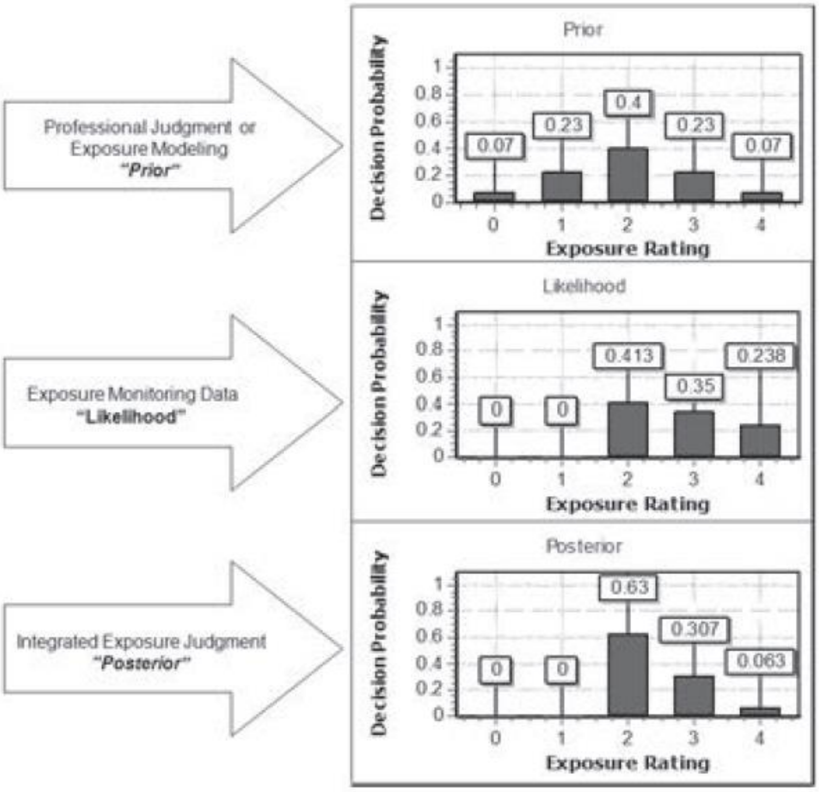

Figure 6. Prior, likelihood and posterior decision charts for industrial hygiene from the BDA.[57]

The main advantage of Bayesian statistical analysis in occupational hygiene is its ability to combine information from experts in the industry with a small number of measurements and still produce results that are conditional on the data, and exact while naturally allowing solid decision-making framework.[58] Classical statistics make the unrealistic assumption that the GM and GSD can be any value, no matter how big or physically impossible it can be. In occupational hygiene, the Bayesian framework, on the other hand, gives the flexibility to limit the analysis to a reasonable parameter space. In the Bayesian estimation of industrial hygiene data, a suitable range of values for GM, GSD, and the P95 can be set. The Bayesian results are also simple to interpret and communicate, even for non-statisticians. For example, there is a greater than 95% chance that the exposure profile falls into category 4, implying that the exposure for SEG or HEG in our case is unacceptable. This is much easier to communicate than saying the point estimate is in category 3 and the confidence limits are in another category, making it difficult to find whether exposures are acceptable or not. The ability of Bayesian statistics to incorporate prior information from expert judgments into a model with current monitoring data is valuable, in the absence of historical data. It is important to include this information in exposure assessments because earlier exposure and exposure control measures influences future exposures among mine workers.

The Bayesian framework may show improvements in worker classification based on occupational exposure categories. Through probabilistic theory, an online Bayesian computation tool called

EXPOSTATS[59] available at https://expostats.ca/site/en/index.html is used to perform occupational exposure compliance assessment. Other tools that can be used to perform occupation hygiene risk assessment are AltrexChimie[60], HYGINIST[61], IHDataAnalyst[62], BW_Stat v3[63], and IHSTAT.[64] Some of these tools are in Excel worksheets while others are web-based and freely downloadable. The IHSTAT was specifically developed for performing BDA in the AIHA.

2.1.3. Expert elicitation

Expert elicitation is the process of synthesizing experts' subjective judgements, opinions or beliefs about an unknown quantity of exposure.[65] Experts elicitation is conducted when there is uncertainty resulting from a lack of data or inadequate data about the quantity of interest. Based on their experience, expertise, insight, and knowledge about the exposure in a job title, experts are expected to give advice or opinions about the exposure. In the Bayesian context, expert judgements are mainly used to generate prior information or distribution about the unknown parameter of interest. The prior distribution is then combined with the current monitoring data to establish exposure compliance. The first thing in conducting elicitation is selecting the experts and providing training, secondly, carrying out the elicitation by getting expert judgements about the exposure, thirdly, a probability distribution such as the lognormal is fitted to the results provided by the experts, and lastly, assessing the quality, reliability, and truthfulness of the elicitation.[66, 67]

The most common expert elicitation is when multiple experts are engaged and then a single probability distribution is fitted to get their combined judgements. Aggregating their knowledge may involve two methods. One is the mathematical aggregation where separate judgments are combined using a mathematical formula also called the pooling rule. Such an example of aggregation can be a simple average or weighted average of the experts' judgements. Another method is behavioural aggregation where experts in a group are asked to discuss their knowledge and experiences about the uncertain quantity of interest and come up with a consensus or a final agreed result.[67, 68]

Few elicitation protocols exist. The cooke protocol uses a mathematical aggregation where experts make judgements about the unknown quantity of interest which is known to the facilitator, and this is compared.[69] Cooke's classical method has been applied in North America with the Great Lakes fisheries[70] and prediction in the rise of sea level and ice melt.[71] The Sheffield protocol which has the SHELF (Sheffield Elicitation Framework) that employs behavioural aggregation is an online tool (https://shelf.sites.sheffield.ac.uk/) freely available for conducting elicitation.72 The SHELF has been used in drug development73 and in understanding response to ASD symptoms.[74] Another protocol is the Delphi technique used previously only for forecasting.[75] Recently the DELPH tool is also used for eliciting judgments from experts. [76] The Delph method employs both mathematical and behavioural aggregation. Another protocol used for structured expert elicitation is the IDEA.[77] The IDEA protocol (Investigate, Discuss, Estimate, and Aggregate). The IDEA protocol has been applied in studies. The US Intelligence Advanced Research Projects Activity assessed the protocol in public health, ecology, and conservation studies.[78,79]

2.1.4. Conceptual Framework

The current way of defining HEGs and the analysis methods employed by the DMRE-CoP for evaluating homogeneity and compliance to the OEL served as the motivation for this PhD research conceptual framework. The research offered an alternative way of constituting HEGs and methods for exposure compliance assessment. Figure 7 presents the current definition of HEGs by the DMRE-CoP. HEG is grouped by a common area and air intake in the DMRE-CoP, regardless of different job titles or labour activities. Underestimating or overestimating the risk of exposure could result from how HEGs are grouped by DMRE-CoP. In this PhD study, it is suggested that HEGs should be classified according to certain job titles because they have a common work activity and exposure. It is easier to detect highly exposed workers with job titles as an alternative group. The DMRE-CoP does not take historical data into account for assessing exposure risk. In this study, historical data are used to update the current data in a Bayesian framework to find the probability of exposure for each exposure category. The robustness of Bayesian statistics, which can give results even with a small sample size, makes it essential to apply them. Since there are thousands of workers in the mining business, sampling every one of them would be impractical and expensive. However, a small sample with the right definition of HEGs should be able to provide representative data, allowing for solid decision making.

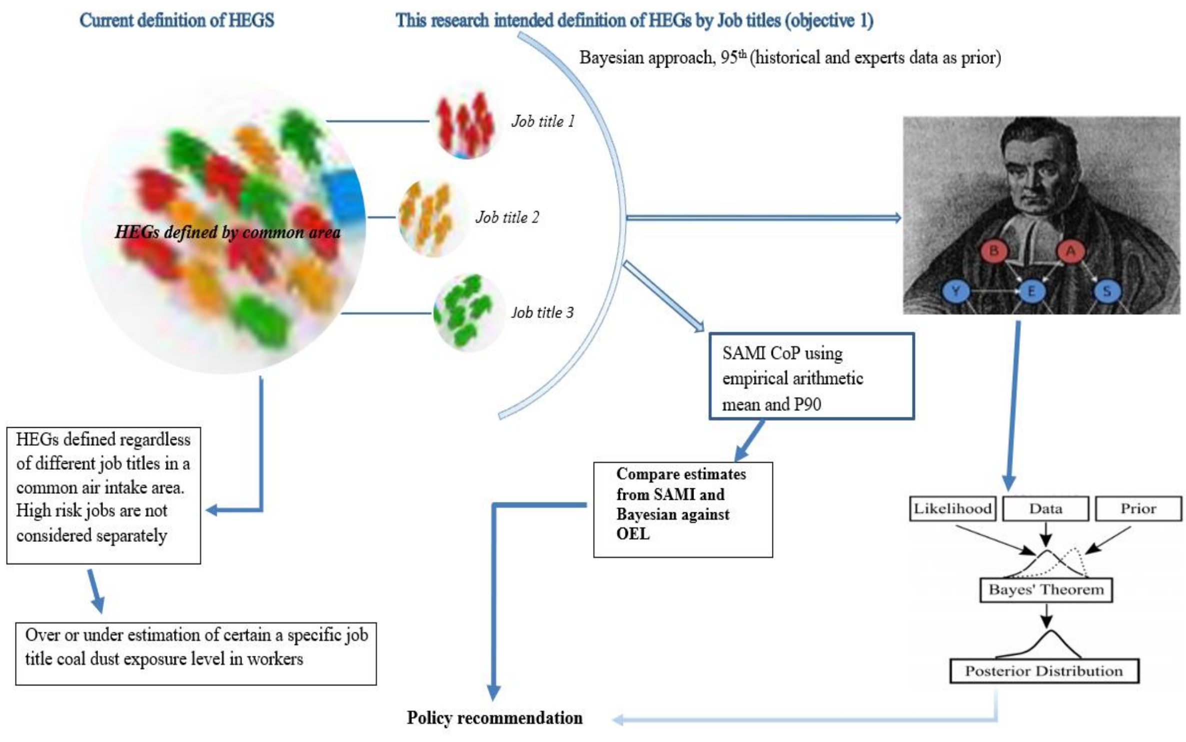

Figure 7. A conceptual framework, illustrating the DMRE's current definition of HEGs and the intended definition, and Bayesian analysis for compliance assessment.

CHAPTER 3: METHODOLOGY

3.1. Study Design and Population

The study included a cross-sectional collection of measurements of workers' exposure to respirable coal mine dust in South African underground coal mines from 2009 to 2018. Data on the exposure to respirable coal dust were periodically gathered from the mines' various geographic sites. Since all the workers in the coal mine were men, only male labourers were included in this study. The coal mining company's occupational hygiene division gathered the information. The survey included workers from 28 HEGs. Within each HEG, there were many job titles. A sample selection was done under DMRE-CoP.[29]

3.1.1. Data Collection

Sampling was done by using a stratified and random method to select 5% of the workers from a HEG.[29] The exposure data collections follow South African National Accreditation System (SANAS) standards. Each selected employee was given a dust sampling pump, which was mounted on their belt and controlled at a flow rate of 2.2 l/min. The calibration was done both before and after sampling to ensure quality assurance. If there was a concentration difference of more than 5% between the time before and after sampling, the sample was rejected. If the coal dust concentration was too high or if workers were not working on the day, more samples were dropped. The average flow rate before and after sampling was 2.24 L/min. If a sampling period was less than 80% of a 12-hour shift, then data was excluded from this analysis in this study.

The size-selective cyclone, which was based on the Higgins Dewell (HD) Cyclone model, was equipped with a mixed cellulose ester (MCE) filter. The HD cyclone is an American model that work at 2.2 lines per minute to give 50% cut at 4 µm per the specified distribution curve. The cyclone is typically used to remove the coarse or non-respirable dust from the sampled air before collecting the respirable dust on a filter substrate. A cyclone was placed in each worker's breathing zone on the collar to measure the amount of respirable dust they would be exposed to. The cyclone's filter was then sampled and gravimetrically analysed using NIOSH procedure 0600.[80] The gravimetric sampler was created because of the Federal Coal Mine Health and Safety Act of 1969, which requires compliance dust sampling.[81] To obtain the eight-hour time weight average dust concentration (TWA8h) concentration expressed in mg/m^3, the gravimetrically calculated coal dust mass (mg) was translated to a 480-minute sample time. The TWA8h concentration refers to the time-weighted average concentration of a substance over an 8-hour work shift. It is a commonly used measure in occupational health and safety to assess worker exposure to potentially harmful substances in the workplace. The TWA8h concentration is calculated by taking the average concentration of a substance over an 8-hour period, weighted by the amount of time workers are exposed to that concentration. This means that short-term exposures at high concentrations are balanced against longer-term exposures at lower concentrations. Each HEG was unique to a single mine or area. Different job titles were present in a particular HEG.

3.1.2. Selection of Experts

An invitation was sent via email, to experts within the coal industry. The experts included general managers, mine ventilation engineers, occupational hygienists, mining engineers, and geologists. Before collecting experts' opinions on exposure based on their past experiences, experts were trained on uncertainty based on their lowest and highest judges if not sure, the P95, and a summary of the tasks that needed to be done. The experts were provided with a description of each job title and shown an example of how to answer the probability of exposure to each DMRE-CoP exposure group to acquaint themselves with the elicitation questionnaire (see Appendix F and G). If an expert became confused or lost during the elicitation process, they were free to consult at any time. A sample size of at least six or more experts is required to conduct elicitation.[82] Eventually, six experts took part in this study.

3.1.3. Experts Elicitation Process

In this study, the IDEA structured expert elicitation protocol was employed to collect data from experts.[77] The IDEA protocol was modified in a way that experts were not allowed to share their individual opinions. The benefit of the modifying the IDEA protocol was that it stops experts from dominating the conversation and potentially skewing the verdict in favour of just one expert. The background information on experts is the first step in the IDEA protocol, which is then followed by an investigation in which experts are asked to individually respond to questions and justify their responses. An expert is shown anonymous responses from each participant during the discussion to settle conflicting understandings of the questions. The second and final estimates from each expert are obtained. The data are aggregated throughout the post-elicitation phase, and the mean of the second round of expert responses is calculated.

The IDEA protocol has the benefit of minimizing bias brought on by overconfidence, availability, and representativeness. A group of professionals were invited to take part in the investigation. The experts were granted time to study the questions before providing their judgement. Expert elicitation was conducted online. Appendix E shows the structured questionnaire used for eliciting expert judgements.

In the four-step process, the credible intervals were the minimum and maximum percentage of the P95 exposure falling in each of the DMRE exposure category. These ranges were standardised using linear extrapolation.[78,79]

The minimum standardised percentage: B-((B-L) X (S/C))

The maximum standardised percentage: B+((U-B) X(S/C))

Where B = the best guess, L = is the minimum percentage, U = is the maximum percentage, S = is the level of credible intervals to be standardised, and C = is the level of confidence (how sure is the expert?).

The responses were standardised to decrease overconfidence. A joint probability distribution was then created by combining the experts' opinions.

3.1.4. Data Management

To ensure consistency, data were double entered during the first data collection. Data cleaning was done including checking for duplicates and outliers. Following guidelines in the DMRE-CoP, extreme values above 10 mg/m^3 and duplicate data were removed. The data set did not have any missing values.

3.1.5. Ethical Consideration

For ethical considerations, the coal mining firm granted permission to use the data with confidentiality and participant privacy according to the Declaration of Helsinki. The Human Research Ethics Committee (HREC) of the University of the Witwatersrand granted ethical clearance for this project (clearance certificate no. M181038; Appendix B). The mining data gatekeeper granted the use of the company's data with a confirmation letter (see attachment in Appendix C).

3.2. Statistical Analysis

3.2.1. Descriptive Analysis

The aims, statistical analysis, and the PhD outcomes are summarized in Table 2, with more information in the results section. Regardless of the underlying distribution of the data, summary metrics like the arithmetic mean and standard deviation were displayed. The geometric mean (GM) and geometric standard deviation (GSD) were tabulated as well. Descriptive summary statistics were computed in Stata version 16.1 (StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC).

Table 5. An overview of the objectives, data collection procedures, and analysis statistical analysis methods, as well as the PhD study outcomes.

No.	Objective	Data collection tool	Variables	Data description	Methods of Analysis	PhD outcomes
1	To compare compliance of HEGs by using DMRE's current practice with international consensus methods (CEN/BOHS-NVvA and AIHA approaches).	This is secondary data collected according to the above procedure in data collection section	HEGs Job titles 8-hour TWA coal dust concentration	Categorical Categorical Continuous	Arithmetic means and Standard deviation. Geometric mean and geometric standard deviation for lognormal data. DMRE-CoP using P90 percentile. International approaches using P95 with 70 and 95% UCL.	Paper 1: Compliance Testing and Homogenous Exposure Group Assessment in the South African Coal Mining Industry.
2	To compare within-group exposure variation by the introduction of job titles.	As above	HEGs Job titles 8-hour TWA coal dust concentration	Categorical Categorical Continuous	DMRE-CoP: both arithmetic mean and P90 must fall within an exposure category. GSD>3 to indicate non-homogeneity. The Rappaport Ratio, the ratio of the most exposed divided by the least exposed using 90 and 95% Analysis of variance	Paper 1: Compliance Testing and Homogenous Exposure Group Assessment in the South African Coal Mining Industry.
3	To compare the posterior probabilities of the P95 of the non-informative and informative prior (from historical data) in the Bayesian framework and	As above	HEGs 8-hour TWA coal dust concentration	Categorical Continuous	DMRE-CoP using P90. Bayesian statistics, using informative prior from historical data and non-informative prior to have	Paper 2: Bayesian hierarchical modelling of historical data of the South African coal mining industry for compliance testing.

	DMRE-CoP relative to the OEL categories.				probability of exposure in each exposure category.	
4	To compare the posterior probabilities of the P95 of the non-informative and informative prior information (from experts' elicitation) in the Bayesian framework compared to the OEL categories.	A questionnaire was designed according to the IDEA criteria for expert elicitation.	Job titles 8-TWA coal dust concentration	Categorical Continuous	Expert elicitation Bayesian statistics, using informative prior from historical and expert data and non-informative prior to have probability of exposure in each exposure category.	Paper 3: Bayesian Hierarchical Framework from Expert Elicitation for Compliance Testing in the South African Coal Mining Industry
DMRE: South African mining industry; OEL: Occupational exposure limit; HEGs: Homogenous exposure group, TWA: Time weighted average, BHF: Bayesian hierarchical model						

3.2.2. Bayesian Inference

Bayesian inference is based on conditional probability distribution where a probability distribution of the current monitoring data vector $Y = (y_i \ldots, \ldots, \ldots, y_n)$ with n being the sample size given the model parameters θ, denoted as the likelihood $p(y|\theta)$. [53, 54] Conditioning on observed data makes Bayesian inference unique from all other statistical approaches. Bayesian inference combines the probability of the likelihood with the prior distribution (unknown quantity before the sample is observed) for θ, $p(\theta)$. The joint probability mass or density function is a product of the prior distribution $p(\theta)$ and the sampling data $p(y|\theta)$. Lastly, the inference is drawn from the posterior distribution as $P(\theta|y)$ where P (θ/y), is obtained through Bayes Theorem:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta} = C^{-1}P(y|\theta)p(\theta) \propto P(y|\theta)P(\theta) \qquad (1)$$

where $p(y)$ is the marginal or unconditional distribution of y. In practice, $p(y)$ is a normalising constant donated by C. Therefore, the posterior distribution will be computed as being proportional to the likelihood times the prior:

$$P(\theta/y) \propto P(y/\theta) x\, P(\theta) \qquad (2)$$

Where the $P(y/\theta)$ is the information from the current monitoring data and P(θ) is the prior information. The prior information may be non-informative, weakly informative, or informative. An informative prior exists when there is reliable external previous knowledge about the event of interest. When there is little or no reliable information available about the event of interest, prior information with little or no influence on the posterior distribution (the information after data is observed) is used.

The inference drawn from the posterior has a confidence limit called the credible interval. The credible interval informs you about the uncertainty estimate of the parameter of interest. The credible interval differs from the confidence interval in frequentist statistics. While the confidence interval is interpreted as the level of confidence in percentages, for example, 95%, that a given parameter is found between the given range, the credible interval is assuring that the probability of a given parameter falling between a given range is say 95%.

Traditionally, statisticians used to dismiss Bayesian statistics because of the difficulty in calculating the posterior distribution. With the development of Markov Chain Monte Carlo (MCMC) simulations, this has recently changed, as it is used for selecting samples from the posterior distribution based on conditional distributions. Examples of MCMC are the Gibbs sampling and Metropolis-Hastings. Gibbs provides parameter estimation with flexibility through simulation. [53, 54]

The MCMC diagnostics are then conducted to show convergence of the posterior distribution. Examples of such diagnostics tests are the trace plot and autocorrelation. A trace plot with a constant or flat distribution of the data points indicates convergence of the MCMC samples. The autocorrelation rules out a strong correlation between successive draws of the sampler. Another method for MCMC diagnostics used is the Gelman-Rubin diagnostics which assesses the difference between the multiple Markov chains. Further details on the MCMC inputs and diagnostics can be found in the Bayesian statistics book.[43,44]

3.2.3. Inferential analysis for objective 1

Objective 1 compared compliance of HEGs by using DMRE-CoP and international consensus methods (CEN/BOHS-NVvA and AIHA approaches).

1. Compliance Testing

The DMRE-CoP

The compliance of 28 HEG exposure profiles against the OEL. The South African OEL of 2 mg/m^3 for coal dust exposure is used for compliance testing. The empirical P90 of the exposure data was calculated and should be below the OEL to say that the HEG is compliant.

The International Consensus

In the CEN/BOHS-NvVA and AIHA approaches, the uncertainty or confidence level is based on the 70% and 95% probability, respectively, and the acceptability criterion for compliance is that the P95 exposure level should be below the OEL. The compliance testing was conducted in EXPOSTATS, the online Bayesian computation tool built-in with Monte Carlo Markov Chain (MCMC) algorithms. [59] In paper one (attached in Appendix F), the P95 of the underlying coal dust exposure data with lognormal distribution was calculated. EXPOSTATS tool 1, which estimates the parameters of the lognormal distribution and compares to the OEL was used to test compliance of the HEGs coal dust exposure data. Figure 8 is a screenshot of the EXPOSTATS tool. Under the section for calculating parameters, the South African OEL of 2mg/m^3 was inputted in the Exposure limit column. The Exposure limit multiplier column is used to modify the exposure limit, in our case the OEL is 2, so in this section, 1 was entered. Credible interval of 95% was inputted, and for the probability of the overexposure risk, 30% probability that 5% of the exposure is above the OEL was entered. The tool automatically conducts the descriptive statistics and proportion of the exposure above the OEL and does risk analysis based on the P95 exposure level as shown in Figure 9.

Calculation parameters

Exposure limit

2

Exposure limit multiplier

1

Credible interval probability

95

Overexposure risk threshold

30

Data

0.23
2.1
3.1
0.11
0.33
2.6

Descriptive statistics

parameter	value
n	6
Proportion censored	0
Minimum	0.11
25th percentile	0.255
Median	1.22
75th percentile	2.48
Maximum	3.1
Proportion >OEL	50 %
Arithmetic mean	1.41
Arithmetic standard deviation	1.34
Coefficient of variation	95 %
Geometric mean	0.722
Geometric standard deviation	4.22

Figure 8: Estimating parameters of the lognormal distribution and comparison to the OEL in EXPOSTATS. Source: Tool1: Data interpretation for one similarly exposed group (shinyapps.io).

Risk analysis based on the 95th percentile

Selection of the critical percentile

95

Risk decision

- Criterion defining overexposure:

  95th percentile ≥ OEL

- Uncertainty management : based on the Bayesian model, the probability that this criterion is met (overexposure risk) is:

  97.4 %

- Uncertainty management : The probability of overexposure (overexposure risk) should be:

  < 30%

- As a consequence, the current situation is declared:

  Poorly controlled

Parameter estimates - Distribution

The geometric mean point estimate and credible interval are: **0.72 [0.22 - 2.4]**

The geometric standard deviation point estimate and credible interval are: **3.8 [2.3 - 12]**

Parameter estimates - 95th percentile

The point estimate is: **6.46**

The credible interval is: [**1.98 - 63.7**]

Figure 9. EXPOSTATS tool 1 risk analysis. Source: Tool1: Data interpretation for one similarly exposed group (shinyapps.io).

The prior distribution installed in EXPOSTATS was weakly informative prior.

- For the GM, a uniform prior distribution that was only marginally informative was applied.[83]
  A uniform prior distribution for mean μ of HEG exposure expressed as the GM was specified as -20 and 20 mg/m^3 for the lower and upper bound, respectively. When expressed as the GM, the values were bounded between 2.10^{-9}*OEL and 5.10^{8}*OEL, with 95% of the distribution lying between 6.10^{-9}*OEL and 2.10^{8}*OEL. It is assumed that 23% of the prior mean of the exposure is found between OEL/100 and 100*OEL.
  Mathematically, this is expressed as $p(\mu) \sim U(\mu_{0,}\mu_1)$ where μ_0 and μ_1 are the lower and upper bound of prior on GM.

- EXPOSTATS used priors estimated by McNally et al, 2014[84] for the GSD. For the prior on the GSD, EXPOSTATS demonstrated that lognormal data generates a GM of 0.84 for σ similar to a GSD of 2.32 for exposure .
  The mathematical expression for the variance is $p(\sigma) \sim LN(\mu^*, \sigma^*)$ where μ^* is the log-mean of the prior lognormal distribution for σ and σ^* is the log-standard deviation of the prior lognormal distribution for σ.

2. Homogeneity assessment of HEG coal dust exposure

The DMRE-CoP

The AM and P90 of the exposure in a HEG must both fall within the same exposure control category for the exposure to be certified as homogenous.

The International Consensus

To confirm homogeneity, 1. a global GSD of less or equal to 3 needs to be established. 2. The R-ratio which is the ratio between the 97.5th and 2.5th percentiles of the distribution of each AM, where a ratio of 2 or less was considered homogenous to calculate the R-ratio point estimate for 95% of the workers. The R-ratio estimate using 90% of the workers, which is less stringent and defined as "the ratio of the 95th to the 5th percentiles of the distribution of individual AM," was also used in this analysis. The ANOVA methods for within-worker variability of exposure in a HEG were not used in this study because the study data does not have repeated measurements of exposure.

3.2.4. Inferential analysis for objective 2

Objective 2 compared within-group exposure variation (homogeneity) by the introduction of job titles.

The homogeneity of exposure from specific job title as an alternative grouping within a HEG was assessed and compared with the DMRE-CoP versus the global GSD. To assess the homogeneity of the job titles, only GSD was used because of the small sample size of the job titles. In the analyses mentioned above, all HEGs that did not show homogenous exposures across all approaches were selected to see if their constituent job titles will indicate homogeneity of exposure. Only HEGs with more than one job title that have three or more samples were selected. The DMRE-CoP approach of both the AM and the P90 having to be within the same exposure band as an indicator of homogeneity was compared to the global GSD.

3.2.5. Inferential analysis for objective 3

Objective 3: To compare the posterior probabilities of the P95 of the non-informative and informative prior information (from historical data) in the Bayesian framework and DMRE-CoP compared to the OEL categories.

Bayesian Analysis for the Occupational Exposure

To produce the posterior GM, GSD, P95, and the posterior probabilities of P95 of exposure in each of the DMRE-CoP exposure categories, current monitoring data was updated using historical coal dust data from the same HEG or expert judgements for the job titles. According to earlier research for occupational exposure assessment [85, 86], five observations were randomly selected from historical data of each HEG for the prior specification. This is important because, as sample sizes get larger, the posterior distribution in Bayesian statistics must be observed to a high degree from the most recent/current data. The posterior distribution in Bayesian statistics is a compromise between information from the prior and the most recent data.

Model Specification Using Current Data

In industrial hygiene compliance testing using the Bayesian framework, exposure data are assumed to follow a lognormal distribution. The current monitoring data of the exposure are updated by prior distribution from historical data or expert judgements. The expert judgements were elicited using the IDEA technique about their opinions on the probability of the P95 of exposure falling in each of the DMRE-CoP exposure control categories and fitted a lognormal distribution in SHELF as in section 3.1.3 above.

If Y_m is the current monitoring data, then the log-transformed, $y_i = \log(Y_m)$ and observations $y_i = (y_{i1}, \ldots, \ldots, \ldots, y_{in})$, where n is the sample size, with the data $y_i \sim \mathrm{Norm}(\mu, \sigma^2)$. The $\exp(\mu)$ is the geometric mean (GM), and $\exp(\sigma)$ is the geometric standard deviation (GSD). The likelihood function of the current monitoring exposure data is presented below:

$$\prod_{i=1}^{n}(y_i|\mu,\sigma^2) = \prod_{i=1}^{n}\frac{1}{y_i\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{(\log y_i - \mu)^2}{\sigma^2}\right\} \tag{3}$$

where y_i is the log-transformed current monitoring data, and n is the number of observations of current monitoring data.

Suppose N_{cat} is the number of South Africa OEL exposure categories, and R is the random variable showing the categories, say $R \in \{1,2,3..,...N_{cat}\}$. The proposed model aimed at classifying the exposure into one of the N_{cat} categories. Therefore, making R a random variable in the model directly produces posterior distribution within the Bayesian hierarchical framework.[83, 86]

Non-informative Prior Specification

A non-informative prior is one with little information about the data. The non-informative prior in this study was motivated by Quick et al, 2017[86]. Appendix D contains the R codes for the non-informative prior distribution.

For the GM, the assumption is that μ is uniformly distributed $\mu = \mathrm{LnGM} \sim \mathrm{Unif}(a_\mu, b_\mu)$ and with the bounds $\mu \in (a_\mu, b_\mu)$ where a and b are the lower and upper bounds of the mean, respectively. Therefore, equation 4 illustrates the assumption that μ is uniformly distributed and the posterior distribution is in equation 5

$$p(\mu) = \frac{1}{b_\mu - a_\mu} \times I\{\mu \in (a_\mu, b_\mu)\} \propto I\{\mu \in (a_\mu, b_\mu)\} \tag{4}$$

$$p(\mu|Y) \propto LN(y_i|\mu) \times I\{\mu \in (a_\mu, b_\mu)\}. \tag{5}$$

For the GSD, the variance σ^2 follows a uniform distribution with $\sigma = \ln GSD \sim \mathrm{Unif}(a_\sigma, b_\sigma)$, and with the bound $\sigma^2 \in (a_\sigma, b_\sigma)$, where a and b are the relaxed lower and upper bounds of the variance, respectively.

Therefore, prior variance σ^2 with uniform distribution is shown in Equation 6 with the posterior distribution in Equation 7:

$$p(\sigma^2) = \{\sigma^2 \in (a_\sigma, b_\sigma)\} \propto 1 \tag{6}$$

$$p(\sigma^2|Y) \propto LN(y_i|\sigma^2) \times \{\sigma^2 \in (a_\sigma, b_\sigma)\} \tag{7}$$

Informative Prior Specification for the Geometric Mean

The informative truncated prior distribution in this study was motivated by Quick et al.2017[86]. Appendix E shows the R codes for the informative prior distribution. For informative prior, if Y_h is the historical data, then the log-transformed data $y_{i0} = \mathrm{Log}(Y_h)$ with n_0 observations, sample mean, $\bar{y}_0 = \sum y_{i0}/n_0$ and sample variance $s_{y0}^2 = \sum(y_{i0} - \bar{y}_0)^2/(n_0 - 1)$.

For prior GM, μ, the historical data follows normal distribution, $y_{i0} \sim \mathrm{Norm}(\mu, \sigma^2)$, then the mean, $\bar{y}_0 \sim \mathrm{Norm}(\mu, \sigma^2/n_0)$, similarly, $\bar{y}_0 - \mu \sim \mathrm{Norm}(0, \sigma^2/n_0$. If $\bar{y}_0$ is random quantity and μ is an unknown fixed value, can change functions of these parameters by putting μ as a random quantity and replace σ^2 with the prior estimate s_{y0}^2 , so that μ takes the form:

$$\mu \sim \mathrm{Norm}(\bar{y}_0, s_{y0}^s/n_0) \tag{8}$$

The full conditional for μ was based on truncated normal prior distributions:

$$\begin{aligned} p(\mu|Y, \sigma^2, \bar{y}_0, S_{y0}^2) &\propto p(Y|\mu, \sigma^2) \times p(\mu|\bar{y}_0, s_{y0}^2) \\ &\propto \exp\left[-\frac{(\bar{y} - \mu)^2}{2\sigma^2/n}\right] \times \exp\left[-\frac{(\bar{y}_0 - \mu)^2}{2s_{y0}^2/n_0}\right] \times I\{\mu \in [a_\mu b_\mu]\} \\ &\propto \exp\left[\frac{(\mu - E[\mu|.])^2}{2V[\mu|.]}\right] \times I\{\mu \in [a_\mu b_\mu]\} \end{aligned}$$

Where $E[\mu|.] = V[\mu|.]\left(\frac{n\bar{y}}{\sigma^2} + \frac{n_0\overline{y_0}}{s_{y0}^2}\right) V[\mu|.] = \left(\frac{n}{\sigma^2} + \frac{n_0}{s_{y0}^2}\right)^{-1}$

This shows that $p(\mu|Y, \sigma^2, \bar{y}_0, S_{y0}^2) \sim \mathrm{Trun} - \mathrm{Norm}(E[\mu|.], V[\mu|.])[a_\mu, b_\mu]$ (9)

Informative Prior Specification for the Geometric Standard Deviation

For GSD, σ^2, the parameter space for the lower bound from BDA was used and the upper bound was let to change iteratively. Appendix E shows the R codes for the informative prior distribution.

To develop prior for σ^2 the expression that:

$$\frac{(n_0 - 1)s_{y0}^2}{\sigma^2} \sim x_{n_{0-1}}^2 \qquad (10)$$

where $x_{n_{0-1}}^2$ is the chi-square distribution with $(n_0 - 1)$ degrees of freedom.

Now let's proof that indeed the above expression

Suppose W:

$$W = \sum_{i=1}^{n} \left(\frac{y_{i0} - \mu}{\delta}\right)^2 \qquad (11)$$

By adding 0 to each term:

$$W = \sum_{i=1}^{n} \left(\frac{(y_{i0} - \bar{y}_0) + (\bar{y}_0 - \mu)}{\delta}\right)^2 \qquad (12)$$

Add and subtract the sample mean by 0 to the quantity in the numerator and square the terms to give:

$$W = \sum_{i=1}^{n} \left(\frac{(y_{i0} - \bar{y}_0)}{\delta}\right)^2 + \sum_{i=1}^{n} \left(\frac{-\mu}{\delta}\right)^2 + 2\left(\frac{\bar{y}_0 - \mu}{\delta^2}\right) \sum_{i=1}^{n} (y_{i0} -) \quad (13)$$

The last term in Equation 13 is 0:

$$W = 2\left(\frac{\bar{y}_0 - \mu}{\delta^2}\right) \sum_{i=1}^{n} (y_{i0} -) \text{ since } \sum (y_{i0} - \bar{y}_0) = n\bar{y}_0 - n\bar{y}_0 = 0 \quad (14)$$

Now W changed to:

$$W = \sum_{i=1}^{n} \frac{(y_{i0} - \bar{y}_0)^2}{\delta^2} + \frac{(y_{i0} - \bar{y}_0)^2}{\delta^2} \qquad (15)$$

Recall that sample variance:

$$s_{y0}^2 = \frac{\sum(y_{i0} - \bar{y}_0)^2}{(n_0 - 1)} \quad (16)$$

Multiple both sides of the above equation by $(n_0 - 1)$:

$$(n_0 - 1)s_{y0}^2 = \sum(y_{i0} - \bar{y}_0)^2 \quad (17)$$

The numerator in the first term was presented as a function of the sample variance:

$$W = \sum_{i=1}^{n}\left(\frac{(y_{i0} - \mu)}{\delta}\right)^2 = \frac{(n-1)S_{y0}^2}{\delta^2} + \frac{n(\bar{y} - \mu)^2}{\delta^2} \quad (17)$$

The right side of the Equation 17 above is now sum of n independent chi-squared random variables, since $x1, x2, .., xn$ are number of observations from a normal distribution $N(\mu, \delta^2)$:

$$\sum_{i=1}^{n}\left(\frac{(y_{i0} - \mu)}{\delta}\right)^2 \quad (18)$$

Therefore,

$$\frac{y_{i0} - \mu}{\delta} \quad (19)$$

is a standard normal distribution. Squaring a standard normal variable equal to a chi-square random variable with 1 degree of freedom

Therefore $\sum_{i=1}^{n}\left(\frac{(y_{i0}-\mu)}{\delta}\right)^2$ is the sum of n independent chi-squared random variables.

The moment generating function of W is the same as that of a chi-square (n) random variable, given as $M_W(t) = (1 - 2t)^{-n/2}$

For $t < \frac{1}{2}$ the second term on the right side of W after the equal sign $\frac{n(\overline{y}_0-\mu)^2}{\delta^2}$ is chi-square (1) random variable, since the sample mean is assumed approximate normally distribution with mean μ and variance δ^2/n

Therefore $Z = \frac{\overline{y}_0-\mu}{\delta/\sqrt{n}} \sim N(0,1)$ is a standard normal variable. By squaring Z , a chi-square random variable with 1 degree of freedom is observed as $Z^2 = \frac{\overline{y}_0-\mu}{\delta/\sqrt{n}} \sim x^1(1)$. And the moment generating function of Z^2 will be $M_{Z^2}(t) = (1-2t)^{-1/2}$

From Equation 10, s_{y0}^2 was a sample random variance from the historical data, while σ^2 is a fixed unknown population parameter. In this study, σ^2 was treated as the random variable, given sample variance s_{y0}^2 from the historical data or expert judgement. Take that $x_{n_{0-1}}^2 \equiv \mathrm{Gam}\left(\frac{n_0-1}{2}, 2\right)$ where the second parameter is the scale parameter. From the gamma distribution,

$$\frac{(n_0-1)s_{y0}^2}{\sigma^2} \sim \mathrm{Gam}\left(\frac{n_0-1}{2}, 2\right)$$

Multiple both side by $(n_0-1)s_{y0}^2$ so that

$$\frac{1}{\sigma^2} \sim \mathrm{Gam}\left(\frac{n_0-1}{2}, \frac{2}{(n_0-1)s_{y0}^2}\right)$$

$$\sigma^2 \sim \mathrm{IG}\left(\frac{n_0-1}{2}, \frac{(n_0-1)s_{y0}^2}{2}\right) \tag{20}$$

For $n_0 > 1$, where IG (a, b) is an inverse gamma distribution with shape parameter a and scale parameter b.

Therefore,

$$\begin{gathered} a = (n_0-1)/2 \\ b = 2/(s_{y0}^2 * (n_0-1)) \end{gathered} \tag{21}$$

The full conditional for σ^2 which was based on truncated inverse gamma prior distributions:

$$p\left(\sigma^2|Y,\mu,\overline{y_0},s_{y0}^2\right) \propto p(Y|\mu,\sigma^2) \times p(\sigma^2|\overline{y_0},s_{y0}^2)$$

$$\propto (\sigma^2)^{-\frac{n+n_0-1}{2}-1} \exp\left[-\frac{n(y-\mu)^2+(n_0-1)s_{y0}^2}{2\sigma^2}\right]$$

$$\times I\{\sigma^2 \in [a_\sigma b_\sigma]\}$$

Therefore, the posterior variance given as the geometric standard deviation is given by:

$$p\left(\sigma^2|Y,\mu,\overline{y_0},s_{y0}^2\right) \sim \text{Trunc} - IG\left(\frac{n+n_0-1}{2},\frac{n(\overline{y}-\mu)^2+(n_0-1)s_{y0}^0}{2}\right)[a_\sigma,b_\sigma] \quad (22)$$

Markov Chain Monte Carlo model inputs

To generate a full posterior conditional distribution, the Gibbs sampler, a Markov chain Monte Carlo (MCMC) algorithm was used.[88] Because it is not computationally intensive, the Gibbs sampler was applied. For the MCMC Gibbs algorithms to simulate multiple draws that converge to the posterior distribution, the number of burn-in iterations was specified. A total of 20 000 burn-in iterations were specified with starting initial values for mean and variance set at 0 and 1, respectively. The Gibbs sampler samples using conditional distribution. For example, if a given parameter has been broken down into sub-parameters, it draws each sub-parameter conditionally on the values of all the others iteratively. To create the marginal posterior distribution, the sub-parameter is conditionally updated multiple times based on the most recent values of all its parts. With the help of the Gibbs sampler, the posterior distribution can be estimated more accurately as it corrects the generated values. The MCMC simulations and Gibbs sampler were programmed directly in R version 4.1.1 (R Core Team, Auckland, New Zealand)[89], and the bayestestR package[90] to describe the posterior distribution of the parameter estimates with 95% credible intervals.

MCMC post-estimation diagnostic tests

MCMC model diagnostic tests were conducted to ensure that simulated draws are close to the posterior distribution. The Gelman-Rubin convergence diagnostic, which assesses the between and within-Markov chain variability for the model parameters to find whether they are stationary[91], was used to perform the posterior convergence. While the within-Markov is the mean of the variance in each sample, the between-Markov is the variance of the posterior means of the samples. Convergence is reached if the test statistics showed by R-hat are less than 1.05. The bulk and tail effective sample sizes were used to prove the posterior quantile reliability.[92] Effective sample sizes of at least 100 are recommended for each chain. In this investigation, the convergence diagnostic test revealed an R-hat of less than 1.05 and an effective sample size of more than 100, showing convergence was successful. The posterior convergence diagnostics of the Markov chains mixing was done using the RStan package[93] within R.

3.2.6. Inferential Analysis for Objective 4.

Objective 4: To compare the posterior probabilities of the P95 of the non-informative and informative prior information (from experts' elicitation) in the Bayesian framework compared to the OEL categories (Appendix F)

Job titles were selected if their mean exposure difference is similar across all HEGs where they were found. This is to make sure that the exposure profile is like the same job title as you would expect in a SEG or HEG. The analysis of variance (ANOVA) F test for the difference in means of job titles exposure data across HEGs was used. If the variation of the means between HEGs is not significantly higher, the job title was selected. Since job title exposure could not assume normal distribution even after log transformation, the Kruskal Wallis test, a non-parametric version of ANOVA was used to assess the median exposure differences.

Establishing Joint Probability Distribution from expert elicitation

The SHELF, an online elicitation tool, was used to produce a lognormal distribution for each elicited expert judgement. The joint probability distribution is a lognormal probability distribution of each of the individual expert judgements.

The SHELF tool can choose the distribution that best matches the expert's elicited judgment. The SHELF tool works by employing least squares processes to generate a distribution that best fits the provided expert's data inputs by minimizing the sum of the squares of the residuals. In the SHELF tool, the quartile method was chosen to elicit our expert's judgment in a continuous quantity.[94]

The quartiles are presented as lower (Q1) and upper quartiles (Q3). Given that our lower (L) and upper (U) limits during the elicitation were 0 and 100%, Q1 is a value between L and M, where M is the median value, while Q3 is a value between M and U. In the expert's elicitation questionnaire, the expert's best guess was collected in percentages to distribute the probability of the P95 exposure being located in each of the exposure control categories. To consider uncertainty in the expert best guess, the probability of finding the P95 exposure in each of the DMRE-CoP exposure categories was required in range. (refer to section 3.1.3). The best guess was the median, minimum percentage was Q1, and maximum percentage was Q3. The cumulative probabilities were set at 2.5% for Q1, 50% for M, and 97.5% for Q3 to capture the expert's 95% certainty. In Table 6 presents an example of expert data that was inputted into the SHELF software for multiple experts to elicit individual distribution from these experts.

Table 6: An example of elicited Expert judgements for one of the job titles in this study for exposure category 1.

	Experts				
	A	B	C	D	E
Quantiles	%	%	%	%	%
2.5%	55	40	35	50	60
50%	61	49	42	79	80
97.5%	80	71	66	91	97
A, B, C, D and E refer to individual experts					

Figure 10 is a snapshot of the SHELF tool where the number of experts is specified and the input is selected, in our case the quantile method. Then the distribution of the data is selected, which in this study is the lognormal distribution.

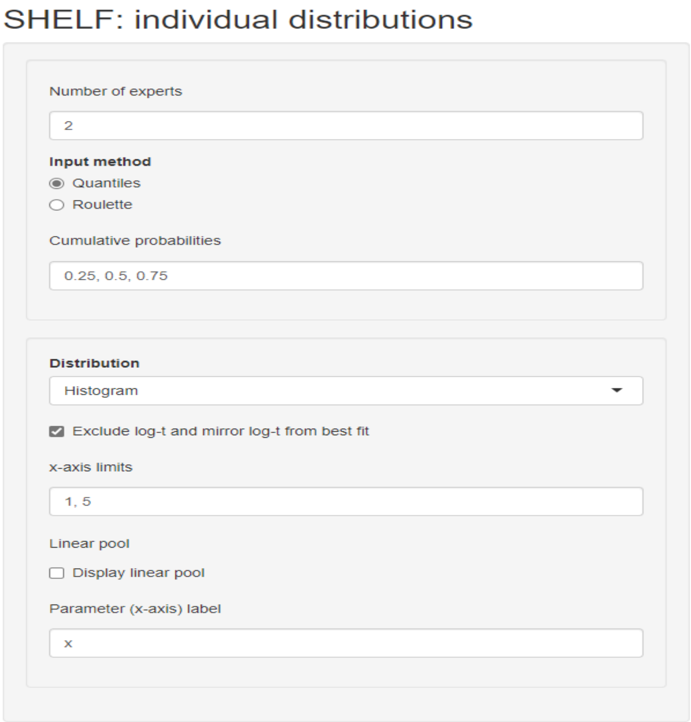

Figure 10: The SHELF multiple expert tools. SHELF: individual distributions (shinyapps.io).

Figure 11 is where the expert's judgements are entered. The L is the lower limit which is 0 and U is the upper limit which is 100. The analysis is then run automatically, and the results can be downloaded in a word document or html format.

Judgements | PDF | Tertiles | Quartiles | Help

Enter the judgements in the table below, one column per expert. Enter lower plausible limits in the first row, upper plausible limits in the last row, and quantile values in between, corresponding to the cumulative probabilities.

	A	B	C	D	E
L	0	0	0	0	0
0.025	55	40	35	50	60
0.5	61	49	42	79	80
0.975	80	71	66	91	97
U	100	100	100	100	100

Figure 11: The SHELF multiple expert tools for entering individual expert's judgements.

Figure 12 shows the results of the lognormal fitted distribution of each expert data and then aggregated to generate combined mean and variance for exposure control category 1. The AM (simple average) of the lognormal mean and variance of the individual expert's judgements was calculated. The use of simple average was on the notion experts equally contributed as the IDEA protocol was modified. The mean and variance were then used as the prior distribution. The Bayesian framework is the same as in objective 3, except the prior distributions and the current monitoring data were from our elicited expert's judgements and job titles (as an alternative grouping), respectively.

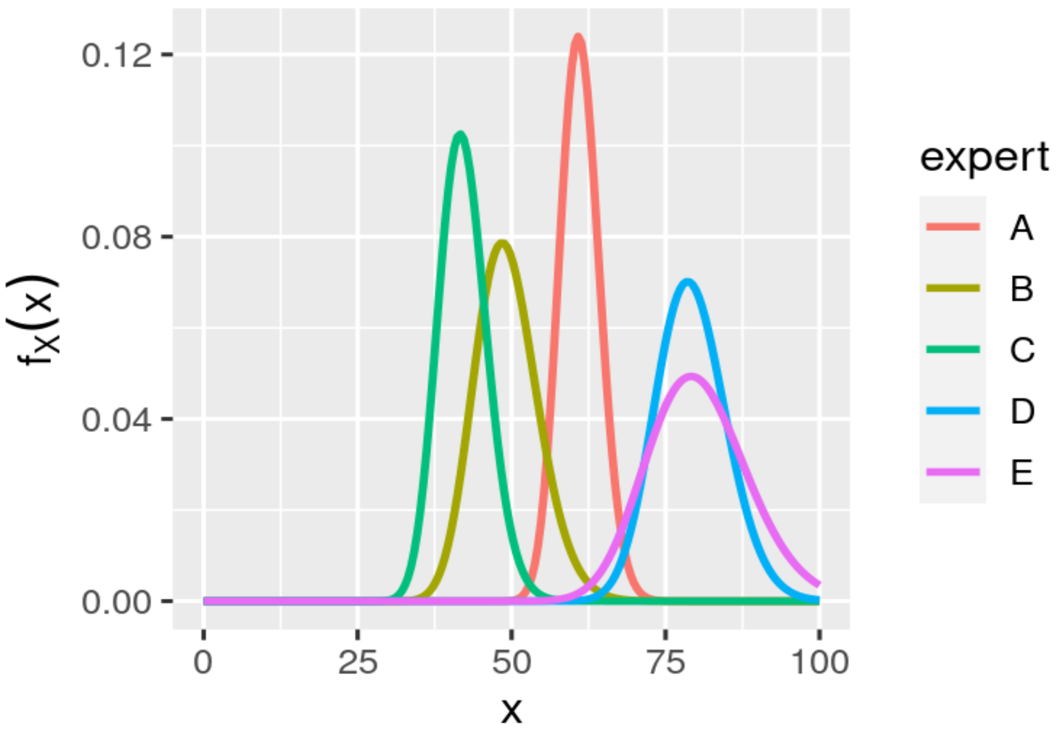

Figure 12. A lognormal distribution fitted to the individual expert elicited judgments. The lognormal mean and variance of the experts were: A: 4.11, 0.003; B: 3.89, 0.011; C: 3.74, 0.009; D: 4.37, 0.005; E: 4.38, 0.010.

The sensitivity analysis of the bounds

In this PhD study, parameter values were specified on the lower bounds of the truncated prior distribution, for μ and σ^2 and P95 using suggestions from BDA.[57, 95] For the non-informative prior, the lower bound of the mean was set at 0 and variance was set at infinity. Value of 0.0001 for the lower value for the prior mean and 1.05 for the variance were specified. The upper bounds were let to vary without any restrictions as motivated by Quick et al.2017[83]. The upper limit was let to vary to avoid a prior from being skewed toward a favourable result. However, putting restrictions in a parameter might misclassify an exposure.

As a result, in published paper 3, two job titles were randomly taken: shuttle car operator and pump attendant.

The lower bound for the GM and GSD in shuttle car operator exposure data was set at 0.00001, with the assumption that coal dust exposure value is never zero. The upper bound was set to be slightly higher than the GM and GSD values calculated from this exposure data. The GM is 1.08, so the upper bound was set to 1.09; the GSD is 2.62, so the upper bound was set to 2.63 to ensure that the exposure estimates are within the limits.

The upper bound of the GM for the pump attendant was set at 0.42, just above the actual value of 0.41, and the GSD at 3.42, also just above the actual value of 3.41. Another analysis was done without imposing any constraints on the parameter space, i.e., without specifying lower and upper bounds.

CHAPTER 4: RESULTS

Introduction

The interpretation of the results is presented in this chapter. In paper 1, various methods for assessing homogeneity and compliance against the current methods used in the DMRE-CoP were compared. In papers 2 (Appendix I) and 3 (Appendix J), Bayesian model was fitted to locate the probability of the P95 in each of the exposure categories. The prior information was obtained from historical data and experts' judgements. Uncertainty, using the credible interval and variability of the exposures, was considered an important indicator of exposure compliance in this study. All three papers have been published, available in the Appendix section.

Summary of Papers from the Objectives

4.1. Paper 1: Compliance Testing and Homogenous Exposure Group Assessment in the South African Coal Mining Industry (Appendix H)

This paper investigated the first two objectives of the PhD study. The paper compared exposure compliance and within-group variation using HEG and job titles data according to DMRE-CoP and the international consensus methods.

This study included 728 workers in 28 HEGs. HEG N exposure level was the lowest (AM: 0.58) while HEG A had the highest AM dust concentration (AM: 2.19 mg/m^3). The highest concentration of GM was found in HEG H (1.70 mg/m^3), and the lowest concentration, 0.20 mg/m^3, was found in HEG R. In 21 (75%) of the 28 HEGs, there was agreement on non-compliance by all the criteria. According to the DMRE-CoP, seven (or 25%) of the HEGs are compliant. The DMRE-CoP and CEN/BOHS-NvVA only concurred on one HEG exposure compliance decision. Agreement of decision was found in 27 (96%) of the 28 HEGS when the 70% and 95% UCL approaches of BOHS/NVvA and AIHA were compared.

A total of 11 (or 39%) of the 28 HEGs were found to be homogenous according to the global GSD. Only seven HEGs met the DMRE-CoP homogeneity criteria. Only four (14%) of the HEGs showed agreement between the DMRE-CoP and GSD indicators. The R-ratio with 95% of the data showed no homogenous HEG and the R-ratio with 90% of the data showed just one homogenous HEG. Only one HEG showed homogeneous agreement across GSD and R-ratio (90%). The non-homogeneity in all the metrics was agreed upon in a total of 14 HEGs (50%).

For the examination of homogeneity in an alternative sub-grouping by job titles, a total of 14 HEGs that were non-homogenous in both the DMRE and GSD criteria were chosen. Following HEG C in terms of job titles were B and D. Nine job titles were homogeneous in HEG B, according to the global GSD. Five out of 10 job titles in HEG B were homogeneous, according to the DMRE-CoP criteria. Most job titles had AM and P90 values that were greater than the parent HEG. While a few job titles and HEGs had lower AM and P90 point estimates than others. The DMRE approach, compared to other approaches, tends to imply greater compliance, and job title sub-grouping showed greater homogeneity than HEGs.

4.2. Paper 2: Bayesian Hierarchical Modelling of Historical Data of the South African Coal Mining Industry for Compliance Testing (Appendix I)

The third objective of this PhD was covered in this paper. The objective was to compare the posterior probabilities of the P95 of the informative prior from historical data, the non-informative prior, and the DMRE-CoP approach.

In the monitoring data, the highest AM was seen in HEG C. While HEG A had the highest AM in earlier data. Comparing the historical and present data, HEGs D and G have the lowest AMs of all the other HEGs. Six HEGs from current monitoring data revealed a high degree of exposure variation, or GSD higher than 3. Five HEGs from the historical data showed considerable exposure variability based on the GSD. Three HEGs under a non-informative prior had low variability, while four HEGs under the informative prior likewise had lower exposure variability.

By employing the non-informative and informative priors to update the current data, the DMRE-CoP P90 values from non-transformed data are significantly lower than the posterior P95 seen in Bayesian statistics. The P90 value for only one HEG (HEG D) was lower (1.62 mg/m^3) than the OEL of 2 mg/m^3. According to the Bayesian framework, the HEGs have P95 values below the OEL. All of the HEGs exposures were below the OEL, according to the posterior median GM. Between the non-informative and informative priors, the median and 95% CrI were comparable across all HEGs. In comparison to the informative prior, the P95 is lower, and the 95% CrI is wider in the non-informative prior, according to five out of nine HEGs. The wider 95% CrI showed that there was significant uncertainty in the non-informative outcomes.

The posterior probability of the exposure falling into category four was lower in HEG D for both Bayesian approaches. All the HEGs in the two prior distributions were at least 90% likely to be in poorly regulated category 4. The probabilities of exposure in lower exposure categories were lower for non-informative priors than for priors derived from historical data.

4.3. Paper 3: Bayesian Hierarchical Framework from Expert Elicitation in the South African Coal Mining Industry for Compliance Testing (Appendix J)

This paper investigated the fourth objective. The posterior probabilities of the P95 from informative prior distributions obtained from historical data and the non-informative prior using the Bayesian framework were compared.

Since a specific job title can be found in more than one HEG, the median difference of exposure of the job titles across HEGs were compared. All the job titles had similar median exposure distributions across the HEGs, without any statistically significant ($p<0.05$) differences. In the current data, six of the nine HEGs had high GSD values larger than 3, which shows high variability, but in the historical data, seven HEGs showed significant exposure variability. For every Bayesian approach, the posterior GM were below the OEL. Certain job titles showed less variation between Bayesian frameworks, according to the GSD. According to the posterior P95, no job titles across the three Bayesian updates of current indicated exposure below the OEL. In comparison to the historical and uninformative priors, the posterior P95 were higher in the Bayesian findings from the expert prior distribution. Compared to the non-informative prior, several job titles had a decreased probability of falling into the highest exposure category 4 in the historical prior distribution. Compared to the other methods, Bayesian framework with informative prior distribution from expert elicitation tend to place HEG in a higher exposure category with increased but with less variability.

Sensitivity analysis using various parameter values and without placing any restrictions on the bounds were done. The sensitivity analysis was done to rule out possible misclassification of exposures caused by using specific parameter values (for methods, see paper 3 in Appendix F. Where no restriction was imposed on the parameter space job titles had probabilities of 100% that they were all in exposure control category 4 (poorly controlled). The parameter specified by this study (99.4%) and the adopted from BDA (99.91%) showed comparable results of locating exposure in exposure category 4.

4.4. Integrated Conclusion

The use of HEG revealed increased exposure variability. The nature of defining HEG using the DMRE-CoP imply that various work activities and job titles are perceived to have comparable exposures. According to the alternative grouping using job titles as SEG within the HEG, one job title is compliant, and exposure distribution is less variable, but the HEG can be non-compliant with significant variability of exposure distribution. There is typically less uncertainty about the probability that the P95 will belong to each exposure-controlled group when updating current data with historical data.

When historical data is not available, prior distribution may be derived from expert judgments or opinions. This can help to improve compliance evaluation and guide decision-making. When all relevant information is considered, using historical exposure data from job titles can produce findings that can support wise exposure control and reduction decisions.

CHAPTER 5: DISCUSSION

5.1. Introduction

To define HEG, the DMRE-CoP uses common and return air take area and additionally common work areas. Unfortunately, there are multiple job titles in each of these HEGs. Due to this approach, it may be challenging to find highly exposed workers or job titles, which increases the risk of exposure overestimation or underestimation. To sample exposure of workers, institutions and bodies in other countries use SEGs. SEGs are defined as workers with the same job titles, tasks, duration, places of employment, and tools. These are different from the strategy employed in South Africa. Remember that presently, it is impractical and expensive to sample every worker, which is the background for the definition of HEG and SEG. The importance of defining a similar exposure profile in a group of workers is highlighted by the fact that taking a small part of workers from HEG or SEG should be able to accurately represent the exposure of all the workers assigned to these HEGs/SEGs. If workers are grouped based on similar exposure, it would help to generalize compliance findings back to the larger population of workers. Since a HEG is constituted by common and return air intake area and work area, including diverse types of activities, there may be a high degree of exposure variability depending on the job titles in that area.

According to Lavoué et al, 2019[55], the CEN and AIHA approaches assume that the data have a lognormal distribution and that P95 should be below the OEL. It is a widespread practice that the probabilities of the P95 of exposure can be grouped according to OEL related exposure categories.[57, 83,86] In a Bayesian framework, historical data were used as a baseline to update current data. The DMRE exposure category was used to group the posterior probability of the P95 HEG exposure. Bayesian statistics is robust even with small sample sizes and still produces results that can confidently influence decisions. Utilizing historical data is important to provide prior knowledge of the current data in case the exposure scenario has not changed substantially. To avoid overpowering the current data, it is also important to consider the weight of the prior data. As in other studies where a small prior sample size generated inferential benefits when the findings were compared to non-informative priors,[86, 96] a small prior sample size has less influence on the current data.

The use of historical data as prior is particularly important as it improves the certainty of the findings. However, in situations where historical data are inadequate or unavailable, prior information can be found using expert judgment. Experts play a key role in occupational exposure assessment decision-making because of their prior knowledge of the work activities. Hewett et al, 2006[57] were the first to use and propose the inclusion of expert judgments in a Bayesian framework for compliance assessment. This study also conducted expert elicitation to derive previous information on dust exposure.

Introducing new robust methods for assessing coal dust overexposure and categorizing HEG in the South African coal mining industry, may improve the identification of highly exposed workers. Using job titles to create a clearly defined, comparable exposure group has the advantages of reducing overestimation and underestimation of exposure, and the number of workers sampled from a job title group can represent the remaining workers who are not selected. Finally increasing confidence in decisions can be achieved by accounting for uncertainty and variability of exposure by using historical data or expert prior knowledge. The published papers that are provided in Appendix H, I, and J addressed the main objectives of the study. In addition, from these findings, future suggestions were provided. . As part of practical applications in future polices, a simple and direct statistical analysis tools were suggested.

5.2. Key Findings

The key findings are discussed according to the conceptual framework (Figure 4). First is the grouping of HEGs and compliance testing; the second is the use of historical data to categorize exposure according to exposure control groups; and lastly is the use of expert judgement in the absence of the historical data using work- or job-title-specific data.

5.2.1. Compliance Testing and Homogenous Exposure Group Assessment

The objectives of this study were to assess compliance and homogeneity among HEGs and alternative groupings by using job titles. A total of 21 out of 28 HEGs are not compliant. The DMRE-CoP revealed that seven HEGs are compliant, but only one showed the same result for the CEN approach. The HEG that agreed with DMRE-CoP and CEN practices had a sample size of six, which is small, especially in a lognormal distribution. A high number of left tail samples were possibly collected, causing compliance with the OEL in the CEN approach. The DMRE-CoP responds to the question of whether actual exposure data is below the OEL (acceptable) or above the OEL (unacceptable), while the CEN and AIHA rely on the probability of overexposure to determine whether a point estimate of the underlying distribution, e.g., P95, exceeds the OEL (uncertainty). It is important to take the uncertainty into account because it could result from sampling errors, improper data, or measurements depending on whether the exposures were checked continuously or intermittently.[97] According to our findings, the DMRE-CoP is more lenient than the CEN and AIHA methods. The DMRE-CoP suggest more compliance of HEGs, as compared to CEN and AIHA. The DMRE-CoP likely underestimate exposure, and misclassified workers at risk of overexposure to coal dust. Another factor that might have contributed to the discrepancy between the approaches is the considerable unpredictability brought on by the fact that many job titles are assigned to the same HEG. Setting up a HEG will require a careful assessment of information, including survey design, procedures, resources or tools, and work strategy for suitable compliance testing using the OEL.[36] For both CEN and AIHA approaches, a total of 27 (96%) HEGs showed non-compliance. This shows that, despite differences in the level of uncertainty, i.e., 70% and 95%, respectively, the approaches are comparable in finding whether the examined HEGs are non-compliant.

This study also evaluated the homogeneity of the exposures by applying DMRE-CoP and CEN, BOHS, and AIHA criteria. The findings prove that using the GSD and DMRE-CoP indicators for homogeneity, 11 and 7 HEGs, respectively, were homogenous, but only 4 HEGs had an agreement between the two techniques. The DMRE-CoP requires that to call exposures homogenous, the AM and the P90parameter estimations must fall within the same exposure category. If the exposure distribution has a long tail or if it is a lognormal distribution, the parameter estimates might not fall into the same exposure category. This would also suggest that it is challenging to show homogeneity in data with a very skewed distribution using the GSD and DMRE-CoP. The DMRE-CoP assumes a normal distribution of the data, but this distribution is not checked. The number of measurements might impact the exposure variability over time, with shorter periods having a similar set of conditions and work practices than over a long period. [98,99] Apart from the GSD criterion, none of the HEG was homogeneous, as shown by R-ratio (95%) showing values all greater than 2. As a result, a ratio of 2 or less may be considered excessively conservative given that only 50% of the HEGs were homogenous across all the other metrics. The global GSD is especially sensitive to sample size because smaller sample sizes might create skewed data as the distribution of the parent data is unknown, which can cause the GSD to be underestimated .[98]

The DMRE-CoP would imply that because a HEG is made up of several job titles and work activities, it is difficult to correctly sample representative samples because of the type of work activities. Thus, a sample of workers with the same job title (within a work area) would easily show direct overexposure detection and strategic control of exposures. Some job titles had a higher P90 dust exposure level compared to its HEG. This suggests that if a HEG is constituted under the current DMRE-CoP guidelines, then a strategic intervention in risk assessment and control measures would be misguided. According to Rappaport et al, 1993[100], categorizing workers by job titles does not always result in decreased variability or compliance, however, the chance of a homogenous exposure is high.

Mulhausen and Damiano, 2006[31] defined a SEG as a population with the same exposure profile due to their shared use of identical materials, processing methods, and job responsibilities. In our findings, workers may not always have a uniform distribution of coal dust exposure, even when they are classified as doing similar work activities as shown by their job titles. There may be an increased variability of exposure among workers with the same job titles because of the differences in exposure times.[101]

5.2.2. Bayesian hierarchical modelling of historical data

In this study, historical data was used to derive prior distribution in assessing compliance in the categorization of the probabilities of posterior P95 according to exposure control groups and compared it with the DMRE-CoP and a weakly informative prior. The use of non-informative and informative priors from historical data revealed that all the posterior GMs were lower than the OEL. This may mean that the level of dust exposure control did not change, according to the GM. However, coal dust exposure compliance is best based on the P95, which is at least 95% of the underlying distribution. When comparing the DMRE-CoP with P90 for compliance and Bayesian use of prior distribution approaches, one HEG indicated exposure below OEL with the DMRE-CoP. This is in line with our earlier research, which found that the DMRE strategy tends to underestimate overexposure risks.[28] The Bayesian techniques did not just rely on a point estimate as for the DMRE-CoP; the framework also considered the uncertainty of overexposure. According to the Bayesian statistics, every HEG showed that their P95 was extremely high and above the OEL. The median P95 had a distribution that was similar across the informative and non-informative prior distributions. When using the non-informative priors, the majority of HEGs showed that P95 was slightly lower and had wider 95% CrIs than informative priors resulting from historical data. This proves how important it is to consider historical data when estimating occupational exposure and risk decisions can be accepted with greater certainty when past data is combined to update current data.

The Bayesian use of prior information, more than 95% of the P95 for the studied dataset were assigned to exposure category 4, which denotes poorly regulated exposure. The posterior P95 of none of the HEGs fell into the lower exposure categories. According to these findings, using historical data as a prior for occupational exposure assessment in a Bayesian framework is good since non-informative priors often place HEGs in a lower exposure category.[86] Reducing the risk of overexposure has an impact on making informed decisions. The classification of HEGs into a lower exposure category by non-informative prior is comparable to a simulation study.[86] Most of the classification of exposure into a lower exposure control category may be attributed to an inaccurate or improper prior used to derive the non-informative prior distributions. Therefore, the results should be interpreted with caution because an infinite integral specified the prior. This may produce incorrect posterior distributions.[102] Therefore, sometimes it is worth deriving different probability density functions for the prior specifications, and results from the Bayesian analysis should be compared before a decision can be reached.

5.2.3. Bayesian Hierarchical Framework from Experts Elicitation

In paper 3 the study focused on job title coal dust exposure data. Expert elicitation was conducted for each of the job titles and the data was used to derive informative prior distribution. The findings were compared with those of non-informative prior and informative prior distribution from historical data. The study showed that posterior GMs from all approaches were below the OEL. This shows no overexposure risk considering the GM. If the non-informative prior approach was compared to the informative prior approach, the P95 values were lower in the non-informative prior. For the non-informative prior and the prior that was based on expert opinions, the 95% CIs of the P95 was wider, showing high uncertainty. Therefore, since there is a great deal of ambiguity, a decision on exposure risk may not be achieved with great confidence. The fact that the data from the experts had a broader 95% CrI is due to the experts' differing opinions on the degree of uncertainty and confidence they have in their responses.

The chances of the P95 falling in a particular DMRE exposure category were investigated. The results from using expert judgments to derive prior distribution showed that more than 95% probability of the P95 fell in the highest exposure category (4), while the non-informative and informative priors from historical data had 75% probability. The results from elicited judgements in this study are different from qualitative assessments (expert elicitation without current monitoring exposure data) which revealed under-prediction of exposure. [103,104] More consistencies can be seen in the outcomes produced using historical data and expert judgments. The low variability of exposures according to the posterior GSD in some job titles derived from prior distribution generated by the expert data was expected to place the posterior probabilities of the P95 in lower exposure categories. These contradictory findings could be the result of incorrect prior that updated the posterior. In such situations, a further sample may be needed before a decision is made on the exposure risk. Our findings also deviate from the earlier research that used HEGs constituted by the DMRE-CoP, for example, in Made et al, 2022 study [105], it was discovered that the chances of the exposures of job titles falling into the higher exposure category were higher than for the HEGs. This explains why grouping according to a job title can result in the correct identification of highly exposed workers than grouping according to HEGs. Exposure determinants, such as specific work activity, workplace conditions, and production volume, should also be considered in exposure assessment.[106]

The lower and upper boundaries in the parameter space were driven by the BDA. It might be good to apply boundaries in the parameter space to prevent the likelihood of the exposure from being categorized in an incorrect exposure category. Unlike the frequentist statistics, the BDA says that the GM and GSD cannot be any random infinite values, there is always a range. To set bounds for GM and GSD, the minimum and maximum values were specified. In this study, however, our thought was that defining limits for the parameter space might lead to an incorrect classification of exposure. To compare the exposure classification outcome based on bounds used and if no bounds were used or if alternative bounds were used, a sensitivity analysis was performed. The results imply that in contrast to placing parameter values for the parameter space, using no bounds may put exposures in the highest exposure category. The application of a different bound suggests outcomes comparable to the one chosen. Naturally, occupational exposure measurement values cannot be at extremes or infinity, which is why boundaries are used. The small sample size might have put the exposure in the highest exposure category if no bounds werespecified.[86]

5.3. Integrated research findings

The constitution of the HEG is an important first step in assessing dust exposure in the mining sector. The exposure results of a small number of workers can be all workers assigned to a HEG if the HEG is defined correctly. By using job titles as an SEG rather than a HEG, it was discovered that highly exposed workers or job titles are overlooked or undetected. Improved certainty of the probability of the exposure falling into any of the exposure categories comes from the feature to combine current data with historical data or expert judgments based on prior knowledge. Table 7 is a compilation of emerging areas of interest for the South African coal mining industry.

Table 7. Emerging areas of interest for the South African coal mining industry

Emerging areas of interest	**The findings**
The use of P95 and P90 for compliance testing to the OEL.	The P90 tends to underestimate overexposure compared to the P95.
The use of the upper credible limit to account for uncertainty surrounding the P95.	The uncertainty set by credible limits allows the estimates to have a high probability enough for decision-makers to be confident in the exposure profile.
Use of the GSD and other methods to assess the homogeneity of exposures	The GSD reflects the variability of the lognormal distribution and allows for a decision on the homogeneity of exposures
SEG based on a specific job title	The HEG constituted according to DMRE-CoP guidelines indicated that some job titles show homogenous exposures even when the parent HEG is not, and some are compliant with the OEL even when the HEG is not, although the opposite can also be true.
Ability to assign a group of workers to an exposure category using historical data	The use of historical data reduces the uncertainty of a group being assigned to an exposure category.
Ability to assign a group of workers to an exposure category using expert judgments	Expert data collected with less bias helps improve uncertainty. Experts have experience and know a lot about the exposure profile. They are useful when historical data are scarce.
Sensitivity analysis to test values in the parameter space	Despite the best practice of the BDA in occupational exposure assessments, misclassification arising from the use of bounds in the parameter space has not been investigated.

5.4. Strengths and limitations of the research

This PhD study also reflects on both the strengths arising from the data collection and the methods employed to produce the findings. Due to confidentiality, the data set did not list the names of the workers but only their job titles, alternative techniques could not be applied, such as within-worker correlation or one-way random effect ANOVA. This could have improved the assessment of variability if multiple exposure assessments for the same worker, in a particular job title, had been obtained over time. The study assessed the HEGs that were based on a shared workspace where many job titles were present at the same time; a specific dust generation for a certain job title could result in different exposures. According to Rappaport et al, 1995[34], the grouping of workers into SEGs or HEGs should consist of individuals performing the same task in the same work area to have a likely uniform distribution of exposure. Then the statistical sampling analysis can take exposure variance into account to make an informed choice about risk management and control.

The availability of historical data enables the study to assess the probability of exposure falling in one of the DMRE exposure categories given previous information from this data. Bayesian statistics naturally allow the combination of historical data and present data within a sound framework for decision-making.[58] The Bayesian inferences are conditionally based on the data, making them exact and simple to interpret. This is due to its ability to produce findings based on a very small sample size; for instance, the probability that the posterior P95 falls into the highest exposure group, can be quantified.

From a Bayesian standpoint, it is sometimes impossible to achieve an informative prior because past data is either unavailable or not comparable and not consistent with the current data. This limitation was reduced by employing expert judgments to drive our prior information. In the expert elicitation, one drawback of the study is that due to insufficient expert data, weighted aggregate was not applied, which would have increased the accuracy of the experts' past estimates. The opinions of experts can also be quite prejudiced, and their responses depend on their familiarity with the duties of a certain job title. Human evaluations may be complicated and prejudiced. Experts occasionally make decisions or judgments about unknown situations by using generalizations. The Rule of Thumb does not combine data to conclude as an algorithmic approach does. [107,108] As a result, biases may be prevalent when using expert judgment. To reduce biases and enhance the cognitive processing of the needed information, expert training is needed. The Bayesian approach, which compares expert prior updates with non-informative priors, offers confidence in deciding whether to employ expert judgment to get prior information. Utilizing informative priors may have potential drawbacks. An inaccurate or inconsistent informative prior can produce posterior probability distribution of exposure that are heavily influenced by the prior data than the current monitoring data. When prior sample size is large, this may overpower the current monitoring data and produce findings that can mislead decision-making. The choice of the prior is crucial, however, there is no better way of deriving a prior distribution. Finally, because the use of parameter values in the parameter space may cause misclassification of exposure, this study investigated the potential bias by conducting a sensitivity analysis. The sensitivity analysis may rule out bias and can assessed by using various bounds.

5.5. Recommendations

After the findings from this study, below are the recommendations.

1. The ability to identify workers who are highly exposed will be easier for exposure assessment by grouping of workers within a HEG according to job titles to constitute a SEG as workers performing the same task are expected to have quite similar exposure distributions, the job titles grouping would also reduce exposure variability. When there are similar distributions of exposures, it will be easy to put them into the correct categories for exposure.
2. To test compliance with the OEL, we recommend using the estimate of the P95. Because occupational exposure data may be very variable, the P95 of the lognormal distribution should be used.
3. When deciding homogeneity, it is best practice to combine all available analysis methods to reduce doubt in findings. As a result, the GSD and the Rappaport index / R-ratio are encouraged in addition to the DMRE-CoP.
4. To classify the probability of exposures into a particular exposure category, we recommend using a Bayesian framework. The use of probability is important because it is easy to translate. For example, if the probability of exposure being found in a poorly controlled exposure category is 100%, this would show urgent decision needs to be made in controlling the exposures.
5. The Bayesian framework should be used with historical data or expert judgment as informed prior information to enhance correct assignment to exposure categories.

5.5.1. Recommendations for future research

The goal of this PhD study was to conduct compliance and homogeneity assessment procedures that could be used to reduce under- or overestimation of occupational exposures in the coal mining industry. This current PhD study was based on limited data; therefore, further research can be conducted if more data is available. Below are the gaps and ideas for future research.

1. In cases of lack of historical data or unavailability of expert knowledge, future research should consider using exposure models[109], for developing priors when assessing exposure using Bayesian statistics.
2. Research on repeated measurements of exposure should be carried out. This is important as it enables analysis of within-worker exposure variability.
3. Further research should be conducted on the DMRE-CoP grouping strategy. For example, apart from the likely problems associated with the current grouping strategy, the recommended minimum number of samples and/or frequency of sampling should be investigated to see if this number is sound enough to conduct compliance testing for all workers.
4. Future research should be done on the economic consequences (e.g., a cost-benefit analysis) of the revised grouping approach recommended by this PhD study.
5. The goal of this PhD study is to reduce overexposure to coal dust; therefore, further research should be conducted to assess how several factors such as ventilation, work process, personal protective equipment, years of experience, etc. can influence coal dust overexposure.

5.6. Scholarly contribution of the PhD study

The literature reviewed on various methodologies and the statistical analysis conducted to generate the results is what constitutes the scholarly contribution. Table 8 describes the knowledge gaps and how this study fills those gaps.

Table 8. Summary of literature reviewed and the PhD scholarly contribution.

Study Objectives	What is known	Knowledge gaps	PhD scholarly contribution
Compliance assessment using HEG and job titles	DMRE-CoP uses the P90 of the measurement data for compliance testing. While international consensus is based on P95 of the lognormal distribution	It is not known if the DMRE-CoP underestimates overexposure or not to prevent increased exposure to coal dust which could cause irreversible lung diseases. Therefore, it is important to compare if the methods agree.	Highlighted the need to incorporate robust methods that are used internationally into the current DMRE-CoP to avoid underestimating exposures.
Homogeneity assessment using HEG and job titles	The DMRE-CoP states that the AM and P90 should fall within the same exposure category for the HEG to be homogenous. While the homogeneity of HEG is determined using several methods including ANOVA, GSD and the Rapaport ratio based on lognormal data.	It is not clear if the HEGs are homogeneous given how they are constituted and if the job title's homogeneity of exposure status is the same as that of the parent HEGs. The comparison of the known methods to see if they agree will help ascertain if the DMRE approach of sampling and statistical analysis is appropriate.	The HEGs showed a high level of heterogeneity in exposures. The need for alternative grouping using job titles within HEGs will improve similar exposure profiles.
Bayesian framework using historical data of HEGs for compliance testing	The DMRE-CoP uses P90 of a present data set for compliance. Yet occupational exposures are assessed using the Bayesian framework because of its ability to incorporate earlier information in the analysis.	Bayesian analysis has been used in exposure data analysis. The P90 does not use a level of uncertainty to give confidence to the decision-makers about exposure risk.	A confident decision can achieve when using historical data in compliance testing. The importance of the informative prior is highlighted with reduced uncertainty when interpreting the probability of exposure in any of the exposure categories.
Bayesian framework using expert judgements for compliance testing	From the first two study objectives (above), it is now known that constituting HEG by job titles improves similar exposure profiles. Since historical data is not always available, expert judgements have been used before occupational exposure assessments.	Expert judgements can be subjective; therefore, it is important to know if the findings are similar when compared to compliance testing using historical data or non-informative prior.	The use of experts in exposure assessment is important as uncertainty is reduced and the probability of locating the posterior P95 is like that of historical data.

5.6.1. New knowledge

The main research question of this PhD study was if alternative grouping and exposure compliance assessment using Bayesian framework would improve the identification of highly exposed workers or group compared to the current guidelines of the DMRE-CoP. According to the findings of this study, a revised within-HEG sub-grouping based on job titles will, in comparison to currently constituted HEGs in the DMRE guidelines, reduce the variability of coal dust exposure and improve compliance testing by identifying overexposed job titles. When evaluating exposure compliance, the Bayesian framework allows the use of past information to guide current exposure profile, and its ability produce exposure probability according to the OEL exposure categories, unlike the DMRE-CoP approach. The Bayesian method allows exposure risk assessment decision making to made in a holistic manner since it can incorporate previous information and even occupational hygienist or expert's opinions in the model to produce a finding that can influence changes with much needed confidence. Therefore, we, encourage the employment of these methods of grouping using job titles and application of Bayesian framework in the DMRE-CoP current guidelines to assess homogeneity or coal dust exposure and robustly document overexposure patterns relative to the OEL exposure categories.

5.7. The policy impact of the PhD study

The results arising from this PhD study have emphasized the need to adjust the current DMRE-CoP guidelines on constitution of HEGs and compliance testing. The goal of continuous coal dust exposure assessments in the coal mining industry should be to inform decision making on overexposure reduction and monitoring of exposure patterns. As highlighted in this study, overexposure to coal dust can cause Coal Mine Dust Lung Disease (CMDLD), which is an irreversible aliment with increased mortality rates and a burden on the health care system. It was observed that non-compliance and heterogeneity of exposure within HEG were evident, this may have health implications and budget over-burn as more resource needs to be directed in controlling exposure. The DMRE-CoP future guidelines will benefit from regrouping HEGs and adopting some of the methods employed internationally as investigated in this study to protect workers' health and increase work productivity.

Implementing compliance assessment strategy such as by use of Bayesian decision-making framework would in future, reduce the need to increase sampling of workers for analysis. Bayesian analysis can produce robust results based on even a small sample size. Therefore, it would impact the burden of 'oversampling' and consequently reduce the economic implications and labour required to conduct sampling of mine workers. One key question that is still unanswered is: how easy is it to apply Bayesian framework used in this study to non-statisticians, occupational hygienists, and health and safety officers? Given Bayesian analysis is computationally intensive, and the choice of prior is challenging. The need for a user-friendly online Bayesian framework statistical tool will allow users to just paste the data and execute the process, which will produce the results without any mathematical understanding of the process. Examples of existing tools which assume lognormal distribution of the data include EXPOSTATS, AltrexChimie[95], HYGINIST[96], IHDataAnalyst[97], etc. The latter, i.e., IHDataAnalyst enables Bayesian decision-making on the assignment of exposure probabilities in each of the OEL exposure control categories. Moreover, it enables the incorporation of expert judgments and historical data as informative priors, but the non-informative prior distribution is set as default. The IHDataAnalyst is a simple online tool that works in Windows 7, 8, and 10. It can accommodate an exposure sample size of up to 2000 workers. The output results can be exported to a built-in text editor and saved as a rich text file (rtf). The IHDataAnalyst allows data entry, statistics, goodness-of-fit analysis, censored data analysis, Bayesian decision analysis, and the generation of the findings. A free version (IHDA-student) is available, which only allows 20 entries.

We compared IHDataAnalyst results with this PhD analysis done for HEG A and HEG D from paper 2 (Appendix I) in R, which requires the development of some complex R scripts or codes (Table 9). For both the informative prior distribution from historical data and the non-informative prior in HEG A, the probability of the P95 exposure being located in OEL exposure category 4 was the same. While for HEG D, the probability of P95 exposure was similar across the two statistical analysis tools. The posterior median P95 were also similar including the results for both informative and non-informative derived prior distributions.

We also tested the comparison by using job titles from paper 3, i.e., pump attendant and shuttle care driver (Appendix F). The results for both job titles were still similar. The minor differences may be attributed to the IHDataAnalyst inflexibility when specifying the parameter space. Since IHDataAnalyst does not freely allow the use of historical data as a prior, we used the prior data for the pump attendant as an example to generate probabilities of the P95 being found in each of the exposure control categories (Figure 13). These probabilities were then inputted as prior distributions from historical data in IHDataAnalyst.

Our findings suggest that both R statistical tool employed in this study and IHDataAnalyst produced similar results. R statistical packages allow you to freely incorporate any parameter values for the parameter space and any exposure control categories in the codes. The IHDataAnalyst is an easy-to-use Bayesian statistical tool, which will undoubtedly increase the use of Bayesian statistics in occupational exposure assessment in industries. Unfortunately, despite the use of IHDataAnalyst as a support tool, it does not freely allow input of any exposure control categories and parameter values for the parameter space. Also, the free version of the IHDataAnalyst only allows entry of 20 data exposure measurements, this limits its application for larger data sets.

Despite the identified limitations, the IHDataAnalyst results do not differ significantly from those of the R statistical package. In addition to this, its user-friendly element is one of its most important contributions to routine exposure assessment. As we look in future for tailor-made solutions, such as the development of new a statistical tool that meets the complete application of Bayesian statistics for occupational exposure assessment, we encourage the use of online Bayesian statistical tools including the IHDataAnalyst for occupational exposure assessment in the South Africa coal mining industry.

Table 9: Comparison of statistical analysis tools used for Bayesian modelling.

Grouping		R Statistical Package				IHDataAnalyst				Compliance	
		Category 1	Category 2	Category 3	Category 4	Category 1	Category 2	Category 3	Category 4	R - P95 (95% CrI)	IHDA - P95 (95% CrI)
HEG A	Non-informative	0	0	0	100%	0	0	0	100%	7.42 (4.48-12.17)	7.07 (4.24-17.50)
	Historical data	0	0	0	100%	0	0	0	100%	-	-
HEG D	Non-informative	0	0.01%	7.73%	92.26%	0	0	2.70%	97.30%	2.58 (1.83-3.92)	3.02 (1.96-5.45)
	Historical data	0	0.00%	2.26%	97.74%	0	0	2.70%	97.30%		
Pump Attendant	Non-informative	0	1.16%	19.14%	79.69%	0	0	10.50%	89.50%	2.61 (1.44, 5.50)	3.1 (1.53-10.7)
	Historical data	0	0.91%	25.99%	73.09%	0	0	10.50%	89.50%	-	-
	Exert judgments	0	0	0.06%	99.94%	-	-	-	-	-	-
Shuttle Car Operator	Non-informative	0	0	0	100%	0	0	0	100%	5.52 (4.08, 8.01)	5.32 (3.67-8.94)
	Historical data	0	0	0	100%	0	0	0	100%	-	-
	Exert judgments	0	0	0	100%	-	-	-	-	-	-

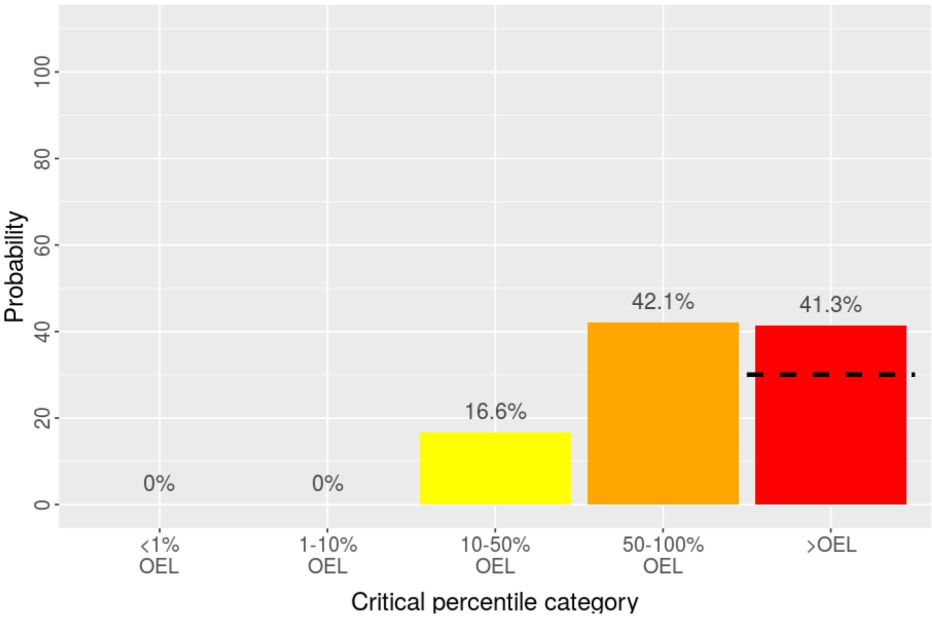

Figure 13: Example of the probability of the P95 exposure falling in the respective AHIA exposure control categories for Pump attendants.

CHAPTER 6: CONCLUSION

Despite the best practices encouraged by DMRE-CoP in South Africa, sampling of so-called HEGs, and compliance testing methods used in the mining industry deviate from those used consensually internationally. This PhD study aimed to find new methods that can help improve the grouping of exposed workers and compliance testing in the DMRE. The findings generated from this study hope to address gaps in the DMRE-CoP sampling strategy and compliance testing in South Africa.

The DMRE-CoP showed that more HEGs exposures complied with the OEL (2 mg/m^3) compared to the international consensus approaches of compliance testing, e.g., CEN, BOHS, NvVA, and AIHA methods. Considering these findings, the DMRE-CoP might have underestimated overexposure. When applying the GSD threshold of three and the DMRE-CoP, subgrouping by job title into a SEG, typically has a higher degree of homogeneity agreement. Our findings suggest that the use of expert judgments to get prior information produce similar posterior probabilities of exposure with less uncertainty when compared to the use of historical data. In general, the findings of this PhD study suggest that future compliance assessments should be conducted using the proposed methods in addition to the current DMRE-CoP guidelines. The study findings encourage subgrouping of workers within a HEG according to job titles, compliance testing based on P95 exposure value. In addition, consider the use of prior information, either based on historical data or expert elicitation, and Bayesian methods.

Moreover, the use of existing or further development of user-friendly statistical tools should be encouraged to improve the evaluation of occupational exposure assessment results in the coal mining industry in South Africa.

REFERENCES

1. World Coal Institute (WCI): The Coal Resource: A Comprehensive Overview of Coal. London: WCI, May 2005. Available from: https://is.muni.cz/el/1423/jaro2017/ESS421/um/World_Coal_Institute_2009.pdf (accessed on 1 December 2019).
2. Board on Earth Sciences and Resources (BESR): "Coal: Research and Development to Support National Energy Policy," 2007. Available from: http://books.nap.edu/openbook.php?record id = 11977&page = 166 (accessed on 1 December 2022).
3. AIDC. Alternative Information and Development Centre. South Africa's coal sector: energy security ownership and climate change. 2017. Available from: https://aidc.org.za/8160-2/ (accessed on 1 December 2022).
4. Guild R, Ehrlich RI, Johnston JR, Ross MH. SIMRAC handbook of occupational health practice in the South African mining industry. Johannesburg: SIMRAC. 2001. Available from: https://www.mhsc.org.za/sites/default/files/public/publications/OH%20Handbook.pdf (accessed on 5 January 2018).
5. Coal and its uses, Environmental resources for tribal energy development, 2015. Available from: http://teeic.indianaffairs.gov/er/coal/restech/ uses/index.htm (accessed on 1 December 2022).
6. Castranova V, Vallyathan V. Silicosis and coal workers' pneumoconiosis. *Environ Health Perspect*. 2000 ; 108(Suppl 4):675.
7. IUPAC. Glossary of atmospheric chemistry terms. International Union of Pure and Applied Chemistry, Applied Chemistry Division, Commission on Atmospheric Chemistry. *Pure Appl Chem*. 1990, 62 (11):2167-2219.
8. Humphries M, Sherlock MF. U.S. and world coal production, federal taxes, and incentives. CRS Report R43011. 2013. Available from: http://www.fas.org/sgp/crs/misc/R43011.pdf (accessed on 7 January 2023).
9. US Department of Energy, E.I.A., Office of Coal, Nuclear, Electric and Alternate Fuels. International energy outlook. 2021. Available from: http://www.eia.gov/forecasts/ieo/coal.cfm (accessed on 7 January 2023).
10. Top Five Coal Producing Countries (Million Tonnes, 2021). 2021. Available from: https://www.globaldata.com/data-insights/mining/the-top-five-coal-producing-countries-million-tonnes-2021/ (accessed on 7 January 2023).
11. IEA. World Energy Balances: 2021. Available from: https://www.iea.org/reports/world-energy-balances-overview (accessed on 7 January 2023).
12. Mineral Council of South Africa. Coal: Key Facts and Figures. 2019. Available from: https://www.mineralscouncil.org.za/sa-mining/coal (accessed on 7 January 2023).
13. Rajen N. Pneumoconiosis in South African coalmines– myth or reality? Occupational Health Southern Africa pp. 6-18. Available from: https://mhsc.org.za/sites/default/files/public/research_documents/Respiratory%20diseases%20among%20South%20African%20coal%20miners.pdf (accessed on 5 January 2018).
14. Statistics South Africa (Stats SA). South Africa's mining production, Gold is dying, Coal is flying. Available from: http://www.southafricanmi.com/south-africa-mining-production-17jul2018.html (accessed on 20 December 2022).
15. Department of Mineral Resources and Energy. South African Coal Sector Report. 2016. Available from: https://www.energy.gov.za/files/media/explained/south-african-coal-sector-report.pdf (accessed on 2 January 2019).
16. Attfield MD, Hodous TK. Pulmonary Function of US Coal Miners Related to past Exposure Estimates. Rev Respir Dis. 1992; 145:805-9.
17. CDC. Pneumoconiosis. Workplace Safety and Health Topics. Available from: https://www.cdc.gov/niosh/topics/pneumoconioses/ (accessed on 2 January 2023).
18. Queensland Government. Hazardous dusts. All airborne dust is hazardous to health. Available from: https://www.worksafe.qld.gov.au/safety-and-

prevention/hazards/hazardous-exposures/hazardous-dusts#::text=Respirable%20dust%20(smaller%20than%2010,and%20lead%20to%20lung%20damage (accessed on 1 November 2022).

19. Almberg KS, Halldin CN, Blackley DJ, Laney AS, Storey E, Rose CS, Go LHT, Cohen RA. Progressive massive fibrosis resurgence identified in U.S. coal miners filing for black lung benefits, 1970-2016. *Ann Am Thorac Soc*. 2018, 15(12), 1420-1426.
20. Hall NB, Blackley DJ, Halldin CN, Laney AS. Pneumoconiosis progression patterns in US coal miner participants of a job transfer programme designed to prevent progression of disease. *Occup Environ Med*. 2020, 77(6), 402-406.
21. International Labour Office. Guidelines for the use of the ILO International Classification of Radiographs of Pneumoconiosis. Geneva: International Labour Office; 2011. Available from: https://www.ilo.org/safework/info/WCMS_108548/lang--en/index.htm (accessed on 15 February 2019).
22. Soutar C, Campbell S, Gurr D, Lloyd M, Love R, Cowie H, Cowie A, Seaton A. Important deficits of lung function in three modern colliery populations: relations with dust exposure. *Rev Respir Dis*. 1993, 147(4):797-803.
23. Attfield MD, Seixas NS. Prevalence of pneumoconiosis and its relationship to dust exposure in a cohort of US Bituminous coal miners and ex-miners. American Journal of Industrial Medicine. 1995 Jan 1; 27(1):137-51.
24. Sarver E, Keles C, Rezaee M. Beyond conventional metrics: Comprehensive characterization of respirable coal mine dust. *Int J Coal Geol*. 2019, 207, 84-95.
25. Petsonk EL, Rose C, Cohen R. Coal mine dust lung disease. New lessons from an old exposure. *Am J Respir Crit Care Med*. 2013, 187: 1178–85.
26. Finkelman RB, Orem W, Castranova V, Tatu CA, Belkin HE, Zheng B, Lerch HE, Maharaj SV, Bates AL. Health impacts of coal and coal use: possible solutions. *Int J Coal Geol*. 2002, 50: 425–43.
27. Pretorius C J, Dekker J, Letsoalo, T Grove, Lottering J. Develop and operation manual for the implementation of the revised code of practice for assessment of personal exposure to airborne pollutants. 2015. Available from: http://www.mhsc.org.za/sites/default/files/Personal%20exposure%20to%20airborne%20pollutants%20Report.pdf (accessed on 10 August 2017).
28. South Africa. Government Notice: Amendment to the Regulations in Respect of Occupational Hygiene; Proclamation No. R989, Regulation Gazette No. 29276; South African Government: Pretoria, South Africa, 2006. Available from: https://www.dmr.gov.za/Portals/0/Resource%20Center/Mine%20Health%20and%20Safety%20Act%20and%20its%20Regulations/MHS%20Regulations_Regulation%2022.9%202%20a_GG%2029276.pdf?ver=2018-03-13-015401-040 (accessed on 17 January 2018).
29. Department of Mineral Resources. Guideline for compilation of mandatory code of practice for an occupational health programme on personal exposure to airborne pollutants. 2018. Available from: https://www.mhsc.org.za/sites/default/files/public/legislation_document/Airborne%20Pollutants%20Guideline_0.pdf (accessed on 18 August 2020).
30. VelocityEHS. IH 101: Introduction to Industrial Hygiene Program Management Part 1 – Similar Exposure Groups (SEGs). 2021. Available from: https://www.ehs.com/2021/03/ih-101-introduction-to-industrial-hygiene-program-management-part-1-similar-exposure-groups-segs/ (accessed on 2 October 2022).
31. Mulhausen J, Damiano J. Establishing similar exposed groups. In Bullock WH, Ignacio JS, editors. A strategy for assessing and managing occupational exposures. 2006. 3rd edn. Fairfax, VA: American Industrial Hygiene Association. pp. 33–44. ISBN 1-931504-69-4.
32. Resources Safety and Health. Coal – Similar Exposure Group (SEG) List. Available from: https://www.rshq.qld.gov.au/resources/documents/mines-resources/similar-exposure-groups.pdf (accessed on 2 January 2023).
33. Bullock WH, Ignacio JS. A strategy for assessing and managing occupational exposures. 3rd edn. Fairfax, VA: AIHA Press. 2008.
34. Rappaport SM, Lyles RH, Koper LL. An exposure assessments strategy accounting for within- and between worker sources of variability. *Ann Occup Hyg*.1995, 39: 469–95.

35. Lyles RH, Kupper LL, Rappaport SM. A lognormal distribution-based exposure assessment method for unbalanced data. *Ann Occup Hyg*. 1997, 41(1): 63-76
36. BOHS-NVvA. (Testing compliance with occupational exposure limits for airborne substances. 2011. Available from: https://www.arbeidshygiene.nl/-uploads/files/insite/2011-12-bohsnvva-sampling-strategy-guidance.pdf (accessed on 4 January 2020).
37. CEN. BS EN 689:2018 Workplace exposure. Measurement of exposure by inhalation to chemical agents. Strategy for testing compliance with occupational exposure limit values. Brussels, Belgium: British Standards Institution. 2018. Available from: https://www.en-standard.eu/bs-en-689-2018-workplace-exposure-measurement-of-exposure-by-inhalation-to-chemical-agents-strategy-for-testing-compliance-with-occupational-exposure-limit-values/ (accessed on 17 January 2019).
38. Maja P. Log-normal Distribution - A simple explanation. How to calculate μ & σ, the mode, mean, median & variance. Available from: https://towardsdatascience.com/log-normal-distribution-a-simple-explanation-7605864fb67c (accessed on 5 March 2023).
39. Krishnamoorthy K. Handbook of statistical distributions with applications. CRC Press. 2016
40. Wonnacott TH, Wonnacott RJ. Introductory Statistics, 5th Edition, John Wiley. ISBN 978-0-471-61518-7. 1990.
41. Hawkins NC, Norwood SK, Rock JC. A strategy for Occupational exposure assessment. *Am Ind Hyg Assoc J*. 1991, 160.
42. Mulhausen JR, Damiano J. A strategy for assessing and managing occupational exposures. Fairfax, VA: AIHA Press.1998.
43. CDC. Mining Topic: Respiratory Diseases. 2021. Available from: https://www.cdc.gov/niosh/mining/topics/respiratorydiseases.html (accessed on 1 November 2022).
44. Starzyński Z, Marek K, Kujawska A, Szymczak W. Mortality among coal miners with pneumoconiosis in Poland. *Int J Occup Med Environ Health*. 1995, 9 (4):279-89.
45. Yi Q, Zhang Z. The survival analyses of 2738 patients with simple pneumoconiosis. *Occup Environ Med*. 1996, 53 (2):129-35.
46. Meijers J, Swaen G, Slangen J. Mortality of Dutch coal miners in relation to pneumoconiosis, chronic obstructive pulmonary disease, and lung function. *Occup Environ Med*. 1997, 54 (10):708-13.
47. Attfield M, Kuempel E. Mortality among US underground coal miners: A 23 year follow up. *Am J Ind Med*. 2008, 51 (4):231-45.
48. Torres RCH, Ibanez Pinilla M, Briceno Ayala L, Checa Guerrero D M, Morgan Torres G, Groot de Restrepo H, et al. Underground coal mining: relationship between coal dust levels and pneumoconiosis, in two regions of Colombia, 2014. BioMed Research International. 2015.
49. Rajen Naidoo, Thomas Robins, Noah Seixas. Respiratory Diseases among South African Coalminers. 2002. Available from: https://mhsc.org.za/sites/default/files/public/research_documents/Respiratory%20diseases%20among%20South%20African%20coal%20miners.pdf (accessed on 1 July 2022).
50. MHSC. A Handbook On: Occupational Health Practice in The South African Mining Industry 2001. Available from: https://www.mhsc.org.za/sites/default/files/public/publications/OH%20Handbook.pdf (accessed on 1 July 2022).
51. Naidoo R N, Robins T G, Murray J. Respiratory outcomes among South African coal miners at autopsy. Am J Ind Med. 2005; 48 (3):217-24.
52. Grové T, Van Dyk T, Franken A, Du Plessis J. The evaluation and quantification of respirable coal and silica dust concentrations: a task-based approach. J Occup Environ Hyg. 2014, 11 (6):406-14.
53. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis (Third ed.). Chapman and Hall/CRC. ISBN 978-1-4398-4095-5.2013.
54. McElreath R. Statistical Rethinking: A Bayesian Course with Examples in R and Stan (2nd ed.). Chapman and Hall/CRC. ISBN 978-0-367-13991-9. 2020.

55. Congdon, P. Bayesian Statistical Modelling. New York: John Wiley & Sons, 2001
56. Ramachandran, G. Retrospective Exposure Assessment Using Bayesian Methods. *Ann Occup Hyg*. 2001. 45:651-67.
57. Hewett P, Logan P, Mulhausen J, Ramachandran G, Banerjee S. (2006) Rating exposure control using Bayesian decision analysis. *J Occup Environ Hyg*. 2006, 3: 568–81.
58. Wilson AG, Fronczyk KM. Bayesian reliability: Combining information. *Qual Eng*. 2017, 29, 119–129.
59. Lavoué J, Joseph L, Knott P, Davies H, Labrèche F, Clerc F, Mater G, Kirkham T. Expostats: a Bayesian toolkit to aid the interpretation of occupational exposure measurements. Ann Work Expo Health. 2019, 63: 267–79
60. INRS. Occupational Health and Safety. 2023. Available from: http://www.inrs.fr/accueil/produits/mediatheque/doc/outils.html?refINRS=outil13 (accessed on 15 March 2023).
61. Scheffers Theo. Theo Scheffers Arbo Consultancy (TSAC). Available from: https://www.tsac.nl/ (accessed on 2 February 2023).
62. Hewett, P. Exposure Assessment Solutions, Inc. 2023. Available from: https://www.easinc.co/ihda-software/ (accessed on 2 February 2023).
63. Belgian Society for Occupational Hygiene. BWstat V3. Available: https://www.bsoh.be/?q=nl/bwstat (accessed on 2 February 2023).
64. AHIA. IHSTAT (macro-free version). An Excel application that calculates various exposure statistics, performs goodness of fit tests, and graphs exposure data. 2013. Available from: https://www.aiha.org/public-resources/consumer-resources/apps-and-tools-resource-center/aiha-risk-assessment-tools/ihstat-macro-free-version (accessed on 2 February 2023).
65. Slottje P, Van der Sluijs JP, Knol AB. Expert Elicitation: Methodological suggestions for its use in environmental health impact assessments. 2008. Available from: http://www.nusap.net/downloads/reports/Expert_Elicitation.pdf (accessed on 7 May 2022).
66. Garthwaite PH, Kadane JB, O'Hagan A. Statistical methods for eliciting probability distributions. *J Am Stat Assoc*. 2005, 100 (470):680-701.
67. Hanea AM, Hemming V, Nane GF. Uncertainty quantification with experts: present status and research needs. *Risk Analysis*. 2022,42 (2):254-63.
68. Ferrell W. Discrete subjective probabilities and decision analysis: elicitation, calibration and combination. In Subjective probability, volume Eds. Wright, G. and Ayton, P. Cambridge Press, New York, 1994.
69. Cooke RM. Experts in uncertainty: opinion and subjective probability in science. Oxford University Press on Demand; 1991.
70. Wittmann ME, Cooke RM, Rothlisberger JD, Rutherford ES, Zhang H, Mason DM, Lodge DM. Use of structured expert judgment to forecast invasions by bighead and silver carp in Lake Erie. *Conservation Biology*. 2015, 29, 187-197.
71. Bamber JL, Aspinall, W. An expert judgement assessment of future sea level rise from the ice sheets. Nature Climate Change. 2013, 3, 424-427.
72. Gosling JP. SHELF: The Sheffield elicitation framework. Elicitation: The science and art of structuring judgement. 2018, 61-93
73. Dallow N, Best N, Montague T H. Better decision making in drug development through adoption of formal prior elicitation. *Pharm Stat*. 2018, 17,301–316.
74. Jurek L, Balthazar M, Gulati S, Novakovic N, Nunez M, Oakley J,O'Hagan A. Response (minimum clinically relevant change) in ASD symptoms after an intervention according to CARS-2: Consensus from an expert elicitation procedure. *Eur Child Adolesc Psychiatry*. 2021.
75. Rowe G, Wright G. The Delphi technique as a forecasting tool: issues and analysis. *Int J Forecast*. 1999, 15(4):353-75.
76. Yousuf MI. Using experts' opinions through Delphi technique. Practical assessment, research, and evaluation. 2007, 12(1):4.

77. Hemming V, Burgman MA, Hanea AM, McBride MF, Wintle BC. A practical guide to structured expert elicitation using the IDEA protocol. *Methods Ecol Evol*. 2018, 9, 169-80.
78. Adams-Hosking C, McBride MF, Baxter G, Burgman M, De Villiers D, Kavanagh R, Lawler I, Lunney D, Melzer A, Menkhorst P, Molsher R. Use of expert knowledge to elicit population trends for the koala (Phascolarctos cinereus). *Divers Distrib*. 2016, 22,249-62.
79. Hanea AM, McBride MF, Burgman, MA, Wintle BC. Classical meets modern in the IDEA protocol for structured expert judgement. *J Risk Res*. 2018, 21, 417-33.
80. NIOSH. Particulates not otherwise regulated, total: Method 0500. In NIOSH Manual of Analytical Methods, 4th ed.; DHHS (NIOSH) Publication No. 94-113; United States Department of Health and Human Services, Centers for Disease Control and Prevention: Cincinnati, OH, USA, 1994. Available online: www.cdc.gov/niosh/nmam/ (accessed on 15 July 2021).
81. CDC. Best Practices for Dust Control in Coal Mining. 2010. Available from: https://www.cdc.gov/niosh/mining/UserFiles/works/pdfs/2010-110.pdf (accessed on 10 January 2023).
82. Cooke RM, Probst KN. Highlights of the Expert Judgment Policy Symposium and Technical Workshop. Conference Summary, 2006. Available online: http://www.rff.org/expertjudgementdocuments/documents/workshopdocs/RFFConf_06-ExpertJudgment.pdf (accessed on 20 July 2022).
83. Banerjee S, Ramachandran G, Vadali M, Sahmel J. Bayesian hierarchical framework for occupational hygiene decision making. Ann Occup Hyg. 2014,58: 1079–93.
84. McNally K, Warren N, Fransman W, Entink RK, Schinkel J, Van Tongeren M, Cherrie JW, Kromhout H, Schneider T, Tielemans E. Advanced REACH Tool: a Bayesian model for occupational exposure assessment. *Ann Occup Hyg*. 2014, 58: 551–65.
85. Carlin BP, Louis TA. Bayesian Methods for Data Analysis, 3rd ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2008.
86. Quick H, Huynh T, Ramachandran G. A method for constructing informative priors for Bayesian modelling of occupational hygiene data. *Ann Work Expo. Health*. 2017, 61, 67–75.
87. Baldos ULC, Viens FG, Hertel TW, Fuglie KO R&D spending, knowledge capital, and agricultural productivity growth: A Bayesian approach. *Am J Agric Econ*. 2019, 101, 291–310.
88. Gelfand AE, Smith AFM. Sampling-based approaches for calculating marginal densities. *J Am Stat Assoc*. 1990, 85, 398–409.
89. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. 2015. Available online: https://www.r-project.org/ (accessed on 10 April 2021).
90. Makowski D, Ben-Shachar MS, Lüdecke D. bayestestR: Describing effects and their uncertainty, existence, and significance within the Bayesian framework. *J Open Source Softw*. 2019, 4, 1541.
91. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992, 7, 457–472.
92. VehtariA, Gelman A, Simpson D, Carpenter B, Bürkner PC. Rank-normalization, folding, and localization: An improved R for assessing convergence of MCMC. arXiv 2019, arXiv:1903.08008.
93. Stan Development Team. RStan: The R Interface to Stan. R Package Version 2.18.2. 2018. Available online: http://mc-stan.org (accessed on 19 March 2022).
94. Barlow RE, Mensing RW, Smiriga NG. "Combination of Experts' Opinions Based on Decision Theory," In Reliability and Quality Control, Basu, A. P. (Ed.), Elsevier Science Publishers (North-Holland), New York, 1986, 9-19.
95. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Anal*. 2006, 1, 515–533.
96. Jones RM, Burstyn I. Bayesian analysis of occupational exposure data with conjugate priors. *Ann Work Expo Health*. 2017, 61, 504–514.

97. Waters M, McKernan L, Maier A, Jayjock M, Schaeffer V, Brosseau Exposure estimation and interpretation of occupational risk: enhanced information for the occupational risk manager. *J Occup Environ Hyg*. 2015, 12 (Suppl. 1): S99–111.

98. Roach SA. Alternative ways of monitoring exposure. In Rappaport SM, Smith TJ, editors. Exposure assessment for epidemiology and hazard control. Chelsea, MI: Lewis Publishers. pp. 3–20. 1991.

99. Buringh E Lanting R. Exposure variability in the workplace: its implications for the assessment of compliance. *Am Ind Hyg Assoc J*. 1991, 52: 6–13.

100. Rappaport SM, Kromhouta H, Symanski E. Variation of exposure between workers in homogeneous exposure groups. *Am Ind Hyg Assoc J*. 1993, 54(11):654-62.

101. Burdorf A. Identification of determinants of exposure: consequences for measurement and control strategies. *Occup Environ Med*. 2005, 62: 344–50.

102. Chakraborty S, Ghosh M, Chakraborty S, Ghosh M. Applications of Bayesian Neural Networks in Prostate Cancer Study. In Handbook of Statistics; Elsevier: Amsterdam, The Netherlands, 2012; pp. 241–262. Available from: https://www.sciencedirect.com/science/article/abs/pii/B9780444518750000099 (accessed on 5 April 2019).

103. Vadali M, Ramachandran G Banerjee S. Effect of training, education, professional experience and need for cognition on decision making in occupational exposure assessment. *Ann Occup Hyg*. 2012, 56, 292–304.

104. Joint ACGIH-AIHA Task Group on Occupational Exposure Databases. Data elements for occupational exposure databases: guidelines and recommendations for airborne hazards and noise. *Appl Occup Environ Hyg*. 1996, 11, 1294–1313.

105. Made F, Kandala NB, Brouwer D. Bayesian hierarchical modelling of historical data of the South African coal mining industry for compliance testing. *Int J Environ Res Public Health*. 2022, 19, 4442.

106. Rajan B, Alesbury B, Carton C, G´erin M, Litske H, Marquart H, Olsen E, Scheffers, T.; Stamm, R.; Woldbaek; T. European proposal for core information for the storage and exchange of workplace exposure measurements on chemical agents. *Appl Occup Environ Hyg.* 1997, 12, 31–3

107. Kahneman D, Slovic P, Tversky A. Judgment under uncertainty: heuristics and biases. Cambridge, UK: Cambridge University Press. 1982.

108. Gilovich T, Griffin D, Kahneman D. Heuristics, and biases: the psychology of intuitive judgment. Cambridge, UK: Cambridge University Press. 2002.

109. Vadali M, Ramachandran G, Mulhausen JR. Exposure modelling in occupational hygiene decision making. *J Occup Environ Hyg*. 2009, 6: 353–62.

APPENDICES

Appendix A: Coal Dust Data Capture Sheet

FULL SHIFT GRAVIMETRIC SAMPLING OF AIRBORNE PARTICULATE CONTROL SHEET

Client name:

Sampling Period: 01/07/2018 - 31/12/2018

PERMISSIBLE DUST EXPOSURE LEVEL PER 8HOUR SHIFT: Coal dust = 2mg/m³
Silica= 0.1mg/m³

Sampling Quarter: Quarter 1 February 2018

Sampling Date	Area Description	Area Number	HEG	Shift Sampled	DME Codes	Filter Number	Instrument Number	Persons Name	Coy or ID Number	Occupation	TWA mg/m³	Silica mg/m³

Remarks

Signatures

COLLIERY VOHE OFFICER	VOHE SUPT.	VOHE MANAGER
Date:	Date:	Date:
SHAFT ENGINEER	SHAFT MANAGER	ENGINEERING MANAGER
Date:	Date:	Date:
MINING MANAGER	GENERAL MANAGER	
Date:	Date:	

Appendix B: Ethics clearance certificate

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R14/49 Messrs F Made, J Lavoue and N-B Kandala

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M181038

NAME: (Principal Investigator)	Messrs F Made, J Lavoue and N-B Kandala
DEPARTMENT:	School of Public Health Medical School University
PROJECT TITLE:	Preventing coal mine dust lung disease: application of the Bayesian Hierarchical Framework for occupational exposure assessment in the South African coal mining industry
DATE CONSIDERED:	26/10/2018
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Professor D Brouwer
APPROVED BY:	CB Penny Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	12/02/2019

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on 3rd floor, Phillip V Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.
I/We fully understand the conditions under which I am/we are authorised to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated from the research protocol as approved, I/we undertake to resubmit to the Committee. **I agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date of the meeting when the study was initially reviewed. In this case, the study was initially reviewed in **October** and will therefore reports and re-certification will be due early in the month of **October** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature

Date 13/02/2019

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG	HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

Office of the Deputy Vice-Chancellor (Research & Post Graduate Affairs)

TO: Messrs F Made, J Lavoue and N-B Kandala
School of Public Health
Medical School
University

E-mail: madefelix@gmail.com

CC: Supervisor: Professor D Brouwer <Derk.Brower@wits.ac.za>
and <HREC-Medical.ResearchOffice@wits.ac.za>

FROM: Iain Burns
Human Research Ethics Committee (Medical)
Tel: 011 717 1252

E-mail: Iain.Burns@wits.ac.za

DATE: 12/02/2019

REF: *R14/49*

PROTOCOL NO: **M181038** *(This is your ethics application study reference number. Please quote this reference number in all correspondence relating to this study)*

PROJECT TITLE: *Preventing coal mine dust lung disease: application of the Bayesian Hierarchical Framework for occupational exposure assessment in the South African coal mining industry*

Please find attached the Clearance Certificate for the above project. I hope it goes well and that an article in a recognized publication comes out of it. This will reflect well on your professional standing and contribute to the Government funding of the University.

MSWorks2000/Iain0007/Clearscan.wps

Appendix C: Data permission letter

Addendum to Donation Agreement between AngloAmerican South Africa and the University of the Witwatersrand as per 19 May 2014

Memorandum of Agreement Data-Sharing and Usage

This agreement establishes the terms and conditions under which the University can acquire and use data from the Donor.

Preamble

The parties agree to cooperate in the pursuit of academic research activities in the field of occupational health and hygiene. The Donor will facilitate such research activities by providing access to relevant data to the University, i.e. the Wits School of Public Health (WSPH) staff and post-graduate students .

1. The confidentiality of data pertaining to individuals will be protected as follows:

a. The data recipient will not release the names of individuals, or information that could be linked to an individual, nor will the recipient present the results of data analysis (including maps) in any manner that would reveal the identity of individuals.

b. The data recipient will not release individual addresses, nor will the recipient present the results of data analysis (including maps) in any manner that would reveal individual addresses.

2. The data recipient will not release data to a third party without prior approval from the Donor.

3. The data recipient will not share, publish, or otherwise release any findings or conclusions derived from analysis of data obtained from the data provider without prior notification to the Donor.

4. Data transferred pursuant to the terms of this Agreement shall be utilized solely for the purposes set forth in the "Donation Agreement".

5. Any third party granted access to data, as permitted under condition #2, above, shall be subject to the terms and conditions of this Agreement. Acceptance of these terms must be provided in writing by the third party before data is released.

Appendix D: Bayesian modelling using non-informative prior distribution

```
#remove all existing variables from work space
rm(list=ls())

##Import data
past.heg.1 <- read.csv("C:/Users /past heg 1.csv", sep=";")

#Install R Packages
install.packages("bayestestR")
install.packages("ggplot2")
install.packages("rstan")
install.packages("insight")
install.packages("dplyr")

library(bayestestR)
library(ggplot2)
library(rstan)
library(dplyr)

#Past Data

#New Data
data_present <- log(present.heg.1$heg_1)
n=length(data_present)

oel=exp(2) ##oel=2 in SA
thres=2 #upper bound for X95

#Prior Specifications
theta=m0; tau2=inf #produce non-informative prior bounds for mu from BDA
# lower bound from BDA and upper bound changed iteratively
mub=c(log(.0001*oel), NA)

##################
#sig2 ~ IG(a,b)
# parameterization of the inverse gamma
a=1/2; b=inf #the same as uniform prior on sig=log(GSD)
#lower bound from BDA and upper bound vary iteratively
sigb=c(log(1.05), NA)^2
#The MCMC algorithm
niter=20000 #20,000 iterations
mu=sig=sig2=rep(NA,niter)

#initialize the parameters
mu[1]=0
sig[1]=sig2[1]=1

for(i in 2:niter){
  #specify upper bound for mu based on most recent value of sigma
  mub[2]=log(thres*oel)-1.645*sig[i-1]
```

```
#update mu
  mumu = (mean(data_present)/(sig2[i-1]/n) + theta/tau2) / (n/sig2[i-1] + 1/tau2)
  musig= 1/(n/sig2[i-1] + 1/tau2)
  #the inverse CDF method to sample mu
  #if mub[1] < mu < mub[2],
  #then pnorm(mub[1]) < pnorm(mu) < pnorm(mub[2])
  #where pnorm(.) denotes the appropriate CDF
  u=runif(1,pnorm(mub[1],mumu,sqrt(musig)),
          pnorm(mub[2],mumu,sqrt(musig)))
  mu[i]=qnorm(u,mumu,sqrt(musig))
  #specify upper bound for sigma^2 based on new value of mu
  sigb[2]=( (log(thres*oel)-mu[i])/1.645 )^2
  #update sig2
  # use the inverse CDF method
  u=runif(1,pgamma(1/sigb[2],n/2+a, (1/b+sum((data_present-mu[i])^2 )/2)),
          pgamma(1/sigb[1],n/2+a, (1/b+sum((data_present-mu[i])^2 )/2)))
  sig2[i]=1/qgamma(u,n/2+a, (1/b+sum((data_present-mu[i])^2 )/2) )
  sig[i]=sqrt(sig2[i])
}
#compute P95 using the samples of mu, sig
postGMP= exp(mu)
postgsd = exp(sqrt(sig))
x95 = exp(log(postGMP) + 1.645 * log(postgsd))

#Calculate the P95 exposure probabilities in each of the OEL categories
Ecat<-ifelse(x95>c(log(oel)),4,
             ifelse(x95>=(log(0.5*oel)) & x95<(log(oel)),3,
                    ifelse(x95>=(log(0.1*oel)) & x95<(log(0.5*oel)),2,
                           ifelse(x95<(log(0.1*oel)), 1, 0))))
table(Ecat)/niter

#Calculate the posterior median of the GM, GSD and P95
median(postGMP)
median(postgsd)
median(x95)
##credible interval
ci_etigm<-ci(postGMP)
ci_etigm
ci_etigsd<-ci(postgsd)
ci_etigsd
ci_eti<-ci(x95)
ci_eti

######End
```

Appendix E: Bayesian modelling using informative prior distribution.

```
#remove variables from the work space
rm(list=ls())

##Import data
past.heg.1 <- read.csv("C:/Users /past heg 1.csv", sep=";")
present.heg.1 <- read.csv2("C:/Users /present heg 1.csv")

#Install R Packages
install.packages("bayestestR")
install.packages("ggplot2")
install.packages("rstan")
install.packages ("insight")
install.packages("dplyr")

library(bayestestR)
library(ggplot2)
library(rstan)
library(dplyr)

#Past Data
prior_data=log(past.heg.1$heg_1)
ybar0 = mean(prior_data)
sy20 = var(prior_data)

#New Data
data_present <- log(present.heg.1$heg_1)
n=length(data_present)

oel=exp(2) ##oel=2 in SA
thres=2 #upper bound for X95

#Prior Specifications
#n0=N;
m0=ybar0; v0=sy20
#mu ~ N (theta, tau2)
theta=m0; tau2=v0/5
#bounds for mu from BDA and upper bound changed iteratively
mub=c(log(.0001*oel),NA)

##################
#sig2 ~ IG(a,b)
# parameterization of the inverse gamma
a=(5-1)/2; b=2/(v0*(5-1)) #the second parameter is inverted
# lower bound from BDA and upper bound vary iteratively
sigb=c(log (1.05), NA)^2

#The MCMC algorithm
niter=20000 #20,000 iterations
mu=sig=sig2=rep(NA,niter)
```

```
#initialize the parameters
mu[1]=0
sig[1]=sig2[1]=1

for(i in 2: niter){
  # upper bound for mu was based on most recent value of sigma
  mub[2]=log(thres*oel)-1.645*sig[i-1]
  #update mu
  mumu = (mean(data_present)/(sig2[i-1]/n) + theta/tau2) / (n/sig2[i-1] + 1/tau2)
  musig= 1/(n/sig2[i-1] + 1/tau2)
  #since we're using a truncated normal,
  #we use the inverse CDF method to sample mu
  #if mub[1] < mu < mub[2],
  #then pnorm(mub[1]) < pnorm(mu) < pnorm(mub[2])
  #where pnorm(.) denotes the appropriate CDF
  u=runif(1,pnorm(mub[1],mumu,sqrt(musig)),
          pnorm(mub[2],mumu,sqrt(musig)))
  mu[i]=qnorm(u,mumu,sqrt(musig))
  #specify upper bound for sigma^2 based on new value of mu
  sigb[2]=( (log(thres*oel)-mu[i])/1.645 )^2
  #update sig2
  #again using the inverse CDF method
  u=runif(1,pgamma(1/sigb[2],n/2+a, (1/b+sum((data_present-mu[i])^2 )/2)),
          pgamma(1/sigb[1],n/2+a, (1/b+sum((data_present-mu[i])^2 )/2)))
  sig2[i]=1/qgamma(u,n/2+a, (1/b+sum((data_present-mu[i])^2 )/2) )
  sig[i]=sqrt(sig2[i])
}

#compute P95using the samples of mu, sig
postGMP= exp(mu)
postgsd = exp(sqrt(sig))
x95 = exp(log(postGMP) + 1.645 * log(postgsd))

#Calculate the P95 exposure probabilities in each of the OEL categories
Ecat<-ifelse(x95>c(log(oel)),4,
             ifelse(x95>=(log(0.5*oel)) & x95<(log(oel)),3,
                    ifelse(x95>=(log(0.1*oel)) & x95<(log(0.5*oel)),2,
                           ifelse(x95<(log(0.1*oel)), 1, 0))))
table(Ecat)/niter

#Calculate the posterior median of the GM, GSD and P95
median(postGMP)
median(postgsd)
median(x95)
##credible interval
ci_etigm<-ci(postGMP)
ci_etigm
ci_etigsd<-ci(postgsd)
ci_etigsd
ci_eti<-ci(x95)
ci_eti

######End
```

Appendix F: Experts general information sheet

Experts Elicitation

Instruction for the experts

Based on the 95th percentiles (P95) value of the eight-hour time weighted average (TWA8h) coal dust concentration distribution for a specific job title, question (a) will ask you to rate (in percentages) where you think the chances are (or the likelihood is) that the P95 would be in each of the occupational exposure category. Instruction: If you think that it is absolutely clear that it will fit in exposure category 4, you fill in 100% and leave the rest of the categories empty; if you think that the highest chance is (most likely) that the P95 will be in exposure category 4, e.g., 80%, but there also a chance that it will be in category 3, say 20%, you fill in 80% for category 4 and 20% for category 3. The total sum of the percentages should be equal to 100. Therefore, you have to distribute your ratings in categories in a way that the total is 100%.

The next question (b) asks you to give the range (minimum-maximum) of the percentages that you answered to question a (your best estimate).

Question (c) is about your level of confidence that the answer(s) you gave to questions an and b are correct. Please bear in mind that this is about your level of certainty and is different from the estimated distribution of the P95 over exposure categories, as was asked in questions (a) and (b).

Table 1 is the description of the exposure categories. Below are the three questions that you can use to put your rating in percentages in Table 2 for each job title. For question a, please make sure your answers are equal to 100%.

Expert elicitation questions for Table 2 (please give all answers in percentages)

a) Based on your experience in the industry, what would be your best possible guess in percentages, of the P95 of exposure concentrations being located in each of the exposure categories?
b) What would be the maximum and minimum percentages in each of the exposure categories?
c) How sure are you that the above answers are correct?

Appendix G: Experts Elicitation Questionnaire

Experts rating. Kindly give a value between 0 and 100% (for example, 10%, 15%, or 60%). No value should exceed 100%.

Job titles	Ratings	SA exposure categories, Coal dust OEL (2mg/m3)				
		P95< 0.1% OEL	P95≥ 0.1 OEL & < 0.5 OEL	P95≥ 0.5 OEL & < OEL	P95≥ OEL	Total (%)
Job title X (This is an example, say for expert 1)	a	40%	5%	20%	35%	100
	b	20% to 70%	0 to 15%	2% to 30%	30 to 90%	
	c	70% sure	50% sure	90% sure	100% sure	
Beltsman	a					100
	b					
	c					
CM Operator	a					100
	b					
	c					
Conveyer belt Attendant	a					100
	b					
	c					
Emico Driver	a					100
	b					
	c					
Electrician	a					100
	b					
	c					
Face Boss	a					100
	b					
	c					
Pump Attendant	a					100
	b					
	c					
Roofbolt Operator	a					100
	b					
	c					
Safety Officer	a					100
	b					
	c					
Shuttle Car Operator	a					100
	b					
	c					

a) Based on your experience in the industry, what would be your best possible guess in percentages, of the P95 of exposure concentrations being in each of the exposure categories?
b) What would be the maximum and minimum percentages in each of the exposure categories?
c) How sure are you that the above answers are correct?

Appendix H: Compliance Testing and Homogenous Exposure Group Assessment in the South African Coal Mining Industry

Annals of Work Exposures and Health, 2021, 1–11
doi: 10.1093/annweh/wxab030
Original Article

Original Article

Compliance Testing and Homogenous Exposure Group Assessment in the South African Coal Mining Industry

Felix Made[1,2,*], Ngianga-Bakwin Kandala[1] and Derk Brouwer[1]

[1]Division of Occupational Health and Epidemiology, School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, 60 York Road, Johannesburg, 2193 Gauteng Province, South Africa; [2]Epidemiology and Surveillance Section, National Institute for Occupational Health, National Health Laboratory Services, 25 Hospital Street, Johannesburg, 2001 Gauteng Province, South Africa

*Author to whom correspondence should be addressed. Tel: +27117126594; e-mail: FelixM@nioh.ac.za

Submitted 12 January 2021; revised 26 March 2021; editorial decision 28 March 2021; revised version accepted 1 April 2021.

Abstract

Objectives: Globally, several strategies for compliance testing and within-group exposure variability have been suggested. This study aimed to evaluate the performance of the South African Mining Industry Code of Practice (SAMI CoP) approach for grouping and compliance testing against international standards.
Methods: A total of 28 homogenous exposure groups (HEGs) with 728 underground coal mine workers' eight-hour time-weighted average coal dust concentration data were obtained. Compliance testing was assessed using 10% exceedance above occupational exposure limit (OEL) for SAMI CoP, and the 95th percentile of the lognormal distribution was computed for the European Standardization Committee (CEN) and American Industrial Hygiene Association (AIHA). Comparison of the homogeneity of the HEGs was done between SAMI CoP which mandates that both the arithmetic mean (AM) and 90th percentile must fall in the same exposure band to certify homogeneity and the global geometric standard deviation (GSD) and Rappaport ratio (*R*-ratio) with specific acceptability criteria. To test the homogeneity of exposure within job titles, eight non-homogenous HEGs that have two or more job titles with three measurements were investigated using GSD and the SAMI CoP criteria.
Results: A total of 21 HEGs out of 28 were non-compliant to the OEL across SAMI CoP, CEN, and AIHA criteria. Compliance to the OEL was observed for seven HEGs according to the SAMI CoP approach, whereas only one HEG was compliant according to both the SAMI CoP and CEN approaches. The GSD criterion and SAMI CoP revealed that 11 and 6 HEGs were homogenous, respectively, and only on 4 occasions, the 2 approaches agreed. The job titles of the majority of non-homogenous HEGs in both SAMI CoP and GSD were actually homogenous. Five out of 10 sub-groups have their AM above that of HEG B. Other HEGs had at least one of their AM and 90th percentile values above that of their respective parent HEGs.

Conclusions: All three approaches mainly confirmed non-compliance of HEGs. SAMI CoP tended to show compliance of HEGs more than CEN. Non-homogenous HEGs had many job titles that were homogenous according to both SAMI CoP and GSD criteria. There was no perfect agreement of homogeneity by all the indicators. For both future constitutions of HEGs as well as a retrospective assessment of high exposure groups, homogeneity can be improved by using job titles.

Keywords: Code of Practice; exceedance criteria; indicators of homogeneity; occupational exposure limit

What's important about this paper

Coal dust exposure level above the occupational exposure limit (OEL) poses a risk of developing lung diseases in the coal mining industry. Using an appropriate strategy for the identification of highly exposed workers is important for evidence-based intervention and prevention management. This manuscript compared three methods of determining the extent of non-compliance with OELs among homogenous exposure groups of coal mine workers in South Africa. The South African Mining Industry Code of Practice method was found to be less conservative than other methods evaluated.

Introduction

In the South African Mining Industry (SAMI), the Mine Health Safety Act (MHSA) 29 of 1996 is a law governing occupational health and safety in the South African (SA) mines. Some of the regulations are that mines must have health and safety management protocols which include workplace exposures risk assessment, medical surveillance to ensure good health and safety of all workers (Guild *et al.*, 2001). Several mechanisms including regular workplace inspections, safety audits, collection, and analysis of occupational disease trends have been put in place.

South Africa is the fourth largest producer of coal in the world, employing more than 86 000 workers in the year 2018 (Mineral Council of South Africa, 2019). Approximately 51% of SA coal mining is done underground and about 49% is produced by open-cast techniques (Department of Energy Report, 2016). The generation of coal dust is inevitable during coal mining activities and since about 5% quartz or silicates found in coal dust, its inhalation is highly associated with the development of Coal Mine Dust Lung Disease (CMDLD) (Finkelman *et al.*, 2002; Stanton *et al.*, 2006; Petsonk *et al.*, 2013). CMDLD risk increases with coal dust overexposure which depends on the duration and type of work activity in the mines (Grove *et al.*, 2014). The current occupational exposure limit (OEL) for respirable coal dust in SA is 2 mg m^{-3} derived from personal eight-hour time-weighted average (TWA8h) coal dust measurement (South Africa, 2006). Since it is assumed that the coal dust above the OEL poses a risk to develop CMDLD, the identification of highly exposed workers is important for risk assessment and risk management purposes. Therefore, appropriate compliance testing strategies and grouping of workers are needed.

For compliance testing, the SAMI Code of Practice (CoP) approach assumes that sample data are normally distributed, and the 90th percentile of the data should be below the OEL (Department of Mineral Resources, 2018), meaning 10% overexposure is implicitly accepted. According to the European Committee for Standardization (CEN) (CEN, 2018) and the British and Dutch Occupational Hygiene Societies' guidelines, non-compliance or overexposure is that the 95th percentile of the exposure distribution should be above the OEL (BOHS-NVvA, 2011). The guidelines assume lognormal distribution of the underlying population distribution and 5% exceedance of the OEL is accepted according to the European standard (CEN, 2018). In addition to this acceptability criterion, the uncertainty is also addressed. The recommendation is that uncertainty should be <30% (BOHS-NVvA, 2011; CEN, 2018). This indicates that non-compliance should be demonstrated with at least 30% certainty so that 5% or less of the exposures in the homogenous exposure group (HEG) are above the OEL (INRS, 2005, 2008; Bullock and Ignacio, 2008). Furthermore, the American Industrial Hygiene Association (AIHA) stipulated that uncertainty should be less than 5% so that there is 95% probability, upper credible limit (UCL) that the true 95th percentile of the exposure is below the OEL (Bullock and Ignacio, 2008). The lognormal distribution assumption is a consensus that is based on the analysis of the variability of (occupational hygiene) exposure data (Esmen, 1979; Rappaport, 2001).

Currently, it is not possible to measure every worker each day, this has led to a prospective grouping approach of exposed workers into so-called HEGs or similar exposure groups (SEGs). According to the SAMI CoP, HEGs in underground mines should be constituted by subdividing each mine into sampling areas based on 'common intake and return air' and subsequently dividing the sampling area into activity areas (Guild *et al.*, 2001; Pretorius *et al.*, 2015; Department of Mineral Resources, 2018). The approach of areas with common air intake results that different job titles will be included in the HEG, which might introduce high heterogeneity of the exposure level in the defined HEG. The SAMI CoP guidance stipulates that for a HEG to be statistically correctly defined, both the arithmetic mean (AM) and the 90th percentile of the exposure data must fall within the same exposure classification band relative to the OEL. However, this differs from the AIHA, CEN (CEN, 2018), and British and Dutch guidelines (BOHS-NVvA, 2011). The strategy for prospectively defining similar or HEGs in these guidelines is based on the assumption that grouping workers performing similar tasks would result in similar exposure profiles. Various methods of retrospectively assessing appropriate grouping of the homogenous or similar exposure group (S)HEG exist. The AIHA and the BOHS guidelines recommend the deployment of analysis of variance (ANOVA) to assess whether the HEG is statistically homogenous (Hawkins *et al.*, 1991; Mulhausen and Damiano, 1998; Bullock and Ignacio, 2008). According to the BOHS guideline, within-worker correlation greater than 0.2 stipulates that the group is not homogenous (BOHS-NVvA, 2011). Another indicator for homogeneity is the so-called *R*-ratio which reflects the ratio of the most exposed worker to the least exposed. It is defined as the 'ratio of the 97.5th to the 2.5th percentiles of the distribution of individual AM', where a ratio of 2 or less has been proposed to define homogeneity (Kromhout *et al.*, 1993; Rappaport *et al.*, 1993). The global geometric standard deviation (GSD) is also used to evaluate the homogeneity of a HEG. A global GSD of more than 3 suggests that the group might not be homogenous (CEN, 2018).

Based on the above literature, it seems important to assess whether the sampling strategy for HEGs in the SAMI CoP yields comparable results as recent consensus guidelines such as those proposed by CEN, the AIHA, and the British and Dutch societies for occupational hygiene. In this study, we aimed to compare compliance to the OEL of each HEG according to SAMI CoP and CEN/BOHS-NVvA and AIHA approaches to establish whether the methods agree. Secondly, we compared the within-group exposure variance of HEGs using the SAMI CoP and GSD criteria, and finally of job titles as an alternative sub-grouping from non-homogenous HEGs.

Methods

Study design and data collection

The primary study was a yearly period survey collection of respirable coal mine dust exposure measurements among male miners working in underground coal mines in South Africa between 2009 and 2018. The data were collected by the occupational hygiene department of the coal mining company. Workers included in the study were from 28 HEGs. The number of employees in each HEG was noted, and then sample selection was done using guidelines from the Department of Mineral Resources. A mixed stratified and random selection sampling frame was used, taking into account that either 5% of the workers assigned to a HEG, or a minimum of five workers should be selected for a measurement campaign. Each selected worker was issued with a dust sampling pump (calibrated at a flow rate of 2.2 l min^{-1}), positioned on the belt of the worker. The calibration was done before and after sampling. If the difference in concentration before and after sampling was more than 5%, the sample was discarded. Other samples were rejected if the concentration of the coal dust were unusually very high and if employees were not performing their usual duties on the day of sampling. On average, the flow rate before and after sampling for the data included in this study was 2.24 l min^{-1}. The shift time was 12 h, and any sampling time less than 80% of the shift time was considered not a representative sample and excluded.

The pump was connected to a size-selective cyclone-based on the Higgins–Dewell cyclone model with a mixed cellulose ester filter. The cyclone was positioned in the breathing zone of the workers on the collar of their overalls (clothes) to measure the respirable dust that could be inhaled by the workers. After sampling, the cyclone's filter was gravimetrically analysed according to the National Institute for Occupation Safety and Health (NIOSH) method 0600 (NIOSH, 1994). The TWA8h concentration was obtained by converting the gravimetrically determined coal dust mass (mg) to a virtual 480 min sampling time and reported for individual workers by their job title and captured in Microsoft Excel spreadsheets.

The HEGs were from nine different mines. Each HEG was specific to a single mine and none of the HEGs was found across different mines, and within a defined HEG, different job titles were included. Ethical clearance for this study was obtained from the Human Research Ethics Committee (HREC) of the University of the Witwatersrand after an agreement between the University and the coal mining company to use the coal dust data while ensuring confidentiality and participants privacy at all times. The clearance number is M181038.

Statistical analysis

Statistical analysis was done using Stata version 16.1 and EXPOSTATS (www.expostats.ca). EXPOSTATS an online Bayesian computation tool that assumes the lognormal distribution of the data (Lavoué *et al.*, 2019). For variance, priors were derived from McNally *et al.* (2014). A weakly informative uniform prior distribution was used for the geometric mean (GM) (Banerjee *et al.*, 2014). Details of the mathematical expression of the Bayesian framework and the prior distribution can be found in Lavoué *et al.* (2019), Appendix A of the Supplementary Material. EXPOSTATS provides estimates of the probability of overexposure for various definitions of overexposure that can be set by the user. The exposure is the TWA8h respirable coal dust concentration over 8 h. A total of 28 HEGs were assessed for compliance. The number of workers, the AM, and GM of individuals HEGs was described.

For compliance testing, the SA OEL for coal dust is used, which is 2 mg m^{-3}. The SAMI CoP criterion for compliance testing is based on the empirical 90th percentile of the exposure data, which should be below the OEL. In the CEN/BOHS-NVvA and AIHA approaches, the estimated 95th percentile of the population is used as an acceptability criterion when compared with the OEL, and the UCL for uncertainty is based on the 70 and 95% UCL, respectively. See Table 1 for the summary of the guidelines.

The within HEG variability was investigated using the SAMI CoP criterion and other indicators of homogeneity which include the GSD and the *R*-ratio. This study adopted the SAMI CoP acceptability criterion for homogeneity where both the AM and 90th percentile must fall within the same SA OEL exposure band. The SAMI CoP OEL exposure bands are A: ≥OEL, B: ≥0.5OEL and <OEL, C: ≥0.1OEL and <0.5OEL, and D: <0.1%OEL. Where bands A, B, C, and D represent poorly controlled, adequately controlled, high controlled, and very highly controlled, respectively. Other approaches for testing 'homogeneity' of the groups were explored with acceptability criteria which include: (i) The global GSD, a GSD of 3 or less was used to confirm homogeneity. (ii) The *R*-ratio (ratio of the AM of the most exposed worker to the AM of the least exposed worker). The *R*-ratio point estimate of 95% of the workers was used, which is the ratio of the 97.5th to the 2.5th percentiles of the distribution of the individual AM, where a ratio of 2 or less was considered homogenous. In this analysis, a less restrictive *R*-ratio point estimate of 90% of the workers was used, which is interpreted as the 'ratio of the 95th to the 5th percentiles of the distribution of individual AM'. See Table 1 for the summary of the guidelines.

We then investigated the constitution of sub-groups by job titles as an alternative approach and tested for homogeneity by comparing the SAMI approach and GSD. All the HEGs that were not homogenous in all approaches in the above analyses were selected. We then restricted the selected HEGs to those with at least two job titles that have three or more measurement values to assess the variability of alternative sub-grouping when using job titles. The SAMI's approach of both the AM and the 90th percentile must have within the same exposure band was employed. Other indicators used to assess the homogeneity were the global GSD, a GSD of less than 3 was considered homogenous. See Table 1 for the summary of the guidelines.

Results

There were a total of 728 workers in 28 HEGs from 9 underground coal mining production sites (Table 2). Most of the HEGs had more than 10 data points. In summary, HEG A had the highest AM dust concentration (AM: 2.19 mg m^{-3}) while HEG N had the lowest exposure level (AM: 0.58 mg m^{-3}). The GM exposure was highest in HEG H (GM: 1.70 mg m^{-3}), while HEG R had the lowest GM exposure concentration of 0.20 mg m^{-3}.

The compliance comparison of SAMI CoP 90th percentile and the CEN/BOHS-NVvA and AIHA 95th percentile with 70 and 95% UCL, respectively, that the true underlying exposure distribution is below the OEL was conducted for all HEGs as shown in Table 1. There was agreement of non-compliance in 21 (75%) of the 28 HEGs across all the 3 criteria. The SAMI CoP indicates compliance in seven (25%) of the HEGs. Only in one HEG where SAMI CoP and CEN/BOHS-NVvA agreed concerning compliance decision, except AIHA which did not agree. Comparison of the 70 and 95% UCL approaches of BOHS/NVvA and AIHA, respectively, showed an agreement of decision in 27 (96%) of the 28 HEGs.

Table 1. Guidelines used to assess compliance and homogeneity.

		Guidelines used by various methods		
		SAMI CoP	CEN/BOHS-NVvA	AIHA
Compliance testing	For data to be compliant	The 90th percentile of the data should be below the OEL	The 95th percentile of the data should be below the OEL	The 95th percentile of the data should be below the OEL
	Assumption of the underlying data	Normal distribution	Lognormal distribution	Lognormal distribution
	Uncertainty	None	70% UCL	95% UCL
		Guidelines used by various methods		
		SAMI CoP	Global GSD	*R*-Ratio
Assessment of homogeneity	For data to be regarded as homogenous	The AM and 90th percentile values must fall within the same OEL exposure band for homogeneity	GSD of 3 or less was used to confirm homogeneity	The *R*-ratio point estimate of 95 and 90% of the workers, where a ratio of 2 or less was considered homogenous

Table 3 presents the comparison of the SAMI approach for assessing homogeneity of HEGs (as constituted by SAMI) to that of various other indicators. The results indicated disagreement among the deployed homogeneity indicators. The global GSD revealed that 11 (39%) of the 28 HEGs are homogenous. In SAMI CoP criteria, only seven HEGs were homogenous. The SAMI CoP and GSD indicators were in agreement for homogeneity in only four (14%) of the HEGs. Utilizing the *R*-ratio with 90% showed just one homogenous HEG, whereas the *R*-ratio with 95% of the data, demonstrated no homogenous HEG. Only one HEG indicated agreement of homogeneity across GSD and *R*-ratio (90%). A total of 14 HEGs (50%) had a perfect agreement of non-homogeneity across all indicators.

A total of 14 HEGs that were non-homogenous in both SAMI and GSD criteria in Table 3 were selected, for the assessment of homogeneity using alternative sub-grouping by job titles (Table 4). After restricting the HEGs to those with at least two sub-groups as job titles that have three or more measurements, eight HEGs were analysed. The job titles that were not included are provided in Supplementary Materials, available at *Annals of Work Exposures and Health* online. HEG C had the highest number of job titles, followed by B and D. The global GSD and SAMI CoP criteria indicated 9 and 5 out of 10 sub-groups by job titles were homogenous in HEG B, respectively.

The comparison of the individual sub-group's AM and 90th percentiles with that of the parent HEGs showed that 5 out of 10 sub-groups had their AM above that of HEG B, and 2 sub-groups had their 90th percentiles value above that of the same HEG. In HEG C, GSD and SAMI CoP each indicated four out of nine sub-groups were homogenous of which three were in agreement for both criteria. Also, four out of the nine sub-groups of job titles had their AM and 90th percentiles higher than for the HEG. Of the remaining six HEGs, SAMI indicated homogeneity in at least one sub-group for three HEGs while the GSD did not at all. Sub-groups in HEG D, F, G, and K had at least one of their AM and 90th percentile values above that of the parent HEG. The AM and 90th percentiles of the sub-groups were above that of the parent values in HEG L. HEG N had lower AM and 90th percentiles than its constituent sub-groups.

Discussion

Appropriate grouping and categorization of workers according to similar exposure are key to improving the identification of highly exposed workers. This study

Table 2. Comparison of P90 percentiles for SAMI CoP and P95 percentiles for CEN/AIHA approaches for compliance testing relative to the OEL for HEGs (mg m^{-3}).

HEGs	Test criteria								
	SAMI CoP					CEN and AIHA		Compliance	
	n	AM	GM	P90	Compliance	P95 (95% CrI)	P95 ≥ OEL (%)[a]	CEN: 70% UCL	AIHA: 95% UCL
HEG A	39	2.19	1.11	5.32	No	10.80 (6.02–23.30)	100	No	No
HEG B	52	1.07	0.60	2.15	No	4.23 (2.72–7.52)	100	No	No
HEG C	64	1.34	0.59	3.39	No	7.77 (4.62–15.10)	100	No	No
HEG D	30	1.14	0.56	2.41	No	5.37 (2.86–13.30)	100	No	No
HEG E	83	1.15	0.67	2.50	No	3.92 (2.86–5.81)	100	No	No
HEG F	17	0.86	0.37	2.74	No	4.14 (1.74–15.8)	94.50	No	No
HEG G	15	0.72	0.43	**1.62**	**Yes**	2.98 (1.43–9.60)	83.50	No	No
HEG H	28	2.16	1.70	3.91	No	6.37 (4.33–11.30)	100	No	No
HEG I	32	1.02	0.57	**1.71**	**Yes**	3.84 (2.26–8.08)	99.50	No	No
HEG J	59	1.70	1.33	3.66	No	4.65 (3.57–6.54)	100	No	No
HEG K	11	0.67	0.24	**1.48**	**Yes**	3.52 (1.13–23.00)	80.20	No	No
HEG L	13	0.66	0.41	**1.35**	**Yes**	3.55 (1.50–14.60)	88.10	No	No
HEG M	13	1.81	1.30	2.78	No	5.01 (2.90–12.60)	100	No	No
HEG N	9	0.58	0.21	2.19	No	2.68 (0.81–20.60)	66.10	No	No
HEG O	34	2.08	1.52	4.08	No	6.33 (4.31–10.90)	100	No	No
HEG P	42	2.15	1.05	4.36	No	13.8 (8.00–27.20)	100	No	No
HEG Q	19	0.95	0.75	**1.76**	**Yes**	2.92 (1.83–6.03)	93.50	No	No
HEG R	18	1.18	0.20	4.25	No	8.26 (2.31–59.90)	98.7	No	No
HEG S	32	1.87	1.45	3.10	No	4.84 (3.47–7.71)	100	No	No
HEG T	17	2.00	1.57	3.99	No	6.18 (3.74–13.70)	100	No	No
HEG U	17	1.13	0.69	2.24	No	5.44 (2.59–17.40)	99.80	No	No
HEG V	17	0.80	0.24	**1.95**	**Yes**	5.54 (1.81–31.00)	96	No	No
HEG W	9	1.27	0.82	3.34	No	4.41 (2.03–18.90)	97.70	No	No
HEG X	14	1.88	1.02	4.67	No	7.65 (3.47–27.40)	100	No	No
HEG Y	10	0.85	0.37	2.79	No	3.81 (1.36–22.80)	86.70	No	No
HEG Z	10	0.76	0.47	2.01	No	2.76 (1.24–41.60)	75.80	No	No
HEG ZA	6	0.72	0.68	**1.26**	**Yes**	**1.41 (0.93–3.91)**	17.80	**Yes**	No
HEG ZB	18	1.42	0.75	4.31	No	5.32 (2.67–15.30)	99.90	No	No

CrI, credible interval; *n*, sample size; P90, 90th percentile; P95, 95th percentile; SAMI CoP, the P90 is ≤2 mg m^{-3}. Note: Bold values are representing compliance.
[a]Probability of overexposure.

Table 2. Comparison of P90 percentiles for SAMI CoP and P95 percentiles for CEN/AIHA approaches for compliance testing relative to the OEL for HEGs (mg m^{-3}).

HEGs	Test criteria								
	SAMI CoP					CEN and AIHA		Compliance	
	n	AM	GM	P90	Compliance	P95 (95% CrI)	P95 ≥ OEL (%)[a]	CEN: 70% UCL	AIHA: 95% UCL
HEG A	39	2.19	1.11	5.32	No	10.80 (6.02–23.30)	100	No	No
HEG B	52	1.07	0.60	2.15	No	4.23 (2.72–7.52)	100	No	No
HEG C	64	1.34	0.59	3.39	No	7.77 (4.62–15.10)	100	No	No
HEG D	30	1.14	0.56	2.41	No	5.37 (2.86–13.30)	100	No	No
HEG E	83	1.15	0.67	2.50	No	3.92 (2.86–5.81)	100	No	No
HEG F	17	0.86	0.37	2.74	No	4.14 (1.74–15.8)	94.50	No	No
HEG G	15	0.72	0.43	1.62	Yes	2.98 (1.43–9.60)	83.50	No	No
HEG H	28	2.16	1.70	3.91	No	6.37 (4.33–11.30)	100	No	No
HEG I	32	1.02	0.57	1.71	Yes	3.84 (2.26–8.08)	99.50	No	No
HEG J	59	1.70	1.33	3.66	No	4.65 (3.57–6.54)	100	No	No
HEG K	11	0.67	0.24	1.48	Yes	3.52 (1.13–23.00)	80.20	No	No
HEG L	13	0.66	0.41	1.35	Yes	3.55 (1.50–14.60)	88.10	No	No
HEG M	13	1.81	1.30	2.78	No	5.01 (2.90–12.60)	100	No	No
HEG N	9	0.58	0.21	2.19	No	2.68 (0.81–20.60)	66.10	No	No
HEG O	34	2.08	1.52	4.08	No	6.33 (4.31–10.90)	100	No	No
HEG P	42	2.15	1.05	4.36	No	13.8 (8.00–27.20)	100	No	No
HEG Q	19	0.95	0.75	1.76	Yes	2.92 (1.83–6.03)	93.50	No	No
HEG R	18	1.18	0.20	4.25	No	8.26 (2.31–59.90)	98.7	No	No
HEG S	32	1.87	1.45	3.10	No	4.84 (3.47–7.71)	100	No	No
HEG T	17	2.00	1.57	3.99	No	6.18 (3.74–13.70)	100	No	No
HEG U	17	1.13	0.69	2.24	No	5.44 (2.59–17.40)	99.80	No	No
HEG V	17	0.80	0.24	1.95	Yes	5.54 (1.81–31.00)	96	No	No
HEG W	9	1.27	0.82	3.34	No	4.41 (2.03–18.90)	97.70	No	No
HEG X	14	1.88	1.02	4.67	No	7.65 (3.47–27.40)	100	No	No
HEG Y	10	0.85	0.37	2.79	No	3.81 (1.36–22.80)	86.70	No	No
HEG Z	10	0.76	0.47	2.01	No	2.76 (1.24–11.60)	75.80	No	No
HEG ZA	6	0.72	0.68	1.26	Yes	1.41 (0.93–3.91)	17.80	Yes	No
HEG ZB	18	1.42	0.75	4.31	No	5.32 (2.67–15.30)	99.90	No	No

CrI, credible interval; *n*, sample size; P90, 90th percentile; P95, 95th percentile; SAMI CoP, the P90 is ≤2 mg m^{-3}. Note: Bold values are representing compliance.
[a]Probability of overexposure.

Table 3. Assessment of homogeneity of HEGs (mg m^{-3}) according to various indicators (GSDs, *R*-ratio, AM, and P90) of homogeneity and their acceptability criteria.

	HEGs																											
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	ZA	ZB
n	39	52	64	30	83	17	15	28	32	59	11	13	13	9	34	42	19	18	32	17	17	17	9	14	10	10	6	18
N	9	13	15	13	17	6	4	10	6	5	6	5	6	4	2	2	2	6	2	1	9	9	5	8	8	4	1	2
SAMI CoP																												
AM	2.2	1.1	1.3	1.1	1.2	0.9	0.7	2.1	1.02	1.7	0.7	0.7	1.8	0.5	2.1	2.2	0.9	1.2	1.9	2.0	1.1	0.8	1.2	1.8	0.8	0.7	0.7	1.4
P90	5.3	2.2	3.4	2.4	2.5	2.7	1.6	3.9	1.71	3.7	1.5	1.3	2.7	2.1	4.1	4.4	1.8	4.2	3.1	3.9	2.2	1.9	3.3	4.6	2.7	2.0	1.2	4.3
Homogeneity decision	Yes	No	No	No	No	No	No	Yes	Yes	No	No	No	No	No	Yes	Yes	Yes	No	No	Yes	No	No	No	No	No	No	No	No
Indicators of homogeneity																												
The global GSD	4.0	3.2	4.8	3.9	2.9	4.5	3.3	2.2	3.2	2.1	5.5	3.8	2.2	5.3	2.4	4.9	2.2	10.8	2.1	2.3	3.5	7.3	2.7	3.4	4.3	2.9	1.4	3.3
Homogeneity decision using GSD	No	No	No	No	Yes	No	No	Yes	No	Yes	No	No	Yes	No	Yes	No	Yes	No	Yes	Yes	No	No	Yes	No	No	Yes	Yes	No
The *R*-ratio (*R* = 95%)[a]	3.5	13	4.9	7.2	2.3	3.5	3.8	4.9	2.8	4.6	3.9	3.6	4.5	8.7	3.3	4.5	3.4	4	3.6	5.1	4.8	5.4	4.1	3.1	5.2	5.7	4.9	3.8
Homogeneity decision using *R* = 95%	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
The *R*-ratio (*R* = 90%)[a]	2.9	8.8	3.7	5.3	2	2.9	3.1	3.8	2.4	3.6	3.3	3	3.5	6.1	2.7	3.5	2.8	3.2	2.9	3.9	3.7	4.1	3.3	2.6	4	4.3	3.8	3.0
Homogeneity decision using *R* = 90%	No	No	No	No	Yes	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No

n, sample size; *N*, number of job titles; P90, 90th percentile; SAMI CoP, both the AM and 90th percentile should fall within the same exposure band for homogeneity. Note: Bold values are representing homogeneity. OEL for coal dust ≤2 mg m^{-3}; SAMI CoP OEL definition exposure band A: ≥OEL, B: ≥0.5OEL and <OEL, C: ≥0.1OEL and <0.5OEL, D: <0.1% OEL.

[a]Probability that *R*-ratio is ≤2 indicates moderate between job titles variability.

Table 4. The assessment of homogeneity of job titles within each HEG using GSD, AM, and P90 percentiles.

	HEG B (*N* = 10)				HEG C (*N* = 9)				HEG D (*N* = 3)				HEG F (*N* = 4)			
	n	AM	90th percentile	GSD	*n*	AM	90th percentile	GSD	*n*	AM	90th percentile	GSD	*n*	AM	90th percentile	GSD
HEG		1.10	2.20	3.20		1.30	3.40	4.80		1.10	2.40	3.90		0.90	2.70	4.50
Job titles																
Bricklayer	3	0.20	0.45	2.74	3	0.66	1.11	2.49	0				0			
CM operator	0				0				0				4	1.10	3.01	5.60
Conveyor belt attendant	4	0.61	0.87	2.08	0				0				0			
Emico driver	7	1.09	2.14	2.54	0				0							
Electrician	5	1.14	1.68	2.34	8	1.00	2.35	2.17	5	2.45	5.39	2.10	0			
Face Boss	5	1.47	2.15	2.24	8	1.22	3.39	3.93	4	1.62	3.86	3.22	5	1.11	2.74	4.85
Fitter	4	1.12	2.15	2.31	3	2.43	3.79	1.57	0				3	0.34	0.88	6.65
Mining operator					6	1.42	4.33	4.20	0				0			
Instrument mechanic	3	0.36	1.01	8.11	4	1.02	3.01	3.03	0				0			
Pump attendant	3	0.24	0.29	1.28	0								0			
Roof bolt operator	6	2.18	8.72	2.99	11	2.05	4.24	5.36	0				0			
Safety officer	0				5	3.12	7.61	2.34	0				3	0.73	1.60	4.53
Shuttle car operator	6	1.78	3.12	2.55	6	0.88	2.77	5.34	6	1.27	2.42	2.67	0			

	HEG G (*N* = 3)				HEG K (*N* = 2)				HEG L (*N* = 2)				HEG N (*N* = 2)			
	n	AM	90th percentile	GSD	*n*	AM	90th percentile	GSD	*n*	AM	90th percentile	GSD	*n*	AM	90th percentile	GSD
HEG		0.70	1.60	3.30		0.70	1.50	5.50		0.70	1.30	3.80		0.50	2.10	5.30
Job titles																
Belts man	4	0.47	1.52	4.52	0				4	0.73	1.35	1.64	0			
Bricklayer	0				3	1.62	3.30	7.64	0				0			
Conveyor belt attendant	3	0.72	0.91	1.69	0				0				0			
Electrician	0				3	0.34	0.70	9.59	0				0			
LHD operator	0				0				0				3	0.15	0.37	4.37
Roof bolt operator	0				0				0				3	0.25	0.55	4.32
Tractor driver	4	0.76	1.62	4.00	0				4	0.77	1.97	10.50	0			

n, sample size; *N*, number of job titles. Note: Bold values are homogenous according to the two methods. SAMICoP OEL definition exposure band A: ≥ OEL, B: ≥0.5 OEL and <OEL, C: ≥0.1OEL and <0.5OEL, D: <0.1% OEL.

investigated compliance according to the SAMI CoP approach, using the 90th percentile as exceedance criterion, and compared this with the 95th percentile exceedance criterion in the CEN/BOHS-NVvA and AIHA approaches. The CEN and AIHA approaches assume lognormal distribution of the parent population, which could improve the interpretation of occupational hygiene exposure data for HEG (Lavoué *et al.*, 2019). There was an agreement of non-compliance in 21 out of 28 HEGs across all approaches. Seven HEGs were compliant according to the SAMI CoP, and only one of the HEGs was considered to be compliant in agreement with CEN. The HEG that was compliant in both approaches had a small sample size (six), therefore in lognormal distributions, it is less likely that a sample in the high tail was collected, this possibly contributed to compliance of the HEG according to the CEN approach. The compliance of a HEG to the OEL represents whether the coal dust exposure level is within the acceptable value. In the SAMI CoP, it answers the question of the exposure being below the OEL (acceptable) or above the OEL (not acceptable), while the CEN and AIHA rely on the probability of overexposure to establish how likely or not likely the point estimate exceeds the OEL (uncertainty). Considering the uncertainty is important because it might arise from how the HEG was constituted, incorrect information or measurement if the exposures were measured continuously or intermittently, and sampling errors (Waters *et al.*, 2015). Our findings indicate that the CEN and AIHA approaches are more conservative or stricter than the SAMI CoP. The findings showed that SAMI CoP for compliance probably underestimated exposure and potentially misclassified workers' overexposure risk when compared with the CEN and AIHA. Also, the high variability due to different job titles defined as belonging to the same HEG might have caused the disagreement in the estimation of compliance by the two methods. Establishing a HEG will require tedious consideration of information which includes the survey design, processes, materials or tools, and work strategy for appropriate compliance testing using the OEL (BOHS-NVvA, 2011). A total of 27 HEGs were in agreement for non-compliance with both CEN and AIHA. This indicates that despite the difference in probability of uncertainty or UCL, the approaches are similar in establishing compliance of the analysed HEGs.

This study further retrospectively assessed the homogeneity of groups of workers deploying SAMI CoP guidelines and CEN, BOHS, and AIHA strategies. Our results (Table 3) show that deploying the GSD and SAMI CoP indicators for homogeneity, 11 and 7 HEGs were homogenous, respectively, but the 2 methods were in agreement in 4 HEGs only. To recall, the SAMI CoP stipulates that the AM and the 90th percentile parameter estimates must fall within the same exposure band to meet the criterion for defining homogeneity. However, an exposure distribution with a long tail, i.e. a lognormal distribution, may have a very wide bandwidth, this could make the estimates not lie within the same exposure band to indicate homogeneity with the SAMI CoP approach. This further would imply that it is difficult to demonstrate homogeneity in data with very skewed distribution using the GSD and SAMI CoP which only assumes a normal distribution of the data, yet the data could be very skewed and that a small fraction of the data would account for the AM and P90 percentiles. External factors could have influenced the variability of exposure distribution of these HEGs, these may include the number of measurements and number of workers which could have had a variable effect on exposure variability over time where shorter periods may have a similar set of conditions and work practices than over a long time. This affects the underlying distribution of exposure (Buringh and Lanting, 1991; Roach, 1991).

The *R*-ratio (95%) values were all above 2, indicating that none of the HEGs was homogenous unlike with the GSD criterion. The disagreement could have been caused by the high sensitivity of the *R*-ratio indicator of homogeneity for moderate variability in exposures. Therefore, the ratio of 2 or less might be too conservative since half of the HEGs were non-homogenous across all the various indicators. The global GSD is also sensitive to sample size as lower sample sizes can lead to the underestimation of the GSD (Buringh and Lanting, 1991), caused by skewed data from a small sample size as the true distribution of the parent data is unknown.

The assessment of homogeneity in an alternative sub-grouping by using job titles showed that the majority of the job titles are homogenous even if their parent HEGs were not (Table 4). This finding suggests that grouping of workers by job titles would require smaller sample sizes, compared with the larger sample size needed for a more heterogeneous exposure group. In this study, we included job titles with three or more observations to capture uncertainty resulting from an observation with extremely high value or extremely low values. According to the SAMI CoP, a HEG is composed of different job titles and work activities, therefore, sampling of a proportion and representative number of workers is difficult because exposure level might vary by job titles. Therefore, total sampling of workers with a similar job title (within a work area) could be cost-effective as direct identification and control of exposure compared with the required sample size of 5 or 5% whichever is greater as stipulated in the SAMI CoP guidelines for the sampling of HEGs. Comparison of each HEG's and job title's AM and the 90th percentile revealed that some job titles have a higher exposure level than their parent HEG. This implies that strategic intervention on risk assessment and control measures would be ignored if only a HEG as constituted by SAMI CoP is considered and not its constituting job titles. However, a previous study reported that grouping workers by job titles do not always lead to reduced variability or uniform exposure (Rappaport *et al.*, 1993). Different job titles might have been affected by a different set of conditions and activities which could have led to increased exposure variability (Buringh and Lanting, 1991). According to Mulhausen and Damiano (2006), HEG or SEG is defined as workers who have the same exposure profile since they use similar materials and process, and perform similar tasks. In the present study, our results demonstrated that even if workers doing similar jobs as suggested by their job title are grouped, sometimes they will not have uniform coal dust exposure distribution. Specific job tasks and time of exposure can affect exposure levels which may result in increased variation of exposure level within and between workers of the same job titles (Burdorf, 2005). Kromhout *et al.* (1993) stated that even grouping strategies based on job titles could still have variable exposure profiles.

One of the limitations of our study is that the data were not repeatedly collected this makes it difficult to use other methods such as within-worker correlation, one-way random effect ANOVA which could improve assessment of variability given several measurements of the same worker in a specific job title could have been taken over time. The present study evaluated the prospective grouping of workers into HEGs which were based on a common work area where several job titles within the same place were present. As reported by Rappaport *et al.* (1995), the observational grouping of workers into HEGs should comprise workers doing the same job, in the same location to have probable uniform exposure distribution and this should be incorporated with statistical sampling analysis which can incorporate variance of the exposure for an informed decision on risk management and control.

Conclusions

Compliance to the OEL was demonstrated more by the SAMI CoP than by the CEN/BOHS-NVvA and AIHA approaches. This may imply that SAMI CoP underestimated overexposure. The GSD criterion (GSD <3) indicated homogeneity in more HEGs than the SAMI CoP approach. Only four HEGs indicated homogeneity from both approaches. Alternative sub-grouping by job titles within non-homogenous HEGs showed that many job titles were homogenous and some had higher exposure distribution than their parent HEGs. Sub-grouping according to job title tends to have more agreement of homogeneity when using the GSD threshold and the SAMI CoP.

Therefore, it is recommended to adjust the present SAMI CoP criterion for compliance testing by acceptance of the 95th percentile of the lognormal exposure data. The future constitution of HEGs can be improved by using job titles and retrospective assessment of exposure distribution by using job title within a HEG (as defined by SAMI) can reveal high exposure job titles.

Supplementary Data

Supplementary data are available at *Annals of Work Exposures and Health* online.

Acknowledgements

Our gratitude goes to the coal mining company in South Africa for granting the data, in particular, Salome Makgato for giving us a full description of the data collection procedures. We thank the National Institute for Occupational Health, National Health Laboratory Services for supporting the first author.

Conflict of interest

The authors declare no conflict of interest relating to the material presented in this article. Its contents, including any opinions and/or conclusions expressed, are solely those of the authors.

References

Banerjee S, Ramachandran G, Vadali M *et al.* (2014) Bayesian hierarchical framework for occupational hygiene decision making. *Ann Occup Hyg*; 58: 1079–93.

BOHS-NVvA. (2011) Testing compliance with occupational exposure limits for airborne substances. Available at https://www.arbeidshygiene.nl/- uploads/files/insite/2011-12-bohs-nvva-sampling-strategy-guidance.pdf. Accessed 4 January 2020.

Bullock WH, Ignacio JS. (2008) *A strategy for assessing and managing occupational exposures*. 3rd edn. Fairfax, VA: AIHA Press.

Burdorf A. (2005) Identification of determinants of exposure: consequences for measurement and control strategies. *Occup Environ Med*; 62: 344–50.
Buringh E, Lanting R. (1991) Exposure variability in the workplace: its implications for the assessment of compliance. *Am Ind Hyg Assoc J*; 52: 6–13.
CEN. (2018) *BS EN 689:2018 Workplace exposure. Measurement of exposure by inhalation to chemical agents. Strategy for testing compliance with occupational exposure limit values*. Brussels, Belgium: British Standards Institution.
Department of Energy. (2016) 2016 Energy price report: coal. Available at http://www.energy.gov.za/files/media/explained/Energy-Price-Report-2016.pdf. Accessed 4 January 2020.
Department of Mineral Resources. (2018) Guideline for compilation of mandatory code of practice for an occupational health programme on personal exposure to airborne pollutants. Available at https://www.mhsc.org.za/sites/default/files/public/legislation_document/Airborne%20Pollutants%20Guideline_0.pdf. Accessed 18 August 2020.
Esmen N. (1979) Retrospective industrial hygiene surveys. *Am Ind Hyg Assoc J*; 40: 58–65.
Finkelman RB, Orem W, Castranova V *et al.* (2002) Health impacts of coal and coal use: possible solutions. *Int J Coal Geol*; 50: 425–43.
Grové T, Van Dyk T, Franken A *et al.* (2014) The evaluation and quantification of respirable coal and silica dust concentrations: A task-based approach. *J Occup Environ Hyg*; 11: 406–414.
Guild RI, Ehrlich JR, Johnston MH *et al.* (2001) Occupational health practice in the South African mining industry. Johannesburg, South Africa: The Safety in Mines Research Advisory Committee (SIMRAC). Available at https://mhsc.org.za/sites/default/files/public/publications/OH%20Handbook.pdf. Accessed 20 August 2020.
Hawkins NC, Norwood SK, Rock JC. (1991) A strategy for occupational exposure assessment. *Am Ind Hyg Assoc J*; 160.
INRS. (2005) *Metropol A1 sheet: exposure assessment strategy and comparison with limit values*. Vandoeuvre, France: National Research and Security Institute.
INRS. (2008) *Metropol A3 sheet: help in diagnosing exceeding/not exceeding the OEL in the assessment of occupational exposure*. Paris, France: National Institute for Research and Security (Metropolitan Area 3A).
Kromhout H, Symanski E, Rappaport SM. (1993) A comprehensive evaluation of within- and between-worker components of occupational exposure to chemical agents. *Ann Occup Hyg*; 37: 253–70.
Lavoué J, Joseph L, Knott P *et al.* (2019) Expostats: a Bayesian toolkit to aid the interpretation of occupational exposure measurements. *Ann Work Expo Health*; 63: 267–79.
McNally K, Warren N, Fransman W *et al.* (2014) Advanced REACH Tool: a Bayesian model for occupational exposure assessment. *Ann Occup Hyg*; 58: 551–65.
Mineral Council of South Africa. (2019) Coal: key facts and figures. Available at https://www.mineralscouncil.org.za/sa-mining/coal. Accessed 9 April 2020.
Mulhausen JR, Damiano J. (1998) *A strategy for assessing and managing occupational exposures*. Fairfax, VA: AIHA Press.
Mulhausen J, Damiano J. (2006) Establishing similar exposed groups. In Bullock WH, Ignacio JS, editors. *A strategy for assessing and managing occupational exposures*. 3rd edn. Fairfax, VA: American Industrial Hygiene Association. pp. 33–44. ISBN 1-931504-69-4.
NIOSH. (1994) *Particulates not otherwise regulated, total: Method 0500*. NIOSH manual of analytical methods. 4th edn. Cincinnati, OH: United States Department of Health and Human Services, Centers for Disease Control and Prevention, DHHS (NIOSH) Publication No. 94-113. Avaliable at www.cdc.gov/niosh/nmam/. Accessed 15 July 2020.
Petsonk EL, Rose C, Cohen R. (2013) Coal mine dust lung disease. New lessons from an old exposure. *Am J Respir Crit Care Med*; 187: 1178–85.
Pretorius CJ, Dekker J, Letsoalo S *et al.* (2015) Develop and operation manual for the implementation of the revised code of practice for assessment of personal exposure to airborne pollutants. Available at https://mhsc.org.za/sites/default/files/public/research_documents/MHSC0311415%20Report.Personal%20exposure%20to%20airborne%20pollutants.pdf. Accessed 10 August 2017.
Rappaport SM. (2001) Interpreting levels of exposures to chemical agents. In *Patty's industrial hygiene*. New York, NY: John Wiley & Sons, Inc. pp. 679–745.
Rappaport SM, Kromhout H, Symanski E. (1993) Variation of exposure between workers in homogeneous exposure groups. *Am Ind Hyg Assoc J*; 54: 654–62.
Rappaport SM, Lyles RH, Kupper LL. (1995) An exposure-assessments strategy accounting for within- and between-worker sources of variability. *Ann Occup Hyg*; 39: 469–95.
Roach SA. (1991) Alternative ways of monitoring exposure. In Rappaport SM, Smith TJ, editors. *Exposure assessment for epidemiology and hazard control*. Chelsea, MI: Lewis Publishers. pp. 3–20.
South Africa. (2006) *Government notice: amendment to the regulations in respect of occupational hygiene, proclamation no. R989, Regulation Gazette no. 29276*.
Stanton DW, Belle BK, Dekker KJ *et al.* (2006) *South African mining industry best practice on the prevention of silicosis*. Gauteng, South Africa: Mine Health and Safety Council.
Waters M, McKernan L, Maier A *et al.* (2015) Exposure estimation and interpretation of occupational risk: enhanced information for the occupational risk manager. *J Occup Environ Hyg*; 12 (Suppl. 1): S99–111.

Appendix I: Bayesian Hierarchical Modelling of Historical Data of the South African Coal Mining Industry for Compliance

International Journal of *Environmental Research and Public Health*

MDPI

Article

Bayesian Hierarchical Modelling of Historical Data of the South African Coal Mining Industry for Compliance Testing

Felix Made [1,2,*], Ngianga-Bakwin Kandala [1,3,4] and Derk Brouwer [1]

1 School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2193, South Africa; ngianga-bakwin.kandala@wits.ac.za (N.-B.K.); derk.brouwer@wits.ac.za (D.B.)
2 Global Biostatistics and Programming, Pharmaceutical Product Development, Thermo Fisher Scientific, Woodmead, Johannesburg 2191, South Africa
3 Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK
4 Département de la Santé Communautaire, Institut Supérieur des Techniques Médicales de Kinshasa, Kinshasa XI, Mont Ngafula, Kinshasa B.P. 774, Congo
* Correspondence: felix.made@ppd.com

Abstract: Bayesian hierarchical framework for exposure data compliance testing is highly recommended in occupational hygiene. However, it has not been used for coal dust exposure compliance testing in South Africa (SA). The Bayesian analysis incorporates prior information, which increases solid decision making regarding risk management. This study compared the posterior 95th percentile (P95) of the Bayesian non-informative and informative prior from historical data relative to the occupational exposure limit (OEL) and exposure categories, and the South African Mining Industry Code of Practice (SAMI CoP) approach. A total of nine homogenous exposure groups (HEGs) with a combined 243 coal mine workers' coal dust exposure data were included in this study. Bayesian framework with Markov chain Monte Carlo (MCMC) simulation to draw a full P95 posterior distribution relative to the OEL was used to investigate compliance. We obtained prior information from historical data and employed non-informative prior distribution to generate the posterior findings. The findings were compared to the SAMI CoP. The SAMI CoP 90th percentile (P90) indicated that one HEG was compliant (below the OEL), while none of the HEGs in the Bayesian methods were compliant. The analysis using non-informative prior indicated a higher variability of exposure than the informative prior according to the posterior GSD. The median P95 from the non-informative prior were slightly lower with wider 95% credible intervals (CrI) than the informative prior. All the HEGs in both Bayesian approaches were in exposure category four (poorly controlled), with the posterior probabilities slightly lower in the non-informative uniform prior distribution. All the methods mainly indicated non-compliance from the HEGs. The non-informative prior, however, showed a possible potential of allocating HEGs to a lower exposure category, but with high uncertainty compared to the informative prior distribution from historical data. Bayesian statistics with informative prior derived from historical data should be highly encouraged in coal dust overexposure assessments in South Africa for correct decision making.

Keywords: non-informative prior; informative prior; 95th percentile; lognormal distributions; exposure category

check for updates

Citation: Made, F.; Kandala, N.-B.; Brouwer, D. Bayesian Hierarchical Modelling of Historical Data of the South African Coal Mining Industry for Compliance Testing. *Int. J. Environ. Res. Public Health* **2022**, *19*, 4442. https://doi.org/10.3390/ijerph19084442

Academic Editor: Luigi Vimercati

Received: 20 January 2022
Accepted: 1 April 2022
Published: 7 April 2022

1. Introduction

South Africa (SA) is one of the largest producers of coal in the world, with an estimated 86,000 workers according to the Mineral Council of South Africa, 2019 [1]. During coal mining, coal dust is generated, and when inhaled, it can cause coal mine lung disease (CMDLD) [2,3]. To lower the risk of coal dust overexposure and potentially prevent life-threatening CMDLD, the occupational exposure limit for respirable coal dust in SA is set at 2 mg/m^3 [1,4]. As part of a continuous process of monitoring overexposure and compliance, the South African Mining Industry Code of Practice (SAMI CoP) stipulates that the identification of homogenous exposure groups (HEGs) is an important proxy for the assessment of personal exposures [5]. HEGs are defined as a group of employees who have similar exposure, such that a sample can be drawn from them for predicting the exposure of all remaining workers [5,6]. In the SAMI CoP, HEGs are constituted by a stepwise process [5,6]; in step one, the mine is subdivided into ventilation districts based on areas with common intake and return air. In the next step (step 2), the area is divided into activity areas (as found in coal mines). The personal exposures in each activity area are then compared to the OEL, which is the eight-hour time-weighted average (TWA8h) coal dust concentration to which almost all workers may be repeatedly exposed without any adverse health effects whatsoever. Each HEG is assigned to exposure classification categories, which are associated with their distance to the OEL [4] (Table 1).

Table 1. The exposure classification categories for classifying the P90 of exposure relative to the OEL and sampling strategy of HEGs in the SAMI CoP.

Category	Description	Statistical Illustration	Exposure Profile	Minimum Frequency
1	Exposures less than 10% of the OEL 10% of the time	P90 < 0.1% OEL	Very highly controlled	No sampling plan for this category. Measurement results that are below 10% of the OEL will be reported under this category
2	Exposures exceed 10% of the OEL and less than 50% of the OEL 10% of the time	P90 ≥ 0.1 OEL < 0.5 OEL	Highly controlled	Sample 5% of employees within a HEG on an annual basis with a minimum of 5 samples per HEG, whichever is greater.
3	Exposures exceed 50% of the OEL and less than OEL 10% of the time	P90 ≥ 0.5 OEL < OEL	Adequately controlled	Sample 5% of employees within a HEG on a 6-monthly basis with a minimum of 5 samples per HEG, whichever is greater
4	Exposures exceed the OEL 10% of the time	P90 ≥ OEL	Poorly controlled	Sample 5% of employees within a HEG on a 3-monthly basis with a minimum of 5 samples per HEG, whichever is greater

The sampling size of each HEG is equal to 5 or 5% of the HEG population, whichever is greater, whereas the sampling frequency is determined by the exposure classification category. The results of each sampling campaign are evaluated independently of the previous data (i.e., historical data) for compliance. This current practice in South Africa shows that HEGs are too heterogeneous concerning exposure levels, resulting in an overestimation or underestimation of exposure for individual workers [7]. Yet, in good practice, previous sampling results can be used to update current data for the categorization of HEGs by using a Bayesian framework to elucidate prior information from the historical data [8]. The framework with broad use of informative prior is highly encouraged in occupational hygiene because it accommodates historical data into the empirical measurement of monitoring exposure data for accurate exposure grouping [8–12].

For compliance testing, the SAMI CoP approach assumes that the sample data are normally distributed, and the 90th percentile of exposure data (P90) should be below the OEL. Overexposure according to the European Standardization Committee (CEN) [13] and the British and Dutch guidelines is that the 95th percentile (P95) of the lognormal exposure distribution should be below the OEL [14]. It is important to note that SAMI CoP is based solely on current data, yet the incorporation of historical data in the Bayesian framework could improve the identification of overexposed HEGs as future risk management of exposure is based on the exposure profile of the data at hand. The SAMI CoP approach is based solely on a point estimate of the P90 of the current data and does not consider uncertainty surrounding the estimate. The Bayesian framework uses the credible interval to describe uncertainty surrounding a parameter. For example, the credible interval is interpreted as the probability of an estimate being found between a certain range, given the data [15]. The SAMI CoP approach would mean that similar exposure in different areas needs to be done repeatedly, but in a Bayesian sense, this can be achieved naturally by using a sample result as a prior distribution from common population distribution [16]. The Bayesian inference is exact and easily understood by anyone. For example, the probability that a person is overexposed to coal dust is 95%. Currently, the Bayesian framework has not been applied for exposure to coal dust in the mining industry, and no study has emphasized the use of historical data in determining overexposure in a routine occupational hygiene assessment. Therefore, the first objective of this study was to compare the posterior P95 of the non-informative and informative application of the Bayesian framework and SAMI CoP relative to the OEL. The second objective was to compare the grouping of posterior probabilities of the P95 exposure according to the South African occupational exposure categories between the two Bayesian approaches. The present paper describes the development of an informative prior, taken from the historical coal dust exposure data, and combines this with the present exposure data to achieve posterior distributions in a Bayesian framework. The decision-making on exposure risk management according to SAMI exposure categories is compared to posteriors derived from a non-informative prior and the strengths and limitations are discussed.

2. Methods

2.1. *Study Design and Data Collection*

This is a cross-sectional study. Respirable coal dust exposure data were collected periodically from different geographic locations of the mines. The population included in this study were only male mine workers who were working in underground coal mines. The data were collected from workers working in HEGs within each mine.

Aligned with the SAMI CoP approach, a mixed stratified and random selection sampling frame was used, considering that "either 5% of the workers assigned to a HEG, or a minimum of five workers should be selected for a measurement campaign" [5]. The sampling collection and sample analysis methods have been described in a previous paper [7]. Briefly, each selected worker was issued with a size-selective cyclone with a mixed cellulose ester filter, which was attached to a dust sampling. The analysis of the cyclone's filter was done according to the standard of the National Institute for Occupation Safety and Health (NIOSH) method 0600 [17].

2.2. *Statistical Analysis*

Statistical analysis was carried out in R version 4.1.1(R Core Team, Auckland, New Zealand), using RStan and bayestestR packages [18–20]. Consistent and similar coal dust historical data of the HEG was used to update current monitoring data to produce posterior geometric mean, geometric standard deviations, P95, and the posterior probabilities of P95 exposure in each of the exposure categories. For the prior specification, we randomly selected a prior sample size of five from each HEG's historical data, as recommended from previous studies for occupational exposure assessment [21,22]. From the occupational exposure perspective, the prior sample size should be between 10% to 40% of the current data to get accurate information on the posterior distribution for decision making. Therefore, the sample size of five was used to keep the focus of the posterior distribution on the current data, as the sample size increased. This is important as in Bayesian statistics, the posterior distribution is a compromise between the information from the prior and the current data, but the distribution must be observed from the current data to a good measure as the sample size increases [15]. For the likelihood function, we took all the available current monitoring data.

2.2.1. Model Specification Using Current Monitoring Data

The model was specified by using the geometric mean (GM), given as $\exp(\mu)$, and geometric standard deviation (GSD), denoted as $\exp(\sigma)$, which are the exponents of the mean and standard deviation, after log transforming the data. The likelihood function was presented as below in Equation (1).

$$\prod_{i=1}^{n} LN\left(y_i \middle| \mu, \sigma^2\right) = \prod_{i=1}^{n} \frac{1}{y_i \sigma \sqrt{2\pi}} exp\left\{-\frac{1}{2}\frac{(logy_i - \mu)^2}{\sigma^2}\right\} \quad (1)$$

where y_i is the log-transformed current monitoring data and n is the number of observations of current monitoring data. The OEL exposure categories were added as a random variable in the model directly to produce the posterior probability distribution of the P95 to each of the categories [21,22]. From the OEL exposure categories (Table 1), the highest category was assumed to have P95 > OEL and the lowest was P95 < 10% of the OEL.

2.2.2. Model Specification Using a Non-Informative Uniform Prior Distribution

In occupational health research, uniform prior distribution is highly encouraged. A current monitoring data vector, $Y = (y_1, \ldots, \ldots, \ldots, y_n)$, where n is the sample size, with the data $Y \sim Norm(\mu\sigma^2)$, where μ represents the log of the geometric means (GM) and σ is the log of the geometric standard deviation (GSD).

Then, $\mu = LnGM \sim Unif(a_\mu b_{\sigma\mu})$, $\sigma^2 = LnGSD \sim unif(a_\sigma b_\sigma)$ where a and b are the lower and upper bounds of the prior distribution, respectively. We took inspiration from Gelman 2006 [23], where, for lower bounds, μ was indicated as 0 and for σ was $-1/2$ and the upper bounds were set to infinity.

2.2.3. Informative Prior Specification from Historical Data

For informative prior, we assumed that the log-transformed historical data with n_0 observations had sample variance $s_{y0}^2 = \sum(y_{i0} - \overline{y}_0)^2/(n_0 - 1)$. If the historical data $y_{i0} \sim \mathrm{Norm}(\mu, \sigma^2)$, then the mean of the historical data $\overline{y}_0 \sim \mathrm{Norm}(\mu, \sigma^2/n_0)$, we put μ as a random quantity and replace σ^2 with the prior estimate s_{y0}^*, so that μ takes the form as shown below in Equation (2).

$$\mu \sim \mathrm{Norm}\left(\overline{y}_0, s_{y0}^*/n_0\right) \quad (2)$$

The full conditional for μ and σ^2 was based on truncated normal prior distributions and truncated inverse gamma prior distributions, respectively [9,22]. In the truncated prior distributions, we placed bounds on μ using the suggestion of Bayesian decision analysis (BDA) from Hewett et al., 2006 [12], which was 0.005. We used 0.001 in this study to make sure it less likely affected the results, while the upper bound was allowed to vary iteratively. The upper bound was allowed to vary to avoid a prior from being unfairly skewed toward a more favourable result.

For σ^2, the lower bound from BDA was used and the upper bound was let to change iteratively.

To develop prior for σ^2, we started with the expression, $(n-1)s_{y0}^2/\sigma^2 \sim x_{n_0-1}^2$, where $x_{n_0-1}^2$ represents the chi-square distribution with $(n_0 - 1)$ degrees of freedom. We put σ^2 as the random variable, given s_{y0}^* from the historical data. Therefore, the variance is given by Equation (3).

$$\sigma^2 \sim IG\left(\frac{n_0 - 1}{2}, \frac{(n_0 - 1)s_{y0}^2}{2}\right) \quad (3)$$

For $n_0 > 1$, where IG (a, b) is an inverse gamma distribution in Equation (4) with parameter a and scale parameter b

Therefore,

$$a = (n_0 - 1)/2b = 2/(s_{y0}^2 * (n_0 - 1)) \quad (4)$$

Further details on the prior specification for μ and σ^2 and the full conditional distributions are available in Supplementary File S1 of the Supplementary Materials of Quick, 2017 [22].

2.2.4. Motivation on Bounds Based on P95

We placed the bound on the P95 as P95 < 2 × OEL with the assumption that the correct OEL exposure category for each HEG would be identified. See Supplementary File S2 of Supplementary Materials for motivation on bounds based on P95 by Quick, 2017 [22].

Markov chain Monte Carlo (MCMC) algorithms to draw full posterior conditional distribution inform of the Gibbs sampler [24] were implemented. The Gibbs sampler was applied because of its easy computational application. The Gibbs sampler samples from a conditional distribution. For example, if a given parameter has been divided into sub-parameters, the Gibbs sampler works by drawing each sub-parameter conditional on the values of all the others iteratively. The sub-parameter is updated conditional several times on the latest values of all the components of the parameter to produce the marginal posterior distribution. We used 20,000 MCMC number of iterations to draw samples from the posterior.

The posterior convergence diagnostic was carried out using the Gelman–Rubin convergence diagnostic, which compares the between and within-Markov chain variability for the model parameters to confirm whether they are stationary [25]. The between-Markov is the variance of the posterior mean of the samples, while the within-Markov is the mean of the variance in each sample. If the test statistics denoted by R-hat is ≤1.05, then convergence is achieved. The reliability of the posterior quantiles was confirmed using the bulk and tail effective sample size [26]. An effective sample size greater than 100 per chain is considered good. The convergence diagnostic test indicated an R-hat of less than 1.05 and an effective sample size of more than 100 in this study (not shown), implying convergence was achieved.

3. Results

Table 1 indicates the exposure classification of SAMI CoP. The highest category is four, which is that P90 exceeds the OEL for exposure to be classified as poorly controlled, and the lowest category is that P90 should be less than 0.1 of the OEL for highly effective control. In Table 2, a summary of the current monitoring data and corresponding historical data from nine HEGs are displayed. HEGs C and G had the highest AM of 2.42 in the current monitoring data compared to the rest of the HEGs. In the past data, HEG A had the highest AM of 2.00. HEGs D and G have the lowest AMs of all the other HEGs in current and past data, respectively. GSD indicated a high variability of exposure with the current monitoring data in HEG B, D, E, F, H, and I (GSD > 3), while in the past historical data, the variability of the exposure was high in HEG B, D, F, G, and I.

Table 2. Summary of coal dust exposure for the current monitoring data and their corresponding historical past data.

Data	Year	N	AM	SD	GM	GSD
Current data						
HEG A	2018	14	2.20	1.96	1.52	2.55
HEG B	2019	21	2.36	1.45	1.58	3.56
HEG C	2018	13	2.42	2.22	1.91	1.96
HEG D	2017	52	0.71	0.66	0.42	3.29
HEG E	2018	35	1.32	1.74	0.62	3.53
HEG F	2018	20	1.24	1.90	0.60	3.80
HEG G	2019	24	2.42	1.70	1.93	2.01
HEG H	2018	38	1.43	1.75	0.74	3.50
HEG I	2019	26	2.04	1.58	1.29	3.67

Table 2. *Cont.*

Data	Year	N	AM	SD	GM	GSD
Past data						
HEG A	2017	20	2.00	1.35	1.51	2.30
HEG B	2018	21	1.93	2.52	0.69	5.96
HEG C	2017	19	1.48	0.95	1.20	2.03
HEG D	2016	53	1.46	1.69	0.76	3.59
HEG E	2017	50	1.18	1.01	0.78	2.78
HEG F	2017	32	0.96	0.82	0.60	3.05
HEG G	2018	40	0.69	0.91	0.29	4.63
HEG H	2017	45	0.90	0.82	0.62	2.44
HEG I	2018	39	1.02	0.91	0.59	3.27

N: sample size; AM: athematic mean; SD: standard deviation; GM: geometric mean; GSD: geometric standard deviation.

Table 3 indicates the SAMI P90, median, and 95% credible intervals of the posterior GM and GSD, and the P95 percentiles for non-informative and informative prior. The SAMI CoP P90 values are much lower than the P95 in the non-informative and informative prior Bayesian approaches. The SAMI approach is the only method that showed that only one HEG (HEG D) had P90 values lower (1.62 mg/m^3) than the OEL of 2 mg/m^3. The posterior median GM indicated that all HEGs exposures were below the OEL 2 mg/m^3. There was high exposure variability in the majority of HEGs, as indicated by GSD greater than 3. Three and four of HEGs under non-informative and informative prior had less exposure variability according to the GSD. The patterns of the medians of the posterior P95 and 95% credible intervals (CrI) from Table 3 are shown in Figure 1. Generally, the median and 95% CrI were similar across all HEGs between the non-informative and informative prior. Five out of nine HEGs in the graph indicated that P95 are lower in the non-informative prior with wider 95% CrI bounds compared to the informative. Overall, there was high uncertainty in the non-informative indicated by the wider 95% CrI (also higher upper bounds) compared to the informative prior distribution.

Table 3. The median (95% credible interval (CrI)) of the posterior GM, GSD, and the P95 percentiles and the SAMI P90.

HEG	SAMI P90	Non-Informative GM	Non-Informative GSD	Non-Informative P95	Informative GM	Informative GSD	Informative P95
		Median (95% CrI)	Median (95% CrI)	Median (95% CrI)	Median (95% CrI)	Median (95% CrI)	Median (95% CrI)
HEG A	4.12	1.47 (0.86, 2.33)	2.67 (2.29, 3.25)	7.42 (4.48, 12.17)	1.56 (1.05, 2.28)	2.59 (2.28, 3.06)	7.50 (4.91, 11.85)
HEG B	4.02	1.40 (0.82, 2.13)	3.01 (2.66, 3.48)	8.67 (5.44, 12.11)	1.24 (0.72, 1.91)	3.12 (2.77, 3.59)	8.14 (5.07, 11.36)
HEG C	3.74	1.89 (1.24, 2.78)	2.33 (2.02, 2.86)	7.55 (5.04, 12.70)	1.63 (1.26, 2.08)	2.21 (1.97, 2.64)	6.03 (4.60, 8.60)
HEG D	1.62	0.42 (0.30, 0.59)	3.01 (2.73, 3.40)	2.58 (1.83, 3.92)	0.46 (0.34, 0.65)	2.99 (2.73, 3.36)	2.83 (2.73, 3.36)
HEG E	3.27	0.62 (0.41, 0.96)	3.11 (2.76, 3.63)	4.04 (2.61, 6.78)	0.62 (0.41, 0.93)	3.11 (2.78, 3.60)	4.02 (2.65, 6.57)
HEG F	2.46	0.58 (0.31, 1.02)	3.23 (2.76, 3.93)	4.01 (2.22, 7.41)	0.63 (0.38, 1.03)	3.16 (2.75, 3.76)	4.20 (2.47, 7.43)
HEG G	4.24	1.93 (1.43, 2.60)	2.34 (2.10, 2.72)	7.79 (5.70, 11.79)	1.66 (1.09, 2.39)	2.70 (2.41, 3.12)	8.53 (5.83, 12.40)
HEG H	4.02	0.74 (0.49, 1.11)	3.09 (2.76, 3.57)	4.76 (3.14, 7.66)	0.79 (0.57, 1.10)	3.01 (2.71, 3.44)	4.88 (3.36, 7.40)
HEG I	4.06	1.19 (0.73, 1.80)	3.09 (2.74, 3.54)	7.72(4.88, 10.98)	1.07 (0.66, 1.62)	3.12 (2.77, 3.59)	6.99 (4.50, 10.21)

P95: 95th percentile; CrI: credible Interval; OEL is ≤2 mg/m^3.

The comparison in the grouping of the HEGs' posterior probabilities of the P95 according to the different OEL categories (see Table 1) is presented in Table 4. In both Bayesian approaches of the prior distribution, HEG D showed the lower posterior probability of the exposure level being in category four, which is poorly controlled compared with the rest of the HEGs. All the HEGs in both prior distributions were in poorly controlled category four with more than 90% and 95% probabilities, respectively. Some of the posterior probabilities of the non-informative prior distribution, although all in category four, were slightly lower than in the informative prior.

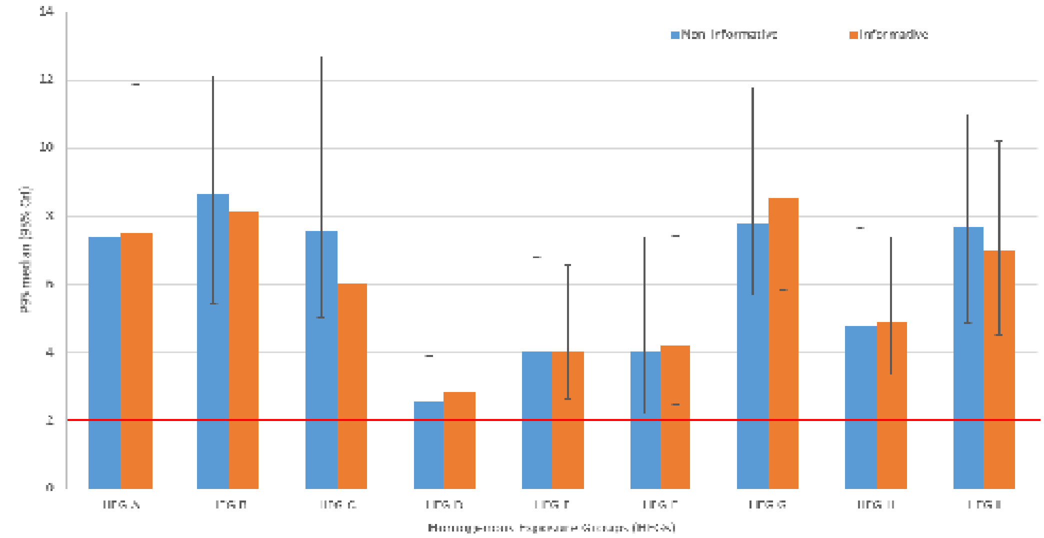

Figure 1. The comparison of the patterns of the posterior median (95% CrIs) of the P95 for non-informative and informative Bayesian framework across HEGs. The red horizontal is the SA OEL of 2 mg/m^3.

Table 4. The estimated exposure category probabilities of the non-informative and informative Bayesian frameworks for the posterior 95th percentile.

HEG	Non-Informative					Informative				
	P95	Category 1	Category 2	Category 3	Category 4	P95	Category 1	Category 2	Category 3	Category 4
HEG A	7.42	0	0	0	100%	7.50	0	00	00	100%
HEG B	8.67	0	0	0	100%	8.14	0	0	0	100%
HEG C	7.55	0			100%	6.03	0	0	0	100%
HEG D	2.58	0	0.01%	7.73%	92.26%	2.83	0	0	2.26%	97.74%
HEG E	4.04	0	0	0.10%	99.90%	4.02	0	0	0.06%	99.94%
HEG F	4.01	0	0.02%	1.07%	98.91%	4.20	0	0	0.27%	99.70%
HEG G	7.79	0	0	0	100%	8.53	0	0	0	100%
HEG H	4.76	0	0	0.01%	99.99%	4.88	0	0	0	100%
HEG I	7.72	0	0	0	100%	6.99	0	0	0	100%

4. Discussion

We used informative prior from historical data to update current monitoring data on lognormal distribution in the Bayesian framework to produce the posterior geometric mean, geometric standard deviation, and the P95. Similarly, a non-informative prior was used, and SAMI CoP was based on P90. The findings were compared. The posterior probability of the P95 percentile exposures was also grouped according to the SAMI exposure category. The use of the past data is important because decision-making on exposure risk management only based on the current data might be misleading. The weight of the historical data in the analysis is important; Symanski et al. [27] used an equal weight for both current and past data. We decided that weight should be unequal, with a small prior sample size to have less influence on the current data. This is consistent with other studies where small prior sample size was thought to produce inferential benefits when the results were compared to non-informative priors [22,28]. The Bayesian framework is also known to be robust with a small sample size [15], so even in the paucity of data, exposure risk analysis can be conducted with relative confidence.

The application of both approaches to the prior distribution indicated that the posterior estimates of GM were below the OEL. This means that the level of coal dust risk control was similar in the past and the present. However, risk mitigation and decision making regarding exposure control should not be based on the central distribution (mean/median) of the data, but on P95, which constitute at least 95% of the underlying distribution. The posterior GSDs were also quite similar across the two Bayesian prior distributions, however, those of the non-informative distributions tended to be somewhat higher. The GSD indicated high variability in the informative prior distribution. The comparison of the SA SAMI using the P90 for compliance and Bayesian prior methods showed that P90 was lower (with one HEG exposure below the OEL) compared to P95. This is consistent with our previous study, where the SAMI approach tended to underestimate overexposure risk [4]. The Bayesian approaches considered the uncertainty of overexposure not just based on a point estimate as to the SAMI CoP. All the HEGs in the Bayesian approach indicated their P95 were very high and above the OEL. The distribution of the median and P95 was similar to the non-informative and informative prior distribution (Figure 1). The majority of HEGs in the non-informative prior indicated that P95 were a bit lower and had wider 95% CrIs, indicating a high uncertainty compared with the informative prior derived from the historical data. This underscores the importance of the use of historical data in coal mining occupational exposure assessment. Decisions on overexposure risk can be made with greater confidence when historical data are brought together to update the current data, as it is only natural that they are part of the current data.

We then compared the posterior probabilities of grouping the P95 in each exposure category between the Bayesian approach from non-informative and informative prior (from historical data) distribution (Table 4). In both approaches, the highest probabilities (greater than 95%) of P95 were observed in category four of the exposure, which indicates poorly controlled exposure. None of the HEGs' posterior P95 was in the lower exposure category. From these results, the use of historical data to update current data in Bayesian statistics for occupational exposure assessment is very important, as non-informative prior tend to assign HEGs' in a lower category. This affects informed decision making with the regard to overexposure risk mitigation. Assigning HEGs to a lower category is similar to a simulation study that showed that non-informative uniform priors group the P95 probabilities in lower exposure categories [22]. The difference with the informative prior from historical data might be because of the possible use of incorrect prior for certain HEGs, resulting from a lack of adequately repeated measurements and sampling of prior data.

As seen from the above, the uncertainty to inform risk management decisions is not low and high variability of exposure is shown by the non-informative prior distribution compared to the informative prior from the historical data. Although this study showed that non-informative priors tend to locate the posterior probabilities of P95 in a lower category and increase variability, its interpretation must be taken with caution as the probability density function used to specify the prior, usually an infinite integral might yield improper posterior distributions [29]. Therefore, the specification of the non-informative uniform prior must be considered. Our findings could indicate that the decision to regard these HEGs as compliant or non-compliant should also consider the variability of the data.

The strength of this study is that the Bayesian analysis naturally allows for combining prior information from historical data with current data within a solid decision-making framework [30]. With the robustness of the findings based on even a small sample size, the Bayesian analysis provided inferences that are based conditionally on the data, which makes them exact and easily interpretable. For example, the probability of the posterior P95 being in category four (poorly control group) can be expressed quantitatively [31]. Regarding the limitations of this study, it is important to recall that HEGs used in this study are created by grouping workers based on the common air intake and return air. This might mean that the HEGs can be too heterogeneous because within a HEG there can be several job titles with different exposure variabilities [7]. As demonstrated earlier [7], HEGs tend to have high variability, which also affects their compliance to the OEL and grouping according to exposure category. From the Bayesian perspective, sometimes historical data are unavailable or are not similar and consistent to the current data, and hence it is not possible to conduct an informative prior.

5. Conclusions

It is clear from the findings that the use of the Bayesian framework with informative prior can inform concise decision making on occupational exposure risk mitigation in the coal mining industry with great confidence. Bayesian analysis from the non-informative uniform prior distribution tends to put HEGs in lower exposure categories than informative prior distribution derived from historical data. The non-informative prior findings also showed high uncertainty and variability, thus a decision on exposure risk assessments would likely be made with less confidence. This makes overexposure risk likely to be underestimated. We recommend increased use of the Bayesian framework with the use of prior information from historical data in the coal mining occupational exposure assessment. This will improve solid decision-making concerning coal dust overexposure risk and compliance.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijerph19084442/s1, Supplementary File S1: full-conditional distributions; Supplementary File S2 bounds based on P95 [32–36].

Author Contributions: Conceptualization, F.M.; methodology, F.M. and N.-B.K.; validation, D.B.; investigation, F.M. and D.B.; formal analysis, F.M.; writing—original draft preparation, F.M.; writing—review and editing, F.M., N.-B.K. and D.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) at the University of the Witwatersrand, South Africa. (Ethical clearance certificate number M181038 and date of approval was 26 October 2018).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: Our gratitude goes to the coal mining company in South Africa for granting the data, in particular, Salome Makgato from the Zibulo Colliery for giving us a full description of the data collection. We also thank Jerome Lavoué for his valuable contribution to this study. We thank the National Institute for Occupational Health and National Health Laboratory Services for supporting the first author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mineral Council of South Africa. Coal: Key Facts and Figures. 2019. Available online: https://www.mineralscouncil.org.za/sa-mining/coal (accessed on 9 April 2020).
2. Finkelman, R.B.; Orem, W.; Castranova, V.; Tatu, C.A.; Belkin, H.E.; Zheng, B.; Lerch, H.E.; Maharaj, S.V.; Bates, A.L. Health impacts of coal and coal use: Possible solutions. *Int. J. Coal Geol.* **2002**, *50*, 425–443. [CrossRef]
3. Grove, T.; Van Dyk, T.; Franken, A.; Du Plessis, J. The evaluation and quantification of respirable coal and silica dust concentrations: A task-based approach. *J. Occup. Environ. Hyg.* **2014**, *11*, 406–414. [CrossRef]
4. South Africa. *Government Notice: Amendment to the Regulations in Respect of Occupational Hygiene*; Proclamation No. R989, Regulation Gazette No. 29276; South African Government: Pretoria, South Africa, 2006.

5. Department of Mineral Resources. Guideline for Compilation of Mandatory Code of Practice for an Occupational Health Programme on Personal Exposure to Airborne Pollutants. 2018. Available online: http://www.gpwonline.co.za/Gazettes/Pages/Published-National-Government-Gazettes.aspx?p=8date (accessed on 19 August 2021).
6. Pretorius, C.J.; Dekker, J.; Letsoalo, T.; Grove, L.J. Develop and Operation Manual for the Implementation of the Revised Code of Practice for Assessment of Personal Exposure to Airborne Pollutants. 2015. Available online: https://mhsc.org.za/sites/default/files/public/research_documents/MHSC0311415%20Report.Personal%20exposure%20to%20airborne%20pollutants.pdf (accessed on 23 September 2021).
7. Made, F.; Kandala, N.B.; Brouwer, D. Compliance Testing and Homogenous Exposure Group Assessment in the South African Coal Mining Industry. *Ann. Work Expo. Health* **2021**, *65*, 955–965. [CrossRef] [PubMed]
8. Sottas, P.E.; Lavoué, J.; Bruzzi, R.; Vernez, D.; Charrière, N.; Droz, P.O. An empirical hierarchical Bayesian unification of occupational exposure assessment methods. *Stat. Med.* **2009**, *28*, 75–93. [CrossRef] [PubMed]
9. Ramachandran, G.; Vincent, J.H. A Bayesian approach to retrospective exposure assessment. *Appl. Occup. Environ. Hyg.* **1999**, *14*, 547–557. [PubMed]
10. Ramachandran, G. Retrospective exposure assessment using Bayesian methods. *Ann. Occup. Hyg.* **2001**, *45*, 651–667. [CrossRef]
11. Banerjee, S.; Ramachandran, G.; Vadali, M.; Sahmel, J. Bayesian hierarchical framework for occupational hygiene decision making. *Ann. Occup. Hyg.* **2014**, *58*, 1079–1093. [PubMed]
12. Hewett, P.; Logan, P.; Mulhausen, J.; Ramachandran, G.; Banerjee, S. Rating exposure control using Bayesian decision analysis. *J. Occup. Environ. Hyg.* **2006**, *3*, 568–581. [CrossRef] [PubMed]
13. CEN. *BS EN 689:2018; Workplace Exposure. Measurement of Exposure by Inhalation to Chemical Agents. Strategy for Testing Compliance with Occupational Exposure Limit Values*; British Standards Institution: Brussels, Belgium, 2018.
14. BOHS-NVvA. Testing Compliance with Occupational Exposure Limits for Airborne Substances. 2011. Available online: https://www.arbeidshygiene.nl/-uploads/files/insite/2011-12-bohs-nvva-sampling-strategy-guidance.pdf (accessed on 4 October 2021).
15. Baldos, U.L.C.; Viens, F.G.; Hertel, T.W.; Fuglie, K.O. R&D spending, knowledge capital, and agricultural productivity growth: A Bayesian approach. *Am. J. Agric. Econ.* **2019**, *101*, 291–310.
16. Gelman, A.; Carlin, J.B.; Stern, H.S.; Dunson, D.B.; Vehtari, A.; Rubin, D.B. *Bayesian Data Analysis*; CRC Press: Boca Raton, FL, USA, 2013.
17. NIOSH. Particulates not otherwise regulated, total: Method 0500. In *NIOSH Manual of Analytical Methods*, 4th ed.; DHHS (NIOSH) Publication No. 94-113; United States Department of Health and Human Services, Centers for Disease Control and Prevention: Cincinnati, OH, USA, 1994. Available online: www.cdc.gov/niosh/nmam/ (accessed on 15 July 2021).
18. Stan Development Team. RStan: The R Interface to Stan. R Package Version 2.18.2. 2018. Available online: http://mc-stan.org (accessed on 19 March 2022).
19. Makowski, D.; Ben-Shachar, M.S.; Lüdecke, D. bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *J. Open Source Softw.* **2019**, *4*, 1541. [CrossRef]
20. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. 2015. Available online: https://www.r-project.org/ (accessed on 10 April 2021).
21. Carlin, B.P.; Louis, T.A. *Bayesian Methods for Data Analysis*, 3rd ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2008.
22. Quick., H.; Huynh., T.; Ramachandran, G. A method for constructing informative priors for Bayesian modelling of occupational hygiene data. *Ann. Work Expo. Health* **2017**, *61*, 67–75. [PubMed]
23. Gelman, A. Prior distributions for variance parameters in hierarchical models. *Bayesian Anal.* **2006**, *1*, 515–533. [CrossRef]
24. Gelfand, A.E.; Smith, A.F.M. Sampling-based approaches for calculating marginal densities. *J. Am. Stat. Assoc.* **1990**, *85*, 398–409. [CrossRef]
25. Gelman, A.; Rubin, D.B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **1992**, *7*, 457–472. [CrossRef]
26. Vehtari, A.; Gelman, A.; Simpson, D.; Carpenter, B.; Bürkner, P.C. Rank-normalization, folding, and localization: An improved R for assessing convergence of MCMC. *arXiv* **2019**, arXiv:1903.08008. [CrossRef]
27. Symanski, E.; Maberti, S.; Chan, W. A meta-analytic approach for characterizing the within-worker and between-worker sources of variation in occupational exposure. *Ann. Occup. Hyg.* **2006**, *50*, 343–357. [PubMed]
28. Jones, R.M.; Burstyn, I. Bayesian analysis of occupational exposure data with conjugate priors. *Ann. Work Expo. Health* **2017**, *61*, 504–514. [CrossRef] [PubMed]
29. Chakraborty, S.; Ghosh, M.; Chakraborty, S.; Ghosh, M. Applications of Bayesian Neural Networks in Prostate Cancer Study. In *Handbook of Statistics*; Elsevier: Amsterdam, The Netherlands, 2012; pp. 241–262.
30. Wilson, A.G.; Fronczyk, K.M. Bayesian reliability: Combining information. *Qual. Eng.* **2017**, *29*, 119–129. [CrossRef]
31. Lavoué, J.; Joseph, L.; Knott, P.; Davies, H.; Labrèche, F.; Clerc, F.; Mater, G.; Kirkham, T. Expostats: A Bayesian toolkit to aid the interpretation of occupational exposure measurements. *Ann. Work Expo. Health* **2019**, *63*, 267–279. [CrossRef] [PubMed]
32. Prior Information, Not Prior Belief. 2015. Available online: http://andrewgelman.com/2015/07/15/prior-information-not-prior-belief/ (accessed on 21 February 2016).
33. Huynh, T.; Quick, H.; Ramachandran, G.; Banerjee, S.; Stenzel, M.; Sandler, D.P.; Engel, L.S.; Kwok, R.K.; Blair, A.; Stewart, P.A. A comparison of the β-substitution method and a Bayesian method for analyzing left-censored data. *Ann. Occup. Hyg.* **2016**, *60*, 56–73. [CrossRef]

34. Lunn, D.; Spiegelhalter, D.; Thomas, A.; Best, N. The BUGS project: Evolution, critique and future directions. *Stat. Med.* **2009**, *28*, 3049–3067. [CrossRef] [PubMed]
35. Mulhausen, J.; Damiano, J.; Pullen, E.L. Defining and Judging Exposure Profiles. In *A Strategy for Managing Occupational Exposures*, 3rd ed.; Ignacio, J.S., Bullock, W.H., Eds.; AIHA Press: Fairfax, VA, USA, 2006.
36. Ramachandran, G. *Occupational Exposure Assessment for Air Contaminants*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2005.

Appendix J: Bayesian Hierarchical Framework from Expert Elicitation in the South African Coal Mining Industry for Compliance Testing

International Journal of *Environmental Research and Public Health*

MDPI

Article

Bayesian Hierarchical Framework from Expert Elicitation in the South African Coal Mining Industry for Compliance Testing

Felix Made [1,2,*], Ngianga-Bakwin Kandala [1,3,4] and Derk Brouwer [1]

1 School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2193, South Africa
2 Global Biostatistics and Programming, Pharmaceutical Product Development, Part of Thermofisher Scientific, Woodmead, Johannesburg 2191, South Africa
3 Department of Epidemiology and Biostatistics, Schulich School of Medicine and Dentistry, Western University, London, ON N6G 2M1, Canada
4 Département de la Santé Communautaire, Institut Supérieur des Techniques Médicales de Kinshasa, Kinshasa XI, Mont Ngafula, Kinshasa B.P. 774, Democratic Republic of the Congo
* Correspondence: felix.made@ppd.com

Abstract: Occupational exposure assessment is important in preventing occupational coal worker's diseases. Methods have been proposed to assess compliance with exposure limits which aim to protect workers from developing diseases. A Bayesian framework with informative prior distribution obtained from historical or expert judgements has been highly recommended for compliance testing. The compliance testing is assessed against the occupational exposure limits (OEL) and categorization of the exposure, ranging from very highly controlled to very poorly controlled exposure groups. This study used a Bayesian framework from historical and expert elicitation data to compare the posterior probabilities of the 95th percentile (P95) of the coal dust exposures to improve compliance assessment and decision-making. A total of 10 job titles were included in this study. Bayesian framework with Markov chain Monte Carlo (MCMC) simulation was used to draw a full posterior probability of finding a job title to an exposure category. A modified IDEA ("Investigate", "Discuss", "Estimate", and "Aggregate") technique was used to conduct expert elicitation. The experts were asked to give their subjective probabilities of finding coal dust exposure of a job title in each of the exposure categories. Sensitivity analysis was done for parameter space to check for misclassification of exposures. There were more than 98% probabilities of the P95 exposure being found in the poorly controlled exposure group when using expert judgments. Historical data and non-informative prior tend to show a lower probability of finding the P95 in higher exposure categories in some titles unlike expert judgments. Expert judgements tend to show some similarity in findings with historical data. We recommend the use of expert judgements in occupational risk assessment as prior information before a decision is made on current exposure when historical data are unavailable or scarce.

Keywords: expert judgments; expert elicitation; exposure control categories; the 95th percentile; historical data

Citation: Made, F.; Kandala, N.-B.; Brouwer, D. Bayesian Hierarchical Framework from Expert Elicitation in the South African Coal Mining Industry for Compliance Testing. *Int. J. Environ. Res. Public Health* **2023**, *20*, 2496. https://doi.org/10.3390/ijerph20032496

Academic Editor: Paul B. Tchounwou

Received: 31 December 2022
Revised: 27 January 2023
Accepted: 29 January 2023
Published: 31 January 2023

1. Introduction

Coal is the second largest energy source in the world, contributing a quarter of the world's energy sources, and the consumption of coal has been on the rise in recent years. By 2040, worldwide coal consumption is expected to rise at a rate of about 1.5% per year [1]. South Africa is the fourth largest coal producer in the world, employing more than 92,000 workers in 2019 [2]. Coal mining areas in South Africa are geographically found in the KwaZulu-Natal, Mpumalanga, and Limpopo provinces. In coal mining activities, coal dust is produced. Inhalation of coal dust is highly associated with the development of Coal Mine Dust Lung Disease (CMDLD). CMDLD is life-threatening to coal mine workers in developing countries [3,4].

A study by Attfield and Seixas [5] defined the risk of developing CMDLD, a result of over-exposure to coal dust during work-life in the coal mining industry which depends on different work activities. Various work activities generate dust in the coal mining industry. Dust also results from poor organization or lack of best decisions in control of exposure to coal dust. Such activities may include inadequate dust disposal techniques, poor transport systems, coal spillage, and dry mine conditions, amongst others. To improve occupational risk assessment and compliance testing, a homogenous exposure group (HEG) was established. The HEG, according to the Code of Practice (CoP) by the South African Department of Minerals and Energy, is constituted by mine, activity area, and incoming air. Within each HEG, different job titles may be found, increasing the variability of exposure to coal dust [6].

The advancement in technology in recent years means that the coal mining industry has adequate control methods for reducing coal dust overexposures among workers. It is important to regularly check compliance with the occupational exposure limit (OEL) and associated exposure control categories (see Appendix A). Bayesian analysis has become popular for evidence-based policy decision-making. For occupational hygiene applications, Bayesian statistics were used to find highly exposed HEG or job titles. The advantage of Bayesian analysis is that earlier or prior data, when available, can be used to update and improve new measurements using Bayesian inference [7]. The prior distribution can be obtained from solid information (informative) from subjective expert data and/or expert elicitation, or from a non-informative source which does not influence the current data.

Expert elicitation is a systematic approach that combines subjective views from experts on exposure where there is uncertainty because of a lack of or insufficient data [8]. An expert's experience and knowledge in a particular subject area are the key factors in the elicitation process. Informed decision-making can be made with higher confidence in industries when experts who routinely see and analyze changes in environmental conditions participate in risk assessment. For instance, the United States Environmental Protection Agency uses an elicitation process to solve uncertain problems [9]. Walker et al. (2001) showed that subjective knowledge about the unknown subject can be elicited from the expert for environmental exposure assessment [10]. In occupational hygiene, expert elicitation has been widely used. For example, expert elicitation was used to develop job exposure matrices to categorize job titles into exposure levels [11]. Job exposure matrices can be used when individual exposure data are unavailable or difficult to obtain due to regulatory or privacy measures. Other studies have used elicitation to assess the health effect of fine particles of air pollution [12,13]. Bayesian application of expert elicitation is based on the subjective probabilistic judgement of the previous or present value of an unknown or uncertain quantity [14]. The Bayesian framework can assign exposure probability distributions by incorporating parameters obtained from the expert elicitation, and this is now easy to apply in occupational hygiene [15]. One other benefit of using the Bayesian framework is that it allows for uncertainty in the assessment of compliance to exposure limits, which improves holistic decision-making. Thus, this study aimed to use the Bayesian framework to compare the posterior (after the data were observed) probabilities of the 95th percentile of the exposure distribution (P95) of the non-informative and informative prior from expert elicitation and subjective expert data.

2. Methods

2.1. Study Design and Data Collection

The study was a survey collection of respirable coal mine dust exposure in underground coal mines in South Africa. The coal mine workers were only males. Workers included in this study were from 10 job titles. The inclusion criteria for the selection of these job titles were based on the availability of previous or historical (past) exposure data and current data. Job titles must have exposure from both the previous and current or the latest year to be included in the analysis. The data from the previous year were used as prior data. Personal sampling of respirable dust was conducted by using cyclones attached to a sampling pump followed by gravimetric analysis [16]. Further details on sampling are available from the earlier paper [6]. Exposures were measured over a full shift and expressed as eight-hour time-weighted average dust concentration (TWA8h concentration).

2.2. Expert Elicitation Process

In this study, a modification of the "Investigate", "Discuss", "Estimate", and "Aggregate" (IDEA) structured protocol was used [17]. The first step in the IDEA protocol is the background information on experts, followed by an investigation where experts are requested to individually answer questions and provide reasons for their answers. During the discussion, an expert was shown anonymous answers from each participant to resolve varying interpretations of the questions. Each expert had to estimate for the second and final time. During the post-elicitation process, aggregation of the data takes place, where the mean of experts' second-round responses is obtained. Many studies have adopted the IDEA protocol seamlessly. The protocol was evaluated in public health, ecology, and conservation studies, and recently in the US Intelligence Advanced Research Projects Activity [18,19]. The advantage of the IDEA protocol is that it minimizes bias resulting from overconfidence, availability, and representativeness. It is also possible to apply for remote elicitation, which is cost-effective. We invited a group of experts to take part in the study. The experts were given time to investigate and understand the questions, and then give their feedback on the questions. The modification from the IDEA protocol is that we did not allow experts to receive feedback on their judgements from other experts. The experts were not encouraged to discuss their results to provide second estimates. We instead adopted and simplified the IDEA protocol four-step elicitation questions in Appendix B. The advantage of the modification is that it prevents more experienced expert(s) from influencing the discussion which may skew the final judgements toward just one dominant expert. The expert elicitation process was done remotely.

In the four-step process we used in this study, the credible intervals (range) were the minimum and maximum probability of exposure in each of the exposure categories. These ranges were standardized using linear extrapolation [18,20].

The minimum standardized percentage: $B - ((B - L) \times (S/C))$

The maximum standardized percentage: $B + ((U - B) \times (S/C))$

Where B is the best possible guess, L is the minimum percentage, U is the maximum percentage, S is the level of credible intervals to be standardized, and C is the level of confidence (how sure is the expert). The distributions were truncated for extreme values if the adjusted credible intervals fall outside the probability bounds of 0 and 100%. The credible intervals were standardized to reduce overconfidence.

The expert's judgements were combined to form a joint probability distribution. The distribution was used as an informative prior, updating current data from respective job titles to produce posterior estimates when compared to the OEL. The individual expert probability distribution can be mathematically or behaviourally aggregated to produce the joint distribution [21]. For this study, we used equal weighting to aggregate the data.

Selection of Experts

Experts were invited through email to take part in this study. The average year of experience was 13 years. The experts were occupational hygienists, mine ventilation engineers, supervisors, mine engineers, and geologists. The experts were found by the lead of occupational health and hygiene in the coal mining industry. The identified experts then recommended other experts through snowballing. All experts who accepted the invitation were asked to voluntarily take part. Experts had a range of experience associated with various responsibilities and/or connections to occupational health and safety in coal mining.

Before the elicitation, experts were trained on the uncertainty, the P95, and an outline of the tasks to be performed. They were also provided with an information sheet to obtain their professional background in the field. Each job title was described in detail to the experts and an example of how the probability of exposure to each exposure category was shown to familiarize them with the elicitation. An expert was allowed to consult at any time if he/she is confused or lost in the elicitation. According to Highlights of the Expert Judgement Policy Symposium and Technical Workshop conference, held in the year 2006 [22], a sample size between six to twelve experts is recommended for expert elicitation. In this study, a sample size of six experts was included.

2.3. Statistical Methods

Statistical analysis was conducted in STATA version 14.1 [23], and R version 4.1.1 (R Core Team, Auckland, New Zealand), using RStan and bayestestR packages for the Bayesian model [24–26]. A summary of the current and historical data was presented as arithmetic mean (AM), standard deviation (SD), geometric mean (GM), and geometric standard deviations (GSD). Each HEG is said to be unique [6], however the same job title and work activity can be found in different HEGs. The median difference in job title exposure across HEGs was assessed by using the Kruskal–Wallis Test, a non-parametric version of the analysis of variance (ANOVA). A job title with a statistically significant *p*-value of less than 0.05 at the 95% confidence interval was not selected for the Bayesian analysis, as this would mean the exposure distribution is different across HEGs. Such a job title is not regarded as having a similar exposure profile.

2.3.1. Producing Joint Distribution from Experts' Elicitation

We used the Sheffield Elicitation Framework (SHELF) tool to produce the joint probability distribution. SHELF is an open-source tool developed by the University of Sheffield [27].

The SHELF tool finds the distribution that best fits the expert's elicited judgement. The SHELF tool works by using lease squares procedures by finding a distribution that best fits the given expert's data inputs by minimizing the sum of squares of the residuals. Recall that occupational exposure data are best fitted by a lognormal distribution. In this study, we used lognormal distribution to produce our prior means and variance for the Bayesian framework. In the SHELF tool, we used the quartile method to elicit our expert's judgement in a continuous quantity [28].

The Quartile methods ask for experts' median and quartiles. The quartiles are presented as lower (Q1) and upper quartiles (Q3). The lower (L) and upper (U) limits during the elicitation were 0 and 100%. The Q1 is a value between L and M, where M is the median value, while Q3 is a value between M and U. In the expert's elicitation questionnaire in Appendices B and C, we collected the expert's best guess in form of percentages to distribute the P95 grouping according to the exposure categories. Additionally, we asked about their minimum and maximum percentages on the probability of the P95 coal dust exposure. This was important to capture the expert's uncertainty when they may not be sure about their best guess. Therefore, in the Quartile methods, the best guess was used as the median, while their minimum and maximum percentages were represented as the Q1 and Q3, respectively. The cumulative probabilities were set at 0.25, 0.5, and 0.75.

A minimum informative non-parametric distribution then spread the mass uniformly between estimates to make sure the joint distribution fully is the expert assessment [29]. The arithmetic mean of the lognormal mean and variance was obtained with the assumption that each expert contributed equally. The mean and variance were then used prior to the Bayesian framework for the exposure compliance assessment.

2.3.2. Bayesian Framework

Consistent and similar coal dust exposure from expert judgements data of the job titles were used to update current monitoring data. The posterior geometric mean (GM), geometric standard deviations (GSD), P95, and the probabilities of finding the P95 exposure in each of the exposure control categories were obtained. For the prior specification, we randomly selected a prior sample size of five out of the six experts' data collected, as recommended from earlier studies for occupational exposure assessment [30,31]. From the occupational exposure perspective, the prior sample size should be from 10% to 40% of the current data so that the posterior distribution learned more from the current likelihood data than from the prior distribution. Therefore, the sample size of five was used to keep the focus of the posterior distribution on the current data, as the sample size of the current data increases. In Bayesian statistics, the posterior distribution is a compromise between the information from the prior and the current data, but the distribution must be seen from the current data to a good measure as the sample size increases [32]. For the likelihood function, we took all the available current monitoring data.

2.3.3. Model Specification Using Current Likelihood Data

The likelihood was specified as μ and σ which is the GM and GSD, respectively. The likelihood function is presented below in Equation (1).

$$\prod_{i=1}^{n} \mathrm{LN}\left(y_i \middle| \mu, \sigma^2\right) = \prod_{i=1}^{n} \frac{1}{y_i \sigma \sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \frac{\left(\log y_i - \mu\right)^2}{\sigma^2} \right\} \quad (1)$$

y_i is the log-normal (LN) current monitoring data, n is the number of observations. The OEL exposure categories were added as a random variable in the model directly to produce the posterior probability distribution of the P95 to each of the categories [30,31].

2.3.4. Model Specification Using a Non-Informative Prior Distribution, and Informative Prior from the Historical Data and Expert Judgements

The non-informative prior follows a uniform distribution with GM and GSD.

For the GM, $\mu = \mathrm{LnGM} \sim \mathrm{Unif}(a_{\mu} b_{\sigma\mu})$, and for the GSD, $\sigma^2 = \mathrm{LnGSD} \sim \mathrm{unif}(a_{\sigma} b_{\sigma})$, where a and b are the lower and upper bounds of the prior distribution, respectively.

For the informative prior, the GM μ takes the form as shown below in Equation (2).

$$\mu \sim \mathrm{Norm}\left(\bar{y}_0, s^s_{y0}/n_0\right) \quad (2)$$

where $\bar{y}_0$ and s^s_{y0} are the prior mean and variance, respectively, and n_0 is the prior sample size.

For the GSD, the variance is given by Equation (3).

$$\sigma^2 \sim \mathrm{IG}\left(\frac{n_0 - 1}{2}, \frac{(n_0 - 1)s^2_{y0}}{2}\right) \quad (3)$$

For $n_0 > 1$, where IG (a, b) is an inverse gamma distribution in Equation (4) with shape parameter a, and scale parameter b

Therefore,

$$a = \frac{n_0 - 1}{2} b = 2/\left(s^2_{y0} \times (n_0 - 1)\right) \quad (4)$$

Further details on the prior specification and full conditional for μ and σ^2 are available at Made et al. [33], and Quick, 2017 [31].

2.3.5. The Sensitivity Analysis for the Parameter Space

In this study, we placed restrictions on the lower bounds of the μ and σ^2 and P95 using the suggestions from Bayesian decision analysis (BDA) [15]. For the upper bound, we took motivation from Quick, 2017 [31], and allowed it to vary iteratively. The upper bound was not placed on any restriction to avoid being unfairly skewed towards a more favorable or unfavorable result. Despite this, we believed that the use of the parameter space might misclassify an exposure to a wrong exposure category. Therefore, we randomly took two job titles, namely shuttle car operator and pump attendant. We changed the parameter space and compared their findings.

In shuttle car operator, the GM and GSD's lower bound was set at 0.00001, assuming that the coal dust exposure cannot be zero. Their upper bound is set to be just above the GM and GSD values calculated in this study data. The GM is 1.08 so the upper bound was placed at 1.09; the GSD is 2.62 and upper bound was placed at 2.63, to make sure the exposures estimate fall within.

For the pump attendant, we placed the upper bound of the GM to be 0.42, just above the actual value of 0.41, while GSD was placed at 3.42 above the actual value of 3.41.

We also ran analysis on these two job titles without placing any restrictions on the parameter space. In other words, no lower and upper bounds were specified.

The full posterior conditional distribution was drawn using the Markov chain Monte Carlo (MCMC) of the Gibbs sampler [34]. The Gibbs sampler was used because of computational flexibility. The Gelman–Rubin convergence diagnostics was used to assess the stationary of the posterior distributions including the reliability of the posterior samples. Please refer to Made et al. [33] for more details on the Gibbs sampler and convergence diagnostics.

3. Results

A summary of the job titles and their corresponding exposures are presented in Table 1. A total of 10 job titles were included in this study. The job titles were chosen on the basis that they have both monitoring (current) data and earlier (historical) data in the HEG they belong to. There are 455 observations in the monitoring data and 108 from historical data. Six job titles in the current data showed high exposure variability according to the GSD (GSD greater than 3), while from the historical data, eight job titles showed high exposure variability according to the GSD. All the job titles showed statistically non-significant differences in median exposure variation across their parent HEGs ($p < 0.05$).

Table 1. Summary of coal dust exposure concentration (TWA8h) in mg/m^3 for the current monitoring data and their corresponding historical past data, and comparison of the median exposure across HEGs.

Data	Year	n	AM	SD	GM	GSD	*p*-Value *
Current data							
Beltsman	2015	18	0.76	0.72	0.48	2.94	0.5694
CM Operator	2018	116	2.08	1.87	1.28	3.30	0.1014
Conveyer belt Attendant	2015	29	0.93	0.75	0.66	2.41	0.0917
Emico Driver	2015	24	1.13	1.02	0.54	5.11	0.3116
Electrician	2015	52	1.55	1.52	0.87	3.75	0.0617
Face Boss	2018	35	1.56	1.76	0.80	3.68	0.8640
Pump Attendant	2016	18	0.74	0.82	0.41	3.41	0.3955
Roofbolt Operator	2018	101	1.77	1.66	1.11	3.03	0.0853
Safety Officer	2018	16	2.78	2.05	2.06	2.45	0.8169
Shuttle Car Operator	2018	46	1.60	1.34	1.08	2.62	0.1615
Historical data							
Beltsman	2009	14	1.07	0.89	0.66	3.31	0.1653
CM Operator	2009	11	2.10	2.05	0.70	8.46	0.1573
Conveyer belt Attendant	2009	7	0.64	0.56	0.42	2.92	0.3012
Emico Driver	2009	6	0.81	0.73	0.43	4.29	0.5319
Electrician	2009	14	1.70	2.03	0.75	5.31	0.3618
Face Boss	2009	22	1.00	0.97	0.46	4.89	0.8495
Pump Attendant	2009	8	0.66	0.69	0.42	2.86	0.3679
Roofbolt Operator	2009	8	1.70	2.58	0.81	3.54	0.5319
Safety Officer	2009	8	0.70	0.56	0.38	4.57	0.4060
Shuttle Car Operator	2016	10	1.3	1.11	0.69	4.59	0.2352

n: sample size; AM: athematic mean; SD: standard deviation; GM: geometric mean; GSD: geometric standard deviation. * *p*-value to test the difference in dust exposure of each job title across HEGs.

Table 2 shows the posterior GM, GSD, and P95 from the Bayesian analysis. The GM from the non-informative and informative prior distribution were all below the OEL. Only one job title had a GM above the OEL in the Bayesian findings generated by informative prior from expert judgments. For variability of exposures, four and five job titles showed high variability according to the GSD in the non-informative and informative prior from historical data, respectively. The posterior GSD from the expert judgments showed that six job titles had a high variability of exposure. Three of the job titles showed consistently high variability of exposures across all the methods.

Figure 1 shows the comparison of the posterior P95 by job titles. The findings from the use of expert judgments as prior indicated higher P95 in all job titles exposure. Only for pump attendants, roofbolt operators, safety officers, and shuttle car drivers, the P95 is lower from historical data than from the non-informative prior distribution. Table 3 estimates the probabilities of grouping exposure in each exposure category. Overall, there were high probabilities (above 75%) of finding exposures in the highest category (4) (above OEL/very poorly controlled). The pump attendants' exposures had a higher probability of being in exposure category 4 from non-informative prior (79.69%) than historical data (73.09%). However, there were higher probabilities of being in exposure category 3 when using historical data compared to the non-informative. The findings from the experts showed more than a 99.9% probability of all the exposures in the poorly controlled group (exposure category 4).

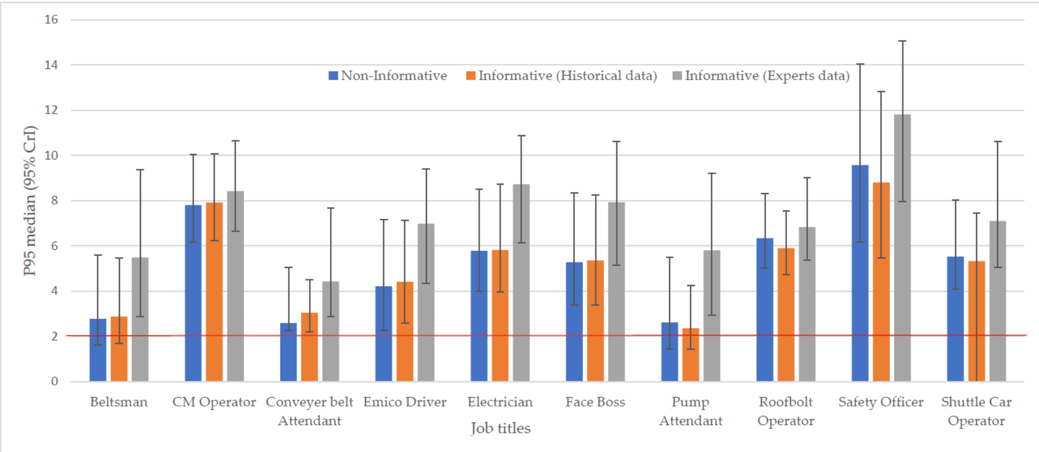

Figure 1. The comparison of the patterns of the posterior median (95% CrIs) of the P95 for non-informative and informative (historical and expert data) distributions of the TWA8h dust concentrations across the job titles. The red horizontal is the SA OEL = 2 mg/m^3.

Table 4 shows the results of the sensitivity analysis. The findings of the parameter space we employed versus putting no restriction on the bound and using different parameter values were compared (see Section 2). Using no parameter space revealed 100% probability of finding the P95 in the highest exposure category. Using parameter values described in the Section 2 showed similar findings with the values inspired by the BDA on the probabilities of locating the P95 in each exposure category. The posterior probabilities when using experts' judgment as prior were similar (99.94% versus 99.91% for exposure category 4). The findings from the non-informative and informative prior from historical data showed slightly higher probabilities of the exposure in category 3 when using different parameter values than values from the BDA. Since we used a lower parameter space for the lower and upper bound, this resulted in lower probabilities of finding exposures in category 4.

Table 2. The median (95% credible interval (CrI)) of the posterior GM, GSD, and the P95 percentiles of the TWA8hdust concentration and the P90 according DMRE CoP (OEL is 2 mg/m^3).

Job Titles	Non-Informative			Informative from Historical Data			Informative from Expert Judgments		
	GM	GSD	P95	GM	GSD	P95	GM	GSD	P95
	Median (95% CrI)	Median (95% CrI)	Median (95% CrI)	Median (95% CrI)	Median (95% CrI)	Median (95% CrI)	Median (95% CrI)	Median (95% CrI)	Median (95% CrI)
Beltsman	0.48 (0.28, 0.83)	2.90 (2.48, 3.64)	2.77 (1.63, 5.60)	0.47 (0.27, 0.80)	3.01 (2.61, 3.66)	2.87 (1.67, 5.46)	0.84 (0.50, 1.38)	3.10 (2.67, 3.58)	5.48 (2.86, 9.36)
CM Operator	1.28 (1.03, 1.59)	2.99 (2.80, 3.23)	7.80 (6.18, 10.03)	1.25 (1.00, 1.57)	3.06 (2.87, 3.29)	7.91 (6.22, 10.06)	1.37 (1.11, 1.70)	3.00 (2.82, 3.22)	8.42 (6.65, 10.66)
Conveyer belt Attendant	0.66 (0.48, 0.94)	2.59 (2.32, 3.03)	2.59 (2.25, 5.06)	0.64 (0.47, 0.87)	2.57 (2.33, 2.95)	3.04 (2.21, 4.50)	0.80 (0.55, 1.21)	2.84 (2.53, 3.30)	4.43 (2.88, 7.67)
Emico Driver	0.51 (0.26, 0.90)	3.58 (3.09, 4.24)	4.21 (2.26, 7.15)	0.56 (0.34, 0.92)	3.47 (3.03, 4.02)	4.41 (2.58, 7.14)	0.91 (0.58, 1.33)	3.41 (3.08, 3.79)	6.98 (4.33, 9.40)
Electrician	0.86 (0.60, 1.22)	3.18 (2.87, 3.57)	5.78 (3.99, 8.52)	0.83 (0.57, 1.18)	3.26 (2.96, 3.64)	5.82 (3.97, 8.43)	1.31 (0.98, 1.69)	3.14 (2.89, 3.39)	8.72 (6.14, 10.87)
Face Boss	0.79 (0.51, 1.20)	3.16 (2.80, 3.64)	5.27 (3.38, 8.34)	0.75 (0.47, 1.16)	3.28 (2.92, 3.74)	5.35 (3.38, 8.25)	1.19 (0.83, 1.64)	3.13 (2.83, 3.45)	7.92 (5.14, 10.61)
Pump Attendant	0.40 (0.22, 0.74)	3.11 (2.63, 3.91)	2.61 (1.44, 5.50)	0.38 (0.24, 0.62)	2.99 (2.60, 3.65)	2.35 (1.43, 4.25)	0.82 (0.48, 1.32)	3.24 (2.80, 3.70)	5.80 (2.94, 9.22)
Roofbolt Operator	1.11 (0.90, 1.39)	2.88 (2.69, 3.12)	6.34 (5.02, 8.31)	1.04 (0.84, 1.29)	2.86 (2.68, 3.10)	5.90 (4.72, 7.56)	1.17 (0.94, 1.47)	2.91 (2.73, 3.16)	6.82 (5.36, 9.02)
Safety Officer	1.98 (1.24, 2.93)	2.59 (2.27, 3.06)	9.57 (6.17, 14.04)	1.57 (0.88, 2.44)	2.84 (2.50,3.36)	8.80 (5.47, 12.83)	2.30 (1.54, 3.18)	2.68 (2.42, 3.00)	11.81 (7.96, 15.07)
Shuttle Car Operator	1.08 (0.81, 1.45)	2.69 (2.46, 3.04)	5.52 (4.08, 8.01)	1.05 (0.80, 1.38)	2.68 (0.80, 1.38)	5.32 (3.98, 7.46)	1.35 (1.03, 1.84)	2.73 (2.50, 3.06)	7.10 (5.06, 10.60)

P95: 95th percentile; CrI: Credible interval.

Table 3. The estimated exposure category probabilities for the posterior 95th percentile of the TWA8h dust concentration of the non-informative and informative Bayesian frameworks (OEL = 2 mg/m^3).

HEG	Non-Informative					Informative from Historical Data					Informative from Experts' Data				
	P95	Category 1	Category 2	Category 3	Category 4	P95	Category 1	Category 2	Category 3	Category 4	P95	Category 1	Category 2	Category 3	Category 4
Beltsman	2.77	0	0.25%	12.07%	87.67%	2.87	0	0.18%	9.52%	90.29%	5.48	0	0	0.06%	99.95%
CM Operator	7.80	0	0	0	100%	7.91	0	0	0	100%	8.42	0	0	0	100%
Conveyer belt Attendant	2.59	0	0	0.38%	99.62%	3.04	0	0	0.40%	99.60%	4.43	0	0	0.02%	99.99%
Emico Driver	4.21	0	0.03%	1.04%	98.93%	4.41	0	0	0.19%	99.81%	6.98	0	0	0	100%
Electrician	5.78	0	0	0	100%	5.82	0	0	0	100%	8.72	0	0	0	100%
Face Boss	5.27	0	0	0	100%	5.35	0	0	0	100%	7.92	0	0	0	100%
Pump Attendant	2.61	0	1.16%	19.14%	79.69%	2.35	0	0.91%	25.99%	73.09%	5.80	0	0	0.06%	99.94%
Roofbolt Operator	6.34	0	0	0	100%	5.90	0	0	0	100%	6.82	0	0	0	100%
Safety Officer	9.57	0	0	0	100%	8.80	0	0	0	100%	11.81	0	0	0	100%
Shuttle Car Operator	5.52	0	0	0	100%	5.32	0	0	0	100%	7.10	0	0	0	100%

Category 1: P95 TWA8h concentration < 0.1 OEL (Very highly controlled); Category 2: P95 TWA8h concentration $\geq$ 0.1 OEL < 0.5 OEL (Highly controlled); Category 3: P95 TWA8h concentration $\geq$ 0.5 OEL < OEL (Adequately controlled); Category 4: P95 TWA8h concentration $\geq$ OEL (Poorly controlled).

Table 4. Sensitivity analysis to compare the parameter values used for the lower and upper bound in this study.

Grouping		Using Parameter Values from BDA				Placing No Restrictions on Bounds				Using Different Parameter Values			
		Category 1	Category 2	Category 3	Category 4	Category 1	Category 2	Category 3	Category 4	Category 1	Category 2	Category 3	Category 4
Pump Attendant	Non-informative	0	1.16%	19.14%	79.69%	0	0	0	100%	0	1.24%	20.85	77.91%
	Historical data	0	0.91%	25.99%	73.09%	0	0	0	100%	0	1.62%	29.11%	69.27%
	Exert judgments	0	0	0.06%	99.94%	0	0	0	100%	0	0	0.09%	99.91%
Shuttle Car Operator	Non-informative	0	0	0	100%	0	0	0	100%	0	0	0	100%
	Historical data	0	0	0	100%	0	0	0	100%	0	0	0	100%
	Exert judgments	0	0	0	100%	0	0	0	100%	0	0	0	100%

Category 1: P95 TWA8h concentration < 0.1 OEL (Very highly controlled); Category 2: P95 TWA8h concentration $\geq$ 0.1 OEL < 0.5 OEL (Highly controlled); Category 3: P95 TWA8h concentration $\geq$ 0.5 OEL < OEL (Adequately controlled); Category 4: P95 TWA8h concentration $\geq$ OEL (Poorly controlled).

4. Discussion

This study aimed to compare the posterior probabilities of grouping the P95 according to the COP exposure control categories or groups. We derived informative prior distributions from expert judgments and historical data for each job title. Since the same job titles were found in more than one HEG or mining area, we included those that were similar in exposure distribution across HEG. This is important, as each HEG could have different exposure control methods that could influence exposure variability. It is only natural that the exposure distribution of a job title should be similar regardless of the area, since they perform the same work activity [6]. This is important, because their findings can be generalized back to the workers that have yet to be selected. In occupational exposure assessment, incorporating expert judgments in a Bayesian framework for exposure categories was first used and proposed by Hewett et al. [15]. The use of expert judgments in this study is necessary because if historical data are unavailable or inadequate, expert knowledge can always be used to answer questions about the quantity of interest. The expert elicitation was conducted using a modified IDEA [17]. We changed the IDEA technique by not allowing expert judgments to be shared among each other. It was good to avoid a more experienced expert having too much influence over the final expert prior to distribution. The confidence level in individual expert judgments was adjusted using the 70% upper credible limit from the Committee of European Normalization approaches [35]. The joint probability lognormal distribution for the prior mean and variance from the individual experts were derived using the SHELF package [27]. The lognormal parameters were generated so that the lower point is the 2.5th percentile and the highest value is the 97.5th percentile. SHELF was used to conduct the elicitation of uncertain quantities in the form of probabilities from a group of experts. Instead of running expert group interaction and sharing information to get a consensus, we only used SHELF to produce the lognormal distribution of the mean and variance of the uncertain quantities. These parameters were used as informative priors in a Bayesian framework.

Our study found that the posterior GM from non-informative priors and the prior derived from historical and expert data were all below the OEL (OEL = 2 mg/m^3), showing no overexposure risk when considering the GM as a parameter to estimate the risk (Table 2). It is important to know that exposure management decisions are based on the P95 of the lognormal exposure distribution [36]. The posterior GSD showed more job titles had very high exposure variability in the expert prior than in the historical and weak priors. The distribution of the median P95 (95% CrI) across all methods is in Figure 1. The P95 estimates the upper bound of the TWA8h exposures and can be achieved for 95% of the workers [37]. Compared to the other two approaches, the P95 was mostly lower in the non-informative prior. The 95% CrIs were more expansive in the non-informative prior and the prior derived from experts' judgments. That means that a risk decision may not be taken with high confidence, since the level of uncertainty is high. The reason the experts' data had a wider 95% CrI may also be due to the different beliefs of experts about their uncertainty and level of confidence.

The posterior probabilities of finding the P95 in a specific exposure control category are shown in Table 3. From the informative prior from historical data, there were at least 75% probabilities of finding the exposures in the highly exposed exposure category 4 similarly to the non-informative prior distribution. The findings from using expert judgments showed more than 95% probabilities of exposures in the highest exposure category. This may show the experts either overstated their beliefs or their exposures were higher. The results from the experts are inconsistent with those from previous studies where qualitative exposure assessments showed underprediction of exposure by experts [38,39]. The results generated by the experts and historical data tend to show more consistency. These higher probabilities of finding the P95 in the highest exposure category when using expert judgments might be attributed to overconfidence on the part of some experts, despite reducing their confidence level by 30%. Some of the GSDs indicated low variability of exposure, but the respective probabilities of locating the P95 were found in higher exposure control categories. These contrasting results might show that incorrect priors updated the posterior P95. Therefore, more sampling might be needed to repeat the process and reach a decision. Our results also contrast with the earlier study, which used HEGs; here, we found that the probabilities of the exposures being in the higher category (poorly controlled exposure) were greater than 95% [33]. This clearly explains why grouping by using a job title can give a truer exposure distribution than grouping according to HEGs. Despite having only some of the data for compliance testing, exposure determinants, e.g., specific activities (often associated with job titles), workplace conditions, and production volume, are highly needed in exposure assessment [40,41]. Therefore, job titles are important to consider as a determinant of exposure variation.

The BDA motivated the lower and upper bounds placed on the parameter space in this study (see Table 4) [15]. The use of bounds in the parameter space is essential to avoid the exposure being classified in an unfavorable exposure category, either in the lowest exposure category, the highest exposure category, or infinity. According to the BDA, the GM and GSD cannot be any values, no matter how large or near-impossible they can be like in frequentist statistics. Thus, the minimum and maximum values were set as bounds for GM and GSD. However, in this study, specifying bounds for the parameter space might misclassify exposure towards the favorable or unfavorable exposure control category. We conducted a sensitivity analysis to compare the bounds adopted in this study with if no bounds are specified, or different bounds are used. The findings suggest that using no bounds tends to place exposures in the poorly controlled exposure category compared to using parameter values in the parameter space. The use of another bound shows comparable results to the one we adopted. It is also natural that exposure measurement values in occupational settings cannot be extreme or at infinity, thus the use of bounds. An earlier study revealed that because of the small sample size of the data, the use of no bound could have placed the exposure in the highest exposure category [31].

We draw strength from this study because Bayesian statistics naturally allow earlier information from historical data or expert judgments to be incorporated into the model with the current data for informed decision-making [7]. Even with a small sample size, the results are robust and easy to interpret, even by non-statisticians. One limitation of the study is that, since we had inadequate data from experts, we could not apply weighted aggregation, which would have improved the accuracy of the prior experts. Experts' judgments can also be very biased, and their responses are dependent on their knowledge of the activities of a specific job title. Human judgments can be complex and biased. Sometimes experts rely on the Rule of Thumb to make decisions or judgments about unknown quantities, including probability assessments. The Rule of Thumb does not synthesize information like the algorithmic process to arrive at a particular judgment [42,43]. Therefore, biases might be common with the use of expert judgments. Expert training is highly encouraged in an expert elicitation process to minimize biases and improve the cognitive interpretation of the information needed. Finally, it is important to note that Bayesian analysis also offers confidence in whether to use experts' judgment as a prior or not, as it can compare expert prior updates with the non-informative prior.

5. Conclusions

Expert judgments are particularly useful when data are scarce or the available data are inadequate. It is well known that experts' judgments can be very biased, but the use of a robust method and a sensitivity analysis may minimize the bias. Our findings suggest that prior expert judgments can produce similar posterior probabilities of exposure when compared to historical data. This study seeks to advance the prevention of overexposure to coal dust. Therefore, mine ventilation engineers, occupational hygienists, and safety officers are encouraged to incorporate our findings into their routine risk assessments. Future occupational exposure compliance assessments should be based on this Bayesian framework, where historical data or, if unavailable, experts' judgments should be used to reach an informed decision.

Author Contributions: Conceptualization, F.M. and D.B.; methodology, F.M. and N.-B.K.; validation, D.B.; investigation, F.M. and D.B.; formal analysis, F.M.; writing original draft preparation, F.M.; writing manuscript and editing, F.M., N.-B.K. and D.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted under the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) at the University of the Witwatersrand, South Africa. (Ethical clearance certificate number M181038 and date of approval was 26 October 2018).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: Our gratitude goes to Salome Makgato from the Zibulo Colliery for giving us a full description of the data collection. We also thank our guess experts for their unwavering support and contributions.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. The Exposure Control Categories for Classifying the P95 of Exposure Compared to the OEL and Sampling Strategy of HEGs (Job Titles) in the SAMI CoP.

Category	Description	Statistical Illustration	Exposure Profile	Minimum Frequency
1.	Exposures less than 10% of the OEL 5% of the time	P95 < 0.1% OEL	Very highly controlled	No sampling plan for this category. Measurement results that are below 10% of the OEL will be reported under this category
2.	Exposures exceed 10% of the OEL and less than 50% of the OEL 5% of the time	P95 ≥ 0.1 OEL and <0.5 OEL	Highly controlled	Sample 5% of employees within a HEG on an annual basis with a minimum of 5 samples per HEG, whichever is greater.
3.	Exposures exceed 50% of the OEL and less than OEL 5% of the time	P95 ≥ 0.5 OEL and <OEL	Adequately controlled	Sample 5% of employees within a HEG on a 6-monthly basis with a minimum of 5 samples per HEG, whichever is greater
4.	Exposures exceed the OEL 5% of the time	P95 ≥ OEL	Poorly controlled	Sample 5% of employees within a HEG on a 3-monthly basis with a minimum of 5 samples per HEG, whichever is greater.

Appendix B Data Collection Questionnaire

Experts Prior Elicitation

Instruction for the experts

Based on the 95th percentiles (P95) value of the eight-hour time-weighted average (TWA8h) coal dust concentration distribution for a specific job title, question (a) will ask you to rate (in percentages) where you think the chances are (or the likelihood is) that the P95 would be in each of the occupational exposure category. Instruction: If you think that it is absolutely clear that it will fit in exposure category 4, you fill in 100% and leave the rest of the categories empty; if you think that the highest chance is (most likely) that the P95 will be in exposure category 4, e.g., 80%, but there also a chance that it will be in category 3, say 20%, you fill in 80% for category 4 and 20% for category 3. The total sum of the percentages should be equal to 100. Therefore, you have to distribute your ratings in categories in a way that the total is 100%.

The next question (b) asks you to give the range (minimum-maximum) of the percentages that you answered to question a (your best estimate).

Question (c) is about your level of confidence that the answer(s) you gave to questions an and b are correct. Please bear in mind that this is about your level of certainty and is different from the estimated distribution of the P95 over exposure categories, as was asked in questions (a) and (b).

Table 1 is the description of the exposure categories. Below are the three questions that you can use to put your rating in percentages in Table 2 for each job title. For question a, please make sure your answers are equal to 100%.

Expert elicitation questions for Table 2 (please give all answers in percentages).

(a) Based on your experience in the industry, what would be your best guess in percentages, of the P95 of exposure concentrations being found in each of the exposure categories?
(b) What would be the maximum and minimum percentages in each of the exposure categories?
(c) How sure are you that the above answers are correct?

Appendix C

Table A2. Experts rating. kindly give a value between 0 and 100% (for example, 10%, 15%, or 60%). No value should exceed 100%.

Job Titles	Ratings	SA Exposure Categories, Coal Dust OEL (2 mg/m^3)				
		P95 < 0.1% OEL	P95 ≥ 0.1 OEL and <0.5 OEL	P95≥ 0.5 OEL and <OEL	P95 ≥ OEL	Total (%)
Job title X (This is an example, say for expert 1)	a	40%	5%	20%	35%	100
	b	20% to 70%	0 to 15%	2% to 30%	30 to 90%	
	c	70% sure	50% sure	90% sure	100% sure	
Beltsman	a					100
	b					
	c					
CM Operator	a					100
	b					
	c					
Conveyer belt Attendant	a					100
	b					
	c					
Emico Driver	a					100
	b					
	c					
Electrician	a					100
	b					
	c					
Face Boss	a					100
	b					
	c					
Pump Attendant	a					100
	b					
	c					
Roofbolt Operator	a					100
	b					
	c					
Safety Officer	a					100
	b					
	c					
Shuttle Car Operator	a					100
	b					
	c					

a: Based on your experience in the industry, what would be your best possible guess in percentages, of the P95 of ex-posure concentrations being in each of the exposure categories? b: What would be the maximum and minimum per-centages in each of the exposure categories? c: How sure are you that the above answers are correct?

Question (c) is about your level of confidence that the answer(s) you gave to questions an and b are correct. Please bear in mind that this is about your level of certainty and is different from the estimated distribution of the P95 over exposure categories, as was asked in questions (a) and (b).

Table 1 is the description of the exposure categories. Below are the three questions that you can use to put your rating in percentages in Table 2 for each job title. For question a, please make sure your answers are equal to 100%.

Expert elicitation questions for Table 2 (please give all answers in percentages).

(a) Based on your experience in the industry, what would be your best guess in percentages, of the P95 of exposure concentrations being found in each of the exposure categories?
(b) What would be the maximum and minimum percentages in each of the exposure categories?
(c) How sure are you that the above answers are correct?

Appendix C

Table A2. Experts rating. kindly give a value between 0 and 100% (for example, 10%, 15%, or 60%). No value should exceed 100%.

Job Titles	Ratings	SA Exposure Categories, Coal Dust OEL (2 mg/m^3)				
		P95 < 0.1% OEL	P95 ≥ 0.1 OEL and <0.5 OEL	P95≥ 0.5 OEL and <OEL	P95 ≥ OEL	Total (%)
Job title X (This is an example, say for expert 1)	a b c	40% 20% to 70% 70% sure	5% 0 to 15% 50% sure	20% 2% to 30% 90% sure	35% 30 to 90% 100% sure	100
Beltsman	a b c					100
CM Operator	a b c					100
Conveyer belt Attendant	a b c					100
Emico Driver	a b c					100
Electrician	a b c					100
Face Boss	a b c					100
Pump Attendant	a b c					100
Roofbolt Operator	a b c					100
Safety Officer	a b c					100
Shuttle Car Operator	a b c					100

a: Based on your experience in the industry, what would be your best possible guess in percentages, of the P95 of ex-posure concentrations being in each of the exposure categories? b: What would be the maximum and minimum per-centages in each of the exposure categories? c: How sure are you that the above answers are correct?

References

1. U.S. Department of Energy, Energy Information Administration, Office of Coal, Nuclear, Electric and Alternate Fuels. International Energy Outlook. 1991. Available online: http://www.eia.gov/forecasts/ieo/coal.cfm (accessed on 1 October 2022).
2. Government Communication and Information System (GCIS): South Africa Yearbook. 2013. Available online: https://www.gcis.gov.za/content/resourcecentre/sa-info/yearbook2013-14 (accessed on 20 April 2022).
3. Petsonk, E.L.; Rose, C.; Cohen, R. Coal mine dust lung disease. New lessons from an old exposure. *Am. J. Respir. Crit. Care Med.* **2013**, *1*, 1178–1185. [CrossRef]
4. Soutar, C.; Campbell, S.; Gurr, D.; Lloyd, M.; Love, R.; Cowie, H.; Cowie, A.; Seaton, A. Important deficits of lung function in three modern colliery populations: Relations with dust exposure. *Am. Rev. Respir. Dis.* **1993**, *147*, 797–803. [CrossRef]
5. Attfield, M.D.; Seixas, N.S. Prevalence of pneumoconiosis and its relationship to dust exposure in a cohort of US Bituminous coal miners and ex-miners. *Am. J. Ind. Med.* **1995**, *27*, 137–151. [CrossRef]
6. Made, F.; Kandala, N.B.; Brouwer, D. Compliance Testing and Homogenous Exposure Group Assessment in the South African Coal Mining Industry. *Ann. Work Expo. Health* **2021**, *65*, 955–965. [CrossRef]
7. Gelman, A.; Carlin, J.B.; Stern, H.S.; Dunson, D.B.; Vehtari, A.; Rubin, D.B. *Bayesian Data Analysis*; CRC Press: Boca Raton, FL, USA, 2013.
8. Slottje, P.; Van der Sluijs, J.P.; Knol, A.B. *Expert Elicitation: Methodological Suggestions for Its Use in Environmental Health Impact Assessments*; Utrecht University Repository: Utrecht, The Netherlands, 2008.
9. U.S. Environmental Protection Agency. Expert Elicitation Task Force White Paper. 2011. Available online: www.epa.gov/stpc/pdfs/eewhite-paper-final.pdf (accessed on 9 January 2022).
10. Walker, K.D.; Evans, J.S.; Macintosh, D. Use of expert judgment in exposure assessment. Part I. Characterization of personal exposure to benzene. *J. Expo. Sci. Environ. Epidemiol.* **2001**, *11*, 308–322. [CrossRef]
11. Fischer, H.J.; Vergara, X.P.; Yost, M.; Silva, M.; Lombardi, D.A.; Kheifets, L. Developing a job-exposure matrix with exposure uncertainty from expert elicitation and data modelling. *J. Expo. Sci. Environ. Epidemiol.* **2017**, *27*, 7–15. [CrossRef]
12. Knol, A.B.; de Hartog, J.J.; Boogaard, H.; Slottje, P.; van der Sluijs, J.P.; Lebret, E.; Cassee, F.R.; Wardekker, J.A.; Ayres, J.G.; Borm, P.J. Brunekreef, Expert elicitation on ultrafine particles: Likelihood of health effects and causal pathways. *Part. Fibre Toxicol.* **2009**, *6*, 19. [CrossRef]
13. Hoek, G.; Boogaard, H.; Knol, A.; De Hartog, J.; Slottje, P.; Ayres, J.G.; Borm, P.; Brunekreef, B.; Donaldson, K.; Forastiere, F.; et al. Concentration-response functions for ultrafine particles and all-cause mortality and hospital admissions: Results of a European expert panel elicitation. *Environ. Sci. Technol.* **2010**, *44*, 476–482. [CrossRef]
14. Ramachandran, G.; Banerjee, S.; Vincent, J. Expert judgment and occupational hygiene: Application to aerosol speciation in the nickel primary production industry. *Ann. Occup. Hyg.* **2003**, *47*, 461–475.
15. Hewett, P.; Logan, P.; Mulhausen, J.; Ramachandran, G.; Banerjee, S. Rating exposure control using Bayesian decision analysis. *J. Occup. Environ. Hyg.* **2006**, *3*, 568–581. [CrossRef]
16. NIOSH. Particulates not otherwise regulated, total: Method 0500. In *NIOSH Manual of Analytical Methods*, 4th ed.; DHHS (NIOSH) Publication No. 94-113; United States Department of Health and Human Services, Centers for Disease Control and Prevention: Cincinnati, OH, USA, 1994. Available online: www.cdc.gov/niosh/nmam/ (accessed on 20 August 2022).
17. Hemming, V.; Burgman, M.A.; Hanea, A.M.; McBride, M.F.; Wintle, B.C. A practical guide to structured expert elicitation using the IDEA protocol. *Methods Ecol. Evol.* **2018**, *9*, 169–180. [CrossRef]
18. Adams-Hosking, C.; McBride, M.F.; Baxter, G.; Burgman, M.; De Villiers, D.; Kavanagh, R.; Lawler, I.; Lunney, D.; Melzer, A.; Menkhorst, P.; et al. Use of expert knowledge to elicit population trends for the koala (*Phascolarctos cinereus*). *Divers. Distrib.* **2016**, *22*, 249–262. [CrossRef]
19. Hanea, A.M.; McBride, M.F.; Burgman, M.A.; Wintle, B.C. Classical meets modern in the IDEA protocol for structured expert judgement. *J. Risk Res.* **2018**, *21*, 417–433. [CrossRef]
20. Bedford, T.; Cooke, R.M. *Mathematical Tools for Probabilistic Risk Analysis*; Cambridge University Press: Cambridge, UK, 2001.
21. Hanea, A.M.; Wilkinson, D.P.; McBride, M.P.; Lyon, A.; van Ravenzwaaij, D.; Singleton, T.F.; Gray, C.; Mandel, D.R.; Willcox, A.; Gould, E.; et al. Mathematically aggregating experts' predictions of possible futures. *PLoS ONE* **2021**, *16*, e0256919. [CrossRef]
22. Cooke, R.M.; Probst, K.N. Highlights of the Expert Judgment Policy Symposium and Technical Workshop. Conference Summary. 2006. Available online: http://www.rff.org/expertjudgementdocuments/documents/workshopdocs/RFFConf_06-ExpertJudgment.pdf (accessed on 20 July 2022).
23. StataCorp. *Stata Statistical Software: Release 14*; StataCorp LP: College Station, TX, USA, 2015.
24. Stan Development Team. RStan: The R Interface to Stan. R Package Version 2.18.2. 2018. Available online: http://mc-stan.org (accessed on 19 March 2022).
25. Makowski, D.; Ben-Shachar, M.S.; Lüdecke, D. bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *J. Open Source Softw.* **2019**, *4*, 1541. [CrossRef]
26. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015; Available online: https://www.r-project.org/ (accessed on 10 April 2021).
27. Oakley, J.E.; O'Hagan, A. *SHELF: The Sheffield Elicitation Framework (Version 4)*; School of Mathematics and Statistics, University of Sheffield: Sheffield, UK, 2019.

28. Barlow, R.E.; Mensing, R.W.; Smiriga, N.G. Combination of Experts' Opinions Based on Decision Theory. In *Reliability and Quality Control*; Basu, A.P., Ed.; Elsevier Science Publishers (North-Holland): New York, NY, USA, 1986; pp. 9–19.
29. Garthwaite, P.H.; Kadane, J.B.; O'Hagan, A. Statistical methods for eliciting probability distributions. *J. Am. Stat. Assoc.* **2005**, *100*, 680–701. [CrossRef]
30. Carlin, B.P.; Louis, T.A. *Bayesian Methods for Data Analysis*, 3rd ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2008.
31. Quick, H.; Huynh, T.; Ramachandran, G. A method for constructing informative priors for Bayesian modelling of occupational hygiene data. *Ann. Work Expo. Health* **2017**, *61*, 67–75.
32. Baldos, U.L.C.; Viens, F.G.; Hertel, T.W.; Fuglie, K.O. R&D spending, knowledge capital, and agricultural productivity growth: A Bayesian approach. *Am. J. Agric. Econ.* **2019**, *101*, 291–310.
33. Made, F.; Kandala, N.B.; Brouwer, D. Bayesian hierarchical modelling of historical data of the South African coal mining industry for compliance testing. *Int. J. Environ. Res. Public Health* **2022**, *19*, 4442. [CrossRef]
34. Gelfand, A.E.; Smith, A.F.M. Sampling-based approaches for calculating marginal densities. *J. Am. Stat. Assoc.* **1990**, *85*, 398–409. [CrossRef]
35. *BS EN 689:2018*; Workplace Exposure—Measurement of Exposure by Inhalation to Chemical Agents—Strategy for Testing Compliance with Occupational Exposure Limit Values. British Standards Institution: Brussels, Belgium, 2018.
36. Ignacio, J.; Bullock, B. (Eds.) *A Strategy for Assessing and Managing Occupational Exposures*, 3rd ed.; AIHA Press: Fairfax, VA, USA, 2006.
37. Banerjee, S.; Ramachandran, G.; Vadali, M.; Sahmel, J. Bayesian hierarchical framework for occupational hygiene decision making. *Ann. Occup. Hyg.* **2014**, *58*, 1079–1093.
38. Logan, P.W.; Ramachandran, G.; Mulhausen, J.R.; Hewett, P. Occupational exposure decisions: Can limited data interpretation training help improve accuracy? *Ann. Occup. Hyg.* **2009**, *53*, 311–324.
39. Vadali, M.; Ramachandran, G.; Banerjee, S. Effect of training, education, professional experience and need for cognition on decision making in occupational exposure assessment. *Ann. Occup. Hyg.* **2012**, *56*, 292–304.
40. Joint ACGIH-AIHA Task Group on Occupational Exposure Databases. Data elements for occupational exposure databases: Guidelines and recommendations for airborne hazards and noise. *Appl. Occup. Environ. Hyg.* **1996**, *11*, 1294–1313. [CrossRef]
41. Rajan, B.; Alesbury, B.; Carton, C.; Gérin, M.; Litske, H.; Marquart, H.; Olsen, E.; Scheffers, T.; Stamm, R.; Woldbaek, T. European proposal for core information for the storage and exchange of workplace exposure measurements on chemical agents. *Appl. Occup. Environ. Hyg.* **1997**, *12*, 31–37. [CrossRef]
42. Kahneman, D.; Slovic, P.; Tversky, A. *Judgment under Uncertainty: Heuristics and Biases*; Cambridge University Press: Cambridge, UK, 1982.
43. Gilovich, T.; Griffin, D.; Kahneman, D. *Heuristics and Biases: The Psychology of Intuitive Judgment*; Cambridge University Press: Cambridge, UK, 2002.