Chapter 5
Research Design and Research Questions

This chapter considers the chosen research design for the current study and evaluates the advantages and disadvantages of the design. On the basis of the rationale for the study discussed in preceding chapters the research questions addressed in the study are then discussed.

5.1 Research Design

This study used retrospective record review as the research design. Retrospective record reviews are often the preferred research design method in clinical-epidemiological research (Gómez de la Cámara, Ciruelo, de la Cruz, Serrano, Pato & Gómez-Reino, 1997) and can be defined as a research method that makes use of existing data recorded for purposes other than the research being undertaken (Hess, 2004; Worster & Haines, 2004).

Retrospective studies have been regarded as inferior to prospective designs (Hess, 2004). This is not necessarily the case, however, if care is taken to include all of the research design elements of a prospective study. Well-designed medical studies have been shown to mirror the results of prospective randomised trials (Johnston, 2002), and if all the necessary elements of the design are given sufficient attention the retrospective design will provide sound scientific results (Hess, 2004).

Various advantages and disadvantages of this method have been highlighted in the research literature (Audet, Goodnough, & Pravin, 1996; Gearing, Mian, Barber, & Ickowicz, 2006; Hess, 2004). An advantage is that the method is inexpensive, since it uses existing records (Lowenstein, 2005). The method also allows for the study of conditions that would not otherwise be ethical or feasible to study. Another advantage is that it is easier to assess conditions where there is a long latency between the exposure and the disease (Pan, Fergusson,

Schweitzer, & Hebert, 2005). Retrospective record reviews can also facilitate the generation of hypotheses that can then be tested prospectively (Hess, 2004).

The disadvantages of the retrospective record review design are, firstly, that it relies on the accuracy of the written records or on the recall of individuals, which can lead to "recall bias" (Audet et al., 1996). Secondly, important data may not be available, especially when institutional regulations restrict access or when there is a conflict in the reasons for recording information, as described in the definition of retrospective record reviews. Thirdly, the control of bias and confounders is more difficult when using this method as randomisation and blinding are often not possible (Hess, 2004; Johnston, 2002), and this may result in difficulty in establishing "cause and effect" from the study (Gearing et al., 2006). Another disadvantage of the method is the difficulty in controlling errors, omissions and idiosyncrasies in the data-extraction process and therefore the reliability of the data (VonKoss Krowchuk, Moore, & Richardson, 1995; Wu & Ashton, 1997). Finding and interpreting information, coding data, and creating ordinal response categories all require training, practice, and judgement. Therefore, the abstraction process requires agreed-to protocols, training, performance monitoring, and performance review to avoid poor research (Gearing et al., 2006).

5.2 Reliability and Validity

The chosen research design will influence the reliability and validity of a study and impact on the degree of replication of the study as well as ability to generalise the results to other research populations.

Validity can be defined as the extent to which an instrument measures what it is assumed to or meant to measure (Leedy & Ormrod, 2005; Roeser et al, 2000; Khoza, 2008). The validity of the results of the current study was ensured by means of an in-depth literature review to ensure that the research design was appropriate for the study and that the research questions were relevant in terms

of previous studies. In audiology, the inter-examiner and intra-examiner consistency The validity of the research was regarded as acceptable because the measures used from the records had been obtained using calibrated equipment, as well as accepted audiological standards by qualified audiologists as described in the data-collection section above. The validity was also ensured by the pilot study that determined the size of the representative sample and identified all the variables to be taken into account in the research. Consultation with two independent statisticians (Steffens, 2008; Fridjhon, 2007) ensured that the statistical analysis was valid and appropriate for the research design.

The reliability of the results of research can be defined as the degree of consistency or agreement between two independently derived sets of scores and the extent to which independent administrations of the same instrument yield the same or similar results under comparable conditions (De Vos, Strydom, Fouche, & Delport, 2005; Leedy & Ormrod, 2005; Khoza, 2008). In this study the reliability was enhanced by the fact that very similar procedures were followed by the audiologists when testing took place in the audiology clinic, thereby increasing the inter-observer reliability as noted in Table 14 (Worster et al., 2005).

The reliability of the coding was addressed in the study as described in Table 14, where an evaluation of the current study made use of trained abstractors (criterion 1) and selection criteria (criterion 2) as well as electronic transfer of data where possible (criterion 4). The reliability of coding was further enhanced by the monitoring of the performance of the abstractor by an independent researcher who checked 10% of the coded files; no errors in coding were identified (criterion 5) (Worster & Haines, 2004; Worster, et al., 2005).

The advantages and disadvantages of a retrospective record review influence the validity and reliability of the study. The advantages and disadvantages of a retrospective review are expanded in the following discussion and highlight the criteria against which a retrospective record review should be measured to improve the reliability and validity of a study (Worster & Haines, 2004; Worster,

et al., 2005). Table 14 then provides the framework of the methods used to improve the reliability and validity of the current study.

Firstly, the availability of the efficiently recorded data, in a well-organised audiology clinic, motivated the choice of this research design since it made it an inexpensive design. Secondly, the choice of this design for the study was influenced by the fact that NIHL has a long latency period between the exposure to noise and the disease, and studying workers who are losing their hearing would make it unfeasible and unethical to study the phenomenon in any other way. Thirdly, the potential for developing a prediction model that could be tested prospectively was also an advantage of using retrospective record review as a research design.

The disadvantages of the method were considered and for the following reasons were felt to be sufficiently controlled to prevent them from reducing the value of the research. Firstly, the researcher as a member of the audiological team when the database was being developed was aware of the standards and controls used during testing and was convinced of the accuracy of the records. There was no need for recall in the data capture since information was clearly recorded in a paper format in each patient file. Secondly, the researcher's credibility with the management at the company where the research was conducted was good as a result of long-term working relationships, and therefore access to all the necessary data was granted for the current study. The records also did not have conflicting reasons for recording information. Instead, the two tests being compared and investigated in this research were both part of the diagnostic test battery for NIHL and therefore ensured that all important data was available. Thirdly, the bias that may have resulted from using this method was excluded by the fact that three audiologists worked in the department at the time of the development of the database, and internal quality controls were performed regularly to confirm the reliability of results. Also the selection of records to be used in the study was made using statistical methods for random selection.

A disadvantage of the choice of design may have been that no blinding of the data extractors was possible. However, the researcher attempted to overcome this disadvantage by training the data capture assistant in the extraction and management of the clinical information (Gómez de la Cámara et al., 1997; Wu & Ashton, 1997) and by using agreed-to protocols and cross-checks of data capture by an independent research team member (Gilbert, Lowenstein, Koziol-McLain, Barta, & Steiner, 1996).

In conclusion, the 12 criteria suggested by Worster et al., (2005) to be applied to a research study design to ensure that reliability and validity are maintained in a research methodology are summarised in Table 14.

Table 14 Current study evaluated against Worster et al.'s (2005) criteria

Method Criterion	Method Criterion Description	Current Study
Abstractors training	Were the abstractors trained before the data collection?	Yes – an intern researcher who assisted with data abstraction for the audiogram was trained by the researcher. The researcher is a qualified audiologist with in-depth experience in the field of NIHL.
Case selection criteria	Were the inclusion and exclusion criteria for case selection defined?	Yes – a trained audiologist made the initial case selection from a randomised list of possible cases in each procedure. Inclusion criteria were used for selection of records.
Variable definition	Were the variables defined?	Yes – the variables were defined as hearing threshold levels for each frequency, DPOAE level for each frequency, age, ethnic group, occupation.
Abstraction forms	Did the abstractors use data abstraction forms?	Yes – DPOAE data was transferred electronically and audiogram data was entered into the spreadsheet with the column headings as guidelines.
Performance monitored	Was the abstractors' performance monitored?	Yes – in 10 % of the cases both data abstractors entered the same data and the accuracy was checked electronically and by an independent researcher. No discrepancies were found.
Blind to hypothesis	Were the abstractors blind to the hypothesis/study	No – because the researcher was the data abstractor and the intern was aware of the study hypotheses. This, however, was felt to be

	objectives?	an advantage since the abstractors were alerted to incorrect data owing to their insights into the hypothesis.
Inter-observer reliability mentioned	Was the inter-observer reliability discussed?	Yes – the three audiologists who worked in the department during the database development regularly retested clients to ensure reliability of results.
Inter-observer reliability tested	Was the inter-observer reliability tested or measured?	No – not formally but as discussed in Question 7 this happened informally.
Medical record identified	Was the medical record database identified or described?	Yes – the mine and the occupational health centre were described in the study.
Sampling method	Was the method of sampling described?	Yes – statistical methods to determine the sample size for a representative sample were performed on the study data using all the variables. Random sampling methods identified the starting record and every fourth record thereafter was selected. If the record met the selection criteria the data was included in the study.
Missing-data management plan	Was the statistical management of missing data described?	Yes – the case was excluded if the audiogram or DPOAE data was missing. If only data required for some analyses was missing, e.g. age or occupation, the record was included in the DPOAE analysis but excluded from the other analyses.
Institutional review board approval	Was the study approved by the institutional or ethics review board?	Yes – Wits Ethics Committee approval and mine management permission were obtained.

5.3 Research Questions

The introductory chapter outlined the researcher's clinical experience that formed the initial questions posed by the rationale for the current study. The review of the literature and the existing knowledge in the related fields further developed the background to the research questions for the study. The research questions appeared to be grouped into two areas namely those:

- relating to the addition to the current knowledge about the characteristics of DPOAE measurements in a noise-exposed population; and
- relating to the potential of developing a prediction tool for use in a noise-exposed population.

These two areas were therefore used in the study to ensure clarity of the investigation and are referred to in the following sections and chapters as “phase one” and “phase two” of the study, respectively.

5.3.1 Phase one

There is currently extensive information available about the audiological characteristics found in noise-exposed workers and, in particular, in gold miners (Abdulla, 1998; Bess & Humes, 1995; Edwards, 2002, Begley, 2006). The audiological information available is in relation to:

- the characteristics of hearing thresholds for the different frequencies on an audiogram;
- the characteristics of hearing thresholds for standard test types such as pure-tone air-conduction testing ; and
- the response characteristics on other tests such as speech recognition thresholds (SRT) and their relationship to pure-tone averages (PTAs); and
- audiological characteristics of different groups, based on age categories, ethnicity and occupation groups.

Therefore, it is hypothesized that similar unique characteristics described above may also be true for the DPOAE responses of gold miners. Specifically, that hearing threshold levels on an air-conduction audiogram would have a strong relationship with the DPOAE measurements in a population of noise-exposed workers. The research question that required investigation was therefore

> Is there a statistical correlation between air-conduction hearing threshold levels and DPOAE emission levels in a noise-exposed gold mining population?

The second hypothesis was that, since the stimulus parameters of the DPOAE test are known to evoke responses from specific positions on the basilar membrane of the cochlea (Chen & Zhao, 2007; Gorga et al., 2007; Sliwinska-Kowalska & Jedlinska, 1998), it could be assumed that if the stimulus parameters varied, the relationship with the hearing thresholds would also vary, possibly according to the proximity of the stimulus frequencies to pure-tones on the audiogram. It could therefore be hypothesised that the use of various stimulus parameters would result in different strengths of relationship between the hearing threshold levels and the DPOAE levels. The retrospective nature of the database available facilitated the possibility of also investigating the effect of using an average of the two emission strengths from a replicated measurement with a single measurement. Results from the comparison of single versus replicated stimulus parameters were expected to confirm or contest any abbreviated clinical practice for the sake of speed of testing when time constraints in a busy practice were considered.

The research question here would be:

> Is the correlation between air-conduction hearing threshold levels and DPOAE levels influenced by the stimulus frequencies used?

Pure-tone averages provide the clinician with an index that relates the thresholds from a number of audiogram frequencies with a the degree of hearing loss (ASHA, 1996). If a relationship existed between the pure-tone average and a group of DPOAE level averages, the DPOAE test could provide an objective tool similar to the pure-tone average that would form the basis of a classification of cochlea functioning. The rationale therefore for investigating groups of DPOAE level averages and their relationship to averages of various groups of pure tones was that if a relationship did exist it would provide the occupational audiologist with a useful clinical tool similar to the degree of hearing loss used in behavioural audiology. The hypothesis was that pure-tone averages were closely related to DPOAE level averages. The research question was:

Is there a statistical correlation between the pure-tone average and a DPOAE level average at f_2 frequencies closest to 500 Hz, 1000 Hz and 2000 Hz?

Fourthly, since the pure-tone average is known to be correlated to the SRT, an average of DPOAE levels at similar frequencies to the pure-tone average could be expected to be related to the SRT. If this was the case, it would provide the clinician with a useful cross-check tool that was objective and did not rely on the response of the person being tested. The research question was therefore:

Is there a statistical correlation between the speech recognition threshold and an average of DPOAE levels at similar frequencies to the pure-tone average in a gold mining population?

Another aspect of phase one that required investigation was related to demographic groupings. Previous studies showed unique hearing threshold level characteristics for different age groups, different occupation types, and different ethnic groups (de Koker et al., 2003; Edwards, 2002). The hypothesis would be that unique characteristics of DPOAE levels also exist for different age categories, different ethnic groups and different occupation types, and that these characteristics are influenced by the stimulus frequencies used for testing DPOAE levels. The research questions were:

Can different DPOAE level characteristics be distinguished for different age categories in a gold mining population and are these characteristics influenced by the stimulus frequencies used for testing?

Can different DPOAE level characteristics be distinguished for different ethnic groups in a gold mining population and are these characteristics influenced by the stimulus frequencies used for testing?

Can different DPOAE level characteristics be distinguished for different occupation types in a gold mining population and are these characteristics influenced by the stimulus frequencies used for testing?

Finally, in phase one, an area that required investigation and that would add to the existing knowledge about the characteristics of DPOAE measurement in this population could be divided into a further two aspects. The first aspect related to the need to replicate the test. As discussed in the section on OAE measurement in Chapter Four, the standard procedure is to replicate a test. However, the researcher's clinical experience appeared to indicate that only one recording seemed necessary to achieve a reliable result. The database contained data from test protocols that had performed one recording and data where the test had been replicated and this therefore facilitated a statistical comparison of single recording and replicated recordings. If the comparison showed that the single recording correlated as closely to the hearing threshold levels as the replicated recording, the busy occupational audiologist could reliably reduce the testing time for DPOAE measurement. The research question was therefore:

> Is the correlation between the hearing threshold levels and the DPOAE levels influenced by the number of recordings made?

The second aspect relating to the measurement of DPOAEs in this population that required investigation was with regard to the difference between the emission level and the noise floor (DP-NF). The discussion in Chapter Four showed that accepted clinical practice requires that the DP-NF is 10 dBSPL or greater (Hall, 2000). The research highlighted the possible use of a less stringent DP-NF difference, without compromising the validity of the results, namely that of DP-NF=6 dB SPL, 3 dB SPL and even 0 dB SPL (Bockstael et al., 2008; Konopka, Pawlaczyk-Luszczynska, Sliwinska-Kowalska, Grzanka, & Zalewski, 2005). The study population had pre-existing, varying degrees of hearing loss, and the hypothesis was that, in the noise-exposed gold mining population, where low intensity emissions predominate, DP-NF would not meet the 10 dB SPL criterion, but that despite this the DPOAE test would be useful in a noise-exposed population. The research question was therefore:

> What are the DP-NF characteristics in a gold mining population?

5.3.2 Phase two

If statistically significant correlations were found between hearing threshold levels and DPOAE levels in the gold mining noise-exposed population, the next logical step for the study would be to use the database to develop a prediction model. A prediction model would provide a valuable tool for the clinician in a difficult testing environment such as in the South African mining occupational audiology setting. The long-term goal would be to provide a method that would allow for compensation for NIHL that was based on an objective test. The discussion in Chapter Four highlighted the fact that prediction models previously reported either used different stimulus protocols or different analysis methods as the basis for the prediction model to predict either categories of hearing loss or hearing thresholds (Balatsouras, 2004; de Waal, Hugo, Soer & Krüger, 2002; Dorn et al., 1999; Gorga et al., 2000; Ma & Zhang, 1999; Seixas et al., 2004; Shaffer et al., 2003.; Shaffer & Dhar, 2006). The various statistical methods for arriving at the previously reported predictions have been used with varying degrees of success, but have not made use of the statistical technique of stepwise regression analysis.

The purpose of a prediction model for the current study is to provide a tool that will allow the South African occupational audiologist in a mining environment to use the results of an objective test (DPOAE) to arrive at frequency specific thresholds of an audiogram. These predicted thresholds can then be used as the basis for the calculation of the PLH that is needed to calculate the amount of compensation due to a worker for NIHL. The PLH, as discussed in Chapter Two, is calculated from tables weighted for speech perception. The COIDA tables (DME, 2003) currently used in South African compensation legislation have been adapted from the Australian table as discussed in Chapter Three on compensation for NIHL. It was hypothesised that, since speech perception is a function of both peripheral and central auditory processes, the measurement of only the functioning of the cochlea with a DPOAE may not provide an accurate prediction of the PLH. This would especially be the case when the weighted tables were used, and, since the current tables seem to provide excessive

weighting in the speech frequencies, the PLH calculations may not be as accurate as required for medico-legal purposes.

With these factors in mind, the requirements of the envisaged tool were therefore:

- that the model development was based on the use of DPOAEs, because of the pre-existing hearing loss in the population which excluded the use of TEOAEs;
- that the model was informed by previous work in the field that found the greatest success of prediction when a multivariate input model was used; and
- that the output of the model provided frequency specific thresholds that could be used in the current practice in the South African context where hearing threshold levels are used to calculate the PLH.

Consultation with a statistician (Steffens, 2008) about potential statistical methods to use to develop a prediction model, as well as an extensive literature and patent search of methods reported in previous studies, concluded that stepwise regression analysis would be a novel way to arrive at a prediction model. The discussion in Chapter Four on the prediction of hearing thresholds from OAEs highlights the reasons for the choice of stepwise regression analysis. Stepwise regression analysis facilitated the development of a model for each frequency of the audiogram. The process of prediction required the use of the DPOAE levels for a specified frequency used as the input variable, multiplied by the constant provided by the analysis. For some audiogram frequencies (the low and the high frequencies in particular) it was found that only one DPOAE level was needed to arrive at the best predicted level, while for other frequencies (mainly the mid frequencies) between four and five DPOAE levels were needed as input variables to arrive at the best predicated threshold. Since three different stimulus frequency procedures were available in the database, a prediction model had to be developed for each stimulus procedure and of course for each audiogram frequency

Once the model was available, the first research question in phase two was:

> How accurately can a stepwise regression analysis prediction model predict the audiogram of a noise-exposed gold miner?

The second research question for this phase related to when a predicted audiogram had been calculated:

> How accurate was the percentage loss of hearing calculated with the tables currently used in the South African compensation legislation?

The research questions and the hypotheses underlying the questions as discussed above are summarised in Table 15.

Table 15 Summary of research questions and related information

Research Question	Hypothesis
Phase one	
What is the correlation between air-conduction hearing threshold levels and DPOAE levels in a gold mining population?	Air-conduction hearing threshold levels and DPOAE levels are strongly correlated in a noise-exposed gold miner.
Is the correlation between air-conduction hearing threshold levels and DPOAE levels influenced by the stimulus frequencies used?	Different DPOAE stimulus parameters will result in different correlation strengths between the hearing threshold levels and the DPOAE levels in a noise-exposed gold miner.
What is the correlation between the pure-tone average and a DPOAE level average at f_2 frequencies closest to 500 Hz, 1000 Hz and 2000 Hz?	Pure-tone averages are closely related to DPOAE level averages.
What is the correlation between the speech recognition threshold and an average of DPOAE levels at similar frequencies to the pure-tone average in a gold mining population?	Average of DPOAE levels at similar frequencies to the pure-tone average will be strongly correlated to the speech recognition threshold.
Can different DPOAE level characteristics be distinguished for different age categories in a gold mining population and are these characteristics influenced by the stimulus frequencies used for testing?	DPOAE levels are distinguishable by age category in a gold mining population and are influenced by the stimulus frequencies used to evoke the DPOAE.

Can different DPOAE level characteristics be distinguished for different ethnic groups in a gold mining population and are these characteristics influenced by the stimulus frequencies used for testing?	DPOAE levels are distinguishable for African and Caucasian ethnic groups in a gold mining population and are influenced by the stimulus frequencies used to evoke the DPOAE.
Can different DPOAE level characteristics be distinguished for different occupation types in a gold mining population and are these characteristics influenced by the stimulus frequencies used for testing?	DPOAE levels are distinguishable by the occupation types in a gold mining population and are influenced by the stimulus frequencies used to evoke the DPOAE.
Is the correlation between the hearing threshold levels and the DPOAE levels influenced by the number of recordings made?	Hearing threshold levels will be as strongly correlated for single recordings as for replicated recordings in a noise-exposed population.
What are the DP-NF characteristics in a gold mining population?	DP-NF differences are < 6 dB SPL for the majority of frequencies measured in a noise-exposed gold mining population.
Phase two	
How accurately can a stepwise regression analysis prediction model predict the audiogram of a noise-exposed gold miner?	The prediction model can accurately predict hearing thresholds to within 5 dB HL in a noise-exposed gold mining population.
How accurate is the percentage loss of hearing calculated from the predicted audiogram?	The accuracy of percentage loss of hearing from the predicted audiogram is good.