Ollscoil na hÉireann, Corcaigh

National University of Ireland, Cork

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Statistical Applications in Health and Social Research

Volume 1 of 1

Thesis presented by

Alan Kenny, BSc

for the degree of

Masters By Research

University College Cork

Department of Statistics

Head of Department: Michael Cronin

Supervisor(s): Kathleen O Sullivan, Damian Conway and Eric Wolsztynsk

2022

Contents

Declaration

The thesis submitted is my own work and has not been submitted for another degree, either at University College Cork or elsewhere.

Acknowledgements

I wish to thank Kathleen O'Sullivan and Damian Conway for their guidance and help throughout the year. I would also like to thank Professor Finbarr O'Sullivan for his help and guidance throughout the PET project. I would also like to thank Dr. Eric Wolsztynski, Dr. Niall Fitzgerald and Paul Keegan for their help.

I would also like to thank my parents, family and friends for their support for the duration of the thesis.

Abstract

The MSc comprised four projects based on different statistical techniques. Each project is an individual chapter in the thesis and has been written independently.

The first project is based on a student-wide survey conducted by University College Cork (UCC) to examine the student experience. As this study was unique to the Irish college system, factor analysis was employed to validate the use of purpose-built scales. Reliability analysis was conducted to determine if the scales were internally consistent, and the results of factor and reliability analysis and summary results were presented.

The second project focuses on ^{15}O Positron Emission Tomography (PET) Imaging Data and its ability to quantify underlying metabolic processes, namely cerebral blood flow in an Alzheimer's disease patient. Kinetic analysis, using a 1-compartmental model, was employed and functional parameters were estimated. The technique was validated through comparisons with a non-parametric approach. This study confirmed that there was lower blood flow in brain regions that are affected by Alzheimer's disease.

The third project is an actuarial study, which explores the process known as graduation by standard table. Using information collected from a population, a set of crude rates are deduced. These rates need to be improved using the graduation by standard table method. Our graduated rates are then tested using five statistical tests to verify they are reliable. This study confirmed that use of the graduation method could improve crude rates to make them reliable.

The fourth project focuses on structural empowerment in an Irish nursing context. Our study concentrates specifically on perceptions of formal power. Univariate analysis was conducted to identify associations between demographic characteristics and perception of formal power. Multivariate regression was used to identify which

combination of demographic characteristics significantly affected a respondent’s perception of formal power. Respondents’ perceptions of formal power were in line with previous studies.

LIST OF TABLES

List of Figures

Chapter 1: STUDENT EXPERIENCE SURVEY

Supervisor: Kathleen O'Sullivan

Disclaimer: Use of this data was granted by the technical working group. The use of this data outside this report requires explicit permission on behalf of the aforementioned group from Kathleen O Sullivan.

1.1 Introduction

The student experience can be described as the degree to which students engage with and experience life in their university. This is dependent on the extent to which the students in higher level education embrace the “vibrant educational and social life that is supposed to represent the best the university has to offer” (McInnis, 1992; 1997). A recent report from University College Cork (UCC) found that more and more universities identify the student experience as being crucial to the future success of their university (University College Cork, 2009). When one considers the benefits a positive student experience brings to a university, it is little wonder that so many universities are putting such emphasis on this. Some of the key benefits outlined in the UCC report were the enhanced ability to attract students, an increase in the reputation of a university, greater levels of funding, recognition from employers, and retention of the allegiance of a university’s alumni. The report explains that the enhanced ability to attract students to a university comes from the fact that the student experience plays a crucial role in marketing and promotional strategies in recruiting students. It also explains why students leaving university with a positive student experience are generally more rounded and happier graduates/people. Rounded candidates is something that the vast majority of employers find very attractive, as the more that graduates satisfy employers’ needs, the more a university will be established as the preferred university for graduate recruitment. In addition, the report tells us how an alumni with positive experiences is more likely to be a potential donor to a university. Furthermore, because of the alumni’s profile and influence in business and in society, their involvement in a university can be very beneficial to the success of a university.

In March 2009, there were 17,026 students attending programmes of study in UCC. Up until recently, there has been only anecdotal evidence of what exactly the student experience was like for these students. In other countries, such as Australia (University of Sydney, 2009) and Britain (University of Oxford, 2008), it is the norm to assess the student experience in universities on a regular basis through the use of purpose-built surveys, with the University of Oxford and the University of Sydney student experience surveys being well developed examples of this. In 1999, the University of Sydney began collecting data on their students' perceptions of teaching and learning within the university, through the use of a survey they constructed for this task, namely, the "Student Course Experience Questionnaire (SCEQ)". The authors of this questionnaire state that the purpose of the SCEQ is "to provide the University community with a basis for strategic, faculty level academic development and curriculum review to further enhance the quality of teaching and student learning" (University of Sydney, 2009). In 2001, the University of Oxford began conducting similar surveys on their student population. Since the introduction of these surveys, it was encouragingly found in both cases that, on a longitudinal basis, students' perception of their student experience had improved, indicating an improvement in quality of teaching and student learning (University of Oxford, 2008).

While there have been some surveys carried out on the UCC population that touch on similar aspects of the aforementioned studies, a survey that focuses on the student experience in UCC as a whole does not exist. It is with this in mind that the UCC Vice President for the Student Experience (VP) set out to develop a survey to assess what the overall feelings were from students regarding their experience in UCC. The VP, through the utilisation of various academic and administrative bodies in the University, including a major input from the Students Union of UCC, succeeded in doing this and a survey on the student experience in UCC was circulated to the students of the university. The VP hoped that the analysis of this survey would tangibly tell what the complete student experience in UCC is really like.

It is our aim in this study to obtain an overall impression of the student experience in UCC. We intend to achieve this through the statistical analysis of the student experience survey. Though this survey was similar to that of the University of Oxford and University of Sydney surveys, in its focus on the quality of teaching and student support in the University, one of the key differences is its focus on the whole social aspect of university life for students. Other important aspects, such as students' financial concerns, were also examined in this survey. This, it was felt, would give a complete picture of what a multifaceted experience would look like. Via an analysis of the many aspects of the student experience identified in the survey, we intend to examine the student experience in UCC.

Conclusions on this survey will be drawn from the results of the statistical data analysis. It is anticipated that the analysis provided will enable the governing body of UCC to make more informed decisions. Subsequent policy changes will focus on the areas where it is clear the University is falling short in providing students with the opportunity to achieve the optimum experience. It is anticipated that the results of this analysis, when compared to other similar universities, will provide a yardstick for performance for all sectors of the university. In addition, the survey, if it is continued, will provide a means of tracking the success of meeting students' needs on a longitudinal basis. It is hoped that this will ultimately improve standards across the university as a result.

In 2021, there were 22,337 students attending programmes of study in UCC (University College Cork, 2021). This is an increase of over 30% of the student population since the original survey was conducted. UCC are continuing to gather information on their students' experience. Instead of directly conducting surveys themselves, they are using www.studentsurvey.ie, this is a national survey asking students across the country about their higher education experience (University College Cork, 2021). Since the survey began in 2013, there has been over 200,000 students that have taken part in the online survey. The goal of this survey is similar to the goal of this original survey, to understand and improve the student experience. This new survey asks students questions relating to their own experience, these include questions relating to their academic, personal and social development, which is online with the original study. In UCC the survey is open to applicants each year for three weeks between February and March. This is earlier than the original study. The survey takes about fifteen minutes to complete, which is quicker than the original study. This new survey is also only open to first years, final years and those in postgrad study, the original survey was open to all students. The results are a combined national result even though the surveys are open to different institutions at different times. There are also incentives to take part in the survey, these were not available in the original survey. Some of the key results of the surveys have been around utilising technology; providing targeted feedback and ways for students to improve, and more contact with employers and greater opportunity for work placement. These themes were seen in the results of the original survey. For example, one of the actions taken by UCC based on the results of the survey was to appoint work placement managers. Work placement managers interact with students interested in work placement or currently on work placement. To support students, work placement managers created online content and workshops so students could develop their employability. For example, some of these workshops focused on communication and IT skills. The work placement managers also engage with employers each year, this is part of a recruitment drive. These measures will help students become more rounded candidates for employment (Deloitte, 2018). In a recent student survey of over 44,000 Irish university students, over 58% of students believed that their university experience contributed to their knowledge, skills, and personal development and their ability to secure jobs or work-related knowledge and skills (Student Survey, 2021). A report by Deloitte in 2018 addresses the challenges of the work environment for students and society in the future. In the report, Deloitte suggest in around 15 years certain traditional jobs or jobs where employment are high presently may not be available or greatly reduced. In their place or substitutes will be jobs that do not exist today. Technology and new skill sets will be key for students, universities are one of the key stakeholders in ensuring graduates are ready. Through engagement with different stakeholders, universities can develop a teaching and learning framework, which will prepare the students as much as possible and also allow them to have the best student experience at university (Student Experience, 2021). Similar surveys continue in the UK and Australia. In the UK, student experience continues to be monitored and influences university life and teaching experiences (Student Experience, 2021). In the University of Sydney they continue to monitor student experience at the university with different surveys. For example, in 2015 the University of Sydney commenced a survey designed specially to get feedback on students' learning experiences, the Unit of Study Survey (USS). This short questionnaire was compiled of just six core questions to provide insight into the students' learning experience. In 2020 the survey was increased by two further questions to understand students' experience of transitioning to online learning due to the Covid19 outbreak. The results showed a positive score for students. A positive result would indicate the university are catering for their students in the online space in a positive way. Student experience is constantly changing, however these surveys allow the college to measure student experience and make policies which influence or shape student experience. The survey results are very powerful tools for moving the University management and policies in the same direction as their students.

1.2 Methods

This section will describe the questionnaire from which the information was gathered. It will also detail the study design and the survey participants, along with a description of the data collection method. Finally, an outline of how the data was analysed is included.

1.2.1 Participants

All students registered in University College Cork (UCC) in March 2009 were contacted via email to partake in the survey. An email was sent to 17,026 UCC students, with an attached link to the questionnaire. A total of 3,356 students participated in the survey, which represents 20% of the student population in UCC.

1.2.2 Instrument – Questionnaire

The instrument used for this survey was an online, self-completed questionnaire. Construction of the instrument involved focus groups, input from the students' union, input from academic and administration groups and through the utilisation of instruments from both the University of Sydney and the University of Oxford Student Course Experience Questionnaires. Students' understanding of the student experience at UCC was identified through focus groups of between eight and nine randomly selected students. Different focus groups highlighted different areas that they perceived to determine a student's experience. A steering group, which consisted of members from the various administration and academic groups in UCC, determined what questions to include in the questionnaire, based on the key areas identified by the focus groups.

A technical working group of members from the Departments of Statistics, Computer Science and Psychology reviewed the questions and finalised the questionnaire for distribution in a pilot study. The pilot study was conducted amongst ten randomly chosen students and, after some minor rewording, the questionnaire was distributed to the student body.

The final questionnaire consisted of 168 questions, which comprised categorical, numerical, and qualitative variables. The specific aspects this study encompassed included: (1) demographics, (2) satisfaction with university life, (3) social aspects and integration, (4) teaching and learning, (5) financial concerns, and (6) pride in UCC.

A detailed description of each section, and how the results for each individual question are presented, are described below.

1.2.2.1 Demographics

The first section in the survey focused on the students' demographics. This consisted of 10 questions that are classified into three categories. Category one focused on demographic details such as gender and nationality. Category two looked at the academic details of the student, such as what qualification the student is studying for, college, student status, and year of study. Category three then examined the respondents' process of entry into UCC. The frequency and percent of each response was reported. A breakdown of the demographics can be found in the Results section of this document.

1.2.2.2 Likert Scales

Throughout the method sections, there is an assessment of the respondents' answers to questions based on Likert scales in this questionnaire. Likert scales were used to capture the respondents' answers to different questions. There were two Likert scales used: a five and a seven-point Likert scale.

For example, the 7-point Likert scale was used for the Satisfaction with University Life (SWUL), (**section 1.2.2.3**). A 5-point Likert scale was used in Social Aspect and Integration (**1.2.2.4**). An example of a 7-point and 5-point Likert scale are outlined in **Table 1.1** and **Table 1.2** respectively.

Table 1.1: 7-Point Likert Scale.

Number	Individual's Response
1	Strongly Dissatisfied
2	Dissatisfied
3	Slightly Dissatisfied
4	Neither Satisfied nor Dissatisfied
5	Slightly Satisfied
6	Satisfied
7	Strongly Satisfied

Table 1.2: 5-Point Likert Scale.

Number	Individual's Response
1	Strongly Disagree
2	Disagree
3	Neither Agree nor Disagree
4	Agree
5	Strongly Agree

The possible response to a particular 5-point or 7-point Likert scale depends on the question. In this study for example, a question asked how socially connected a student felt in UCC, the question was assessed on a 5-point Likert scale, students could choose 1 very isolated to 5 not isolated.

For questions that used the 5-point Likert scale in **Table 1.2**, a respondent is said to be in broad agreement if they choose any of strongly agree, agree or neither agree nor disagree with an item, i.e., if they chose 3 or higher. Respondents are in agreement if they choose any of strongly agree or agree with an item, i.e., if they chose 4 or higher. Furthermore, the results are reported as a percentage of respondents for each question unless started otherwise, percentage broad agreement is the proportion of students who indicate that they strongly agree, agree or neither agree nor strongly disagree with an item. Percentage agreement is the proportion of students who indicate that they strongly agree or agree with an item.

For questions that used the 7-point Likert scale in **Table 1.1**, a respondent is said to be broadly satisfied if they choose any of strongly satisfied, satisfied, slightly satisfied or neither satisfied nor dissatisfied with an item, i.e. if they chose 4 or higher. Furthermore, the results are reported as a percentage of respondents for each question unless started otherwise, percentage broadly satisfied is the proportion of students who indicate that they strongly satisfied, satisfied, slightly satisfied or neither satisfied nor strongly dissatisfied with an item. Percentage satisfied is the proportion of students who indicate that they strongly satisfied, satisfied, or slightly satisfied with an item.

1.2.2.3 Satisfaction with University Life (SWUL)

To determine a students' perception of their satisfaction with university life, a scale was utilised. The SWUL scale measures students' overall assessment of satisfaction with life in university and consisted of 5 questions in the questionnaire. It is a modification of the well utilised 'Satisfaction with Life Scale', developed by Denier, Emmons, Larsen and Griffin (1985) as a multi-item scale to measure respondents' overall life satisfaction. The original scale did not assess satisfaction with other life domains such as health or finances but did allow respondents to weight different aspects of their life in whatever way they chose. The modified SWUL scale items were slightly reworded for the university setting. The questions as they appeared in the questionnaire can be seen in **Table 1.3.**

Table 1.3: Items Contributing to Satisfaction with University Life Scale.

Items
In most ways my life in university is close to my ideal.
The conditions of my life in university are excellent.
I am satisfied with my life in university.
So far, I have gotten the important things I want in my life in university.
If I had my time in university again, I would change almost nothing.

The respondent was asked to rate the items using the 7-point Likert Scale **(1.2.2.2).** An overall score is calculated by summing the individual responses, giving a minimum overall score of 5 and a maximum score of 35. A high score reflects a high level of satisfaction with life in university. A respondent must have answered all items in the scale.

Those who have a low overall score of 5-9 are described as extremely dissatisfied, scores of 10-14 are dissatisfied, and 15-19 are slightly dissatisfied with their life in university. Respondents with an overall score of 20 are said to be neutral. Those with overall scores of 21-25 are slightly satisfied, 26-30 are satisfied, and 31-35 are extremely satisfied with their university life. A respondent is said to be broadly satisfied with a score between 20 and 35 and satisfied with a score between 21 and 35.

1.2.2.4 Social Aspects and Integration

This section focuses on how satisfied respondents were with social aspects of life in university and how socially connected they feel. For each question, respondents had the option to choose "not applicable", if a respondent chose this option they were not included in the results.

Respondents were asked to rate how satisfied they were with the social aspects of their life, their level of satisfaction with social events organised for students, and how satisfied they were with security and safety on campus, on a 7-point Likert scale. It ranged from 1 to 7, 1 being "Very dissatisfied" to 7 "Very satisfied".

Respondents were also asked how socially connected they felt in UCC, on a 5-point Likert scale. Students were asked to state whether they felt very isolated, isolated, neither integrated nor isolated, socially integrated, or very socially integrated, which ranged from 1 to 5 respectively. A respondent was deemed broadly socially integrated with a score of 3 or higher and socially integrated with a score of 4 or higher.

1.2.2.5 Teaching and Learning Instrument

The section focuses on students' perceptions of several dimensions that encompass teaching and learning. The 'Teaching and Learning' instrument comprised three defined scales: the Good Teaching Scale, the Appropriate Workload Scale, and the Motivation Scale.

Good Teaching Scale (GTS)

The GTS was first used in a student experience questionnaire by the University of Sydney in 1999 and comprised of six items (http://www.itl.usyd.edu.au.sceq). It measures students' perceptions of the teachers' abilities to contribute to student learning and the scale is characterised by teaching practices, which include providing students with useful and timely feedback on their progress, motivating students to do their best work, making an effort to understand difficulties that students have, clear explanations, and making subject matters interesting. The six items that comprise the scale are shown in **Table 1.4.**

Table 1.4: Items Contributing to the Good Teaching Scale.

Items
My teachers/lecturers normally give me helpful feedback on my progress.
The teachers/lecturers in this programme motivate me to do my best work.
My teachers/lecturers make a real effort to understand any difficulties I have.
My teachers/lecturers are extremely good at explaining things.
The teachers/lecturers work hard to make their subjects interesting.
My teachers/lecturers put a lot of time into comments (orally and/or in writing) on my work.

The scale consists of six items on a 5-point Likert scale. Each item is recalibrated to a -100 to +100 scale, where -100 represents strongly disagree, -50 represents disagree, 0 corresponds to neutral, 50 represents agree, and 100 represents strongly agree. A score above 0 indicates a positive response. A respondent's scale score is computed by averaging over all six items. Only those who responded to all six items had a score computed. Higher scores indicate that students perceive that good teaching practices are present, while a low score reflects that these practices occur less frequently.

Appropriate Workload Scale (AWS)

The University of Sydney also used the AWS as part of their initial survey in 1999 (http://www.itl.usyd.edu.au.sceq). The scale shows the degree to which respondents feel that the workload associated with their course is manageable. It centres on the academic pressure respondents feel they are under, the weight of their academic workload, and the length of time available to fully understand new material. The AWS scale consists of four items, as seen in **Table 1.5.**

Table 1.5: Items Contributing to the Appropriate Workload Scale.

Items
There is a lot of unwanted academic pressure on me as a student*
My academic workload is too heavy*
I am generally given enough time to understand things I have learned
The volume of work necessary for my programme of study means that it cannot all be thoroughly comprehended*

*Statements that had to be reverse scored.

The scale consists of four items on a 5-point Likert scale, each item is recalibrated as per GTS. Three of the four items are negatively phrased. The responses to which must be reverse scored so that items which infer a negative attitude are reverse scored so that disagreement, which infers a positive attitude, receives a positive score. A respondent's scale score is computed by averaging over the items (once the negatively phrased items have been reverse scored). This is computed for respondents who answered all items. Higher scores on this scale indicate a perception of reasonable workloads.

Motivation Scale (MS)

The MS was introduced into the University of Oxford's student experience questionnaire in 2003 (http://ceq.oucs.ox.co.uk/questionnaire.cfm). The scale measures the extent to which all aspects of the course are motivating the students. This includes the students' feelings as to the levels of intellectual stimulation, motivation, encouragement for further learning, involvement with a community of scholars, and expectations and intellectual standards involved in their course. The AWS scale consists of nine items, as seen in **Table 1.6.**

Table 1.6: Items Contributing to the Motivation Scale.

Items
My programme of study is intellectually stimulating.
I find my programme of study motivating.
My programme of study has stimulated my enthusiasm for further learning.
My programme of study has stimulated my interest in the field.
I feel that I benefit from being in contact with active researchers/scholars.
The academic expectations of me on my programme of study are too high.*
Intellectual standards at UCC are set too high.*
I feel part of a community of scholars who are committed to learning.
Being selected to study at UCC is a source of motivation to me.

*Statements that had to be reverse scored.

The scale consists of nine items on a 5-point Likert scale, each item is recalibrated as per GTS. Two of the nine items are negatively phrased. The same approach for negatively phrased questions was adopted as per AWS. Higher scores indicate that students are more likely to have a higher quality approach to learning, outcome of learning and degree result.

For the three scales in the 'Teaching and Learning Instrument', the individual questions were reported in terms of the percentages broad agreement and agreement as defined in **1.2.2.2.** Responses of 'not applicable' were excluded from these summary statistics. Other aspects of teaching and learning are described in the next section.

1.2.2.6 Additional Teaching and Learning Aspects

Additional aspects of teaching and learning consist of the student learning community, programme of study components, timetable, workload and assessment, work experience, and transferable skills. These six aspects are outlined in this section.

Student Learning Community

This aspect comprised of a set of six questions that examine a respondent's level of commitment to study, self-assessment of attendance and preparation for lectures and the extent to which respondents utilise Blackboard, an e-Educational learning platform, and whether they felt teaching staff made sufficient use of this platform.

Students' commitment to study: a respondent was asked to describe their level of commitment to their studies by stating whether they described their commitment as (1) bad, (2) poor, (3) average, (4) good, or (5) very good. A respondent is deemed as being broadly committed to study with a response of 3 or higher and committed to study with a response of 4 or higher.

A respondent's attendance and preparation for lectures was examined in terms of three questions. Respondents were asked to answer each question by stating whether they (1) never, (2) rarely, (3) sometimes, (4) usually, or (5) always performed these activities. A respondent attendance was included as 'sometimes to always' with a response of 3 or higher and 'usually to always' with a response of 4 or higher.

The responses to this question are reported in terms of percentages 'sometimes to always' and percentages 'usually to always'. Percentages 'sometimes to always' is the sum of the 'sometimes', 'usually' and 'always' ratings, expressed as a percentage of the total number of respondents for the question. Percentage 'usually to always' is the sum of the 'usually' and 'always' ratings, expressed as a percentage of the total number of respondents for the question.

To determine the extent to which respondents considered Blackboard to be an important learning resource, respondents were asked to answer the question by stating whether they considered Blackboard to be (1) very important to (5) very unimportant. A respondent deemed Blackboard broadly important with a response of 3 or lower and usually to always with a response of 2 or lower. Respondents could also answer 'not applicable' to the question. Percentage broadly important and percentage important were the statistics used to summarise the responses to this question. Percentage broadly important is the proportion of students who indicated they did not consider Blackboard to be somewhat unimportant or very unimportant. Percentage important is the proportion of students who indicated they considered Blackboard to be a somewhat important or very important learning resource. The summary statistics did not include responses of ‘not applicable’.

To assess the extent that respondents felt the teaching staff made sufficient use of Blackboard, respondents were asked to answer the question by stating whether they felt that the teaching staff made sufficient use of Blackboard: not at all, somewhat, or very much. Although respondents could answer ‘not applicable’, these responses were not reported.

Programme of Study Components

This aspect focuses on the respondents’ level of satisfaction with various facets of their programme of study, such as the respondents’ most and least satisfactory teaching experience, their satisfaction with promptness of feedback, and how the feedback helps with future exams and assignments and their teaching, tutorial, and research supervision components of their programme. Respondents were asked to rate their level of satisfaction on a seven-point Likert scale, ranging from strongly dissatisfied to strongly satisfied. Respondents could answer ‘not applicable’, these were not used as part of results.

The final facet contained in this section determines if respondents feel that all of the information they receive in their programme of study will prove valuable in their careers. Respondents answer in terms of not at all, somewhat, or very much so. They could also answer ‘not applicable’, which was excluded from the summary statistics. The result for this question is given in terms of percentages for each response.

Timetable

This aspect, which contained three questions, examined respondents’ assessment of their timetable as follows: if respondents had sufficient time to travel between lectures, if they had long gaps in their timetable, and if so, were these gaps disruptive? Respondents were asked to answer each question by selecting one of the following options: never, rarely, sometimes, usually, or always. The responses were summarised using percentage ‘sometimes to always’ and a percentage ‘usually to always’.

Percentage ‘sometimes to always’ is the sum of the ‘sometimes’, ‘usually’ and ‘always’ ratings, expressed as a percentage of the total number of respondents for the question. Percentage ‘usually to always’ is the sum of the ‘usually’ and ‘always' ratings, expressed as a percentage of the total number of respondents for the question.

Workload and Assessment

This aspect examines the satisfaction level respondents had with certain academic components of their programme of study and consisted of three questions. The first of these questions focused on whether respondents felt that standards applied in grading assignments were consistent across teachers/lecturers and departments. Respondents could answer this question by one of three responses: not at all, somewhat, or very much so. They could also answer ‘not applicable’, which was excluded from the summary statistics. The result for this question is given in terms of percentages for each response.

Respondents were also asked to what extent they felt lecturers were available to them and to what extent they felt the deadline for assignments were student friendly. Respondents were asked to answer these questions by choosing one of the following responses: not at all, somewhat, moderately, very much or greatly. Respondents could also answer ‘not applicable’, which was excluded from the summary statistics. The results for these questions are given in terms of percentages for each response.

Work Experience

Respondents were asked if they had or will be offered work experience as part of their programme of study in UCC. Respondents were asked to answer the question by giving a yes or no response. The proportion of students who responded yes expressed as a percentage of the total number of respondents is reported.

Respondents were then asked if they have had work experience, to what extent they felt it was valuable. Respondents were asked to answer these questions by choosing one of the following responses: not at all, somewhat, moderately, very much or greatly. Respondents could also answer ‘not applicable’, which was excluded from the summary statistics. The results for these questions are given in terms of percentages for each response.

Finally, respondents were asked if work experience should be a part of their programme of study. Respondents were asked to answer the question by giving a yes, no or don't know response. The percentage of respondents who answered yes, no' or don't know are presented.

Transferable Skills

This aspect of the questionnaire asked the respondents to rate the likelihood that they would make use of training in nine listed areas if it were available in UCC. Respondents were asked to answer the question by stating whether they would definitely, very likely, likely, neither likely nor unlikely, unlikely, very unlikely or definitely not avail of this training. Respondents could also answer 'not applicable' to the questions. Percentage broad likelihood and percentage likely were reported as summary statistics. Percentage broad likelihood is the proportion of students who indicated they were not unlikely, very unlikely, or definitely not to avail of training within the different listed areas. Percentage likely is the proportion of students who indicated they are definitely, very likely or likely to avail of each individual skills training. Responses of 'not applicable' were excluded from the summary statistics. A respondent was deemed to be broadly likely to avail with score of 4 or higher and likely to avail with a score of 5 or higher.

1.2.2.7 Financial Concerns

This section of the questionnaire, consisting of six questions, concentrated on the financial concerns of the respondents. Respondents were asked to rate their level of agreement which each statement on a 7-point Likert scale ranging from very strongly disagree to very strongly agree. Percentage broad agreement and percentage agreement are reported as summary statistics. Percentage broad agreement is the proportion of respondents who indicated they did not mildly disagree, strongly disagree, or very strongly disagree with each statement. Percentage agreement is the proportion of students who indicated they mildly agree, strongly agree, or very strongly agree with each statement.

1.2.2.8 Pride

The final section of the questionnaire determined the respondents' pride in being a member of UCC. Respondents were asked to answer, giving a yes or no response. The percentage of students who felt proud to be a member of UCC is reported.

1.2.3 Study Design

The design of this study is categorised as a descriptive study, consisting of quantitative and qualitative questions. A descriptive study aims to "describe the occurrence of variables, underlying dimensions in a set of variables or in the relationship between and among variables" (Heppner, Kivlighan & Wampold, 1998). As such, there is no underlying hypothesis in a descriptive study.

Quantitative research "asserts that the major constituents of the world are physical objects and processes" (Heppner et al., 1998). There are two specific types of design: survey and classification. In general, as stated in Heppner et al. (1998), "the basic aim of survey research is to document the nature or frequency of a particular variable within a certain population". In this case, the attitudes towards a broad aspect of the student experience in UCC are recorded. The population for this survey includes all current students of UCC in the academic year 2008-2009. Therefore, specific information has been gathered via a questionnaire, with the aim of describing the student experience in UCC and the various aspects that make up the student experience as a whole.

Classification research is used for identifying underlying dimensions or groupings in a set of variables (Heppner et al., 1998). As the specific instruments (i.e. SWUL, Teaching and Learning Instrument) in the questionnaire have not been previously used in an Irish university context, these need to be validated and the items constituting the instruments require simplification by examining factors that underlie these scales.

In contrast, qualitative research is defined as "an inquiry process of understanding based on distinct methodological traditions of inquiry that explore a social or human problem" (Cresswell, 1994). Both kinds of research are evident in the questionnaire, with a mix of both quantitative and qualitative questions.

1.2.4 Procedure

Students who followed the link provided in the email were directed to the online survey. The survey was set up on the online survey engine UCCASS (Unit Command Climate Assessment and Survey System). The survey remained open for one month, during which three reminders of the survey's deadline were sent out. Advertisements were placed around the campus, in the student newspaper 'UCC Express', and on the 'Blackboard' e-Education platform to encourage participants to complete the questionnaire. Three prizes comprising a laptop computer, a holiday, and an iPod were offered in a raffle as incentives to complete the questionnaire.

For each survey question, the categories and their assigned numeric code were recorded. The data was extracted from the online database in a comma separated value format (.csv) which is readable by Microsoft Excel. The data was then exported in numerically coded format and imported into Microsoft Excel. An advantage of using an online self-completed questionnaire is the low cost and the ability to reach many students at one time. The ease of data collection also makes it convenient for manipulation with suitable software. To ensure confidentiality, the student numbers, which were recorded for the prize draw purposes, were kept separate from the survey responses at all times. Students were also permitted to withdraw from the survey at any time.

1.2.5 Data Preparation

Data analysis was carried out using PASW 17.0 (Predictive Analytic Software, formerly SPSS) for Windows. The data was received in Microsoft Excel format, with the responses to qualitative questions removed. This was then imported into PASW for analysis. Variable labels were added to each variable, i.e. the statement from the questionnaire was attached to the relevant item. Value labels were also added to each variable, to record the meaning of coded values of categorical variables, which were obtained from a coding frame. For example, in question 1 'What gender are you?' the responses, Male = 1 and Female = 2. For an ordinal variable, for example question 16 'Rate the extent to which you are satisfied with the social aspects of your life', with responses ranging on a 7-point Likert scale, coded as follows: 1=very dissatisfied, 2=dissatisfied, 3=slightly dissatisfied, 4=neither satisfied nor dissatisfied, 5=slightly satisfied, 6=satisfied and 7=very satisfied. These variables are based on ranking and are given an assigned numerical code; however, the code has no mathematical meaning, besides ordering the responses.

The variable 'Area of study' was recoded into a new variable, showing which of the four colleges in UCC the student was a member. A new variable was created called 'College'. Responses of 1 (Arts, Celtic Studies, Social Sciences) in area of study was coded as 1, College of Arts, Celtic Studies and Social Sciences. Responses of 2 and 3 (Law and Commerce) were coded 2, College of Business and Law. Responses of 4, 5, and 6 (Science, Engineering and Food Science and Technology) were recoded to 3, College of Science, Engineering and Food Science. Finally, responses of 7 (Medicine and Health) were coded 4, College of Medicine and Health. Finally, respondents who answered 'not applicable' to any of the questions were coded as a missing value and therefore did not contribute to the analysis.

1.2.6 Data Analysis

This section will explain the methods used to report the responses to individual questions from the questionnaire. We will detail the use of statistical methods such as the Chi-square test to verify the respondents were representative of the UCC population as a whole; factor analysis to validate the use of scales in an Irish university context; reliability analysis to confirm the internal consistency of the scales; and a description of the formulation and method of reporting overall scale scores.

1.2.6.1 Representative Sample

To compare the demographic distribution of the sample of respondents to that of the UCC population, Chi-square goodness-of-fit tests were conducted (Sheskin, 2004).

The goodness-of-fit test summarises the discrepancy between observed values and the values expected under the model in question. The measure of discrepancy that is commonly used is the Chi-square statistic, given as follows:

$$x^2 = \sum_{All\_cells} \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the observed values and E_i is the expected values.

The null hypothesis testing procedure formed in this case states that the distribution of the observed sample is the same as the distribution of the expected population for each demographic variable obtained from the questionnaire. If the test statistic is contained in the critical region, the null hypothesis is rejected. A 5% significance level is used for this testing procedure.

The assumptions of the Chi-square test are that the data is categorical or nominal in nature, that the data consisted of independent samples, and that the expected cell count be greater than or equal to five for variables with two categories (Sheskin, 2004), while variables with three or more categories must have a cell count of at least one.

The demographic variables examined were gender, nationality, what qualification a respondent is studying for, college, status, year of study, and process of entry. If it is determined that the observed sample is representative of the overall student population, then further analysis can be performed.

1.2.6.2 Validation of Scales

Factor analysis was employed to all items in the SWUL scale (consists of 5 items) and Teaching and Learning Instrument (which contains the 19 items that GTS, AWS and MS consist of) to determine if they were applicable in a UCC context. Factor analysis is used to simplify complex sets of data and reduce a large number of variables to several key factors that underlie the relationships in the data. For this study, we hope to validate that all items load on their original scales.

Factor analysis is an interdependence technique, that is, the items are not classified as dependent or independent but examined as a whole (Hair, Black, Babin, Anderson & Tatham, 2006; Kim & Mueller, 1978; Kline, 1994). Factor analysis aims to simplify correlations among a large group of items by defining sets of items that are highly interrelated. These sets of interrelated items are the factors, which can then be interpreted to determine the relationship between the items. Factor analysis summarises the information contained in the data by defining factors that are representative of the original data with minimal loss of information. Factor analysis assumes that the original items are linear combinations of the underlying factors and that these factors are responsible for the covariation among the variables. The factor analysis model is as follows:

$$X_j = \lambda_{j1} f_1 + \lambda_{j2} f_2 + ... + \lambda_{jm} f_m + e_j$$

Where $j = 1, 2, 3, \dots, n$. $\{\lambda_{ij}\}$ are the factor loadings, so that λ_{jk} is the loading (correlation) for the j^{th} item on the k^{th} factor. e_j describes the residual variance specific to the j^{th} item. The factors {fm} are assumed to be uncorrelated, though this is dependent on the factor rotation method used.

The basic assumption of factor analysis is that some underlying structure does exist in the set of items which explains the interrelationship between them (Hair et al., 2006; Kim and Mueller, 1978). Sample size is also important when employing factor analysis. A rule of thumb states that there should be a least five observations for each item being analysed. It is also desirable to have a minimum of three items representing each factor. Assuming that the conceptual requirements for the items included in the factor analysis has been met, the next step is to ensure that the data is sufficiently intercorrelated to produce representative factors. Measures of intercorrelation are used to determine if it is appropriate to employ factor analysis to the data.

To determine whether that data is appropriate for factor analysis, a measure of sampling adequacy is used (Hair et al., 2006; Pett, Lackey & Sullivan, 2003). This measure quantifies the degree of intercorrelation between items and is calculated using the Kaiser-Meyer-Olkin (KMO) statistic. The KMO statistic ranges from 0 to 1 and can be calculated for each individual item or the sum of all the items (the overall KMO). The statistic is based on correlations and partial correlations of the overall KMO statistic and is given as follows:

$$\frac{\sum_{i\neq j}\sum r^2{}_{ij}}{\sum_{i\neq j}\sum r^2{}_{ij}+\sum_{i\neq j}\sum a^2{}_{ij}}$$

Where r_{ij}^2 is the correlation coefficient and a_{ij}^2 is the partial correlation coefficient.

The following guidelines are used to interpret the KMO values (Pett et al., 2003): above 0.9 is marvellous, a value in the 0.8's is meritorious, the 0.7's is middling, the 0.6's is mediocre, the 0.5's is miserable and below 0.5 is unacceptable. Small KMO values indicated that there may be a problem with sampling. The partial correlations should not be large if the data is likely to factor well. For this study, a KMO value of 0.5 or above is deemed appropriate to carry out factor analysis.

The overall KMO statistic is first calculated and interpreted based on the guideline range of values. If the overall KMO value is less than 0.5, a KMO statistic for each individual variable is computed. A variable with a KMO value less than 0.5 is dropped and this process continues until the overall KMO values rises above 0.5.

Another method used to measure the strength of the relationship between variables is the Bartlett test of sphericity (Hair et al., 2006). The test is used to determine if correlations are present among the variables. The null hypothesis assumes that the correlation matrix (R) is equal to the identity matrix (I), where each variable is correlated perfectly with itself but not correlated with the other variables. The null hypothesis states that $H_o: R = I$, while the alternative hypothesis states that $H_a: R = I$. That is, the items in the data are uncorrelated. A true null hypothesis implies that there is no correlation among the items, thus indicating that the data is not suitable for factor analysis. It is necessary to reject the null hypothesis at a statistically significant level to proceed with factor analysis.

The next step is to choose a factor extraction method. There are many different methods that can be used to obtain factor solutions (Kline, 1994):

- Principal component analysis (PCA) is the most common form of factor analysis and seeks a linear combination of items to form factors such that the maximum variance is extracted from the items.

- Common factor analysis (CFA) seeks the smallest number of common factors which can account for the common variance in a set of variables. CFA differs to PCA, the difference arises in the calculation of the communalities for the correlation matrix. In this case the diagonals elements are not 1's.
- Maximum likelihood factoring (MLF) obtains factors by explaining as much of the variance in the population correlation matrix as possible, whereas PCA uses the observed correlation matrix.
- Alpha factoring yields factors based on maximising the reliability.

In this study, PCA is used to extract factors, as it is the most common method and accounts for the total variance of the items. PCA separates the data into clusters which are highly correlated based on their characteristics or similar characteristics. These clusters (or factors) are plotted on a two-dimensional PCA plot. The clusters are then ranked in level of importance, the principal component axis one is the highest ranking. The level of priority between clusters depends on the distance between clusters in the PCA plots.

PCA seeks a linear combination of items (Hair et al., 2006; Mardia, Kent & Bibby, 1979). This linear combination forms a factor that explains the maximum variance in the items. PCA is obtained by the eigen-decomposition of the correlation matrix. Once the eigenvector-eigenvalue pairs are calculated, the eigenvectors are ordered based on their associated eigenvalues. The eigenvector with the highest eigenvalue is the first principal component in the data set. Therefore, the components explain the magnitude of variance in decreasing order. The eigenvectors represent the degree of correlation that the data has with a given component. PCA is an iterative process that extracts one component at a time. The first component explains as much of the variability as possible. This process is repeated, with the remaining components of the original data, until all of the variance is explained. One can go on to extract components until all the variance is explained, although this may lead to components which explain very little of the variance. Therefore, is it necessary to employ a set of criteria to determine how many factors to extract.

The two primary methods used to determine the number of factors to extract are the latent root criterion and the scree plot. The latent root criterion is one of the most commonly used methods to determine how many factors should be extracted and is based on a factor's eigenvalue (Hair et al., 2006). A factor's eigenvalues is defined as the sum of squares of the factor loadings of each factor and represents the amount of variance in the sample accounted for by a factor. Therefore, only factors that have eigenvalues greater than 1 are considered significant enough to extract.

The second method is to examine a scree plot (Hair et al., 2006; Kline, 1004). A scree plot is a graph of eigenvalues plotted against the factor number. The factors are plotted in order of extraction. A sharp change on the slope of the line indicates how many factors should be extracted.

After determining the number of factors to extract, it is necessary to interpret what these factors mean. However, initial factor extraction may not produce interpretable factors, therefore it is necessary to rotate the factors.

During factor analysis, a factor loadings matrix is computed (Hair et al., 2006; Kline, 1994). A factor loading is the correlation between an item and a factor in a given data set. The higher the factor loading, the more representative the item is of the factor. Therefore, we wish to exaggerate the factor loadings and make the larger loadings larger and the smaller loadings smaller so that each item is associated with a minimum number of factors. The factor loadings matrix is rotated to facilitate an easier interpretation of the derived factors by identifying the rotation that produces values as close to -1, 0 and 1 as possible.

There are two different techniques used to rotate factors (1) orthogonal rotations and (2) oblique rotations (Hair et al., 2006, Kline, 1994). When factors are rotated orthogonally, they are rotated so that they are always at right angles to each other and are therefore uncorrelated. This rotation technique creates a simple structure. Oblique rotations allow the factors to take up any position in factor space and are not constrained like orthogonal rotations. The oblique rotation techniques allow for correlations to exist between the factors.

As we wish to produce uncorrelated factors, an orthogonal rotation technique is used. There are three main orthogonal rotation methods: (1) quartimax, (2) varimax, and (3) equimax (Hair et al., 2006).

Quartimax rotation rotates a factor so that it loads high on one factor and low on all other factors. Quartimax minimises the number of factors used to explain each item. Varimax rotation maximises the sum of variance loadings in the factor matrix, each factor will have either a large or small factor loading on each item. Finally, equimax rotation is a compromise between varimax and quartimax. For this study, we will use the varimax rotation method as it is the most common of the orthogonal rotation methods.

Before interpreting the factors, we must determine which items load on which factor by examining the rotated factor loading values. We evaluate the rotated factor loadings in terms of their practical significance. Hair et al. (2006) provide the following criteria for assessing the significance of the factor loadings. Values of 0.5 or above are practically significant, 0.4 or above are more important and values of 0.3 or above meet the minimal level. For this study, an item should have a value of at least 0.4 on its primary factors and less than 0.4 on all other factors for this item to be retained. If an item does not load on any factor, it should be analysed on its own. Items with a higher loading are considered to be more important and to have greater influence on that factor. In addition, a factor requires a minimum of three items to load on that factor to justify retaining it.

After it has been determined, which factor an item loads on (if any), this factor solution can then be given a meaningful label. Items with higher loadings are given more emphasis when labelling a factor. A label is given to the extracted factor which represents the underlying dimension of a given factor. Items can then be brought together to formulate a score for that factor.

1.2.6.3 Reliability and Internal Consistency of Scales

After factor analysis is carried out, the internal consistency of each derived factor is computed. If a scale has an acceptable reliability level, then an overall score can be created.

Reliability Analysis

Reliability is an assessment of the degree of consistency between multiple measurements of an instrument (DeVellis, 1991; Marsden & Wright, 2010). We want all items to measure the same construct; whether it is a uni-dimensional scale or a single dimension of a multi-dimensional instrument. This type of analysis is used infrequently in scale development. It is important to prove that multiple items are equivalent measures of the same underlying score. The goal of designing a reliable instrument is for scores on similar items to be related (internally consistent), but for each to contribute some unique information as well. There are four common reliability estimation methods, given as follows:

- Internal consistency is a measure based on the correlation among the items contained in the same construct.
- Split-half reliability is a method used when a test is split in two and the scores obtained from both segments are compared.
- Test-retest reliability is a method used when a test is administered to a set of individuals and the same test is administered again at a later date. The score from both tests are compared.
- Inter-rater (inter-observer) reliability is a method used when the researcher uses more than one observer or individual to rate the same item. The scores are compared.

Internal consistency uses the correlations among all items in a construct rather than comparing scores and is the method of reliability estimation used in this study. When items form a scale, they should be internally consistent. DeVellis (1991) states that scales are meant to measure a "single phenomenon". We want the items to be highly correlated. A coefficient used for measuring internal consistency is the Cronbach's alpha coefficient, given as follows:

$$\alpha = \frac{k}{k-1}.(1-\sum_{i=1}^{k}\frac{\sigma_i^2}{\sigma_y^2})$$

where k is the number of items, σ_t^2 is the variance of the i^{th} item and σ_y^2 is the variance of the total score. A multiple item construct is internally consistent if there is a high level of inter-correlation between the items. So, if the Cronbach's alpha coefficient lies within the acceptable range, then the construct has a high level of internal consistency and an overall score can be computed.

Guidelines for interpreting the Cronbach's alpha values, outlined by DeVellis (1991), are given as follows: values less than 0.6 are unacceptable, values between 0.60 and less than 0.65 are undesirable, values between 0.65 and less than 0.70 are minimally acceptable, values between 0.70 and less than 0.80 are respectable, values between 0.80 and less than 0.90 are very good; if values of 0.90 or above are obtained, the researcher may consider shortening the scale.

For this study, a value of at least 0.7 is necessary to determine the reliability of a construct. Once reliability is attained, overall score can be calculated.

1.2.6.4 Descriptive Analysis of Scale Scores

If the SWUL scale and the Teaching and Learning Instrument, which consists of the GTS, AWS and MS, are internally consistent, an overall score can be created for each scale. A measure of centrality is used to describe each summated score. Histograms are examined to determine whether it is appropriate to report the mean or the median value. The associated measure of dispersion is also reported.

1.3 Results

In this section, the data obtained from the survey is examined and the results of the analyses are presented. Firstly, a profile of the respondents will be outlined, before the results from factor analysis and reliability analysis, which were employed on the SWUL Scale and the Teaching and Learning Instrument, are presented. The results from the overall scale scores and the individual questions from these instruments are also presented. The results from the analysis on the remaining areas of the study, such as, social aspects and integration, additional teaching and learning aspects, financial concerns, and pride in UCC will then be given.

1.3.1 Profile of Respondents

A profile of the respondents was conducted in terms of demographic details, academic details, and the respondents' profile of entry into UCC. Chi-square tests were conducted to determine whether each sample obtained through the survey was representative of the overall UCC population. The null hypothesis states that the distribution of survey population for each demographic characteristic is the same as the overall UCC population.

Demographic details relating to gender and nationality are summarised in **Table 1.7.**

Table 1.7: Demographic Characteristics

Characteristic	Option	UCC Population	%	Survey	%	P - Value
Gender (n=3224)[1]	Female	10056	59.1	2116	65.6	<0.001
	Male	6970	40.9	1108	34.4	
Nationality (n=3216)	Irish	14215	83.5	2865	89.1	<0.001
	Other EU	1146	6.7	178	5.5	
	Non-EU	1665	9.8	173	5.4	

[1]Number of Respondents

Two-thirds of the respondents to the survey were female (**Table 1.7**). The majority of respondents were Irish (89.1%), while 5.5% of respondents were from other EU countries, with 5.4% from non-EU countries. A 5% level of significance is used for the Chi-square test. It states that the distribution of the survey population and the distribution of the overall UCC population for each demographic characteristic are statistically significantly different ($p<0.001$). Although the differences were statistically significant, they are not practically significant. The characteristics relating to the academic aspects of the respondents are summarised in **Table 1.8.**

Table 1.8: Academic Characteristics

Characteristic	Option	UCC Population	%	Survey	%	P – Value
Studying for which qualification (n=3221)[1]	Bachelor's Degree	11930	76.4	2629	81.6	<0.001
	Taught Masters	1428	9.1	178	5.5	
	Higher/Post Graduate Diploma	1038	6.7	136	4.2	
	PhD	930	6	103	3.2	
	Diploma	47	0.3	82	2.5	
	Certificate	47	0.3	40	1.2	
	Research Masters	189	1.2	20	0.6	
	Other	NA	NA	33	1	
College (n=3190)	Arts, Celtic Studies & Social Sciences	5318	31.2	1158	36.3	<0.001
	Science, Engineering & Food Science	3890	22.8	844	26.5	
	Business & Law	3415	20.1	612	19.2	
	Medicine & Health	2903	17.1	576	18.1	
Status (n=3187)	Full-time	15524	91.2	3020	94.8	<0.001
	Part-time	1328	7.8	148	4.6	
	Distance Learner	174	1	19	0.6	
	Neither First nor Final Year	5246	33.6	1143	35.9	
	Final Year	2874	18.4	592	18.6	

[1] Number of Respondents

Table 1.8 Continued: Academic Characteristics

Characteristic	Option	UCC Population	%	Survey	%	P – Value
Status (n=3187)	Postgraduate	3722	23.8	316	9.9	<0.001
Year (n=3183)	First Year	3781	24.2	1132	35.6	<0.001
	Neither First nor Final Year	5246	33.6	1143	35.9	
	Final Year	2874	18.4	592	18.6	
	Postgraduate	3722	23.8	316	9.9	

n = Number of Respondents

The majority of respondents were studying for a Bachelor's degree (81.6%) and studying full-time (94.8%). Approximately equal proportions of respondents were in first year (35.6%) or neither in first nor final year (35.9%). Over a third (36.3%) of respondents were from the College of Arts, Celtic Studies and Social Sciences, while 26.5% of respondents were from the College of Science, Engineering and Food Science. Fewer than 20% of respondents were in the College of Business and Law (19.2%) and the College of Medicine and Health (18.1%). The p-value is < 0.01; therefore, although the differences were statistically significant, they were not practically significant.

Table 1.9: Process of Entry Characteristics

Characteristic		UCC Population	%	Frequency	%	P - Value
Process of Entry (n=3181)[1]	CAO-Irish School Leaver	13384	78.6	2469	78	<0.001
	Non-EU Student Degree Application	806	4.7	95	3	
	Visiting Student Application	645	3.8	94	3	
	CAO-Non-Irish EU Student Application	634	3.7	71	2.2	
	Other	1557	9.1	452	14	
Special Programmes (n=243)	Mature Student Application	NA	NA	111	46	<0.001
	UCC PLUS+ Application	NA	NA	111	46	
	Disability Student Application	NA	NA	21	8.6	

[1]Number of Respondents

Table 1.9 shows that, of 243 who responded to special programmes entry, 45.7% applied as mature students and 45.7% applied through the UCC PLUS++ programme, with 8.6% applying through the disability student application. The p-value is < 0.01; therefore, although the differences were statistically significant, they were not practically significant.

1.3.2 Satisfaction with University Life (SWUL)

The Satisfaction with University Life scale is an overall assessment of satisfaction with life in university and contains five items. Factor and reliability analysis were performed to validate the use of the SWUL in a UCC context.

1.3.2.1 Factor Analysis of SWUL

Factor analysis was performed to validate the SWUL scale by examining if the individual items load on the same factor. Prior to factor analysis being performed, it is necessary to determine if the data is suitable for factor analysing. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test for sphericity determines whether the data given is suitable. The results of these tests are summarised in **Table 1.10.**

Table 1.10: Preliminary Tests and Measures for Factor Analysis.

Measure of Sampling Adequacy (MSA)	
Kaiser – Meyer – Olkin (KMO)	0.877
Bartlett Test for Sphericity	< 0.001

The KMO measure has a value of 0.877, which exceeds the acceptable level of 0.5 (**Table 1.10**). Bartlett's test for sphericity gives a statistically significant result ($p<0.001$). This indicates there are significant correlations between the five items in the scale. Both the KMO and Bartlett's test for sphericity indicate that factor analysis can be performed on the five items.

Factor analysis incorporating a principal component extraction was performed on the five items. Latent root criterion states that only factors that have eigenvalues greater than or equal to one are retained. The initial factor extraction results, with associated eigenvalues and the percentage of variance explained are given in **Table 1.11.**

Table 1.11: Eigenvalue and Percentage of Variance Explained per Factor

Factor	Eigenvalues	% of Variance Explained
1	3.623	72.47
2	0.509	10.175
3	0.353	7.061
4	0.284	5.684
5	0.231	4.613

The principal results for the initial factor analysis revealed the presence of a single factor with an eigenvalue of 3.628, which is greater than 1 (**Table 1.11**). This factor explains 72.5% of the variance.

Figure 1.1 presents the scree plot, which is a plot of the eigenvalues against the factor numbers.

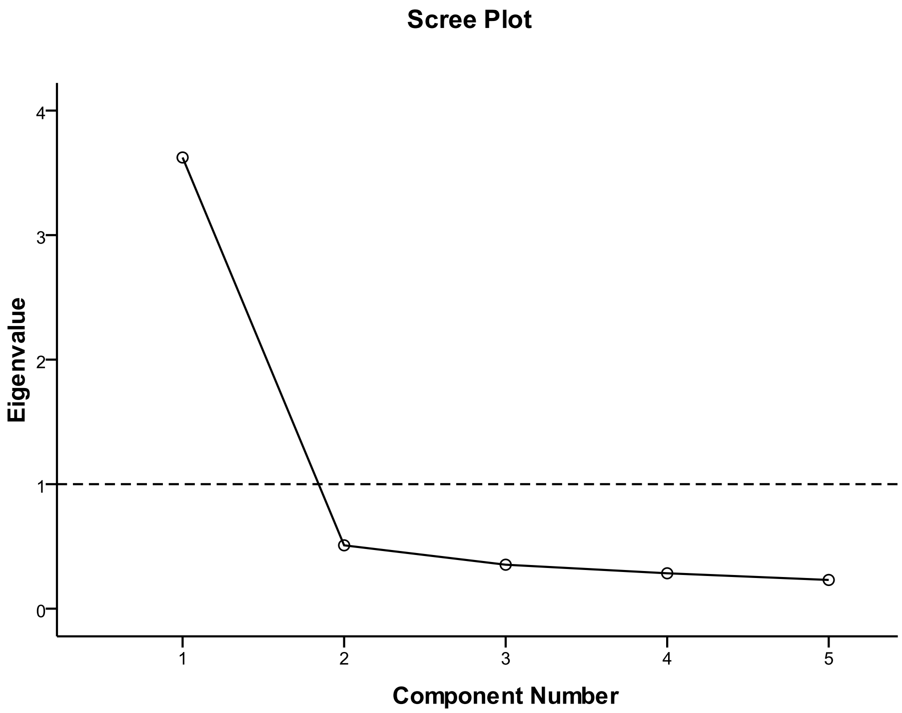

Figure 1.1: Scree Plot for Factor Analysis of SWUL Scale

As illustrated in **Figure 1.1**, the scree plot levels off after the first factor. All remaining factors are below the recommended level of 1. Therefore, only one factor is extracted.

Once the number of factors to extract is determined, a factor loading matrix is examined to interpret the factor. The factor loadings of factor 1 are calculated and presented in **Table 1.12**.

Table 1.12: Factor Loadings on the Five Items of the SWUL Scale.

Question	Item	Factor 1
11	In most ways my life in university is close to my ideal.	0.873
12	The conditions of my life in university are excellent.	0.848
13	I am satisfied with my life in university.	0.897
14	So far, I have gotten the important things I want in my life in university.	0.841
15	If I had my time in university again, I would change almost nothing.	0.795

Table 1.12 shows that each item had a substantial loading on the extracted factor. All items attained a factor loading ranging from 0.795 to 0.897. Therefore, this scale is unidimensional, as all items load on one factor, and it can be labelled as "Satisfaction with University Life". The SWUL examines the conditions of the respondents' life and satisfaction with their life in university.

1.3.2.2 Internal Consistency of SWUL

To ascertain if the items forming the SWUL are internally consistent, a Cronbach's alpha was computed. The Cronbach's alpha value for the resources subscale is 0.9, which is within the recommended level outlined by DeVellis (1991). Therefore, a score, as described in the methods section, can be calculated ranging from 5 to 35.

1.3.2.3 Description of Satisfaction with University Life Scale

The overall score of the SWUL scale was formed as a summated score of the five items. A histogram of the data (see **Appendix 1.A**) indicated that the data was skewed, so the median and interquartile range were reported. The mean is also presented in order to make comparisons with other universities. These descriptive statistics are given in **Table 1.13**, where the percentage of respondents who were satisfied and broadly satisfied are also presented.

Table 1.13: Descriptive Statistics for Overall SWUL Scale

Scale	Median (IQR)	Mean	% Satisfied	% Broadly Satisfied
SWUL (n=3015)[1]	25 (20 – 29)	24.2	73.4	77.9

n = Number of Respondents

The median value is 25, with an interquartile range of 20 to 29. A scale score of 25 indicates that the respondents are "slightly satisfied" with university life. Approximately, 73% of respondents were satisfied with university life, while 77.9% of respondents were broadly satisfied.

The percentage of respondents who agreed and broadly agreed with each of the five items in the SWUL scale are summarised in **Table 1.14**.

Table 1.14: Response to Each Item of the SWUL Scale.

Question	No. of Respondents	% Agree	% Broadly Agree
I am satisfied with my life in university.	3048	78.5	86.2
The conditions of my life in university are excellent.	3057	72.1	83.2
In most ways my life in university is close to my ideal.	3062	68.4	79.7
So far, I have gotten the important things I want in my life in university.	3040	66.9	80.6
If I had my time in university again, I would change almost nothing.	3044	48.4	59.5

The highest percentage (78.5%) of students agreed they were satisfied with their life in university and 86.2% broadly agreed. Approximately equal proportions of respondents agreed that the conditions in their university were excellent (72.1%) and that in most ways their life in university was close to their ideal (68.4%). Approximately two-thirds (66.9%) of respondents agreed that so far they have gotten the important things they wanted in their university life. Just under half of respondents (48.4%) agreed that, if they had their time in university again, they would change almost nothing.

1.3.3 Social Aspects and Integration

Respondents rated their level of satisfaction through various aspects such as safety on campus, social life, and social events organised for them, as presented in **Table 1.15**.

Table 1.15: Satisfaction with Social Aspects of University Life

Question	No. of Respondents	% Satisfied	% Broadly Satisfied
Please rate how satisfied you are with security and safety on campus.	2961	80.4	91.8
Rate the extent to which you are satisfied with the social aspects of your life.	3025	74.1	82.0
To what extent are you satisfied with the social events organised for students.	2837	56.9	79.6

Over 80% were satisfied and 91.8% of respondents were broadly satisfied with security and safety on campus (**Table 1.15**). Nearly three quarters of the respondents were satisfied with the social aspects of their life and 82% were broadly satisfied. Just over half of the respondents were satisfied with the social events organised for students, with 79.6% broadly satisfied.

Respondents were asked how socially integrated they felt in UCC, with the responses summarised by percentages socially integrated and broadly socially integrated (Table 1.16).

Table 1.16: Social Integration in UCC.

Question	No. of Respondents	% Socially Integrated	% Broadly Socially Integrated
How socially connected do you feel in UCC?	2967	57.1	89.6

Approximately 57% of respondents felt socially integrated. Almost 90% of respondents felt broadly socially integrated (89.6%) (see **Table 1.16**).

1.3.4 Teaching and Learning Instrument

There are 19 items in the "Teaching and Learning Instrument" (questions 57 to 62 and 67 to 79). Factor analysis was applied to all 19 items to validate the use of this instrument in a UCC context.

1.3.4.1 Factor Analysis of Teaching and Learning Instrument

Prior to factor analysis being performed, it is necessary to determine if the data is suitable for factor analysing. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test for sphericity determines whether the data given is suitable. The results of these tests are summarised in **Table 1.17**.

Table 1.17: Preliminary Tests and Measures for Factor Analysis

Measure of Sampling Adequacy (MSA)	
Kaiser–Meyer–Olkin (KMO)	0.892
Bartlett Test for Sphericity	<0.001

The KMO measure has a value of 0.892, which exceeds the acceptable level of 0.5 (**Table 1.17**). Bartlett's test for sphericity gives a statistically significant result ($p<0.001$). This indicates there are significant correlations between the 19 items in the instrument. Both the KMO and Bartlett's test for sphericity indicate that factor analysis can be performed on the 19 items.

Factor analysis incorporating a principal component extraction, with a varimax rotation, was performed on the 19 items. Latent root criterion states that only factors that have eigenvalues greater than or equal to one are retained. The initial factor extraction results, with associated eigenvalues and the percentage of variance explained for the first five factors are given in **Table 1.18**. A full list of eigenvalues and percentage of variance explained by each of the 19 factors is given in **Appendix 1.B**.

Table 1.18: Eigenvalue and Percentage of Variance Explained per Factor.

Factor	Eigenvalues	% of Variance Explained
1	6.09	32.07
2	2.92	15.39
3	1.78	9.39
4	0.999	5.26
5	0.817	4.30

The principal results for the initial factor analysis revealed the presence of 3 factors with eigenvalues greater than 1, ranging from 1.78 to 6.09 (**Table 1.18**). The fourth factor is on the broader of the cut-off point with an eigenvalue of 0.999. It was determined that a 3-factor solution was better than a four factor solution. The cumulative percentage of variance explained by the first three factors amounts to 56.9%.

Figure 1.2 presents the scree plot, which is a plot of the eigenvalues against the factor numbers.

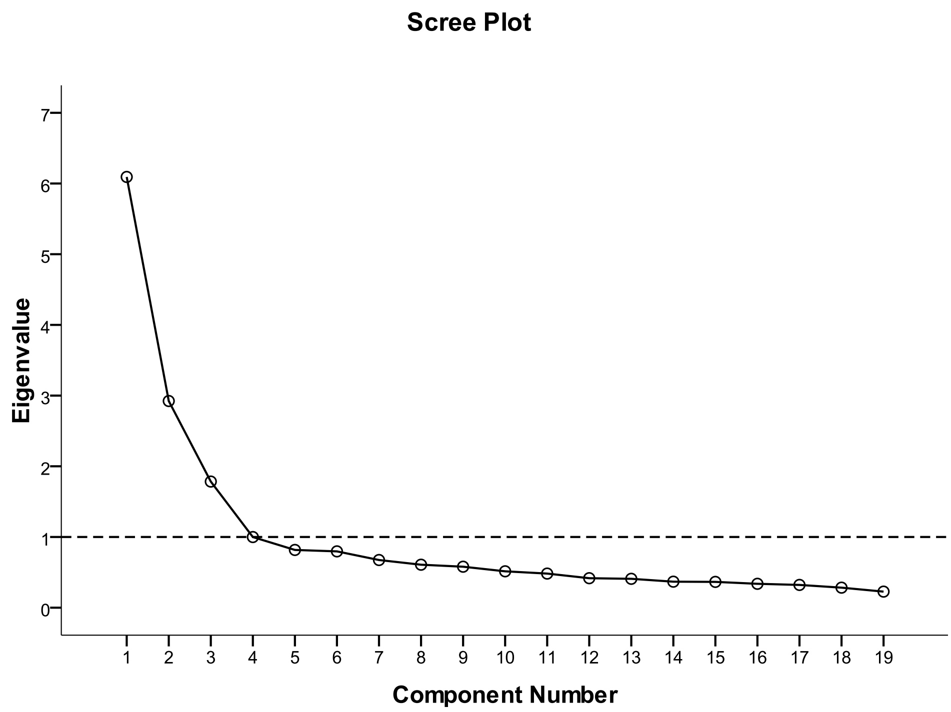

Figure 1.2: Scree plot for factor analysis of the Teaching and Learning Instrument

The scree plot tends to level off at approximately an eigenvalue of 1 (**Figure 1.2**). Three factors were extracted.

After determining the number of factors to extract, a rotated factor loading matrix was examined to determine which items load on which factors. The rotated factor loading for each individual item is presented in **Table 1.19**. Factor loadings of less than 0.4 are suppressed.

Table 1.19: Rotated Factor Loadings for the Teaching and Learning Instrument.

Item	Statement	Factor		
		1	2	3
Question 71	My programme of study is intellectually stimulating.	0.800		
Question 72	I find my programme of study motivating.	0.838		
Question 73	My programme stimulated my enthusiasm for further learning.	0.831		
Question 74	My programme has stimulated my interest in the field.	0.833		
Question 75	I feel that I benefit from being in contact with active researchers/scholars.	0.650		
Question 78	I feel part of a community of scholars who are committed to learning.	0.502		
Question 57	My teachers/lecturers normally give me helpful feedback on my progress.		0.766	
Question 58	The teachers/lecturers in this programme motivate me to do my best work.		0.726	
Question 59	My teachers/lecturers make a real effort to understand any difficulties I have.		0.747	
Question 60	My teachers/lecturers are extremely good at explaining things.		0.624	
Question 61	The teachers/lecturers work hard to make their subjects interesting.		0.641	
Question 62	My teachers/lecturers put a lot of time into comments (orally and/or in writing) on my work.		0.790	
Question 67*	There is a lot of unwanted academic pressure on me as a student.			0.751
Question 68*	My academic workload is too heavy.			0.818
Question 69	I am generally given enough time to understand things I have learned.			0.498
Question 70*	The volume of work necessary for my programme of study means that it cannot all be thoroughly comprehended.			0.685
Question 76*	The academic expectations of me on my programme of study are too high.			0.799
Question 77*	Intellectual standards at UCC are set too high.			0.671
Question 79	Being selected to study at UCC is a source of motivation to me.			

*These items are reversed scored.

Items 71 to 75 and item 78 attained factor loadings ranging from 0.502 to 0.838 on factor 1. These six items represent the students' motivation within their programme of study and therefore factor 1 is labelled the "Motivation Scale".

Items 57 to 62 attained factor loadings ranging from 0.624 to 0.970 on factor 2. These six items represent the students' perception of good teaching practice and therefore factor 2 is labelled the "Good Teaching Scale".

Items 67 to 70 and 76 to 77 attained factor loadings ranging from 0.498 to 0.814 on factor 3. These six items represent the students' perception of their workloads and therefore factor 3 is labelled "Appropriate Workload Scale".

Item 79 did not attain the required factor loading on any of the factors (factor 1, 0.370; factor 2, 0.384 and factor 3, -0.049).

1.3.4.2 Reliability Analysis of Teaching and Learning Instrument

To ascertain if the items forming the derived scales are internally consistent, a Cronbach's alpha is computed for each. The Cronbach's alpha value for the GTS, AWS, and MS are presented in **Table 1.20**.

Table 1.20: Cronbach's Alpha for GTS, AWS, and MS

Scale	Cronbach's Alpha
Good Teaching	0.864
Appropriate Workload	0.812
Motivation	0.874

The Cronbach's alpha values for the GTS, AWS, and MS range from 0.812 to 0.874. These values are within the recommended level, outlined by DeVellis (1991). Therefore, a score, as described in the methods section, can be calculated for each scale.

1.3.4.3 Description of the Teaching and Learning Instrument

Histograms were examined to determine the appropriate measure of centrality to summarise each scale score (see **Appendix 1.C**). The MS was skewed and, as a result, the median and interquartile range were reported. The GTS and the AWS were not skewed, and hence the mean and standard deviation were reported. The descriptive statistics, along with the percentage agreement and broad agreement for each, are presented in **Table 1.21**.

Table 1.21: Descriptive Statistics for Teaching and Learning Scale.

Scale	Mean (S.D.)	% Agree	% Broadly Agree
GTS	5.96 (41.9)	41.9	71.5
AWS	7.34 (37.2)	44.1	70.6
MS	41.67 (16.7-58.3)[1] (36.1)[2]	66.7	86.8

[1]The median and IQR were reported for this variable.

The mean score values for the GTS and AWS are 5.96 and 7.34, which suggest low positive perceptions of the presence of good teaching practice and reasonable workloads, respectively. The median value for the MS is 41.67; this implies a more moderately high perception of being motivated by their programme of study.

Under half of the respondents agree that good teaching practices are present (41.9%) and that they have reasonable workloads (44.1%), but these percentages increase to 71.5% and 70.6%, respectively, when including students who neither agree nor disagree. Approximately two-thirds of respondents stated that they are positively motivated by their programme of study, while 86.8% of respondents broadly agree.

Good Teaching Scale (GTS)

The percentage of respondents who agreed and broadly agreed with each of the six items in the GTS are summarised in **Table 1.22**. This scale measures the students' perception of the teacher's ability to contribute to student learning.

Table 1.22: Response to each Item of the GTS

Item	No. of Respondents	% Agree	% Broadly Agree
The teachers/lecturers work hard to make their subjects interesting.	2588	55.8	81.8
My teachers/lecturers are extremely good at explaining things.	2602	52.9	84.7
The teachers/lecturers in this programme motivate me to do my best work.	2604	40.4	72.4
My teachers/lecturers make a real effort to understand any difficulties I have.	2534	39.1	71.3
My teachers/lecturers normally give me helpful feedback on my progress.	2564	34.4	61.7
My teachers/lecturers put a lot of time into comments (orally and/or in writing) on my work.	2507	28.9	57.0

As seen in **Table 1.22,** just over half of respondents agreed that the lecturers work hard to make their subjects interesting (55.8%) and that their lecturers are extremely good at explaining things (52.9%). These percentages increase to 81.8% and 84.7% for those who broadly agreed. Approximately 40% of respondents agreed that their lecturers motivate them to do their best work and a similar percentage agreed that their lecturers make a real effort to understand any difficulties they have. These figures increase to 72.4% and 71.3% for respondents who broadly agreed. Less than one third (28.9%) of respondents agreed that their lecturers put a lot of time into comments in their work and just over one third (34.4%) give helpful feedback on their progress. These percentages increase substantially for broad agreement (57.0% and 61.7%).

Appropriate Workload Scale (AWS)

The percentage of respondents who agreed and broadly agreed with each of the six items in the AWS are summarised in **Table 1.23**. This scale measures the extent to which the workload given is perceived by students to be manageable.

Table 1.23: Response to each Item of the AWS.

Item	No. of Respondents	% Agree	% Broadly Agree
I am generally given enough time to understand things I have learned.	2562	51.4	75.8
The volume of work necessary for my programme of study means that it cannot all be thoroughly comprehended.*	2559	47.0	69.8
My academic workload is too heavy.*	2571	34.8	58.7
There is a lot of unwanted academic pressure on me as a student.*	2556	31.7	61.8
The academic expectations of me on my programme of study are too high.*	2534	23.4	54.1
Intellectual standards at UCC are set too high.*	2536	15.4	42.3

*These questions are negatively phrased.

Half of the respondents agreed that they are generally given enough time to understand things they have learned, but also, just under half (47.0%) of the respondents agreed that all the work they do could not be thoroughly comprehended, because of the volume of work necessary for their programme. This increased substantially for broad agreement (75.8% and 69.8% respectively). Approximately one third (34.8%) of respondents agreed that their workload is too heavy, and a similar percentage (31.7%) agreed that there was a lot of unwanted academic pressure on them as students. This increased to 58.7% and 61.8% for broad agreement. Under one-quarter of respondents agreed that the academic expectations on them on their programme of study are too high and 15.4% of respondents agreed that intellectual standards at UCC are set too high. These percentages increase substantially for broad agreement.

Motivation Scale (MS)

The percentage of respondents who agreed and broadly agreed with each of the six items in the MS are summarised in **Table 1.24**. This scale measures the extent to which students feel their programme of study is intellectually stimulating, motivating, encourages further learning, involves a community of scholars including active researchers, and for which the expectations are realistic.

Table 1.24: Response to Each Item of the MS.

Item	No of Respondents	% Agree	% Broadly Agree
My programme of study is intellectually stimulating.	2580	79.4	91.9
My programme of study has stimulated my interest in the field.	2549	78.3	91.2
I feel that I benefit from being in contact with active researchers/scholars.	2377	66.6	89.5
My programme of study has stimulated my enthusiasm for further learning.	2581	66.4	86.3
I find my programme of study motivating.	2578	64.4	84.9
I feel part of a community of scholars who are committed to learning.	2521	45.3	76.9

Just less than 80% of respondents agreed that their programme of study is intellectually stimulating, with a similar percentage (78.3%) agreeing that their programme of study has stimulated their interest in the field (**Table 1.24**).

Approximately 90% broadly agreed with these two statements. Approximately two thirds of the respondents agreed that they find their programme of study motivating (66.6%), a similar percentage agreed that their programme of study has stimulated their enthusiasm for further learning (66.4%) and that they benefit from being in contact with active researchers/scholars (64.4%). These percentages increase substantially for broad agreement. Approximately 45% of respondents agreed that they felt part of a community of scholars who are committed to learning, while over three quarters broadly agreed with this statement.

1.3.5 Additional Teaching and Learning Aspects

The additional aspects of teaching and learning consist of the student learning community, programme of study components, timetable, workload and assessment, work experience and transferable skills.

1.3.5.1 Student Learning Community

The student learning community examines the respondents' levels of commitment to study, self-assessment of attendance and preparation for lectures, their perceptions of Blackboard, and whether study at UCC is a source of motivation.

Respondents' description of their level of commitment to their studies is presented in **Table 1.25** and is summarised by percentages good to very good and average to very good.

Table 1.25: Commitment to Studies

Question	No. of Respondents	% (Good – Very Good)	%(Average – Very Good)
How would you describe your level of commitment to your studies?	2779	62.0	92.7

Approximately two thirds (62%) of respondents described their commitment level as good to very good (**Table 1.25**). This increases substantially to 93%, when respondents with an average commitment to study were taken into account.

Respondents were asked to assess their lecture attendance and whether they read the subjects both before and after given lectures. The percentage ranges shown are usually to always and sometimes to always (**Table 1.26**).

Table 1.26: Lecture Commitment

Question	No. of Respondents	% (Usually – Always)	% (Sometimes – Always)
Do you attend lectures?	2739	90.2	97.8
Do you spend time reading around your subject after lectures?	2717	23.6	61.9
Do you read in advance of lectures?	2722	14.9	49.2

There was a high level of lecture attendance, with 90.2% of students attending lectures usually to always (**Table 1.26**). A low percentage of respondents read in advance of lectures (14.9%) or spend time reading their subjects after lectures (23.6%). These percentages increased substantially when the respondents who sometimes perform these activities were included in the range.

Respondents were asked to what extent they felt that Blackboard was an important learning resource. Results were summarised by percentages important and broadly important, as presented in **Table 1.27**.

Table 1.27: Importance of Blackboard as a Learning Resource.

Question	No. of Respondents	% Important	% Broadly Important
To what extent do you consider Blackboard to be an important learning resource?	2451	86.0	89.8

The vast majority of respondents felt Blackboard to be an important learning resource, with 86% of respondents considering it to be important and 90% feeling that it is a broadly important learning tool (**Table 1.27**).

Respondents were asked if they felt that the teaching staff made sufficient use of Blackboard. They could respond with either very much so, somewhat or not at all, as summarised in **Table 1.28**.

Table 1.28: Sufficient use of Blackboard

Statement	Response	%
Do you feel teaching staff make sufficient use of Blackboard? (n=2458)[1]	Very Much So	23.5
	Somewhat	57.6
	Not at all	18.9

[1]Number of respondents.

Fifty-eight percent of respondents felt that teaching staff somewhat made sufficient use of the Blackboard learning resource (**Table 1.28**). While 23.5% felt that Blackboard was used very well, 18.9% of students felt that the teaching staff did ‘not at all’ make sufficient use of Blackboard.

The final aspect to be examined in this section was to determine if being selected to study at UCC was a source of motivation for respondents. This question was originally in the MS but as it did not load substantially to any factor it was decided that this question should be analysed separately. Results were summarised by percentages agree and broadly agree, as presented in **Table 1.29**.

Table 1.29: Studying at UCC is Motivating.

Question	No. of Respondents	% Agree	% Broadly Agree
Being selected to study at UCC is a source of motivation to me.	2515	54.8	74.7

Just over half of the respondents agreed that being selected to study at UCC is a source of motivation (**Table 1.29**). Nearly three quarters (74.7%) broadly agreed with this statement.

1.3.5.2 Programme of Study Components

Programme of study components focus on the respondents' level of satisfaction with various facets of their programme of study, such as, the respondents' most and least satisfactory teaching experience, their satisfaction with promptness of feedback and how the feedback helps with future exams and assignments and their teaching, tutorial, and research supervision components of their programme.

The results of analysis for each component were summarised by percentages satisfied and broadly satisfied, as presented in Table **1.30**.

Table 1.30: Satisfaction with Programme of Study Components.

Question	No. of Respondents	% Satisfied	% Broadly Satisfied
How would you rate the most satisfactory teaching you have experienced this year?	2430	93.9	97.7
How satisfied are you with your academic experience in UCC?	2735	78.7	86.2
If teaching is part of your programme how satisfied are you with it?	1706	73.3	88.1
If tutorials are part of your programme how satisfied are you with them?	2203	68.7	80.6
If research supervision is part of your programme how satisfied are you with it?	1102	56.9	81.9
How satisfied are you with how this feedback helps with future assignments?	2568	51.1	67.0
How satisfied are you with how this feedback helps with examinations?	2472	49.2	68.4
How satisfied are you with the promptness of feedback on assignments?	2602	39.8	50.5
How would you rate the least satisfactory teaching you have experienced this year?	2384	12.9	22.7

Ninety-four percent of respondents were satisfied with their most satisfactory teaching experience (**Table 1.30**). Over three-quarters (78.2%) of respondents were satisfied with their academic experience in UCC and 86.2% were broadly satisfied. A similar percentage (73.3%) of respondents were satisfied with the teaching in their programme if it was part of their programme, with 88.1% being broadly satisfied. The respondents were 12.9% satisfied with their least satisfactory teaching experience, this increased to 22.7% for broad satisfaction. Over two-thirds (68.7%) of respondents were satisfied with tutorials if they were part of their programme.

Over half (56.9%) of the respondents were satisfied with research supervision where research was a part of their programme of study. A similar percentage (51.1%) of respondents were satisfied with how feedback helped in relation to future assignments and examinations, while 67% were broadly satisfied. Less than half (49.2%) were satisfied with how feedback helped with examinations, while 68.4% were broadly satisfied. Less than 40% (39.8%) were satisfied with the promptness of feedback on assignments.

1.3.5.3 Timetable

The responses to travel time between lectures, gaps in the timetable and whether these gaps are disruptive were summarised by percentages usually to always and sometimes to always (**Table 1.31**).

Table 1.31: Timetable

Question	No. of Respondents	% (Usually to Always)	% (Sometimes to Always)
Are you provided with sufficient time to travel from lecture theatre to lecture theatre between classes?	2461	70.1	92.7
Are there long gaps between items on your lecture timetable?	2493	40.3	79.1
If you do have such gaps between lectures, do you find them disruptive?	2300	39.1	65.0

Seventy percent of respondents felt they were usually to always provided with sufficient time to travel from lecture theatre to lecture theatre between classes, while 92.7% felt they sometimes to always had sufficient time (**Table 1.31**). Forty percent of respondents felt they usually to always had long gaps between items on their lecture timetable and almost double that percentage (79.1%) felt they sometimes to always had long gaps between items on their lecture timetable. Almost 40% (39.1%) found these gaps between lectures usually to always disruptive while 65% found them sometimes to always disruptive.

1.3.5.4 Workload and Assessment

Respondents' perceptions of the grading of assignments and examinations, lecture availability and assignment deadlines were examined. Respondents were asked if they felt that standards applied in grading were consistent across staff and departments. These results are shown in **Table 1.32**.

Table 1.32: Consistent Grading

Question	Response	%
Do you feel that standards applied in grading assignments and examinations are consistent across teachers/lecturers and departments? (n=2455)[1]	Very Much So	15.2
	Somewhat	50.8
	Not at all	34.0

[1]Number of respondents.

Half of the respondents felt that the grading standards of assignments and examinations were somewhat consistent across teachers/lecturers and departments (**Table 1.32**). Just over a third (34.0%) of respondents felt that grading standards were not at all consistent. Fifteen percent of respondents felt grading standards were very consistent across lecturers and departments.

The extent to which respondents felt that lecturers were available to them is summarised in **Table 1.32**.

Table 1.23: Lecturers Availability

Statement	Response	%
To what extent do you feel lecturers are available to students? (n=2487)[1]	Greatly	7.4
	Very Much	31.2
	Moderately	36.3
	Somewhat	21.7
	Not at All	3.4

[1]Number of respondents.

Less than 10% of respondents felt that their lecturers were greatly available to them (7.4%) (**Table 1.33**). Approximately one-third (36.3%) of respondents felt that lecturers were moderately available to them, while 31.7% felt the lecturers were very much available to them. Twenty percent felt that lecturers were somewhat available, while 7.4% of respondents felt they were greatly available and 3.4% of respondents felt they were not available at all.

Respondents were asked whether they felt the schedule of assignment deadlines were student friendly (see **Table 1.34** below).

Table 1.34: Assignment Deadlines

Statement	Response	%
To what extent do you feel the schedule of deadlines for assignments are student friendly? (n=2455)[1]	Greatly	5.5
	Very Much	28.1
	Moderately	35.6
	Somewhat	19.1
	Not at all	11.7

[1]Number of respondents.

Just over a third (35.6%) of respondents felt that deadlines for assignments were moderately student friendly. Twenty-eight percent felt that the schedule for assignment deadlines was very much student friendly. Nineteen percent felt that deadlines were somewhat student friendly, 11.7% felt they were not at all student friendly. Five and a half percent of the respondents felt that the schedule of deadlines for assignments were greatly student friendly.

1.3.5.5 Work Experience

Respondents were asked whether they had or would be offered work experience as part of their programme of study in UCC. Respondents were asked to answer the question by giving a yes or no response. Less than half (47.2%) of the respondents had or will be offered work experience as part of their programme.

Respondents who have had work experience were then asked to what extent they felt it was valuable. Results are summarised in **Table 1.35**.

Table 1.35: Value of Work Experience

Question	Response	%
If you have had work experience, to what extent do you feel it was valuable? (n=802)[1]	Greatly	56.0
	Very Much	25.4
	Moderately	8.6
	Somewhat	7.9
	Not at All	2.1

[1]Number of respondents.

The majority (56%) of respondents felt that the work experience they participated in was greatly valuable (**Table 1.35**). One-quarter of respondents felt the work experience was very valuable, while 8.6% felt it was moderately valuable. It can also be seen that 7.9% of respondents felt work experience was somewhat valuable, while only 2.1% felt that work experience was not at all valuable.

Finally, respondents were asked if work experience should be part of their programme of study (see **Table 1.36**).

Table 1.36: Work Experience as part of Study

Question	Response	%
Should work experience be a part of your programme of study? (n=2509)[1]	Yes	78.5
	Don't know	13.4
	No	8.1

[1]Number of respondents.

A high percentage (78.5%) of respondents said yes, that work experience should be available as part of their study (**Table 1.36**). Thirteen percent did not know, while 8.1% thought it should not be.

1.3.5.6 Transferable Skills

Respondents were asked, in the event of training being available in the listed areas in UCC, to rate the likelihood that they would avail of it. The responses are summarised by percentages likely and broad likelihood, as presented in **Table 1.37**.

Table 1.337: Transferable Skills

Question	No. of Respondents	% Likely	% Broad Likelihood
Presentation Skills	2477	64.8	77.1
Study Skills	2477	51.7	63.0
IT Skills	2469	50.3	63.5
Time Management	2486	50.1	63.0
Writing Skills	2449	49.2	59.9
Communication Skills	2465	45.7	61.5
Teamworking	2471	37.1	54.6
Interpersonal Skills	2466	36.9	53.2
English Language Skills	2244	27.1	39.6

Two-thirds (64.8%) of respondents were likely to avail of training in the development of their presentation skills if it was made available to them (**Table 1.37**). Over half (51.7%) of respondents were likely to avail of study skills (51.7%), and a similar percentage (50.3% and 50.2% respectively) were likely to avail of IT skills and time management skills if they were made available to them. Less than half (49.2%) of the respondents were likely to avail of training in writing skills (49.2%) and a similar percentage (45.7%) were likely to avail of communication skills if they were made available to them. Over a third (37.1%) would avail in training for team work skills while a similar percentage (36.9%) of respondents were likely to avail of training for interpersonal skills if they were made available to them. Only 27.1% of respondents were likely to avail of English language skills training.

1.3.6 Financial Concerns

This section of the questionnaire examined the financial concerns of the respondents. Six questions, such as, whether discounts and free services are a valuable feature of student life, how students felt about debt level on graduation, their present level of debt and concerns over money shortages and whether they missed college due to financial concerns. Results are presented as percentages agree and broadly agree (**Table 1.38**).

Table 1.38: Financial Concerns

Question	No. of Respondents	% Agree	% Broadly Agree
Discounts and free services for students are a valuable feature of student life.	2492	83.9	92.7
I am consistently short of money.	2499	63.4	75.2
I am concerned about my likely level of debt on graduation.	2489	47.9	63.8
I worry about my present level of debt.	2501	44.7	59.8
I regularly miss lectures/tutorials and other study commitments because I need to work.	2494	17.3	29.9
I regularly miss lectures/tutorials and other study commitments because I cannot afford childcare/after-school care.	2474	4.1	22.7

Eighty-four percent of respondents agreed that student discounts and free student services are a valuable part of student life, while 92.7% were in broad agreement (**Table 1.38**). Two-thirds (63.4%) of respondents agreed that they were concerned about money shortages. Almost 45% agreed they were worried about their present level of debt, and a similar percentage (47.9%) agreed they were worried about their likely level of debt after graduation. A low percentage (17.3%) of respondents agreed they regularly missed college, as they needed to work, while 29.9% broadly agreed. Only 4% of respondents agreed they regularly missed lectures and other study commitments as they were unable to afford childcare services but this jumps to 22.7% broadly agreed with this.

1.3.7 Pride in UCC

The final section of the questionnaire determined the respondents' pride in being a member of UCC. Respondents were asked to answer giving a yes or no response. From this, 94% of respondents felt proud to be a member of UCC.

1.4 Discussion

In this section, we will discuss the results given in the previous section and make inferences based on these findings. We will also compare our study with other studies conducted, such as studies undertaken in the University of Sydney, Australia, and the University of Oxford, in the UK. For our discussion, the results will be divided into the categories given in the Results section. Limitations and future recommendations of this study will then be outlined.

1.4.1 Profile of Respondents

The first section of the survey to be discussed is the demographics. A comparison was made between the actual demographical of UCC and the sample of students who completed the survey, the goal was to determine if it was an acceptable representation of the university population. The results of the Chi-squared test for goodness-of-fit deduced that there were significant differences between the demographics of the survey and the demographics of the university. We can be confident in this result, as all the demographic characteristics are significantly different, but they are not practically different. The comparison was distorted due to the large sample size. For example, the demographic breakdown of respondents and the actual UCC population have very similar results, except for the year of study students were in. There was a substantial response by the first years but a very low level of postgraduates by comparison. This poor return could be due to the fact that some of the questions on the questionnaire were aimed at the undergraduates rather than the postgraduates. The response rate for this survey was approximately 20%, a similar student experience survey in 2020 was 31%, although this was the highest response to a survey they had received to date.

1.4.2 Satisfaction with University Life

The factor and reliability analysis results based on the SWUL scale are shown in section **1.3.2**, which states that there is one common factor and that internal consistency exists. This was important from an analysis point of view, as it was essential to determine if this scale was applicable in an Irish university context.

In the results, a high proportion of respondents (78%) were broadly satisfied with university life in UCC with 73% satisfied. Examining the scale and the individual items of the scale, there was a very positive response for four out of the five items in the scale. The remaining item, “If I had my time in university again I would change almost nothing”, had a low response broadly agree (48%). It was felt with the array of clubs, societies, and course codes, respondents would like to try something new if they had the opportunity.

The difference in respondents who were broadly satisfied and satisfied with university life was 5%. However, if we examine the responses to each of the items in the SWUL scale, the difference in responses between broad agreement and agreement to the item is in the region of the 10%. Given the magnitude of difference for each of the individual items of the scale, we could expect the magnitude to be the same for the overall score. The discrepancy occurs due to the scoring method, they are not like for like, each item is based on a 7-point Likert scale while the SWUL score is a numeric score from 5 to 35.

Finally, comparing UCC to other international universities that have also conducted student experience surveys on an annual basis, University of Oxford and the University of Sydney were chosen as benchmarks. In comparison with the University of Sydney, we compared this survey to their initial year of the survey, which was 1999. They did not include a Satisfaction with University Life scale, but instead included just one item asking students to gauge their satisfaction on a Likert scale. The results for Sydney (University of Sydney, October 2009) were as follows: 67% satisfied, 88% broadly satisfied. In comparison with the University of Oxford, who conducted their first survey in 2001 and who also did not include a Satisfaction with University Life scale but just one item, their results (University of Oxford, October 2009) were as follows: 81% satisfied, 93% broadly satisfied. Therefore, for a result of 73% satisfied and 78% broadly satisfied, satisfactions levels are in line with the results of both universities in their first year of surveying, especially with Sydney. The difference between broadly satisfied and satisfied are much closer for our study compared to the other studies, this is due to the other studies calculating overall satisfaction based on one question with a single Likert scale, while this study uses a numeric scale to capture overall satisfaction The definition of broadly satisfied is wider and captures more respondents in their definition. This study could consider asking a single question based on a single Likert scale to capture overall satisfaction.

1.4.3 Social Aspects and Integration

All responses from the social aspects of this survey were very positive, reflecting how students do like to interact with one another through the wide variety of services, clubs/societies, and social areas in UCC. The University of Sydney ran a pilot for three years dealing with the "On Campus Experience", students were asked to rate their level of agreement with six items dealing with social and personal development outside classes. The students in Sydney responded positively to aspects such as interacting with others (87% broadly agree) and participation in clubs and societies (70% broadly agree). This is quite comparable to the response UCC students gave in this survey.

1.4.4 Teaching and Learning Instrument

The teaching and learning instruments consist of three scales: the Good Teaching Scale (GTS), the Appropriate Workload Scale (AWS) and the Motivational Scale (MS). The GTS and AWS are both used by the University of Sydney and the University of Oxford, while the MS was only incorporated by the University of Oxford. Similar to the SWUL scale, factor and reliability analysis indicated three distinct factors exist in an Irish context.

Good Teaching Scale (GTS)

There was a very positive response of individuals being broadly satisfied; approximately 72%. Looking at the breakdown of the scale into its individual items, the majority of responses were pertaining to the lecturer's ability to explain and to make the subject interesting, with close to 85% and 82% broadly agreeing, respectively. However, there was a very low response in relation to lecture feedback, ability to understand difficulties, and lectures putting time into oral and written comments on student work. These items were looking at lecturer feedback, ability to understand difficulties and putting time into oral and written comments on students' work. In order to find the underlying reasons behind this response, it is important to look at perceptions of lecturers and students. It was reported that students are 97% satisfied with lecturer availability, and we also know that 90% of students attend lectures regularly and 93% are committed to their study. Therefore, this would be evidence showing to the contrary, that students attend lectures, and lecturers are available to meet students. However, when asked how satisfied they were with feedback on assignments and exams, they were 40% and 49% satisfied, respectively.

Comparing UCC with the University of Sydney and the University of Oxford, both universities included this scale in their surveys from the beginning. The results for Sydney were as follows: 39% agree, 74% broadly agree. The results for Oxford were as follows: 42% agree and 77% broadly agree. So, for UCC to have a result of 42% agree and 72% broadly agreeing in our first year conducting the survey, the outcome in UCC is very similar to the two other universities.

Appropriate Workload Scale (AWS)

The response for this scale was, again, very positive overall, with 71% broadly agreeing. Looking at the individual items of the scale, there was a positive response for respondents feeling they have time to understand the material. However, students did feel the standard and level expected of them was too high. In particular, they felt there was academic pressure, heavy workload, high expectations and high intellectual standards. Looking at the "Information from programme important for career" question, approximately 85% of students felt that the material covered was relevant to their career.

Comparing UCC to the University of Sydney and the University of Oxford, again, both universities included this scale in their surveys from the beginning. The results from the University of Sydney in 1999 were as follows: 25% agree and 56% broadly agree. The results from Oxford University in 2001 were as follows: 21% agree, 45% broadly agree. The overall result for UCC was as follows: 44% agree, 71% broadly agree.

Therefore, UCC performed better than the other two universities in their initial years of the survey at this aspect.

Motivational Scale (MS)

The third and final factor for the Teaching and Learning scales was the MS. From the results we can see that there was, again, a very positive response for this scale, with 87% broadly agreeing. One item in the scale received a low percentage agreement, which was 'students feeling like part of community of scholars who are committed to learning'. The result for this item was as follows: 45% agree and 77% broadly agree. 70% of respondents of the studentsurvey.ie survey felt the institute emphasised spending a significant amount of time on academic work. Overall, there is high agreement with this statement, which can be linked to pride of UCC. Eighty four percent of respondents of the studentsurvey.ie national results stated they would probably / definitely go to the same institution they are now attending.

1.4.5 Additional Teaching and Learning Aspects

1.4.5.1 Student Learning Community

The vast majority of students (90%) attended lectures and close to 93% reported that their level of commitment to their studies was average-very good. A study conducted in 2008 which surveys students in UCD programmes of study, showed an overall lower attendance rate of 75% (Kelly,2012). The study indicated that self-assessment questionnaires could lead to a bias in responses, this could be the case for the high level of attendance rates in this survey. Despite this high level, a very low proportion (49%) of students, read in advance or after lectures; a lack of available material or heavy workloads could contribute to lack of advanced reading.

Feedback on the Blackboard facility was very positive. Ninety percent of responses found blackboard important. There was 58% of students responding that lecturers somewhat made sufficient use of Blackboard, with 24% responding that lecturers 'very much so' make sufficient use. Blackboard is used by lecturers to post reading material, notes, and any other course related material, as well as to post assignments and test results. The majority felt staff used Blackboard sufficiently.

1.4.5.2 Programme of Study Components

This section is focused more on the students' level of satisfaction with academic aspects such as teaching experiences, tutorial satisfaction, and feedback from assignments and exams. Firstly, there was an extremely positive response from the students when asked to rate their most satisfactory teaching experience. There was a very high response from students, 90% of whom regularly attend lectures and 98% of them feeling broadly satisfied. Also, when asked to rate how satisfied they were with the academic experience as a whole, and how satisfied they were in general with the teaching in UCC, the students responded with 86% and 88% broadly satisfied, respectively. These results are slightly higher than the 80% of respondents who were asked to evaluate their entire educational experience at their institution as good / excellent (student survey, 2021).

In this section, aspects such as how information in programmes of study are valuable in careers, how being at UCC is a motivation in itself, and financial concerns of the students will be discussed. When students were asked if they felt the information received in their programmes of study were valuable for their working careers, they gave a very positive response. Of those who answered the question, 86% responded to some degree of agreement. This is a good result, as it shows students have a strong knowledge of the field they are studying (in order to know it will be valuable). While 58% of respondents to the 2020 results believed their experience at the institution contributed to their knowledge, personal development and skills, which would help them to acquire a job or work related knowledge.

1.4.5.3 Timetable

Seventy percent of students found they have sufficient time to get from lecture theatre to lecture theatre between classes, however this could be a result of students finding the gaps in their timetable are too long. An attempt to amend the current timetable should be employed by running skill courses for students to utilise time between classes. Approximately half the students would have liked to participate in extra courses to acquire new skills in writing, IT, communication and presentation skills, so these courses could be run during certain days to optimise students' time spent between lectures. Another study found a similar result, 55% of respondents felt their institution emphasised quite a bit or very much learning supports services including writing support, learning support. The institution area now providing services for students to increase these skills, ensuring these services are available at peak times will allow students to maximise the service and their time learning these new skills.

1.4.5.4 Workload and Assessment

This section will discuss the students' response to topics related to assignment deadlines, consistent grading, and availability of lecturers. The first topic to be focused on is in relation to assignment deadlines. This question dealt with the students' perception of the assignment deadlines being manageable throughout the year. There was a very positive response to this question, with 88% of students responding with some level of understanding. This follows from and strengthens the point made by item 3 in the AWS, where students felt they had enough time to comprehend the material in their course. However, it is slightly in contrast to item 4 of the AWS, where students felt the volume of work means they can't comprehend the material. So overall, there was a mixed view in that respect.

The next topic to be looked at is the question dealing with students' perception of grading of their assignments and tests across departments. From the results of this question, we can see that 66% of students felt grading was consistent. For the other 34%, this could be explained by the response given to the GTS previously. In items 1 and 6 of that scale, dealing with lecturers' feedback and lecturers putting time into comments on students work, there was a low response here too. So, if students are not receiving adequate feedback throughout the year, they may not be fully aware of their own understanding and the marking schemes incorporated by the lecturer. Again, to reiterate the point about Blackboard being used as a facility to post results, this was an easy way to access them. Finally, with regards lecturers' availability, there was a very positive response to this question, with 97% of students responding to some level of availability. Student responses to faculty interaction was very low in the 2021 student survey, four questions were asked about different types of interactions between students and staff of institution with the highest response of 19% of respondents often / very often discussing course topics, ideas or concepts with staff outside of class (Student Survey, 2021). Less than half of the respondents believed that lecturers provided feedback on draft or work in progress and less than half believed the responses were prompt on completed assignments and exams (Student Survey, 2021). The original results show a more positive level of interaction between students and staff.

1.4.5.5 Work Experience

Work experience is becoming a more and more important feature in student CVs, as the economic decline has had a very negative effect on the job markets. Students are aware of this trend and have expressed an interest in doing work experience as part of their degree programme. We know that 47% of respondents partook in work experience. This is slightly higher than the 40% which often / very often blended academic learning with work experience (Student Survey, 2021). There was a very positive response to this section, with 98% feeling that work experience was valuable (56% responded “Greatly”), and 79% felt it should be a part of their course. This section can again be linked to students’ perception that information gained in their programmes of study is valuable to their careers. With work experience, students get a “hands-on”, close-up look at how the material they studied in class is used in industry. I believe this has a direct impact on the responses to transferable skills, as respondents who are unaware of work experience and the value it adds may not see the benefits in a number of transferable skills. Work experience is not only valuable to the students but also to future employers when seeking graduates.

1.4.5.6 Transferable Skills

The most popular skill students wanted to possess was presentation skills, 77% of respondents opting for this choice. Other key skills which are listed in decreasing order of popularity are study, IT, time management, writing, and communications skills. With respect to work experience prospects and information valuable to student’s careers; presenting, time management and IT skills are all perceived to be very important. From the AWS, we can see that as students feel they have a lot of work and high expectations placed upon them, study skills and time management would be very important.

Students felt less likely to avail of team working, interpersonal, and English language skills training if offered to them. They responded 55%, 53%, and 40% broadly likely to avail of these skills, respectively. Looking at the results for Social Aspects and Integration, it was concluded that students like to interact with one another through the wide network of clubs and societies available to them. Therefore, there may be less of a need for assistance with team working and interpersonal skills, as they obtain this through their own social networks. University is a multi-cultural organisation and close to 15% of the student population are non-Irish. English language skills training should be available to help these students maximise their time in UCC, both academically and socially.

Comparing these results to those in the University of Sydney, they used a slightly different scale, the “Generic Skills Scale”, which looks at the extent to which students perceive their studies to harvest skills which can be applicable in a broader context. The students responded with 87% broad agreement, which illustrates the understanding of these important generic skills required for career and industry needs. Across all transferable skills, UCC students are lower in a broad sense and the results are positive but could be higher. The importance of these skills should be promoted. Promoting these skills or giving students more exposure to them could be through mandatory modules as part of programmes of study or building in assignment and exams to incorporate the core skill sets here. Some key factors identified by Anderson and Higson were, the ability to work under pressure, good written and verbal communication skills, the ability to plan and think strategically and good self and time management (Anderson & Higson, 2008). Transferable skills are still important to employers today. In a survey conducted in 2018, the National Survey of Employers Views of Irish Higher Education Outcomes, 70% of employers believed they would receive candidates who have the relevant knowledge, workplace or transferable skills and attitude. Institutions must continue to expand their programmes for broadening students’ skill sets as 30% of employers were not confident of finding candidates that meet their needs.

1.4.6 Financial Concerns

In relation to financial concerns of students, a high proportion (75%) of them broadly agree that they are consistently short of money, however this did not cause them to miss lectures/tutorials. Very few students with children agreed they missed lectures/tutorials because they cannot afford childcare/after school care. This could be a consequence of a valuable resource to students, discounts, and free services provided by the university and surrounding businesses.

However, over one fifth of the respondents broadly agreed to missing lectures/tutorials because they cannot afford childcare/after school care. The increase in number is substantial, 18.5% but that could be made up of two types of respondents, respondents who are impacted by affordability problems and respondents who are not impacted but answered neutral as it did not have an effect on them. However, it is pivotal to maintain and promote this essential support for students, and the university must guarantee these services are kept intact at a low cost to the students and retain low prices in campus stores. This is an aspect of student life and the student experience that will contribute to the students' overall perception of satisfaction with university life, which is vital in attracting new students to UCC.

1.4.7 Pride in UCC

Almost 94% of students are proud to be members of UCC. This is a very positive response and a good sign for the university. High levels of satisfaction with university and the majority of students feeling socially integrated could influence this sense of pride respondents have towards UCC. Aside from the social aspect, many respondents were in broad agreement that that there are good teaching practices in place and an appropriate workload at the university. Students were asked about their entire educational experience in their institution, 80% deemed it good to excellent and 84% said if they had their time over again they would probably / definitely go to the same institution they are now, again, this is lower than the overall 94% who are proud to be a member of UCC but still a very strong response showing loyalty to their current institution (Student Survey, 2021).

1.4.8 Limitations

This survey had a number of limitations that may have hindered the collection of responses in some cases. The first limitation to be reviewed is the use of an online survey. It is far more efficient to gather information online and in electronic format, as this cuts down on both time and costs. For online surveys, the typical response rate is between 20% and 30% (Lefever et al., 2006). The response rate for the survey was approximately 20%. Lefever (2006) outlines the main pitfalls of online surveys as faulty email addresses and technical difficulties with the internet. Faulty email addresses are more of a concern for industry due to input error or out of date emails on record. The email addresses used in this study were issued and recorded by UCC. The addresses should be assumed to be working correctly and the survey was issued to all correctly. The low response rate could be due to students not using their accounts, simply missing the mail if too much junk mail is sent to the address or ignoring the email through lack of interest. A study by Porter and Whitcomb in 2005 showed response rates are impacted by student characteristics such as demographics, engagement and personality types. For example, females who are socially engaged are more likely to respond. Another study explored how the survey is presented to students can have an impact on response rate. The study by Dommeyer in 2000 compared the survey response rate for 150 students. Their study compared two methods of providing the survey, one method, the survey was embedded in the email vs the survey being attached. The response rate was significantly higher with the embedded survey in the email. Incentives could be used to increase the response rate. A study by Church in 1993 showed a positive increase in response rate if an incentive was provided initially with the survey.

Another limitation of the survey may be the time it was issued, this could have a direct impact on the low responses rate. There are a number of factors that can impact on the response rate, these can include fieldwork or other major events occurring within the institution (Irish Survey of Student Engagement National Report 2020 | Higher Education Authority, 2021). The questionnaire was issued in April 2009. At this time, students were more focused on preparing for exams. The survey run by Studentsurvey.ie each year is released over a three week period in February and March, they received a higher response rate closer to 30% (Student Survey, 2021). This is nearly 10% higher than the response rate from our study. Moving away from a busy time of the academic year with end of term exams, could allow students to take time to participate in the survey. Another report suggests collecting survey information at multiple times during the year to cater for different teaching practices in programmes of study (American Institute of Research 2015). Different programmes of study could have different busy periods, for example, programmes of study with midyear exams or busy with continued assessment would mean the survey should target different segments of the student base at different times. This could have a positive impact on response rates as students would have free time to participate (a known factor to reduced response rates is field work or major events) (Irish Survey of Student Engagement National Report 2020 | Higher Education Authority, 2021).

The third limitation is in relation to the length of the questionnaire. The UCC questionnaire had 168 questions. A questionnaire this large may have discouraged students from partaking in the survey, or from finishing the questionnaire. This was evident, as there was a decline in the number of respondents for the latter part of the questionnaire. There was a 23% drop off in the number of respondents between the beginning and the end of the questionnaire. Another study showed that question length has a negative impact on response rate but did not significantly affect the accuracy rate, where accuracy rate is a measure of response quality (Kato and Miura, 2021). The results of this study are reliable, however poor response rates are directly related to survey quality. The length of instrument should be addressed to maximise the uptake levels and improve the overall quality of the survey responses. This could introduce a non-response bias, if the level of respondents and non-respondents is very large then the results could be skewed by respondents more likely to participate in surveys. In our analysis, people who partially answered questions were dropped from the analysis entirely, this reduces the sample size of responses and could distort the results for a question as the data is removed. This study had an initial low response rate and large level of drop off, every effort should be made to use as much available data as possible. A method to address this is outlined in the future recommendations section.

Another limitation of the study is the question on missing lectures / tutorials due to childcare affordability. Broad agreement is noted in over a fifth of respondents while agreement is 4%. The question should only ask respondents with children to answer this question. Only asking respondents with children would remove neutral answers impacting the broad agreement %. It does not read well that over a fifth of students in UCC broadly agree with regularly missing classes due to childcare affordability. For a future study, the wording or approach should be adjusted, for example there should be an option of does not apply. Significant attention should be paid to the instructions provided to students (American Institute of Research, 2015) and the survey should provide as much clarity as possible (Popham 2013).

1.4.9 Recommendations

This section will outline the recommendations to University policy change and the questionnaire.

1.4.9.1 University Recommendations

In general, students were satisfied with their experience of university life in UCC. The study identified different aspects of the student experience which could be improved, for example, teaching and learning and students' financial concerns. These areas can be improved through policy change, some changes are easier to implement compared to others.

Students identified some areas of concern around the satisfaction with their programme of study, and the promptness and level of feedback for assignments. Students were not satisfied with the promptness or level of feedback. A potential remedy is to provide guidance for lecturers on suitable turnaround times for feedback. A study by Nottingham Trent University in 2013 stated feedback should be returned within three weeks of submission, they believed there was a balance between the time returning the feedback with the level quality and detail contain in it. Trent University state that satisfaction levels are low nationally for students with regard to feedback and that promptness is a key element. Lecturers could also issue feedback differently, initially they could provide generic feedback to the class on common mistakes, and this can be used by students to see possible pitfalls while they wait to receive their own personal feedback.

Students expressed the importance of Blackboard, however felt it is not utilised enough. The medium by how lecturers and students interact impacts the turnaround times with online systems to provide feedback being more beneficial (Nottingham Trent University, 2013). Lecturers should try to utilise Blackboard where possible as this allows them to interact with all students at once. Providing specific feedback on assessments to each student takes time, however the lecturer could provide generic information to all students on areas or topics where there is the most cause for concern. To utilise Blackboard more lecturers should start providing feedback on assignments to their students. Utilising online tools such as Blackboard could increase turnaround times (Nottingham Trent University, 2013), this could be a step to utilising an online asset like Blackboard. Covid19 created an unpreceded situation for students and lecturers with online learning and tools like Blackboard and zoom becoming more important. A survey held in 2021, which focused on the impact of Covid 19 and online learning, found that 84% of respondents had adequate online opportunities and support. To students, recorded lectures were the most important online tool and a feature that should remain when onsite campus resumes. A blended learning environment could become the new normal for universities for the foreseeable future.

Students who took part in work experience had a greater appreciation for work experience and its benefits. Crebert et al, (2004) findings showed students recognised the contribution from university in their skill development, however they greatly valued their work experience and the learnings from their work placement. Some of the skills highlighted in the study were, importance of teamwork and collaborative learning. A study by Anderson and Higson (2008) identified that key transferable soft skills and competencies are integral to graduate employability. Some of the skills they identified were written skills, oral presentation skills and teamwork skills. The study highlighted that the respondents' university programme of study had helped them with their written skills, however additional work was required to increase the skill set for oral or presentation skills. Group-based assignments aided their team work skill set. Work experience shows students the importance of those skills in the working environment. The working environment allows students to identify and develop these skills through experience. These skills are important for employers, as they want to employ more rounded individuals (Deloitte Report, 2018). To increase the students understanding and provide more rounded candidates for employers UCC could consider the following.

UCC should explore making work experience mandatory for all programmes of study. A study in 2012 by Reddy and Moore showed the positive impact on students who partake in work placement. Reddy and Moore recommended that universities should ensure that placement in any form should be promoted to all students. UCC could consider setting up a working committee to oversee a work experience programme for each programme of study. Partaking in work placement may not be achievable for every course but UCC could provide help or guidance to set up summer internships for programmes of study with different companies. Since this study was conducted, UCC have appointed work placement managers to organise work placement programmes for students, to provide support to students and also to organise workshops for transferable skills. Anderson and Higson (2008) suggested that employers held work placement (even part-time work) as important in the recruitment process. Another study supported the fact that employers desire candidates who have transferrable skills in addition to educational and technical requirements (Collins-Nelson et al 2021). This vital experience would be a big step in developing candidates. UCC could hold events or workshops to try and create working experience. Here they could focus on different transferrable skills and promote the importance of transferrable skills for students. Workplace managers should run workshops tackling different transferrable skills, for example some transferrable skills require more support than others, one study noted that written skills were well developed through their programme of learning, however felt oral presentation skills were less developed (Anderson and Higson, 2004). Graduates can have a competitive advantage in the labour market, this is through four key areas, intellect, personality, meta-skills and job specific (Finch, Peacock, Levallet and Foster, 2016). These include soft skills such as communication and teamwork. Lecturers could include exercises or assignments that focus on transferrable skills. Collins-Nelson et al (2021), showed co-curricular programmes play an important role in transferrable skill development, an example of co-curricular activities would be group assignments. Group assignments can help build team based skillset or working as part of team while lecturers could also have presentations as part of continued assessment. Exposing each student to these skills will allow the students to understand the importance of these skills in their working life.

To get buy in from the staff, the executive summary of the report should be distributed. Since the original study was conducted, an executive summary is produced and readily available publicly. A report by the American Institute of Research (2015), stated that there is an increase in the usefulness of the results of student survey when communicated and used by teachers. The rich data can be used at different levels across individual programmes of study or schools in the university. It is much easier to fix or improve a problem if people believe there is something to fix, the students have spoken and albeit in general their satisfaction levels are high there are other key areas that need to be improved. Further student experience surveys should be conducted, and the responses monitored to determine if the policy changes were effective. For example, the response rate of a shorter survey should be examined and compared, or the same survey could be released but provide incentives for completion of the full survey such as, incentives like gift vouchers for the cinema or opportunities to win a prize like an Apple watch. There are pros and cons to using incentives to increase response rates, incentives increase the cost to the researcher, but some incentives can increase response rates while others are still unproven (Survey Monkey, 2021). Cash incentives can increase the responses rates for online surveys but increase costs to the researcher.

1.4.9.2 Questionnaire Recommendations

There are several recommendations for the previous limitations mentioned and for the survey as a whole. For a survey of this size, it is necessary to use the online method. Although, in order to optimise the response rate, the questionnaire should be shortened as there was over a fifth of respondents dropping off through the survey. Cape and Phillips (2015) stated that questionnaire fatigue was impacted by both physical length and cognitive length of a survey. The average length of a survey was 23 minutes and they deemed this too long. The duration and number of questions in the survey should be considered when constructing further surveys. Other studies have said that the average length of a study should be between 10 to 15 minutes and not exceed 20 to 28 minutes (Revilla and Ochoa, 2017) (Revilla and Höhne, 2020). When constructing the new questionnaire, an assessment of the rate of drop off should be conducted on the existing survey to see if it is a particular question or group of questions which led to the drop off or if it was gradual. For future questionnaires Google Analytics would be a powerful tool to monitor and measure users as they navigate through the questionnaire. It would provide analysis which would show pain points, i.e. areas where students stop due to a particular questions or question grouping but also show the time being spent before a respondent drops off. This can also be used to monitor the average length of completion of a survey. Users of the survey should also be given an indication of how long it will take at the beginning before they start the questionnaire.

The survey should be issued earlier in the year, for example February, to not coincide with fieldwork or other major events occurring within the institution (Irish Survey of Student Engagement National Report, 2021). By not going too early in the academic year, this allows first year students time to encompass the student experience and formulate an opinion. It is important that they are accommodated, as they account for about 35% of the student population. It is also recommended that the survey be promoted through alternative media, such as the Students' Union social networking websites (e.g. Facebook, Twitter, Instagram). A study that investigated the impact on social media and response rates to a survey found use of social media doubled the expected return for the survey (Ali et al., 2020). This finding would suggest that social media should be used as part of the marketing strategy of the survey to promote awareness and increase response rates. Since the original survey was conducted, the power of social media has become extraordinary. These platforms should be used heavily to promote the survey, there should be incentives for completing the survey and students should be provided a link to access the survey. Providing easier access to the survey could positively impact the uptake levels of the survey. The national Student Experience Survey used across each third level institution use social media adverts to build awareness.

In our analysis, people who partially answered questions were dropped from the analysis completely. For example, the SWUL had five questions in the scale, a respondent must have answered all items in the SWUL scale to be included. The scoring methods are defined by the authors and have been psychometrically validated. In future studies, the researcher should apply the scales as they exist today. The response rate for this study is low, every effort should be kept to use as much available data as possible. Instead of dropping responses who do not respond to each question of an instrument, perhaps the researcher could do some additional analysis to include partial responses as part of the scoring scale. For example, instead of coding the score to the SWUL scale on an integer scale and adding them, one could code the score between 0 and 1 and get a weighted average on people who answered more questions. This could be applied to each of the instruments, which require all questions to be answered to formulate a valid score. This would allow the researcher to use as much information available, which could be beneficial if the response rate is low. Graphics could be used to demonstrate the level of drop off.

Some questions may not be applicable to all respondents, and so them answering the question does not give an accurate response. A rewording of some questions or change to answers could be considered for future questionnaires. The results showed 77% of respondents said they did not miss lectures/tutorials or other study commitments because they cannot afford childcare/after-school care. This may not be a concern to all candidates, as they would not all have children. This question should only be answered by respondents who have children, so the results are not skewed by individuals who are not impacted by missing lecturers/tutorials because of affordability problems for childcare. A comparison between the existing results and results of future studies should be conducted.

The SWUL scale comprises of five questions, the difference in each response was in the magnitude of 10% between respondents being in agreement or broad agreement. The overall SWUL score is a numeric value based on each response of the scale for each individual respondent. The difference between a respondent being satisfied and broadly satisfied differed by 4.5%. The items scored on the SWUL scale and SWUL score are not like for like, one is a 7-point Likert scale and the other is a summated score. The anomaly occurred due to the definition given to broadly satisfied being narrower than a respondent being in broad agreement with a statement i.e. more respondents are used to determine the latter. In the Sydney and Oxford studies, they did not use the same scoring method to capture overall view of satisfaction, instead they asked one question on satisfaction based on a single Likert scale. The SWUL scale is psychometric scale and the existing scoring outlined by the authors should be calculated, however in further analysis the researcher could ask one question to capture overall satisfaction based on a single Likert scale. The current definition is narrow and would mean UCC are reporting lower levels of broad satisfaction compared to other institutions even on a like for like basis. Updating the definition would allow UCC to show a higher level of satisfaction and be used a positive aid for marketing the college.

It is recommended that the same scales; Satisfaction with University Life (taking into account the point above), Appropriate Workload, Good Teaching and Motivation scales be re-used in the next study. The wording of some scales was altered, for example the wording of the SWUL had to be altered slightly to suit the UCC context. For validation of the alterations made, the same scales should be reused so that the same analysis can be performed. It should result in similar findings. Only until the altered scales are tested in this way, can they be considered for future use.

1.5 Conclusion

The study was the first of its kind for UCC and the aim of the study was to get an overall impression of the student experience in UCC. The overall student experience in UCC was positive for students. These results provide senior management with information to help understand the current student experience and identify areas, which are good and areas in need of further development. Some of the headline takeaways were that students were proud to be members of UCC and the vast majority were satisfied with university life at UCC. Students felt UCC demonstrated good teaching practices and the programmes of study were motivational for students. In addition to teaching and learning experience at UCC being positive, the social side of UCC was also well received. These results can help drive the student experience in UCC, because in Ireland, student experience is becoming a much more competitive state for universities.

One issue identified in the study was the dissatisfaction with the timelines and level of feedback from assignments. This issue is not unique to UCC, and can be addressed immediately by UCC management. The study also showed a willingness by students to avail of training for transferrable skills. Transferrable skills are very important for both employers and students. Students who are well rounded have a competitive advantage in the labour market. An initial skill programme should be put in place with longer strategy developed to encompass work placed training for all students.

The study was a success as it provided senior management with a baseline view of the student experience at UCC. The study identified good and bad components of the student experience at UCC and outlined some of improvements for the inaugural study. Future surveys should incorporate the changes and further studies will provide better information to senior management to make decisions that reflect the UCC population. Offering the best student experience will help UCC attract the best candidates.

Since the original student experience study UCC have continued to monitor student experience. The results of the survey are driving change and providing insight into key areas of the student experience at university. Changes identified in this study have been implemented to date, for example, the introduction of work placement managers to oversee activities for work experience and workshops for transferrable skills. In conclusion, the surveys are providing rich insight into student experience in UCC and UCC are using this feedback to drive change and improvement in each of the areas of student experience.

1.6 References

Ali, S., Foreman, J., Capasso, A., Jones, A., Tozan, Y. and DiClemente, R., 2020. Social media as a recruitment platform for a nationwide online survey of COVID-19 knowledge, beliefs, and practices in the United States: methodology and feasibility analysis. BMC Medical Research Methodology, 20(1).

America Institute of Research Air.org. 2021. [online] Available at: <https://www.air.org/sites/default/files/Uncommon-Measures-Student-Surveys-Guidance-Nov-2015.pdf

Andrews, J. and Higson, H., 2008. Graduate Employability, 'Soft Skills' Versus 'Hard' Business Knowledge: A European Study. Higher Education in Europe, 33(4), pp.411-422.

Anderson, P. and Novakovic, Y., 2017. Listening to student views on the transition from work placement to the final year. Accounting Education, 26(4), pp.377-391.

Cape, P. , & Phillips, K. (2015) Solutions, W., 2021. [PDF] White Paper. Questionnaire Length and Fatigue Effects: The Latest Thinking and Practical Solutions - Free Download PDF. [online] Silo.tips. Available at: <https://silo.tips/download/white-paper-questionnaire-length-and-fatigue-effects-the-latest-thinking-and-pra>

Collins-Nelsen, R., Koziarz, F., Levinson, B., Allard, E., Verkoeyen, S. and Raha, S., 2021. Social context and transferable skill development in experiential learning. Innovations in Education and Teaching International, pp.1-10.

Coplin, W.D. (2003). Ten things employers want you to learn in college: The know-how you need to succeed. Berkeley, CA: Ten Speed Press.

Corkery, E., Cronin, F., Grannell, A., Dool, R., Kenny, A., O'Sullivan, K., Fernandez, F., Murphy, E., Carroll, M., Hourihane, R., Madden, J., McCarthy, D., Murphy, S., O'Keeffe, J., O'Sullivan, S. & Poirier, J. (2010). Statistical consulting project utilising a team-based learning environment: A student's perspective. INTED2010 Proceedings.

Cresswell, J., (1994). Research design: Qualitative and quantitative approaches. Sage Publications.

Crebert *, G., Bates, M., Bell, B., Patrick, C. and Cragnolini, V., 2004. Developing generic skills at university, during work placement and in employment: graduates' perceptions. Higher Education Research & Development, 23(2), pp.147-165.

Department of Education and Skills. (2009). Retrieved October 29, 2010, from http://www.education.ie/servlet/blobservlet/web_stats_08_09.pdf

DeVellis R.F. (1991). Scale development: Theory and applications. Thousand Oaks, CA: Sage Publications.

Diener, E., Emmons, R., Larsen, R., & Griffin, S. (1985). The satisfaction with life scale. Journal of Personality Assessment 49 (1), 71-75.

Finch, D., Peacock, M., Levallet, N. and Foster, W., 2016. A dynamic capabilities view of employability. Education + Training, 58(1), pp.61-81.

Hair, J., Black, W., Babin, B., Anderson, R. & Tatham, R. (2006). Multivariate Data Analysis (6th ed.), London, Pearson Prentice Hall International.

Heppner, P.P., Kivlighan, D.M., & Wampold, B.E. (1998). Research design in counseling (2nd ed.). California: Wadsworth.

Higher Education Authority. 2021. Irish Survey of Student Engagement National Report 2020 | Higher Education Authority. [online] Available at: <https://hea.ie/2020/11/24/irish-survey-of-student-engagement-national-report-2020/

Kato, T. and Miura, T., 2021. The impact of questionnaire length on the accuracy rate of online surveys. Journal of Marketing Analytics, 9(2), pp.83-98.

Kelly, G., 2012. Lecture attendance rates at university and related factors. Journal of Further and Higher Education, 36(1), pp.17-40.

Kim, J. & Mueller, C.W. (1978). Introduction to factor analysis: What it is and how to do it (Volume 13). Sage Publications.

Kline, P. (1994). An easy guide to factor analysis. NY: Routledge.

Mardia K., Kent J. & Bibby J. (1979). Multivariate analysis. Academic Press Inc.

Marsden, P.V. & Wright, J.D. (2010). Handbook of survey research (2nd ed.). Emerald Group Publishing Limited.

McGann, K. and Anderson, G. 2018. National Survey of Employers Views of Irish Higher Education Outcomes

Nottingham Trent University| www.ntu.ac.uk/cadq|Mar-13

Pett, M.A., Lackey, N.R. & Sullivan, J.J. (2003). Making sense of factor analysis: The use of factor analysis for instrument development in health care research. Thousand Oaks, CA: Sage Publications.

Porter, S. and Whitcomb, M., 2005. Non-response in student surveys: The Role of Demographics, Engagement and Personality. Research in Higher Education, 46(2), pp.127-152.

Reddy, P. and Moores, E., 2012. Placement year academic benefit revisited: effects of demographics, prior achievement and degree programme. Teaching in Higher Education, 17(2), pp.153-165.

Revilla, M. and Höhne, J., 2020. How long do respondents think online surveys should be? New evidence from two online panels in Germany. International Journal of Market Research, 62(5), pp.538-545.

SAGE Journals. 2021. Ideal and Maximum Length for a Web Survey - Melanie Revilla, Carlos Ochoa, 2017. [online] Available at: <https://journals.sagepub.com/doi/abs/10.2501/IJMR-2017-039>

Sheskin, D.J. (2004). Handbook of parametric and non-parametric statistical procedures (3rd ed.). United States of America: Chapman & Hall/CRC.

The 1994 Group. (2007). Enhancing the student experience. Retrieved October 29, 2010, from http://www.1994group.ac.uk/documents/public/SEPolicyReport.pdf

Universitiesuk.ac.uk. 2021. Student experience. [online] Available at: <https://www.universitiesuk.ac.uk/policy-and-analysis/Pages/student-experience.aspx>.

University College Cork. (2009). University College Cork: strategic plan for the student experience 2009-2012. Retrieved November 26, 2009, from http://www.ucc.ie/en/strategicplanning/strategic/Strategic%20Plan%20for%20the%20Student%20Experience%202009-2012.doc

University of Oxford. (2009). Retrieved November 26, 2009, from http://ceq.oucs.ox.ac.uk

University of Sydney. (2009). Student Course Experience Questionnaire. Retrieved November 26, 2009, from http://www.itl.usyd.edu.au/sceq/

Www2.deloitte.com. 2021. [online] Available at: <https://www2.deloitte.com/content/dam/Deloitte/br/Documents/technology-media-telecommunications/preparing-tomorrow-workforce-for-4IR_C.pdf>

Appendix 1.A

Histogram for SWUL Scale

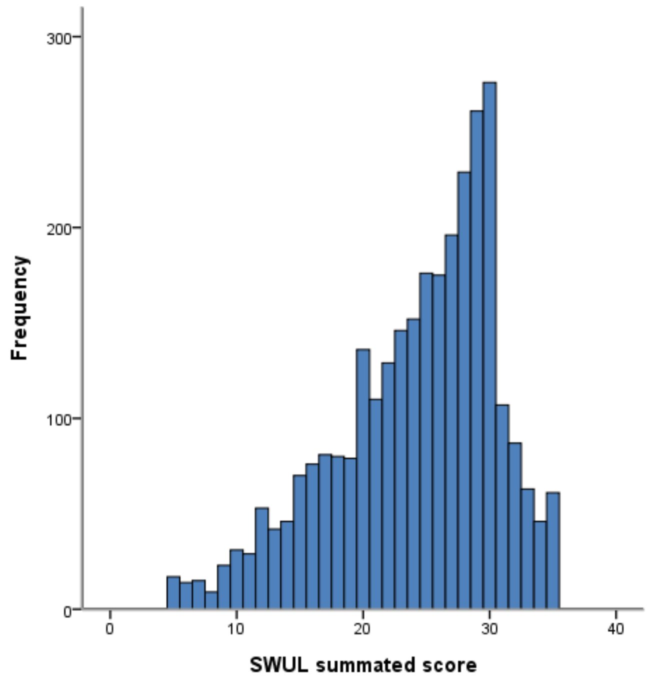

Figure A1.1: Histogram of the SWUL Summated Score

Appendix 1.B

Table 1.40: Eigenvalues and Percent Variance Explained for all Factors of Teaching and Learning Scale.

Factor	Eigenvalues	% of Variance Explained
1	6.09	32.07
2	2.92	15.39
3	1.78	9.39
4	.999	5.259
5	.817	4.299
6	.796	4.192
7	.674	3.549
8	.608	3.200
9	.580	3.051
10	.514	2.708
11	.482	2.539
12	.417	2.195
13	.408	2.150
14	.368	1.937
15	.365	1.921
16	.338	1.778
17	.322	1.695
18	.283	1.491
19	.227	1.196

Appendix 1.C

Histograms for scales in Teaching and Learning Instrument.

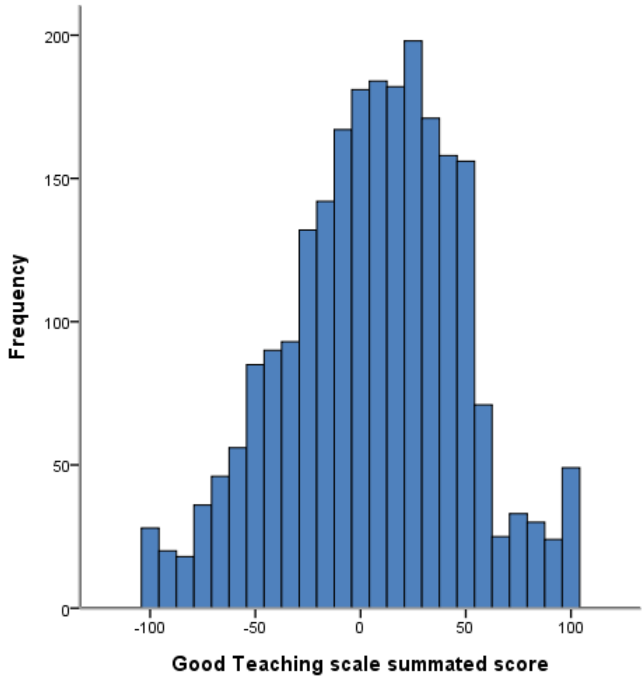

Figure A1.2: Histogram of the Good Teaching Summated Score.

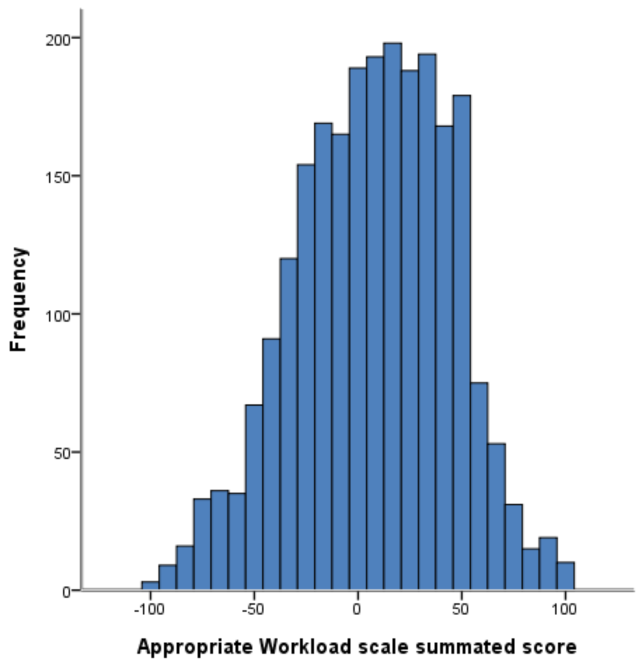

Figure A1.3: Histogram of the Appropriate Workload Summated Score.

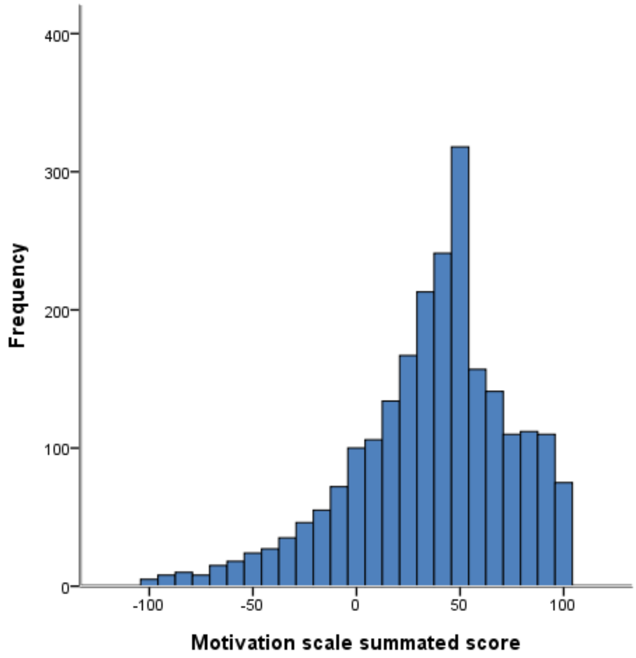

Figure A1.4: Histogram of the Motivation Summated Score.

Chapter 2: Quantification of Dynamic ^{15}O Data from PET Imaging of Alzheimer's Disease

Supervisor: Prof. Finbarr O'Sullivan and Eric Wolsztynski

Disclaimer: The anonymised image data in this chapter was generated at the University of Washington PET Imaging Center on a GE Advance scanner. Use of these data was supported in part by Science Foundation Ireland under the MI-2007 grant. Work on the data was supervised by Finbarr O'Sullivan, Department of Statistics, UCC. The use of these data outside of this thesis requires explicit permission from investigators at the University of Washington. Details can be had from Professor O'Sullivan.

2.1 Introduction

Dementia is the cause of a steady decline in a person's mental functioning. Alzheimer's disease is a common form of dementia that causes nerve cell death and tissue loss. Dementia is an irreversible disease, with patients affected by a range of symptoms. Patients experience progressive and fatal neurodegenerative disorder shown through cognitive and memory deterioration, reduced ability to conduct daily duties along with a number of behavioural disturbances (Cummings, 2004).

The cause of what initiates the disease has not yet been determined, however Alzheimer's patients tend to build up a high amount of tangles and plaque substances known as beta-amyloid in the brain. The formation of these substances can develop up to twenty years prior to symptoms. These substances affect neurons in the brain, causing them to function less efficiently. The neurons cannot then communicate with one another and evidently die. As the amount of tangles and plaque substances increase in the brain, the more neurons are affected and thus the brain physically shrinks (US National Institute of Health). Around 44,000 people in Ireland suffer from dementia, with that number to increase to 104,000 by 2036 (Alzheimer's Society of Ireland).

Ireland's population, along with a number of other countries, are living longer due to better quality in health care. Approximately one in twenty people over the age of sixty are affected by dementia, with this number rising to one in four being affected over the age of eighty. Approximately sixty-six percent of cases are Alzheimer's. Although there is no cure for Alzheimer's, there are drugs to combat the disease. However, early detection of the disease may enable earlier treatment or at a minimum allow the patient to plan for the future. Confirming the presence of Alzheimer's is a lengthy process of elimination which could stretch to nearly three years of cognitive and diagnostic tests.

Medical advances have been great in the area of molecular imaging. Using this technique, abnormalities in the brain can be detected, which indicate the presence of the disease, by using Positron Emission Tomography (PET), the most common molecular imaging technique to detect the disease (Society of Nuclear Medicine, 2009).

Alzheimer's is a disease and diseases are a biological process. PET imaging is both sensitive and informative of biological processes. Here, PET scans detect the disease by identifying abnormalities in the patient before symptoms are expressed (Phelps, 2000). PET provides the examiner with a very powerful tool to study Alzheimer's disease. Multiple scans of a patient allow the user to examine the disease over time, monitoring the disease and how it evolves over time will provide insight into the state of the disease. Given that there is no cause or cure, understanding the disease and how it behaves could lead to a cure or better drugs to fight Alzheimer's disease. The field of Alzheimer's research is so active at present trying to understand the causes of the disease and develop better techniques to treat it (Haense, Herholz, Jagust & Heiss, 2009; Tolboom et al., 2010).

PET measures the kinetics of radiotracers in living tissue. PET scans can measure body functions, for example blood flow, oxygen use, and glucose metabolism (O'Sullivan F, 2006). A tracer is injected intravenously into a patient; this tracer is a positron emitting radioisotope. In general, there are two main tracers used for PET scans/ radio labelled water ($^{15}O - H_2O$) is used for studies when cerebral blood flow is of interest. ^{18}F-Flourodeoxyglucose or ($^{18}F - FDG$) is another tracer which is used in studies, if the examiner was interested in assessing glucose metabolism (Terry, Katzman, Bick & Sisodia, 1999).

Radioactive tracers, by their nature, decay or breakdown over time, as they break down they emit positrons. When the tracer is administered to a patient, these positrons are emitted in the body. Annihilation occurs when a proton and an electron collide; here the energy is emitted through two photons in the form of gamma rays.

The gamma rays are then detected in the PET scan and recorded (Phelps, 2000), whereby three-dimensional images are then created using the data.

The scans produced by PET imaging can be either static or dynamic. A static scan will produce results from a point in time while a dynamic scan produces results over time. The dynamic scan provides more information to the user compared to the static scan. Dynamic scans involve using multiple scans, where each scan is known as a frame for the dynamic PET image. However, there are issues with using dynamic scans; in particular, there are issues with the tracer. As such, the scans used in the dynamic scan need to have increasing duration in order to compensate for the tracer decay. The tracer can also leave the body through the kidneys and the transient retention in membrane lipids (O'Sullivan et al., 2009). Other issues arise due to the random nature of radioactive decay, which makes it hard to establish the origin of the gamma rays due to noise. The measurement process is also hard to interpret, due to the multivariate nature of the process. Modelling techniques exist which allow for corrections for tracer decay and timing delays, which occur in the measurement process. Kinetic analysis by compartmental model is the most common technique used.

The data produced from dynamic PET scan is very large, analytical software have allowed users to view this large data in an image. The users can conduct analysis of different areas / regions. These areas / regions are known as regions of interest (ROI) sampling. Selecting multiple ROI's and comparing them is an important aspect of the study, this can be used to analysed the effect the disease on different parts of the brain. The ROI's can be analysed using compartmental models and a parametric analysis.

This study examines a PET scan of a brain infected with Alzheimer's disease. There are known abnormalities in the brain, which indicate a presence of Alzheimer's. Cerebral hypo-perfusion is one of these characteristics (Nishimura et al., 2007). In essence, this is a decrease in cerebral blood flow; studies have shown that a decrease in cerebral blood flow leads to an increase in cerebral atrophy. An increase in cerebral atrophy is a feature of dementia of the Alzheimer type (Jobst et al., 1992; Meyer et al., 2000).

The aim of this current study is to quantify biological parameters for different regions of the brain, one of which can be used to analyse cerebral blood flow in different regions of the brain. Using the ROI sampling, the user can analyse areas more susceptible to the disease and compare them to areas less susceptible. A comparison of the kinetic parameters obtained via the parameter method and non-parameter method. The study uses free software to conduct statistical analysis. An important aspect of this study is demonstrating how free software can be used in analysing PET data.

Using the Kety-Schmidt model, a parametric analysis of the data was conducted to estimate values for the blood flow (Kety-Schmidt, 1948). A non-parametric residue analysis was then carried out on our data to validate our results (O'Sullivan et al., 2008).

Alzheimer's disease affects close to 55,000 people in Ireland, with that figure increasing to 113,000 by 2036 (Alzheimer's Society of Ireland, 2021). This figure has increased since the original study in 2010. Research and fighting the disease has continued with international funding increasing year on year. One of the key tools for fighting the disease is the use of PET techniques (Peretti et al., 2019; Milestones, Alzheimer's Disease and Dementia).

The Alzheimer's research field is expanding, however these new studies continue to investigate the impact on cerebral blood flow using methods and tools outlined in this study. The Kety-Schmidt model is still the method used for estimation in similar studies which utilise PET and MRI imaging, (Koopman et al., 2017). A study by Deo et al (2014) compared patients with Alzheimer's disease and patients without the disease. Using a control group allowed them to compare the results of different parameters, for example cerebral blood flow was found to be lower in Alzheimer's patients.

Another study investigated the impact of cerebral blood flow and tau pathology (Peretti et al., 2019). Tau is a protein, in Alzheimer's patients, tau acts in a different manner to healthy patients. A glucose-based tracer such as (Flortaucipir. ^{18}F) is used as the tracer in the study by Peretti (2019). Low level of cerebral blood flow still leads to the possibility of the disease being present. The software used in the original study is still freely available for use.

2.2 Methods

This section explains how our data was acquired and prepared for our analysis. We describe here two exploratory techniques used in the analysis, the parametric and non-parametric residue analysis, and how they were validated.

2.2.1 Data Acquisition

This study focuses on a patient suffering from Alzheimer's disease. Blood flow quantification in the brain is an important part of studying Alzheimer's disease, with a PET scan using a water tracer being the most effective way of analysing this.

A PET scan was taken of the patient using a GE Advance PET machine. As mentioned above, water tracers are best for analysing blood flow and perfusion. In this study, we are using a radiolabelled water tracer ($^{15}O - H_2O$), which is constructed using a nuclear cyclotron or generator. It is a normal molecule of water, except that there is a radioactive oxygen atom instead of a non-radioactive isotope. This radioactive atom allows the PET machine to obtain the images of the body. After the tracer is administrated into the blood, it has a half-life of 2.08 minutes.

$$half - life = e^{-\alpha t} = \frac{1}{2}$$

The half-life of the isotope is modelled using an exponential function, $e^{-\alpha t}$, this implies that, α , the decay constant of the isotope is

$$\alpha = \frac{\log 2}{2.08}$$

The scan lasted five minutes, a dynamic PET image consists of a sequence of continuous scans. Each scan increases in duration, this highlights the distribution of the tracer throughout the region. Each sequence scan is referred to as a frame in the dynamic PET image. There were 42 frames in this study. The dynamic PET image was viewed in A Medical Imaging Data Examiner (AMIDE). AMIDE is a freely available software and is a very important tool in PET studies for viewing and analysing imaging data sets (Loening & Gambhir, 2003). AMIDE produce a three-dimensional view of the images. Magnet Resonance Images scans (MRI scans) of the patient were also available and used in this study.

The images shown through AMIDE comprise a number of voxels in each frame and our PET values are the values of each voxel.

$y_i(t)$ is the raw measurement at voxel i at time t. We assume that $y_i(t)$ is approximately Poisson.

To adjust our measurements for time and radioactive decay, we need to divide by duration of the scan and multiply by the half-life.

$$z_i(t) = \frac{y_i(t)}{D} * half - life$$

Where D denotes the overall duration of the scan. PET values adjusted for time and decay, $z_i(t)$, are referred to as time activity curves. Adjusting the raw data like this allows multiple scans taken at different times to be analysed at a single point in time. Statistical tests can be conducted on the data to get a full picture of what has occurred.

2.2.2 Co-Registration of PET images and MRI scans

The data was uploaded onto AMIDE in order to view the data as an image.

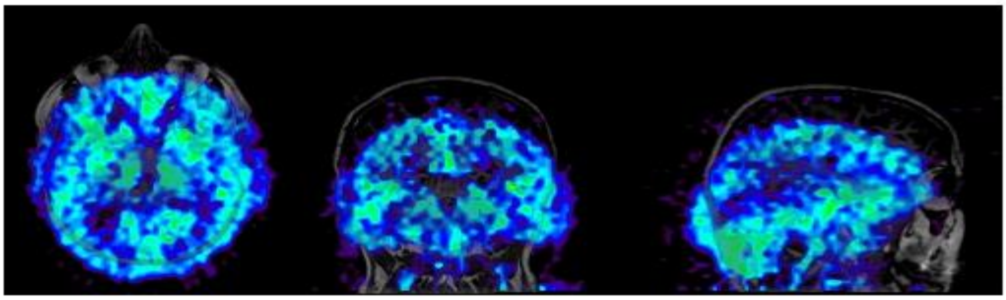

Figure 2.1 Coloured PET Water Dynamic Images with Black and White MRI Images Superimposed.

Figure 2.1 shows one frame of the PET water dynamic scan on top of one frame of the MRI scan. There were 42 frames in total. Each one representing a snapshot of the tracer activity, the frames are not equal in length. The PET scan part of the image is the coloured part, the CT scan is the black and white part of the scan. There are two problems with the combined image as it stands, the PET image is not aligned to the MRI scan and the PET image has a lot of noise. The scans need to be aligned and the noise needs to be reduced.

The images must be aligned correctly, to do this, features or landmarks must be identified in both the PET and MRI scans to align them on top of one another. To get the best results, each frame was examined to determine which frame was best to use for the co-registration. On review of all the frames, a different frame to the one used in **Figure 2.1** was deemed to have more landmarks, this was used to co-register the PET and MRI scan. This can be highlighted by comparing the nose or front of head in image 3 of scans **Figure 2.1** and **2.2**. Co-registration of images is an iterative process to get the best result. Co-registration can be conducted on any frame but once co-registration is complete it applies to all frames.

The fiducial mark option and alignment wizard in AMIDE were used to co-register the images. At least three landmarks common to both must be chosen, the fiducial marks are applied to both images. Apply the Alignment Wizard and the images will be superimposed on to one another. The result is shown in **Figure 2.2.**

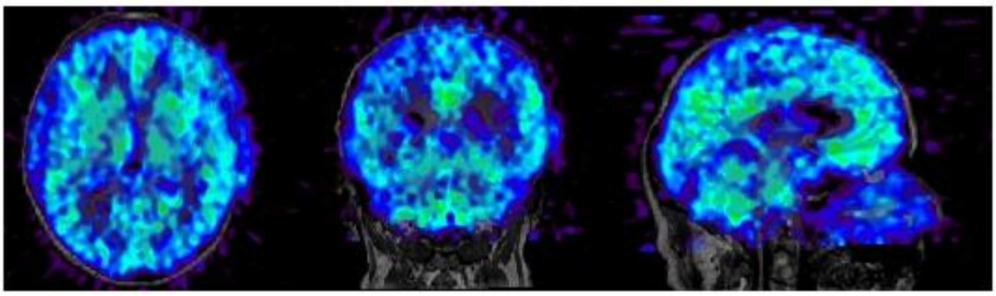

Figure 2.2: Rotated Coloured PET Water Dynamic Image Co-Registered with MRI Scan.

Figure 2.2 is not that clear, this is known as noise in the image. A Gaussian filter was applied to the water dynamic data set to clear or remove as much of this noise as possible. Reducing or removing the noise gives the user a clearer image, this allows the user to choose regions of interest with more accuracy and certainty. The result is illustrated in **Figure 2.3**.

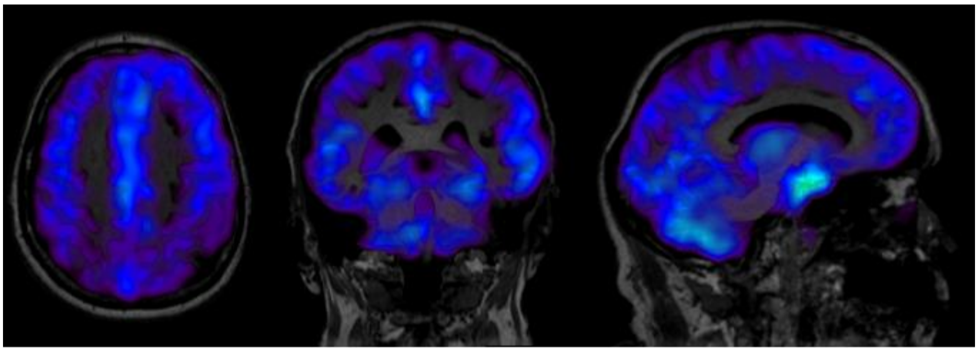

Figure 2.3: Rotated Gaussian Filtered PET Water Dynamic Co-Registered with an MRI Scan.

This analysis was conducted on the Gaussian filtered data set. Using an anatomic reference (The Whole Brain Atlas), regions of interest were identified. These regions were then located on the MRI scan and analysed. Locations of all regions of interest can be found in the Results section.

2.2.3 Principal Component Analysis

Principal Component Analysis (PCA) is used in dynamic PET imaging to summarise the kinetic information contained in the image (Pavlopoulos, Thireou, Kontaxakis, & Santos, 2007; Razifer et al., 2009; Thireou et al., 2003). PCA provides a visual inspection of the data. PCA denosies the PET image and it retains meaningful information of the cellular process in the image. PCA can be used to identify areas with unique uptake patterns and highlight interesting information in the first few components.

Principal Component Analysis (PCA) is an exploratory tool used to explain the internal structure of a dataset, which best explains the variance with losing as little information as possible. PCA is a data reduction technique and tries to explain as much of the variance in as little components as possible (Hair et al., 2005). PCA is used to reduce the information contained in a large number of variables into a smaller set of components. The smaller component can be analysed to provide a meaningful understanding of the underlying structure of the data. This technique can be employed if the dataset is correlated, the PCA will provide a smaller set of uncorrelated components.

PCA is an iterative process, the first step is to extract the first principal component, which explains as much of the variance as possible. The second step of the PCA process is to select another component from the components left in the original data set. This component will explain as much variance as it possibly can. The process continues until there are no components remaining and all the variance is explained. The analysis was conducted in R, this determined the number of components. AMIDE was used to generate principal component images.

As described in **Section 2.1**, the PET data $z_i(t)$ is represented by the voxel values. For PCA to work (i.e. explain the variance in as few components as possible), the subspace closest to Z is required, where $Z = (z_1, z_2 \dots z_T)T$.

Suppose we attempt to approximate a T-dimensional vector z in terms of a k-dimensional subspace:

$$z = \beta_1 e_1 + \beta_2 e_2 + \dots + \beta_k e_k + noise$$

Where e is a k-dimensional subspace $e = (e_1, e_2 \dots e_T)T$ and without loss of generality we get:

$$e_j{}' e_k = \begin{cases} 0 \ if \ i \neq k \\ 1 \ if \ i = k \end{cases} \tag{1}$$

In general, for a simple regression model of the type,

$$z = x\beta + noise$$

Where $x = [e_1, e_2 \dots e_k]$, the least squares estimate is given by the following,

$$\beta = (e'e)^{-1} e'z$$

As a result of (1), this simplifies to:

$$\beta_j = e_j' z \qquad \text{j = 1,2,...,k}$$

We have the Total Sum of Squares (TSS) = Regression Sum of Squares (RegSS) + Residual Sum of Squares (RSS). (Weisberg, 2005)

Giving the following equation:

$$z'z = \sum_{j=1}^{k} \beta_j^2 + \left| z - \sum_{j=1}^{k} \beta_j \ e_j \right|^2$$

We want to see how close the k-dimensional subspace is to the distribution of the data. We use the following Loss function.

L[e_1 ,..., e_k] = E[distance(z, k-dimensional subspace)] =E[RSS(e_1 ,..., e_k)]

$$= \mathrm{E}(z'z) - \sum_{j=1}^{k} e_j' \sum e_j$$

Where $\sum$ is the covariance matrix of z, $E(z) = 0$

If we minimise the loss function with respect to e, we get a subspace which is very close to the distribution of z. To do this we must choose e to maximise the RegSS. Thus, to make the expected distance small, choose $(e_1, e_2 \dots e_k)$ to maximise

$$\sum_{j=1}^{k} e_j' \sum e_j$$

Subject to (1)

This can also be done by letting the vector e be equal to a linear combination of eigenvectors of the covariance matrix, $\sum$;

$$e_j = \Gamma c_j = \gamma_1 c_{j1} + \gamma_2 c_{j2} + \cdots + \gamma_T c_{jT} \tag{2}$$

Where $\Gamma = [\gamma_1 \dots \gamma_T]$ and thus the covariance matrix is expressed as $\sum = \Gamma D \Gamma'$

Where D is the diagonal matrix;

$$\begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_T \end{pmatrix}$$

Where

$$\Gamma'\Gamma = \Gamma\Gamma' = I$$

$$e_k' e_j = c_k \Gamma' \Gamma c_j = c_k' c_j = \begin{cases} 0 \ if \ i \neq k \\ 1 \ if \ i = k \end{cases}$$

i.e. e is orthogonal

Since the vector e is a linear combination of eigenvectors from (2)

$$e'_k e_j = c'_k \Gamma' \Gamma c_j = c'_k c_j = 0$$

And finally

$$e'_j \sum e_j = c'_j \Gamma' \sum \Gamma c_j = c'_j D c_j = d_j$$

The maximum possible value for this quantity has to be bounded by total variance, which is equal to the sum of the diagonal of the covariance matrix ∑. But

$$\sum_{j=1}^{k} c_j{}' D c_j \leq tr(\Sigma) = \sum_{j=1}^{T} d_j$$

And with

$$\hat{c}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \quad \hat{c}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \quad ,\cdots, \quad \hat{c}_k = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix}$$

$$\sum_{j=1}^{T} \hat{c}_j{}' D \hat{c}_j = d_1 + d_2 + \cdots + d_T$$

This produces the optimal value for the plane and,

$$\hat{e}_j = \Gamma \hat{c}_j = \gamma_j$$

The span $\gamma_1, \gamma_2, \ldots, \gamma_k$ is the best k-dimensional subspace approximation for z.

To get the optimal results for the PCA method, the subspace that is closest to z must be achieved.

2.2.4 Cluster Analysis

Cluster analysis is a statistical technique that examines a data set to see if it contains a set of distinct groups or clusters. In the context of PET imaging, cluster analysis was used to separate time activity curves into a number of clusters with similar characteristics. Cluster analysis classifies the observations into groups based on particular common characteristics. Thus, the clusters should be homogeneous within the groups but heterogeneous across the groups (Hair et al., 2005). The two types of clustering employed in this study are hierarchical clustering and k-means clustering.

2.2.4.1 Hierarchical Clustering

Hierarchical clustering begins by assuming each observation in the data set is in their own cluster. Therefore, at the beginning we have N clusters, given N elements in our data set, the aim of hierarchical cluster analysis is to condense all these clusters into 1 cluster.

Hierarchical clustering is based on the distance between clusters and it uses the Euclidean distance, d_{ij}, between each cluster. For the case of N clusters, each point is a cluster. The two clusters that are closest together are combined. These two clusters form a new cluster and cannot move to a new cluster. Let one of these cluster be, C_j, and the other be, C_k, where $C_j \cap C_k = \emptyset$. We get the Euclidean distance for a (nxp) data matrix z, with $z_1 \ldots z_n$ rows. The distance between these points is, d_{ij}, where

$$d_{ij} = \sqrt{\sum_{k=1}^{p} (z_{ik} - z_{jk})^2}$$

There are now N-1 clusters remaining. It is an iterative process, and it continues until there is just one cluster left. Cluster analysis is best illustrated using a dendrogram. Dendrograms show graphically all steps of the cluster analysis as they converge to one cluster (Lorr, 1983).

2.2.4.2 K-means Clustering

K-means can be described as a non-hierarchical clustering method. K-mean clustering is another form of clustering but differs to hierarchical clustering because the user actually picks the number of clusters to be used. The last method used distance between clusters and, once the object was clustered, it was stuck there. That is not the case here, however; objects are combined due to the mean value and once an object is clustered it can move out of the cluster due to cluster means changing.

The process begins by letting the first K values in the data set act as temporary estimates of the K cluster means. Now the next object from the data is assigned to a cluster with a similar mean. The mean of this cluster now changes, as a new mean value is used in the calculation. The process continues in this fashion until all the objects are assigned to clusters. This process is again iterative, and objects can combine with different clusters whose mean values are closer to theirs. The process continues like so until no more objects can move from their cluster or until it hits K iterations.

2.2.5 Kinetic Analysis

"PET imaging has several advantages such as high sensitivity and ability for quantitative measurement" (Watabe et al., 2006). The images viewed using AMIDE software are generated by the PET data captured by the PET cameras, these images comprise of various signals. A mathematical framework was developed to isolate the components of these signals (Watabe et al., 2006). There are many well-established models in PET; for instance, the quantification of blood flow (Kety and Schmidt, 1948) and cerebral metabolic rate for glucose. However, in order to optimise our PET data, we must understand some of the fundamental principles of PET. PET data represents the tracer concentration at a certain time t. When we want to analyse the PET data, we must assume there are physiologically separate pools of tracer substance namely "compartments".

"Compartmental analysis forms the basis for tracer kinetic modelling in Positron Emission Tomography (PET)" (Gunn et al., 2001). Kinetic analysis allows us to obtain important metabolic information from dynamic PET data. Kinetic analysis is based on a convolution between an arterial input function and a tissue residence. The concentration of the tracer of a particular voxel at time t is the PET data at time t, which has been attenuated, time delayed and decay corrected. Each voxel can be viewed as a different compartment or vascular space, which would be made up of plasma, arteries, and veins, along with extra-vascular space such as tissue and muscles.

Suppose the tracer is located in a particular compartment, say, the plasma, C_p and it is free to move into another compartment, say, the extra-vascular space, C_e. The rate of interaction between these compartments is known as a kinetic constant and each compartment has a specific kinetic constant.

There are, however, a number of assumptions to using compartmental models when interpreting PET data, (1) linearity, the total concentration of tracer in the tissue can be described as a linear sum of the concentration in each compartment. (2) Tracers act in a time invariant fashion. (3) Tracers that pass from one compartment to another are instantaneously mixed in the compartment. Finally, (4) the concentration of the tracer is small enough not to influence physiological processes and molecular interactions of the system under study (Gunn et al., 2001).

In this study, a water tracer ($^{15}O - H_2O$) is used, the number of compartments is based on the type of tracer used. A one compartmental model is used in the event of a water tracer (Watabe et al., 2006).

2.2.5.1 The One Compartment Model

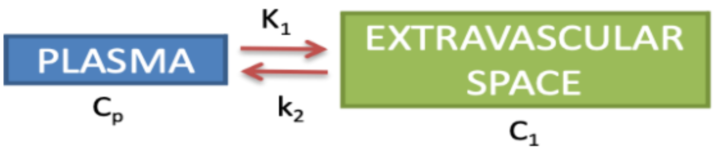

Figure 2.4: Kety-Schmidt 1-Compartmental Model for the ^{15}O Tracer. Arrows Indicate Transfer Constant between Compartments.

Where;

- The function $C_p(t)$ has two interpretations, it can be known as the Arterial Input Function (AIF) or Plasma Input Function (PIF). $C_p(t)$ represents the activity in plasma over time. A plot of our AIF can be found in **Appendix 2.C**.
- The function $C_e(t)$ is known as the retention. $C_e(t)$ represents the activity in the extra-vascular space.

- K1 is said to be the delivery of the tracer to the tissue or the rate at which material goes from plasma$C_p(t)$ into extra-vascular space $C_e(t)$.
- K2 is said to be the delivery of the tracer back to the plasma $C_p(t)$ from the extra-vascular space $C_e(t)$.

The total tissue concentration, $C_T(t)$, represents the activity in the total tissue concentration.

Figure 2.4 shows the one compartmental model. The first order differential of the extra-vascular space,

$$C_e^{'} = K_1 C_p(t) - k_2 C_e(t)$$

Using Fourier methods it becomes,

$$C_e(t) = K_1 \int_0^t e^{-k_2(t-s)} C_p(s) ds$$

$C_e(t)$ is as a function of time. The total tissue concentration

$$C_T(t) = V_B C_P(t) + K_1 \int_0^t e^{-k_2(t-s)} C_p(s) ds$$

Thus, the one compartmental model is,

$$C_T(t) = V_B C_p(t) + K_1 C_e(t)$$

This model is used to estimate a number of important parameters. Consider a function, a residue function, $R(t)$. $R(t)$ acts like a survival function where,

$R(t)$ decrease as t increases

$$R(0) = 1$$

$$R(t) = 1 - \int_0^t h(\tau) d\tau$$

And h is a function, such that $\int_0^t h \le 1$

$$C_e(t) = \int_0^t R(t-s) C_p(s) ds$$

Substituting into original expression for

$$C_T(t) = V_B C_p(t) + K_1 \int_0^t R(t-s) C_p(s) ds$$

Thus, in the case of the one compartmental model we get that the residue function,

$R(T_e) = e^{-k_2(t)}$ The residue function is also known as the impulse response, as it resembles the memory of the tissue. This leads to further parameters which can be estimated in the one compartmental model. These parameters are the flow, flux, volume distribution (V_D) and mean transit time (MTT). In the one compartmental model, the MTT is represented as,

$$MTT = \int_0^t R(t)dt = \int_0^\infty \tau h(\tau) d\tau = \frac{1}{k_2}$$

Where $h(\tau)$ is a residence time density, it is a probability density function of the length of time a tracer atom stays in the tissue.

The volume distribution, (V_D) is given by

$$V_D = K_1 MTT = \frac{K_1}{k_2}$$

Flow is represented as,

$$K_1 = Flow$$

Flow is also represented as,

$$Flow = \frac{V_D}{MTT}$$

Finally, Flux can be calculated as follows,

$$Flux = K_1 R(T_e)$$

Where $R(T_e)$ is the extraction at the end of the study.

Please note, the kinetic model works on the basis that the data is homogeneous data.

2.2.6 Non-Parametric Residue Analysis

In this study, parametric and non-parametric residue analysis were investigated. The non-parametric analysis focuses on tissue residue. The residue is related to tracer residence in the tissue, and it is a scaled survival function (O'Sullivan et al., 2008).

The non-parametric residue analysis is not reliant on specific compartmental modelling assumptions. This technique utilises regularised B-spline approximation of the residence time distribution. Statistical evaluations of specific parametric models can be determined. Comparisons can be made between different key functions of the residue, i.e. tracer flow, VB and VD.

There are benefits to using a non-parametric residue analysis over using a parametric analysis. The difference is due to early representation of the structure of the tracer residence as a result of the tracer being a large function of the vascular supply network. In addition, the parametric analysis is a homogeneous model, where a non-parametric analysis can provide valid kinetic quantification in heterogeneity studies or studies where the accuracy of a compartmental model approximation of the tissue residue is unknown.

2.2.7 Correlation between Parametric and Non-parametric Residue Analysis

To validate the results between our parametric and non-parametric residue analysis, we investigated the correlation between the two techniques. A parametric and non-parametric measure of statistical independence between the variables was applied to our estimates. The non-parametric measure we used was Spearman's rank correlation coefficient test,

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

Where, d_i^2 is equal to the distance between the two estimates and n was the number of estimates.

Pearson's correlation coefficient was the parametric measure we used, which measures the strength and direction of a linear relationship between two variables.

Pearson's coefficient is usually used with symmetric data. Pearson's coefficient is formulated as follows,

$$\rho_{X,Y} = corr(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y},$$

where X and Y are two random variables, with expected values $\mu_x$and μ_y and standard deviations $\sigma_x$andσ_y. E represents the expected value function.

2.3 Results

In this section, the PET data is examined. The results of our exploratory techniques, parametric and non-parametric analysis, are presented. The exploratory techniques PCA and Cluster analysis are presented first. Using these results, blood regions of interest were located, and parametric and non-parametric analysis were conducted. Using an anatomic reference (The Whole Brain Atlas), regions of interest were identified and located for the study.

Alzheimer's disease causes cerebral blood flow to decrease in areas infected by the disease. The difference in activity in different grey and white matter regions will be an important aspect of our results, in particular, analysing the grey/white matter ratio and comparing these results with published results.

2.3.1 Methodological Approach

Principal Component Analysis (PCA)

Principal Component Analysis is a data reduction technique. The process determines how many components explain as much of the variance without loss of information. The data was analysed using R and the number of components can be seen in **Figure 2.5**.

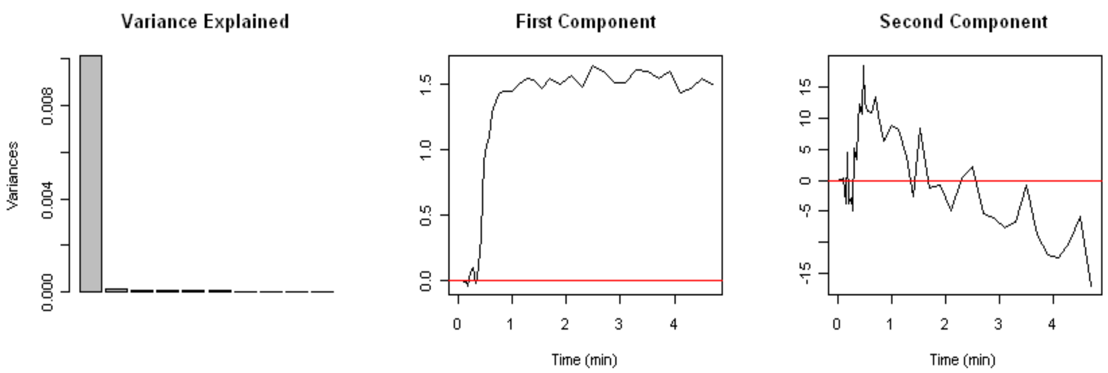

Figure 2.5: Results of PCA in R. 1st plot indicates that the majority of our variance is explained in the 1st component. The 2nd and 3rd plots show the activity attributed to each component.

Figure 2.5 illustrates that the first component explains the majority of the variance, 90.93%, while the second component explains a small amount of the variance, 0.9%, and less variance again is explained from there on. We used AMIDE to get a visual representation of this information.

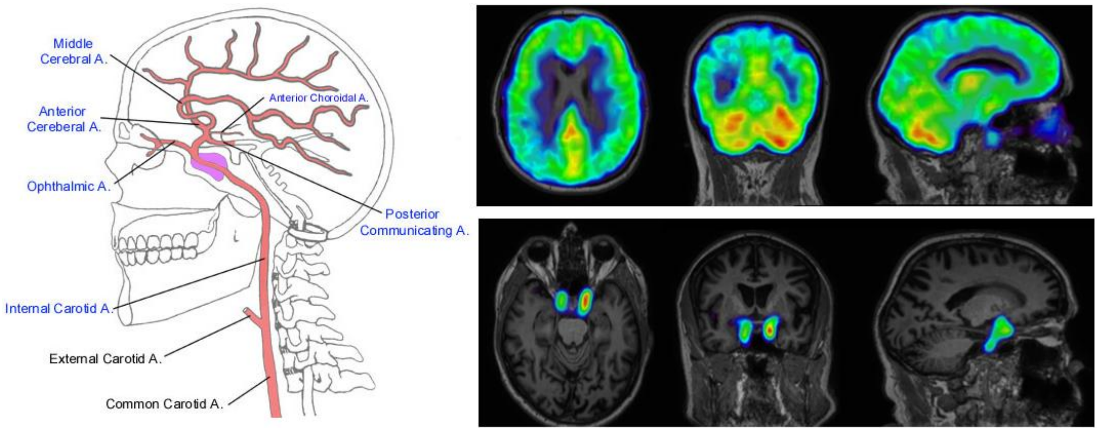

Figure 2.6: Map of Arteries. 1st and 2nd principal component co-registered with the MRI scan.

Figure 2.6 illustrates the information shown by the 1st and 2nd component, whereby the red areas depict blood flow, the green area depicts white matter, while the outline is the grey matter. The 1st component shows a lot of activity, while the 2nd component shows activity in one particular region. Using the map of arteries (Loyola University of Chicago School of Medicine) we feel this region is a combination of the anterior cerebral artery and the ophthalmic artery. Finally, using these images, we could select blood regions to run our parametric and non-parametric analysis on.

When we looked at the third, fourth and fifth component onwards we found no new information.

Cluster Analysis

The second exploratory technique investigated was cluster analysis, both hierarchical and k-means cluster analysis were applied. The two techniques were applied to a small grey and white matter region. **Figure 2.7** depicts the dendrograms, which show two clusters in the grey matter region and three in the white matter region.

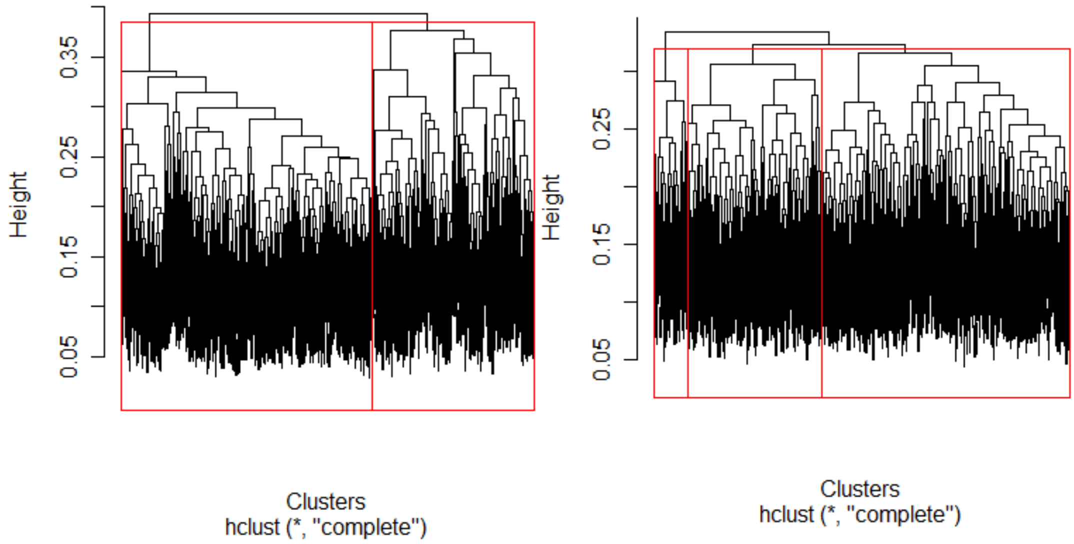

Figure 2.7: Dendrogram of small grey matter region split into two clusters and dendrogram of small white matter region split into three clusters.

Using the designated clusters for each region, hierarchical and k-means clustering were applied. The difference in the number of group of clusters could be due to the different structures of the regions.

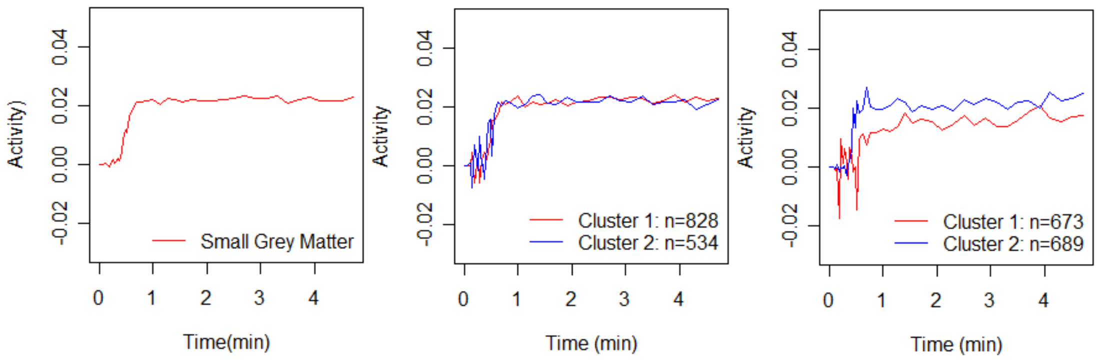

Figure 2.8: The activity of the small grey region is shown. The three cluster solutions for both hierarchical and k-means are shown.

Plot 1 in **Figure 2.8** shows the mean time activity curves (TAC) of the small grey matter region. The curve peaks and levels off at an activity level of 0.2 around the 30 second mark. An activity level of this amount would suggest it is a tissue region. Plots of the activity of curves under the two methods of cluster analysis are presented too. The numbers in each cluster differ depending on the technique used. Activity levels in the hierarchical method are similar to the TAC, while cluster 2 shows higher activity than cluster 1 when using the K-means method. This would suggest that there are different levels of activity in the tissue.

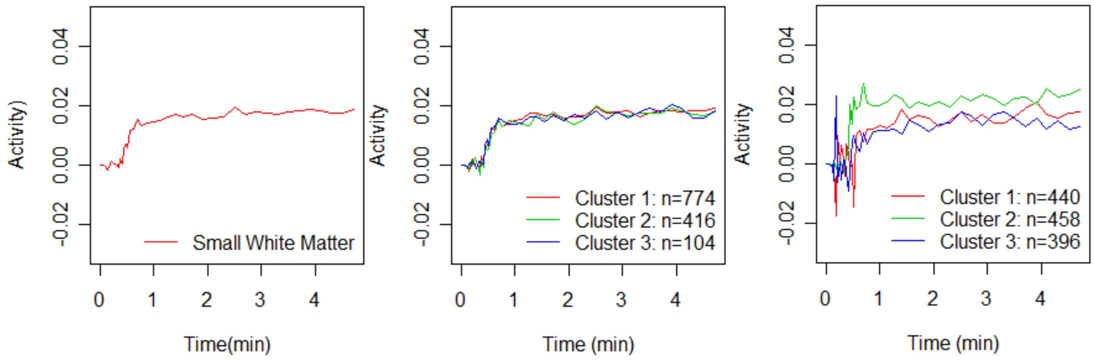

Figure 2.9: The activity of the small white region is shown. The three cluster solutions for both hierarchical and k-means are shown.

Plot 1 in **Figure 2.9** shows the mean time activity curves (TAC) of the small white matter region. The curve peaks and levels off at an activity level just under 0.2, around the 30 second mark. An activity level of this amount would suggest this is a tissue region. Plots of the activity of curves under the two methods of cluster analysis are presented too. There are three clusters identified in this region. The numbers in each cluster differ depending on the technique used. Activity levels in the hierarchical method are similar to the TAC, while cluster 2 shows higher activity than cluster 1 and 3 when using the K-means method. Although cluster 2 rises higher than the curve in the original TAC, the activity still resembles a tissue region compared to an artery. This would suggest that there are different levels of activity in the tissue.

Kinetic Analysis

Kinetic analysis was conducted on the data. Kinetic parameters were obtained through non-linear least squares, the results for each parameter can be found in appendix **2.B.** Boxplot of the estimates are provided in **Appendix 2.C** to determine the symmetric nature of the data. In general, where a boxplot is deemed to be symmetric the mean value is used otherwise the median value is used.

On review of the box plots, some regions may not be symmetric. Regions with a large number of outliers could be deemed not symmetric. To fully determine symmetry, further tests should be conducted, for example a simple visual inspection of a histogram. **Figure 2.26**, in **Appendix 2.C**, region 1 shows a number of outliers, which would insinuate the data is not symmetric while one region 4 would appear symmetric. The parametric values produced in the Results section are the mean values for each ROI. Also included in the results is a detailed analysis of activity levels, model fits and estimates for the parametric and non-parametric analysis. Illustrations of each region of interest are also supplied.

2.3.2 Hippocampus Regions and Grey Matter from the Cerebellum

The region of the brain that is most susceptible to Alzheimer's disease is the hippocampus region, located on both sides of the brain. From here on, they will be referred to as the hippocampus left and right regions. The hippocampus is a grey matter region. This region was compared with grey matter from the cerebellum, to look at the difference in activity of an area susceptible to early-stage Alzheimer's disease and an area that is not. Time activity curves for the three regions are illustrated in **Figure 2.10**.

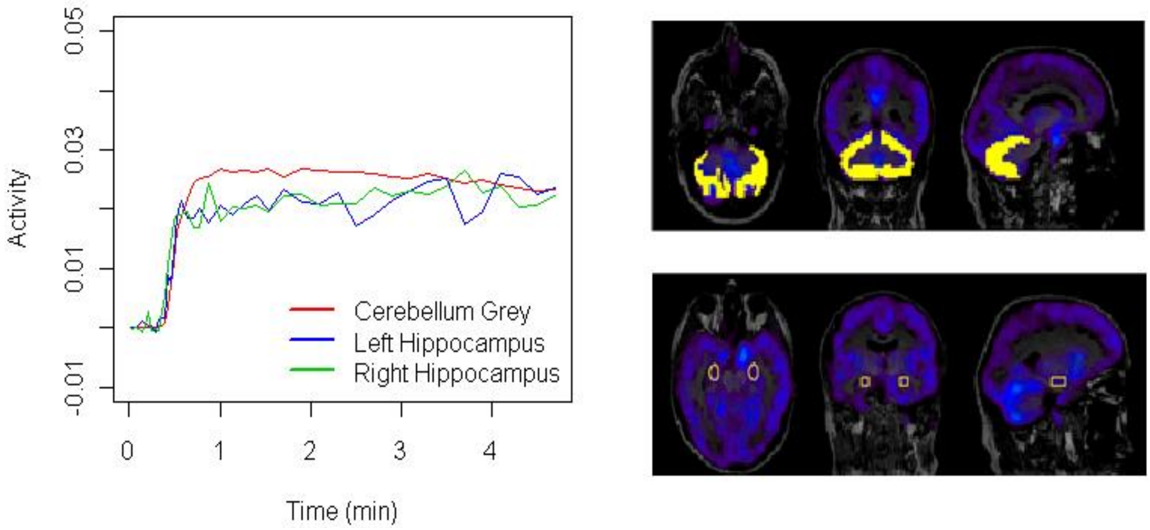

Figure 2.10: Time Activity Curves of the cerebellum grey matter region with the hippocampus left and right region. PET water dynamic image with MRI scan superimposed showing the cerebellum grey region and the left and right hippocampus regions.

The activity of the grey matter region from the cerebellum is higher than both sides of the hippocampus regions. This is what we would expect when dealing with an Alzheimer's patient. The area most susceptible to the disease is showing reduced activity compared to another region of the brain. As we are unaware to the extent of the Alzheimer's disease or the level of activity between the regions for a healthy brain, we could conclude that the activity in the hippocampus region is decreasing at a faster rate compared to the cerebellum grey due to the TACs.

The cerebellum plot is also less volatile than both hippocampus regions, due to the size of the regions. The model fits of these regions are illustrated in **Figure 2.11**.

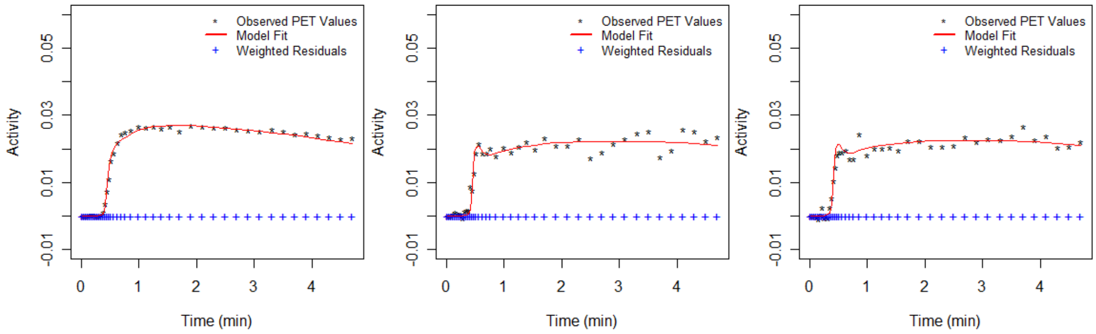

Figure 2.11: Model fits of the observed PET values for the cerebellum grey matter, hippocampus left and right. The residuals for each region are illustrated as well.

The model fits the data well with the cerebellum showing a very smooth fit. The model for both the hippocampus left and right region fits the observed PET values well. The weighted residuals are centred around zero, which is what we would expect. A further check using a Q-Q plot of the data could be conducted to ensure there is no pattern in the residuals.

The box plots in **Appendix 2C** show a symmetric nature, though the hippocampus region does have some outliers. Further checks of the residuals and normality could be conducted to ensure there is no sign of a pattern. If the variance does not appear to be constant, then it could affect the results.

The estimates for the model parameters, δ, VB, K1and k2, along with the number of voxels in each region, are provided in **Table 2.1**.

Table 2.1: Parametric analysis of the cerebellum grey, left and right hippocampus region.

Region	δ ± s.e	V_B ± s.e.	K_1 ± s.e.	k_2± s.e.
Hippocampus (Left) (n=108) *	0.0291±0.0254	0.1944±0.0368	1.0653±0.0455	0.1891±0.0281
Hippocampus (Right) (n=117)*	0.0163±0.0239	0.1997±0.0494	1.0995±0.0509	0.1995±0.0292
Cerebellum Grey (n=1,109)*	0.0102±0.0226	0.1031±0.0331	1.5576±0.0275	0.3172±0.0119

*n is the number of voxels per region.

The parametric results for K1 and k2 are larger in healthier areas. For example, the K1 for the cerebellum grey region is 1.557, compared to 1.0995 and 1.5576 in the hippocampus left and right region, respectively. The values are presented in table, with ±to allow for standard error either side. All the table of estimates in the results will be provided in this format.

Higher levels of K1 and k2 insinuate a higher level of activity in this area and is where the tracer moves between components. The equivalent parametric and non-parametric results for Flow, VB and VD are provided in **Table 2.2**.

Table 2.2: Parametric and non-parametric values of the Cerebellum grey, left and right hippocampus region

Region	Flow	VB	VD
Cerebellum Grey	1.5576±0.0275[1]	0.1031±0.0331	4.9105±1.236
	2.8328±0.0249[2]	0.0841±0.0008	1.9296±0.0151
Left Hippocampus	1.0653±0.0455[1]	0.1944±0.0368	5.6335±3.267
	2.5762±0.0310[2]	0.1419±0.0030	1.3300±0.0170
Right Hippocampus	1.0995±0.0509[1]	0.1997±0.0494	5.5113±2.634
	2.4051±0.0341[2]	0.1357±0.0030	1.5529±0.0185

[1]Parametric Analysis.

[2]Non-parametric Analysis.

The parametric values for flow in the cerebellum grey region are higher than both hippocampus regions. In essence, the area of the brain less susceptible to the disease has a level of flow. The non-parametric values differ to the parametric values substantially, however the values adhere to the parametric findings.

The parametric estimates for fractional blood volume value are lower in the cerebellum grey region compared to both hippocampus regions. The parametric and non-parametric values for VB in the cerebellum differ greatly, however the hippocampus regions are quite similar. Again, the findings of the non-parametric analysis adhere to the parametric analysis.

The parametric estimates for the distribution volume show the tracer concentration in the tissue to be higher than the plasma. They have values ranging from 4.91 to 5.63. The non-parametric estimates are lower but show a higher concentration of the tracer in tissue than plasma.

2.3.3 Grey and White Matter Regions

An important aspect of our study is to look at the difference in activity levels in grey and white matter regions. Three different areas of each are examined: a small region, approximately 1200 voxels, medium region, approximately 2300 voxels, and finally a large region, roughly 3500 voxels. In addition to these 6 regions, a further region was examined, a white region of the cerebellum, as it is not at immediate risk to early-stage Alzheimer's.

The activity in grey matter regions should be higher than white matter regions. The difference in grey and white matter is that grey matter consists of neural cell bodies where white matter does not. White matter gets its colour due to the presence of myelin. Grey matter gets its colour from capillary blood vessels. Due to the presence of these capillary blood vessels, we would expect higher blood flow.

The grey / white matter ratio is another important aspect of our study. We will look at the ratio between the different regions.

Figure 2.12 illustrates the TAC's for the grey matter regions.

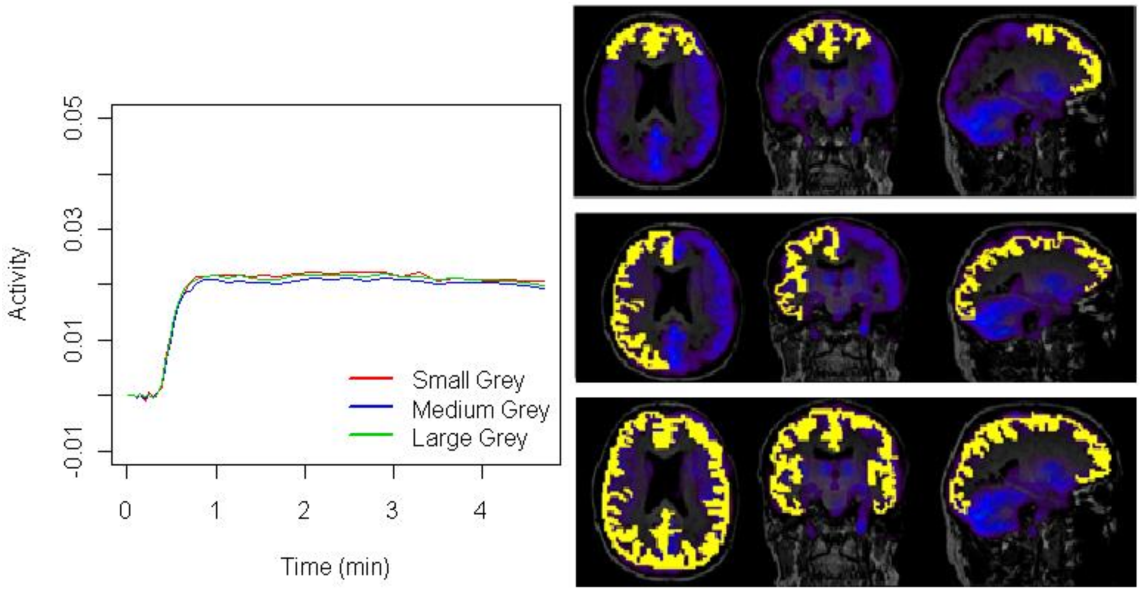

Figure 2. 12: Time Activity Curves of three grey matter regions, small, medium and large. PET water dynamic image with MRI scan superimposed showing three grey matter regions, small, medium and large.

The activity levels of all three grey matter regions are similar and smooth. The activity in the small grey region is the highest and lowest in the medium region. The curves peak and level off just about 0.2 after 30 seconds. Compare this to the hippocampus regions or grey matter from the cerebellum (**2.3.2**) which are so much more volatile.

Figure 2.13 shows the white regions, where the white matter regions are subsets of one another. For this reason, there is only one set of PET images provided.

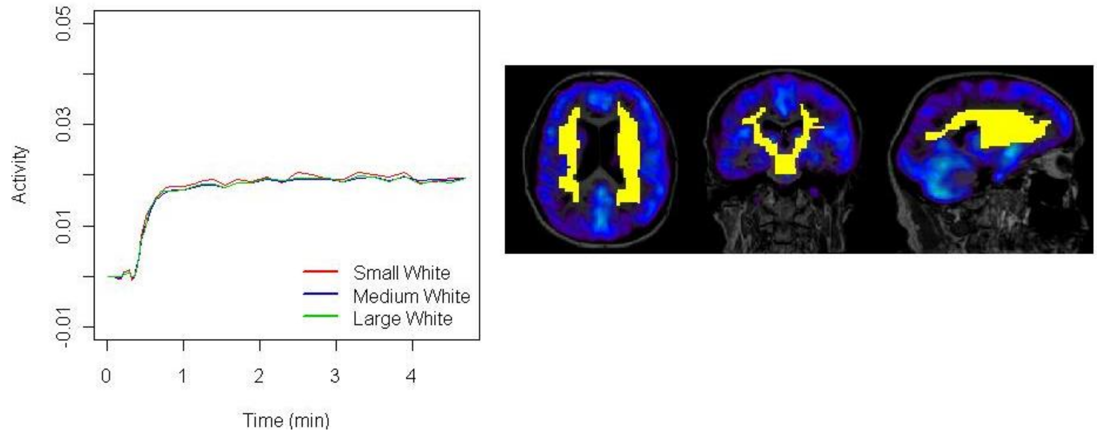

Figure 2.13: Time Activity Curves of three white matter regions, small, medium and large. PET water dynamic image with MRI scan superimposed showing the three white matter regions, small, medium and large.

The activities of the three regions peak and level off just under 0.02 after 30 seconds. The curves are very similar, with the small region showing the highest activity. The other two regions are practically the same.

The time activity curves show a higher level of activity in the grey matter regions, compared to the white matter regions.

The model fits for the grey matter regions are shown in **Figure 2.14**.

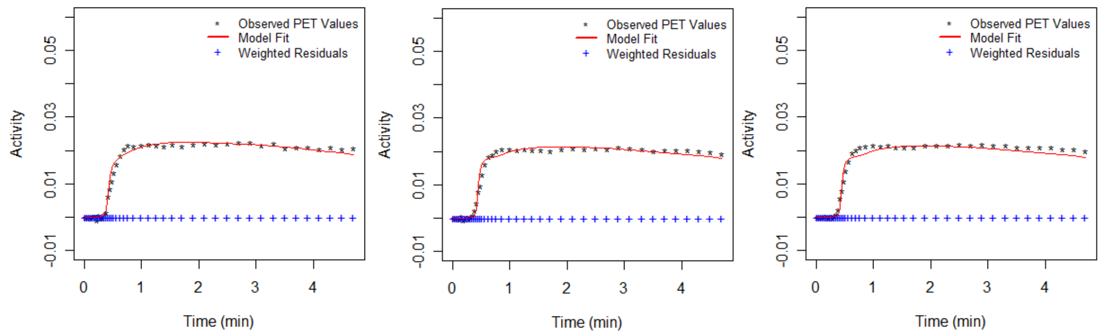

Figure 2.14: Model fits of the observed PET values for the three grey matter regions. The residuals for each region are illustrated as well.

The model fits the data well and again the weighted residuals are what we would expect. The box plots in **Appendix 2C** show a symmetric nature, though the large grey matter region does have some outliers. Further checks of the residuals and normality could be conducted to ensure there is no sign of a pattern. If the variance does not appear to be constant, then it could affect the results.

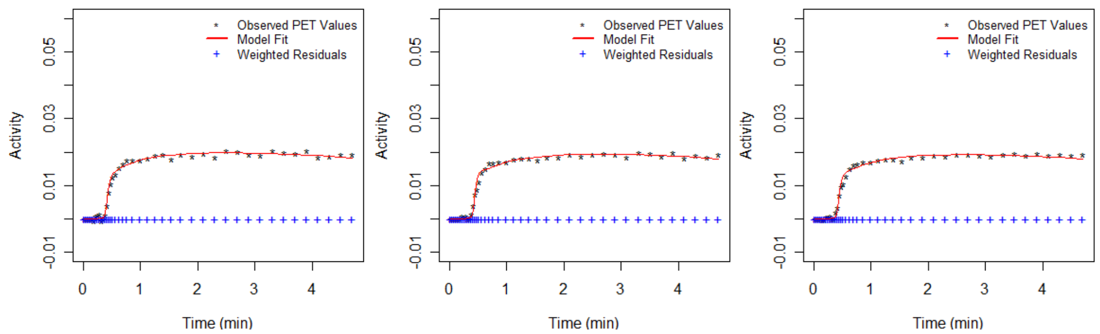

Figure 2.15: Model fits of the observed PET values for the three white matter regions. The residuals for each region are illustrated as well.

The box plots in **Appendix 2C** show a symmetric nature, where the three regions have some outliers, though still have a symmetric shape. Further checks of the residuals and normality could be conducted to ensure there is no sign of a pattern. If the variance does not appear to be constant, then it could affect the results.

The estimates for the model parameters, δ, VB, K1and k2along with the number of voxels in each region are provided in **Table 2.3**.

Table 2.3: Parametric analysis of the three grey matter regions.

Region	δ ± s.e	V_B ± s.e.	K_1 ± s.e.	k_2± s.e.
Small Grey Matter (n = 1,294*)	0.0186±0.0247	0.0896±0.0322	1.2500±0.0246	0.2832±0.0125
Small White Matter (n = 1,362*)	0.0127±0.0252	0.0822±0.0242	1.0208±0.0192	0.2232±0.0118
Medium Grey Matter (n = 2,326*)	0.0105±0.0171	0.1209±0.0237	1.1724±0.0247	0.2794±0.0143
Medium White Matter (n = 2,291*)	0.0126±0.0131	0.1000±0.0147	0.9761±0.0140	0.2125±0.0093
Large Grey Matter (n= 3,736*)	0.0100±0.0165	0.1265±0.0239	1.2200±0.0249	0.2843±0.0140
Large White Matter (n= 3,486*)	0.0063±0.0144	0.0877±0.0161	0.9729±0.0139	0.2121±0.0090

N is the number of voxels.

The parametric results for each grey region are higher than its white counterpart. These are in line with the time activity curves. Parametric results for flow, fractional blood volume, and distribution volume are calculated.

Table 2.4: Parametric and non-parametric analysis of the three grey and white matter regions.

Region	Flow	VB	VD
Small Grey Matter	1.2500±0.0246[1]	0.0896±0.0322	4.4138±2.1002
	2.0026±0.0159[2]	0.1262±0.0007	1.2469±0.0137
Small White Matter	1.0208±0.0192[1]	0.0822±0.0242	4.5735±1.7364
	1.9152±0.0148[2]	0.1392±0.0015	1.3220±0.0148
Medium Grey Matter	1.1724±0.0247[1]	0.1209±0.0237	4.1961±1.8452
	2.0306±0.0128[2]	0.1239±0.0005	1.2205±0.0089
Medium White Matter	0.9761±0.0140[1]	0.1000±0.0147	4.5934±1.9634
	1.9730±0.0159[2]	0.1191±0.0001	1.1741±0.0104
Large Grey Matter	1.2200±0.0249[1]	0.1265±0.0239	4.2912±0.8967
	2.0876±0.0100[2]	0.1197±0.0004	1.2845±0.0076
Large White Matter	0.9729±0.0139[1]	0.0877±0.0161	4.5870±1.4686
	2.0087±0.0136[2]	0.1213±0.0007	1.2417±0.0091

[1] Parametric Analysis

[2] Non-parametric Analysis

The results displayed in **Table 2.4** are coherent with the activity curves in **Figure 2.13**. The parametric values for flow in each of the grey regions are higher than the white regions. These results show blood flow to be higher in areas of grey matter. The non-parametric values are approximately equal to the parametric. This would suggest blood flow to be equal in both white and grey matter regions.

The parametric estimates for fractional blood volume are all approximately equal to the non-parametric values except for the large white and grey matter parametric result. Here, there is a difference in VB between the grey matter region (0.127) and the white region of (0.088).

The parametric estimates for the distribution volume show the tracer concentration in the tissue to be higher than the plasma. All the values are similar and range from 4.196 to 4.59. The non-parametric estimates are much lower but show a higher concentration of the tracer in tissue than plasma.

Figure 2.16 illustrates the time activity curve for the cerebellum white matter region, the model fit on the observed PET values, and the region of interest.

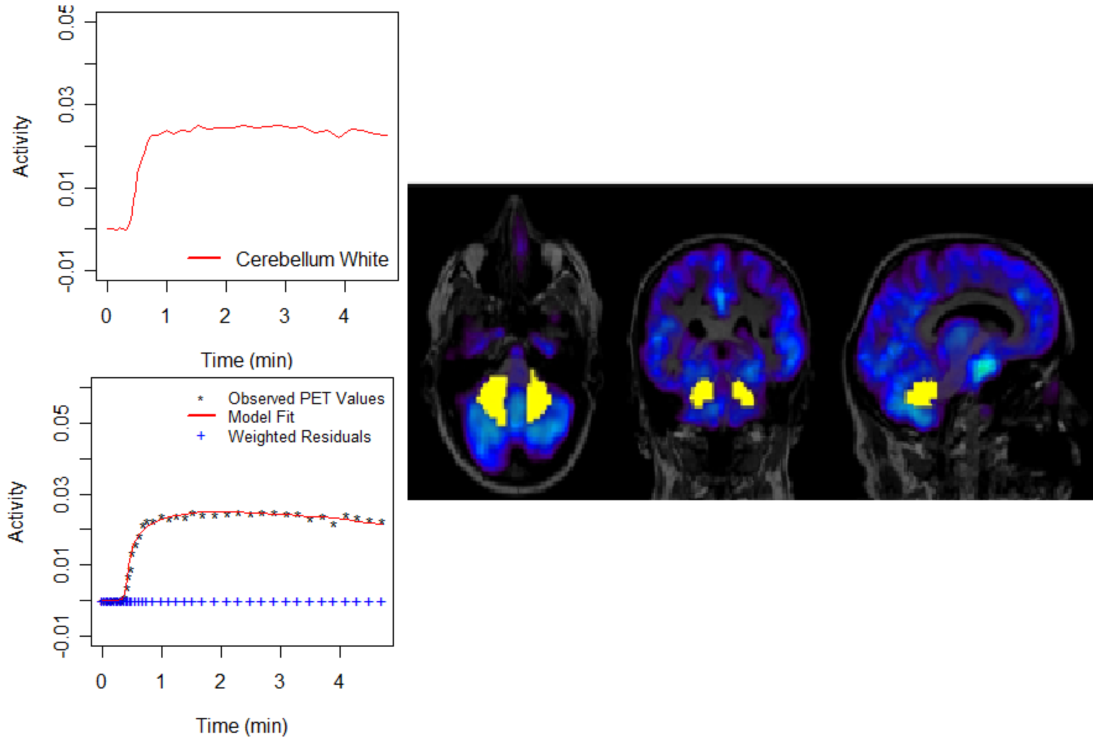

Figure 2.16: Time Activity Curve for the cerebellum white matter region and the model fit of observed PET values. PET water dynamic image with MRI scan superimposed showing the cerebellum white matter region.

The time activity curve peaks and levels off above 0.02 after 30 seconds. The activity in this white matter region is higher than the white matter regions discussed above. Further inspection of the parametric and non-standard parametric values will indicate how it compares to the grey matter regions. The parametric and non-parametric results are given in **Table 2.5**.

Table 2.5: Parametric and non-parametric estimates for the cerebellum white matter region.

Cerebellum White	Parametric	Non-parametric
Flow	1.3829±0.0246	2.8469±0.5677
VB	0.0427±0.0322	0.7121±0.0171
VD	5.1697±2.346	2.0030±0.2892

The parametric flow value for the cerebellum white region is higher than the 3 grey regions discussed above, but it is lower than the cerebellum grey region we discussed in **2.3.2**. This area of the brain appears to have more blood flow than regions discussed in this section, however the grey matter region in the cerebellum was still higher. The non-parametric values were higher than the parametric values and the values adhered to the findings of the parametric values.

The parametric estimates for fractional blood volume in the cerebellum white region were lower than all the white and grey regions above and the cerebellum grey region. Here, there is a difference in VB between them both; the grey matter region (0.1265) compared to the white region (0.1265).

The parametric estimates for the distribution volume show the tracer concentration in the tissue to be higher than the plasma. All the values are similar and range from 4.196 to 4.59. The non-parametric estimates are much lower, but show a higher concentration of the tracer in tissue than plasma.

2.3.4 Grey / White Matter Ratio

The grey / white matter ratio is approximately 2 using a water tracer. That result is based on a normal brain. The grey / white matter ratios for particular regions from our analysis are listed in **Table 2.6**.

Table 2.6: Grey / white matter ratio for a select number of regions.

Grey / White Matter Ratio	Ratio
Cerebellum Grey / Small White Region	1.58
Hippocampus Left / Small White Region	1.04
Grey / White Small Region	1.22
Grey / White Medium Region	1.21
Grey / White Large Region	1.26

We would expect the values listed in **Table 2.6** to be lower than the recorded 2.5 if Alzheimer's disease was affecting most of the brain. As there is a decrease of activity in the grey matter regions, the estimates of the overall faction would also decrease. The ratio for the hippocampus region shows the lowest value, which we would expect, as this area is most susceptible to the disease

2.3.5 Blood Regions

Using the results of our PCA, we were able to identify three regions of interest. The first two regions, blood region 1 and 2, were chosen from the 1st principal component. The third, blood region 3, was chosen from the 2nd principal component. These regions of interest are illustrated in **Figure 2.17**. It is worth noting that these regions are very small, with the voxel size for region 1, 2, & 3 being 36, 40, & 33, respectively.

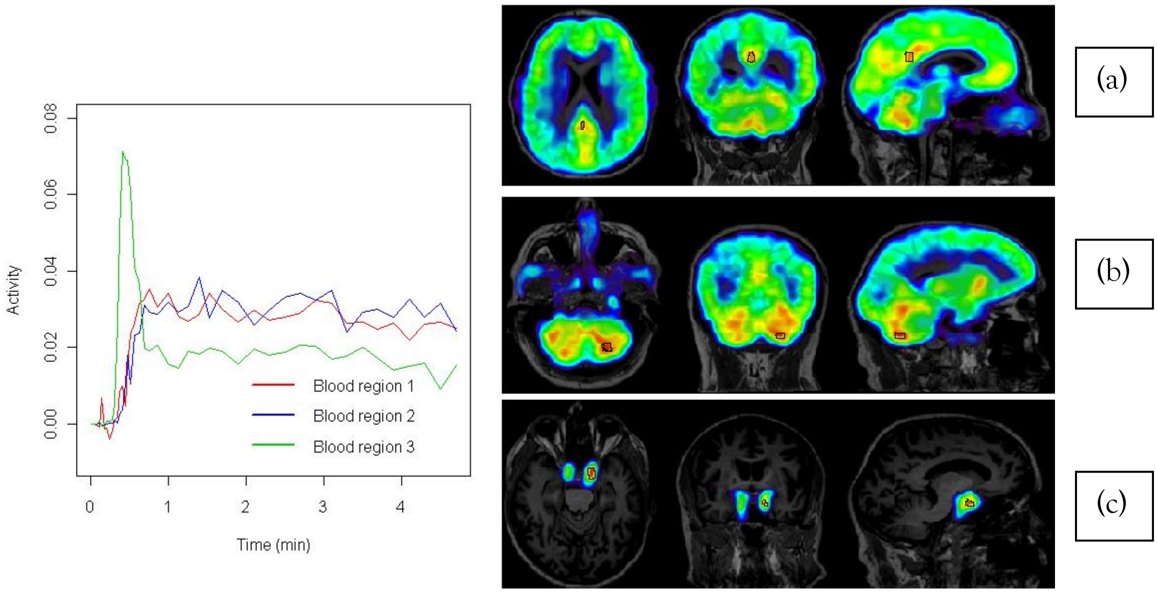

Figure 2.17: Time Activity Curves of the three blood region regions. (a) and (b) are the 1st principal component image with MRI scan superimposed showing blood regions 1 and 2. (c) is the 2nd principal component image with the MRI scan superimposed showing blood region.

The time activity curves show levels of activity higher than the grey and white matter regions explored in **2.3.3**. Blood region 3 peaks around 0.07 in the first minute and tails off under 0.02. This shape is similar to an arterial input function (AIF). A plot of the AIF can be found in **Appendix D**. Blood regions 1 and 2 show higher activity than both the grey and white matter tissue; both peak around 0.35 and tail off. Because these blood regions do not inherit the same shape as the AIF, these regions could be regions of tissue with high blood flow. After the initial spike in blood region 3, the activity levels in the blood regions 1 and 2 are higher.

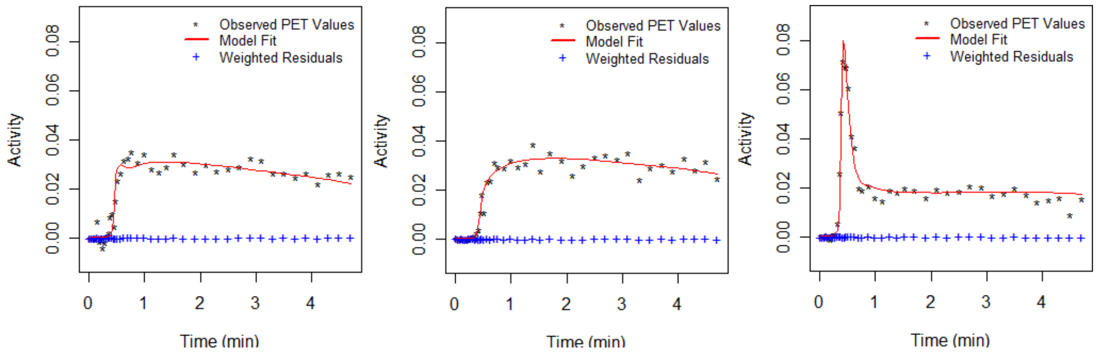

Figure 2.18: Model fits of the observed PET values for the three white matter regions. The residuals for each region are illustrated as well.

The models in **Figure 2.18** fit the data well and the weighted residuals are centred around zero, which we would expect. The box plots in **Appendix 2C** show a symmetric nature; blood region 3 region does have some outliers though. Further checks of the residuals and normality could be conducted to ensure there is no sign of a pattern. If the variance does not appear to be constant, then it could affect the results.

The estimates for the model parameters, δ, VB, K1and k2, along with the number of voxels in each region, are provided in **Table 2.7**.

Table 2.7: Parametric analysis of the three blood regions.

Region	δ ± s.e	V_B ± s.e.	K_1 ± s.e.	k_2± s.e.
Blood Region 1 (n=36)	0.0254±0.0358	0.2438±0.0612	1.8068±0.0842	0.3705±0.0359
Blood Region 2 (n=40)	0.0183±0.1646	0.0633±0.2809	1.8844±0.1284	0.3074±0.0349
Blood Region 3 (n=33)	0.0446±0.0034	0.9766±0.0372	0.6117±0.0562	0.1111±0.0600

The results of the parametric analysis for the blood regions are quite surprising, K1 estimate for the blood region 3 is much lower than expected. Compared to the TAC's, one would expect the value to be much higher.

The parametric and non-parametric results are provided below in **Table 2.8**.

Table 2.8 : Parametric analysis of the three blood regions.

Region	Flow	VB	VD
Blood Region 1	1.8068±0.0842[1] 3.3930±0.0919[2]	0.2438±0.0612 0.1333±0.0029	4.8767±1.365 1.7226±0.0125
Blood Region 2	1.8844±0.1284[1] 4.0212±0.0951[2]	0.0633±0.2809 0.0734±0.0113	6.1301±2.458 2.0022±0.0261
Blood Region 3	0.6117±0.0562[1] 4.3889±0.0160[2]	0.9766±0.0372 0.5515±0.0101	5.5059±3.864 0.6846±0.0158

[1]Parametric Analysis.

[2]Non-paramertirc Analysis.

The parametric values for flow, show region 3 to be lower than the other regions. The non-parametric values show region 3 to be higher than the parametric value. Furthermore, the non-parametric value for region 3 is higher than region 1 and 2.
A higher value in blood region 3 would be expected given the TAC. The shape is similar to that of an artery, so flow should be high.

The parametric estimate for fractional blood volume is highest in region 3 (0.9766), compared with (0.243) and (0.0633) in region 1 and 2, respectively. The non-parametric analysis yields the same results. One note is that the non-parametric result for region 2 is lower than the parametric but not for the other regions.

The parametric estimates for the distribution volume are very high in each blood region. These regions were identified using principal component analysis. The estimates show the tracer concentration in the tissue to be higher than the plasma. The non-parametric estimate for region 2 (06846) would suggest there is more tracer in the plasma than tissue, which is what one would expect for a region of this nature.

2.3.6 Grey and White Matter Regions near the Blood Regions.

A region of grey and white matter was chosen in the proximity of the blood regions to see if there was a difference in the activity levels.

Six regions in total were used in the analysis, three white and three grey regions close to the 3 blood regions investigated in the section above.

The naming convention used was as follows: grey tissue near blood region 1 is known as GT BR1; grey tissue near blood region 2 is known as GT BR2; and finally grey tissue near blood region 3 is known as GT BR3, with the white regions are known as WT BR1, WT BR2, and WT BR3.

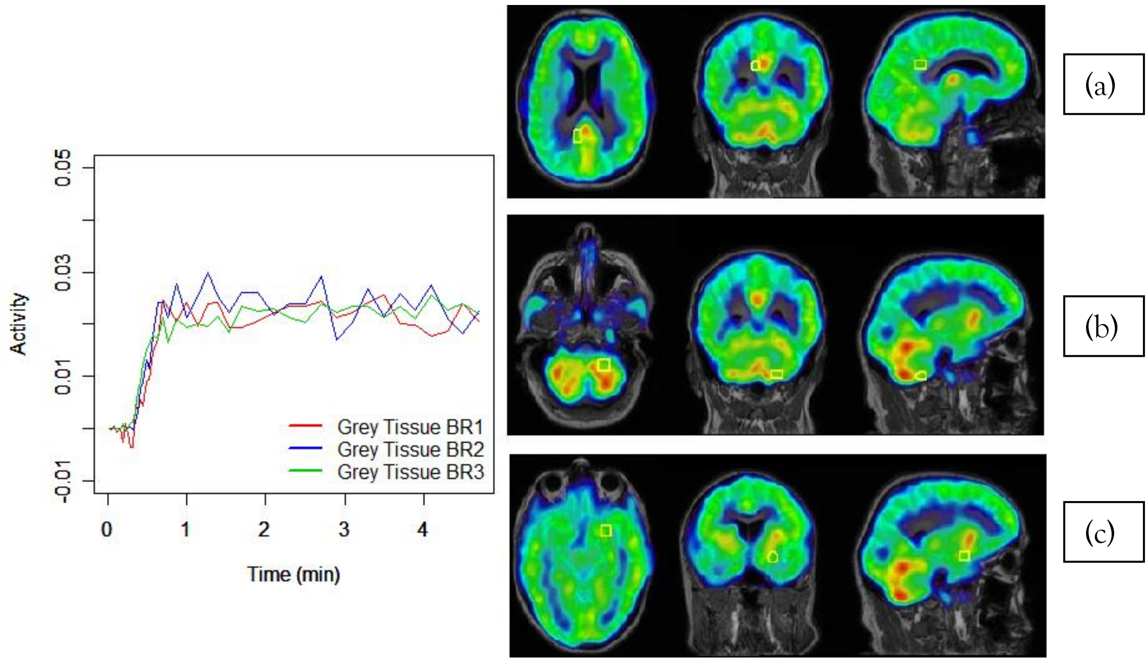

Figure 2.19: Time Activity Curves of the three grey regions near blood regions. a represents grey region near blood region1, b represents grey region near blood region 2 and c represents grey region near blood region2. All regions are on the 1st principal component.

The activity curves are very noisy, which could be due to the small number of voxels used in the analysis. Unlike some of the TAC's in earlier sections, the peak is not in the first minute. GT BR2 shows the highest level of activity, with GT BR3 showing the lowest initial activity.

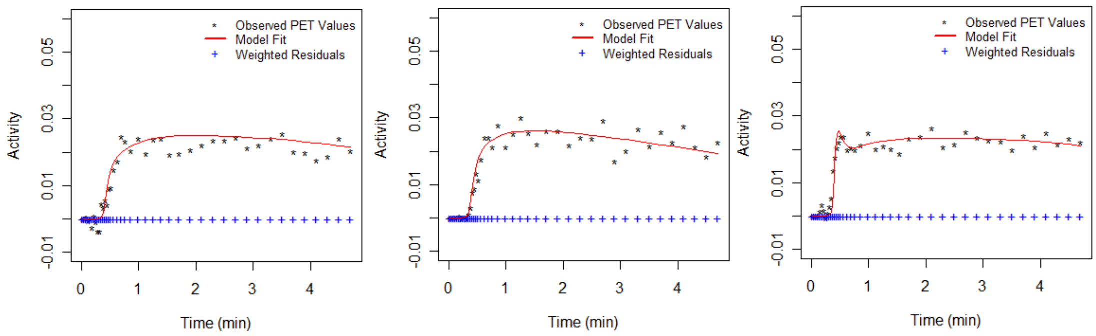

Figure 2. 20: Model fits of the observed PET values for the three grey regions near blood. The residuals for each region are illustrated as well.

The models fit the data well and the weighted residuals are close to 0 which is good.

The curves for the white regions are less noisy than the grey but noisier than previous curves. Again, this could be due to the small number of voxels.

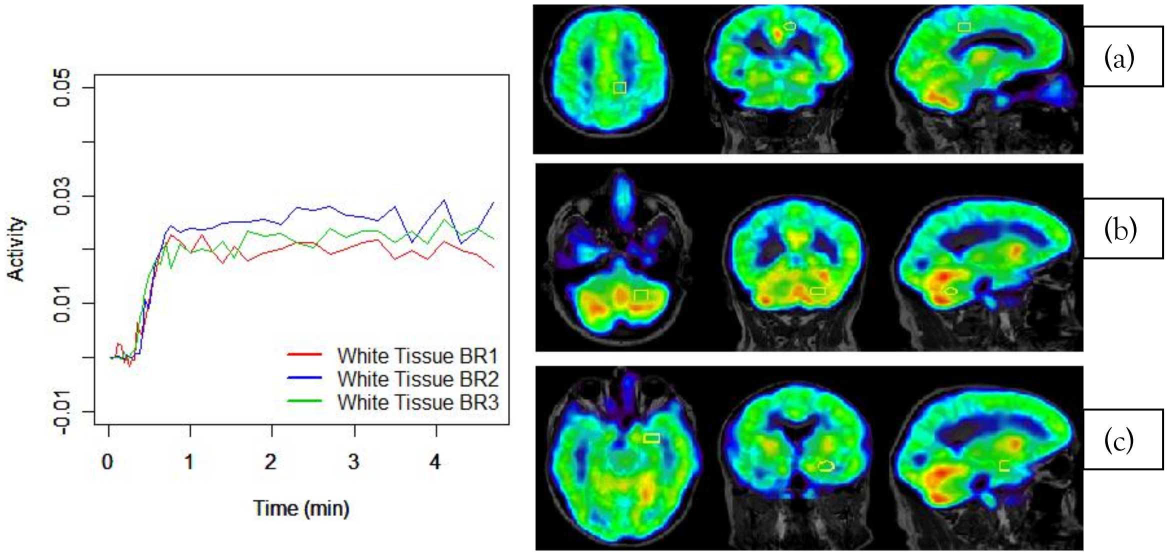

Figure 2.21: Time Activity Curves of the three white regions near blood regions. a represents white region near blood region 1, b represents white region near blood region 2 and c represents white region near blood region 2. All regions are on the 1st principal component image with MR scan superimposed.

WT BR2 experiences the highest activity, while WT BR1 experiences the lowest. The initial peak is above 0.02 activity inside the first minute but drops and rises for all the curves.

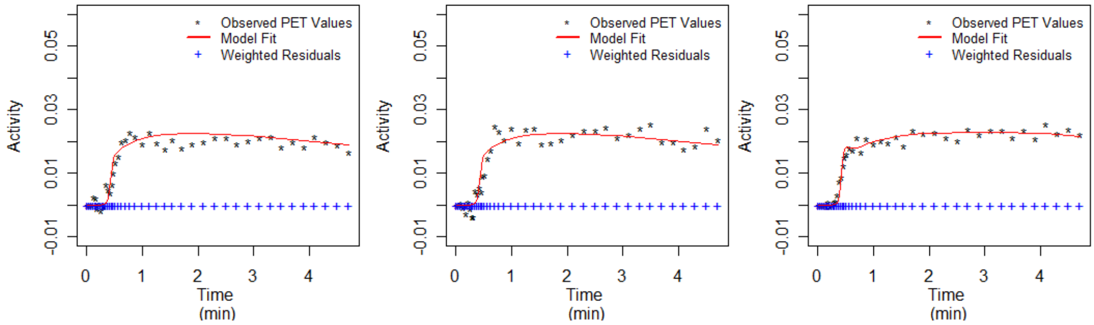

Figure 2.22: Model fits of the observed PET values for the three grey regions near blood. The residuals for each region are illustrated as well.

The model fits for WT BR1, BR2, and BR3 are good. The weighted residuals are centred around zero, which is what one would expect.

The box plots in **Appendix 2C**, exhibit a symmetric nature for the three grey and white regions close to the different blood region. Further checks of the residuals and normality could be conducted to ensure there is no sign of a pattern. If the variance does not appear to be constant, then it could affect the results.

Table 2.9: Parametric analysis of the three blood regions.

Region	δ ± s.e	V_B ± s.e.	K_1 ± s.e.	k_2± s.e.
Grey Tissue BR1 (n=98)	0.0074±0.0587	0.1057±0.0746	1.2475±0.0663	0.2706±0.0349
White Tissue BR1 (n=112)	0.0163±0.0330	0.1523±0.0468	1.1452±0.0547	0.2927±0.0333
Grey Tissue BR2 (n=105)	0.0042±0.0521	0.1377±0.0979	1.5267±0.1018	0.3446±0.0469
White Tissue BR2 (n=119)	0.0117±0.0196	0.0922±0.0536	1.3599±0.0373	0.2177±0.0171
Grey Tissue BR3 (n=93)	0.0257±0.0159	0.2476±0.0487	1.1718±0.0530	0.2476±0.0487
White Tissue BR3 (n=113)	0.0163±0.0241	0.1464±0.0388	1.1124±0.0365	0.1922±0.0204

The parametric estimates for tissue near blood region are quite similar. GT near BR2 showed the highest estimate for K1 while WT near BR3 showed the lowest. The parametric and non-parametric results are displayed below.

Table 2. 10: Non-parametric analysis of the three grey regions near blood regions.

Region	Flow	VB	VD
Grey Tissue BR1	1.2475±0.0663[1] 2.7668±0.0544[2]	0.1057±0.0746 0.0971±0.0017	4.6101±1.349 1.7501±0.0272
White Tissue BR1	1.1452±0.0547[1] 2.3887±0.0414[2]	0.1523±0.0468 01474±0.0025	3.9125±0.876 1.5603±0.0207
Grey Tissue BR2	1.5267±0.1018[1] 3.2407±0.0314[2]	0.1377±0.0979 0.0829±0.0018	4.4304±3.256 2.3752±0.0307
White Tissue BR2	1.3599±0.0373[1] 3.2684±0.0556[2]	0.0922±0.0536 0.0734±0.0014	6.2466±2.563 1.9906±0.0159
Grey Tissue BR3	1.1718±0.0530[1] 2.1425±0.0374[2]	0.2476±0.0487 0.1690±0.0045	4.7326±1.891 1.5429±0.0305
White Tissue BR3	1.1124±0.0365[1] 2.3949±0.0289[2]	0.1461±0.0388 0.1105±0.0021	5.7877±1.865 1.3999±0.0220

[1]Parametric Analysis.

[2]Non-parametric Analysis.

The parametric estimates for blood flow are highest in grey tissue near blood region 2. The lowest value was discovered in white tissue near blood region 3. The non-parametric values are higher than the parametric values. The white tissue near blood region 2 had the highest flow value, with the grey region having a very similar result. The grey tissue near blood region 3 had the lowest.

The parametric estimates for fractional blood volume are highest in grey tissue close to blood region 3 (0.247), and the lowest is in the white tissue near blood region 2 (0.0922). The non-parametric results were lower but yielded the same outcome.

The parametric estimates for distribution volume showed a higher level of tracer in the tissue compared to the plasma. This was the same for the non-parametric estimates.

2.3.7 Validation of Kinetic Analysis Technique

The study has presented estimates for blood flow, fractional blood volume VB and distribution volume VD using a parametric and non-parametric analysis. The validation of the Kinetic Analysis Technique was conducted with and without blood region 3.

Using Spearman's correlation test, the results for each estimate under the two techniques are examined to see if there is a correlation between one another.

The null hypothesis is that the two estimates are not correlated, and the alternative is that they are. A correlation test was employed to the parametric and non-parametric functional estimates. A significantly positive correlation would imply that the kinetic analysis employed is a valid method of estimating functional metabolic parameters.

An analysis of the whole brain, left and right hemisphere was conducted at the beginning of the analysis but not included in the results. The results can be found in **Appendix 2.A**.

2.3.7.1 Validation of Kinetic Analysis Technique Including BR3

The results of the correlations of the estimates of the parametric and non-parametric results are provided in **Figure 2.23** and illustrate scatter plots for flow, VB and VD estimates.

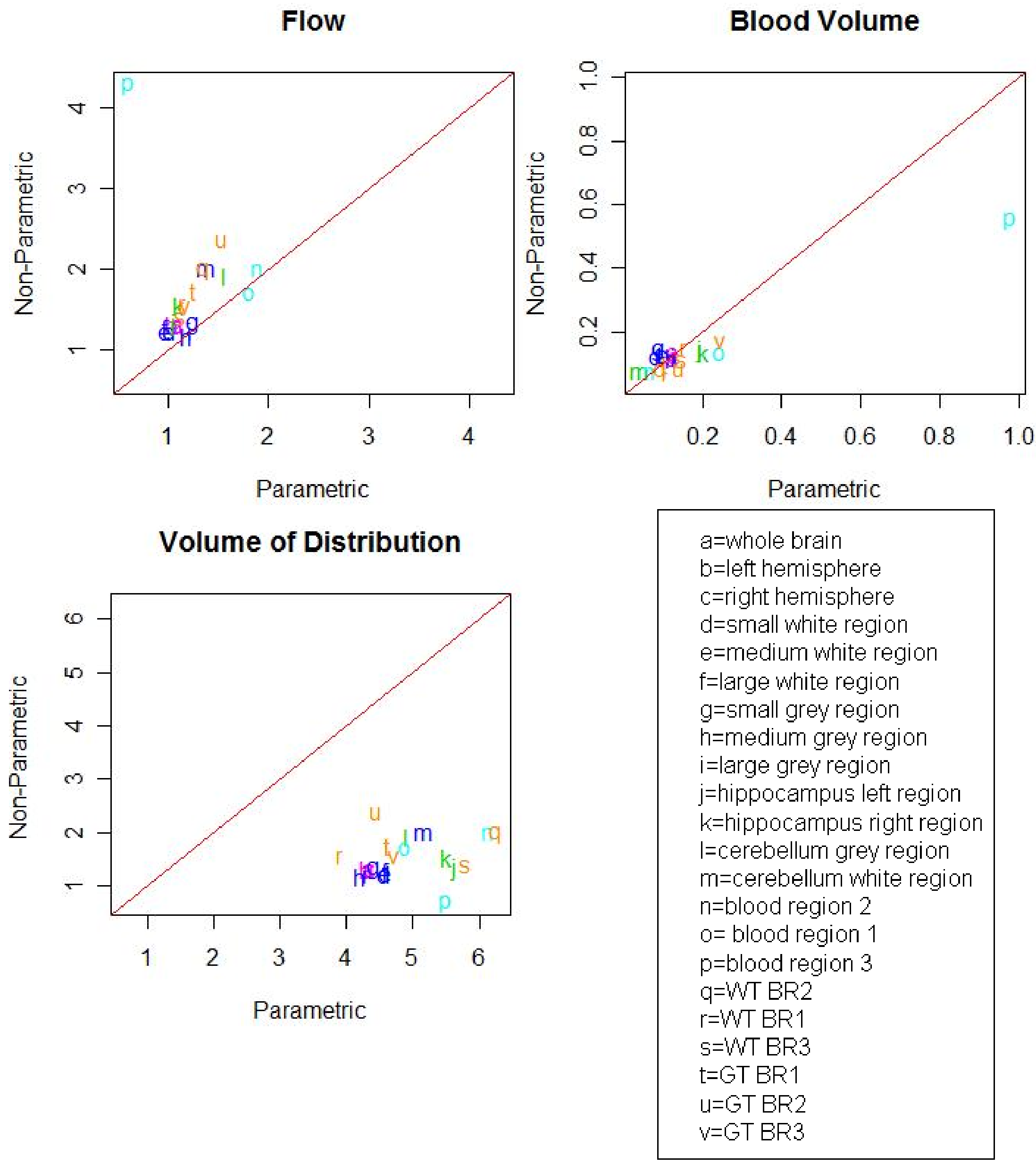

Figure 2.23: Correlation plots of all regions for Flow, VB and VD including blood region3.

The estimates for the parametric and non-parametric analysis were included in the Spearman test.

Blood region 3 (or P in plot) appears to be an outlier for flow and VB. The remaining points are clustered but appear to be a positively upwards relationship. For VD, P is clustered with the other points. Overall, as the relationship is not strictly linear, Spearman's correlation method was used to quantify the magnitude and direction of the relationship as summarised in **Table 2.11**.

Table 2.11: Correlation coefficients for the different parameters with blood region 3 included.

Region	Spearman
Flow	0.485*
VB	0.648*
VD	0.435*

*Represents statistically significant at the 5% level.

The correlations between the functional parameters obtained using the parametric and non-parametric approach are provided in **Table 2.11**. With a 5% level of significance level, the correlation coefficient for all estimates is significant. This implies that the estimates are positively correlated and that the parametric approach is a valid method for interpreting PET data.

2.3.7.2 Validation of Kinetic Analysis Technique Excluding BR3

The results of the correlations of the estimates of the parametric and non-parametric results are provided in **Figure 2.24 and** illustrate scatter plots for flow, VB and VD estimates.

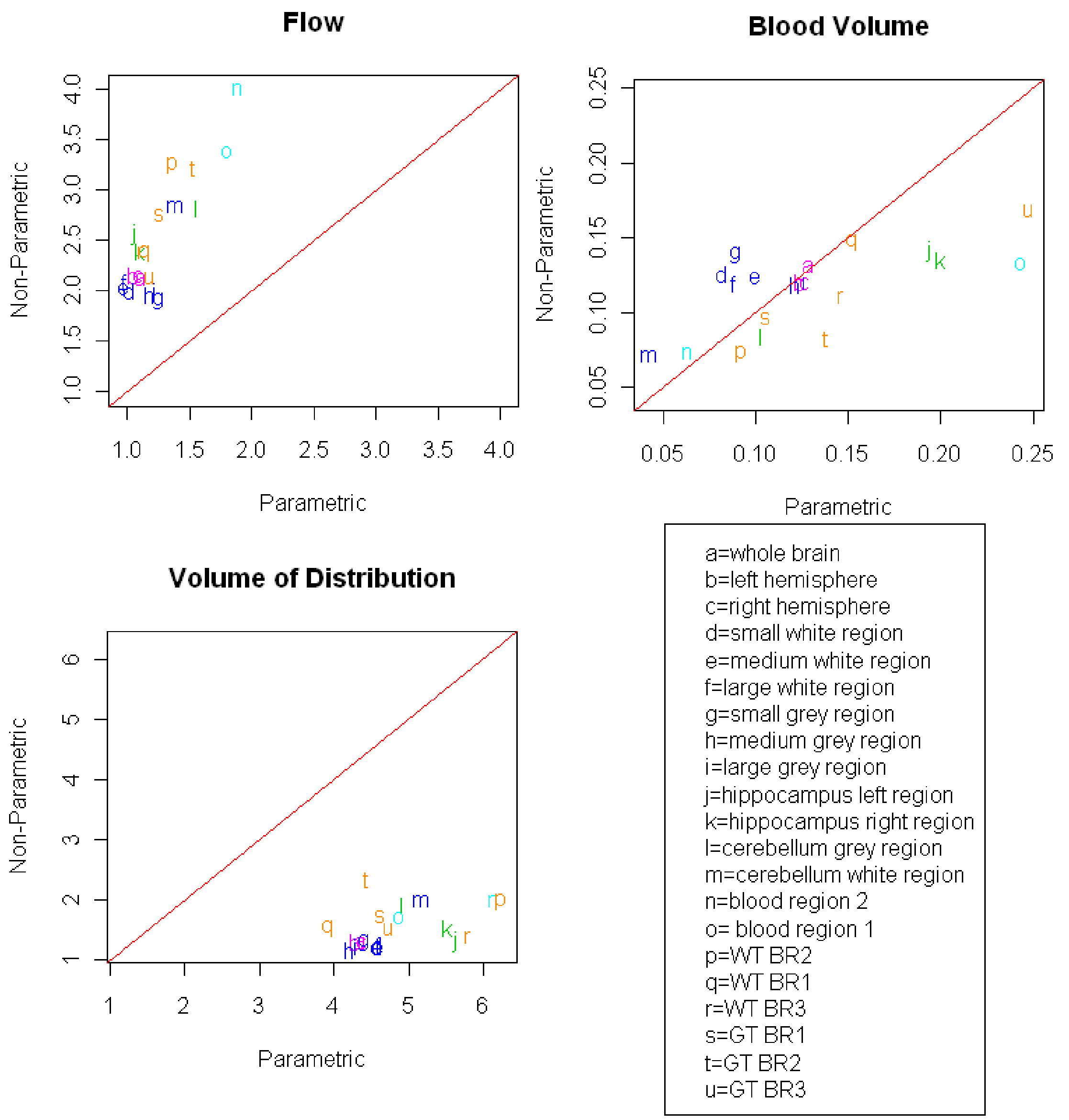

Figure 2.24: Correlation plots of all regions for Flow, VB and VD except blood region 3.

With omission of blood region 3, the outlier is removed from the plots. The graphs for flow and VB have changed compared to **Figure2.23**. The plots have a positive increasing relationship. The plot for VD is rather unchanged after the removal of BR3.

The results of the Spearman test in the table below now show the estimates of flow to be correlated with the removal of blood region 3.

Table 2.12: Correlation coefficients for the different parameters without blood region 3.

Region	Spearman Correlations
Flow	0.6025*
VB	0.5950*
VD	0.5571*

*Represents statistically significant at the 5% level.

The correlations between the functional parameters obtained using the parametric and non-parametric approach are all significantly correlated (**Table 2.12**). With a 5% level of significance level, the correlation coefficient for flow, VB and VD are significant. This implies that the estimates are positively correlated and that the parametric approach is a valid method for interpreting PET data.

2.4 Discussion

The following section will discuss the important features of the results.

2.4.1 Methodological Approach

Exploratory data analysis was used to determine the structure and classification of PET data prior to performing kinetic analysis. Principal component analysis and cluster analysis were used.

PCA analysis determined the behaviour of the tracer in both the tissue and blood regions. PCA explained the variance in as little components as possible and identified two principal components. The first component represents the tracer activity in the tissue. This was interpreted as the uptake of the tracer by the tissue. The second component was the tracer activity in the blood regions, it was interpreted as the tracer activity into the blood stream.

Cluster analysis establishes the difference in the tracer activity in a region, it would group homogeneous tissue together based on tracer levels. Two forms of cluster analysis were used, hierarchical and k-mean clustering. Two regions were examined, a small grey and a white matter region were examined. Both clustering techniques were used. The grey matter region had 2 clusters, while the white region had 3 clusters. Both clustering techniques yield the same number of clusters overall, however the activity curves were different due to the techniques used. The clusters had different levels of activity.

The activity of each cluster in the grey region are higher than the white region under the hierarchical techniques. However, under the k-means techniques one of the three clusters activities is higher in the white region than the grey. Grey matter is supposed to have higher levels of activity than white. There could be some levels of white and grey matter in the region just marked white or grey, which could have affected the results.

Identifying the areas of the brain is very difficult, as it is a very complex organ and the images are quite small so may not ensure the exact area is getting analysed (Gao, Tay & Nowinski, 2006). Cluster analysis can show different tissue structures based on the tracer activity. Studies conducted after this study in 2012 by Huisman et al used the same approach and AMIDE software for manually drawing and identifying ROIs. Their study did not conduct PCA or cluster analysis, however these methods are still used. A study by Deo et al used cluster analysis utilising a programme randomise and another report (Segmentation of Alzheimer's Disease in Pet Scan Database using Mathlab, Meena & Raja, 2013) used two different forms of cluster analysis k-means and Fuzzy C-Means on the data.

This study wanted to perform analysis of PET data using free software. Exploratory techniques such as cluster analysis and principal component analysis can be performed using available software.

2.4.2 The Hippocampus Regions and Cerebellum Grey Region

This section of the analysis was conducted on parts of the brain most susceptible to Alzheimer's disease and another region which is not at risk of early Alzheimer's. The estimates for each region were presented using parametric and non-parametric analysis.

The estimates for blood flow were highest in the area of the brain less susceptible to the disease compared to both hippocampus regions. The estimates under the non-parametric analysis were higher than the parametric estimates. However, in conjunction with the parametric results, the area less susceptible to the disease is the highest again. Reduced blood flow in susceptible areas is consistent with past studies (Nishimura et al., 2006). While the results in the parametric and non-parametric analysis differed significantly, other studies (O'Sullivan et al., 2009) have suggested that when there is uncertainty with heterogeneity or other factors in the model, the non-parametric approach is more reliable. Heterogeneity can occur when choosing areas of the brain. The brain is a very complex and complicated organ. Heterogeneity arises as identifying the exact tissue can be difficult.

The estimates for the fractional blood volume in the more susceptible regions are almost double the less susceptible area. The non-parametric values were lower and gave the same magnitude of results.

The estimates of distribution of volume for both the parametric and non-parametric analysis show more of the tracer in the tissue than plasma. To note, there are large differences in the values of the two techniques, this could be due to the uncertainty of heterogeneity in the regions selected.

The findings of O'Sullivan et al. (2009) would contribute to the results seen for fractional blood volume and distribution volume. However, although fractional blood volume is expected to be lower in Alzheimer's patients (Harris et al., 1996), the findings in this study would contradict these findings. Given these areas are tissue regions, it is not a surprise then that there is more tracer in the tissue than plasma.

The study conducted by Deo et al in 2014, used different ROI in the brain for a patient with Alzheimer's. Their study and results are significant as they analyse and compare patients with and without the disease. Although the average flow figures for this study are higher, it is hard to make an assessment as their ROIs are different. Our study showed flow is lower in areas more susceptible to the disease, which is consistent with details in their study.

2.4.3 Grey V's White Matter

Grey and white matter regions were identified and analysed using the parametric and non-parametric methods. Estimates for flow for the grey matter regions were higher than white for each region, although the non-parametric estimates were very similar.

Parametric estimates for fractional blood volume are higher for grey matter regions compared to the white matter regions. The parametric and non-parametric estimates for distribution volume differ significantly but both show more tracer in the tissue, when compared to plasma.

Results provided by (Tofts, 2003) suggested blood flow in grey matter regions is higher than white matter regions, which is consistent with the findings of the parametric results. Similar to the reasons described in the previous section, the non-parametric results would be preferred. Isolating exact grey and white matter regions is difficult, and this could lead to heterogeneity issues in the data, meaning the estimates for the parametric analysis may not be reliable.

A similar study conducted by Huisman et al, (2012) this study used a water tracer and kinetic analysis. The results presented suggested that blood flow in grey matter regions is higher than white matter regions, which is consistent with the parametric findings in this study.

2.4.4 Blood Regions

There were two blood regions extracted using the 1st principal component and one blood region extracted using the 2nd principal component. Three regions in total were examined. The TAC produced and the plot AIF are very similar to results from a study for similar regions by O'Sullivan et al (2010). The study by O'Sullivan et a (201)) used a different tracer, FDG, and produced similar graphs. In addition, three white and grey matter regions close to the blood regions were analysed also. The parametric and non-parametric estimates for flow are much higher than other flow estimates calculated for other regions, except for blood region 3, which is much lower. The TAC of blood region 3 is very similar to the AIF. Given this, we would expect a high level of flow. This result would be further evidence then that the parametric results are not reliable.

The results for all regions other than region 3, show volume distribution of the tracer to be higher in the tissue when compared to the plasma. The non-parametric estimate is what we would expect, with more tracer in the plasma when compared to the tissue.

The non-parametric results for flow are much closer to what we would expect and much higher than the parametric. This is particularly true for the volume distribution estimates. As discussed earlier, heterogeneity could lead to the parametric estimate not being reliable.

The parametric and non-parametric estimates for flow show grey matter near blood regions to have higher activity to white matter regions. The non-parametric values are higher than the parametric values. All volume distributions have higher level of tracer in the tissue than plasma. The areas are close to blood regions so one would expect more tracer in the tissue but not significantly, the non-parametric estimates are closer to what we would expect. The parametric estimates appear very high, given the fact that the blood region is so close. The parametric issues discussed could have contributed to the differences, the non-parametric approach and results would be preferred.

2.4.5 Validation of Kinetic Analysis

A validation of the kinetic analysis techniques was conducted with and without blood region 3 given it is an outlier.

Including BR3, the comparison of the parametric and non-parametric results showed positive correlation, a 5% level of significance. This would imply that the kinetic analysis is an appropriate method to measure functional parameters.

Excluding BR3, the comparison is still correlated but the magnitude the of coefficient correlation increases upwards towards 1. This would imply that the kinetic analysis is an appropriate method to measure functional parameters.

A paper on the Potential Benefit by Application of Kinetic analysis of PET in Clinical Oncology, by Takesh in 2012 explores the application of kinetic analysis in a variety of clinical applications. Kinetic analysis can aid and enhance the PET techniques to enhance results which can be used during the preclinical and clinical period.

2.4.6 Limitations and Recommendations

One of the main limitations of the study was conducting the analysis on one data set. Conducting the analysis on multiple lives would provide more room to analyse results. In addition, a control life (e.g. a brain with no impairment) would have given more perspective to the results of the impaired life. Results from a controlled life would give a much better insight into the level of deviations between estimates for regions suspected to have the disease and not have. This would show the true impact of the disease but possibly also insight into what stage of the disease. The study by Deo et al (2014), uses a control group, in this case patients without Alzheimer's disease along with patients with the disease. Deo et al (2014) also use multiple lives. In their results they can compare the findings to see the implication of the disease on key metrics. Another study used multiple lives instead of one life (Huisman et al, 2012). This study explored flow and other characteristics in healthy volunteers.

The analysis of PET data can be advantageous for preclinical and clinical testing (Takesh, 2012), however there are limitations with using the kinetic analysis technique here. Deviations between the parametric and non-parametric approaches mainly depend on the level of heterogeneity of the tissue region. An assumption (and a limitation) of the 1-compartmental parametric model is that it assumes that the tissue used in the analysis is homogenous. However cluster analysis has shown that tissue regions are not always homogenous. Violation of this model assumption can lead to model misfit and misleading parameter estimates, which can in turn impact the accuracy and reliability of the physiological parameters (such as flux, flow, etc.) derived from these. The care and precision required to determine ROI is very important.

In this study, cluster analysis was performed on tissue samples, however they were not homogenous. Cluster analysis should be performed on a blood region to get a better understanding of the structure. Blood region 3 would be a region of interest to apply cluster analysis as it was so close to the AIF function. In a future study, a different cluster analysis could be used or different software like Mathlab would yield different results (Meena & Raja, 2013). Another software Randommize was used in another study (Deo et al).

This study just used one tracer, a water tracer. Other studies have used the FGD tracer in addition to using a water tracer (Deo et al, 2012) (Peretti et al, 2019). The studies by Deo et al (2012) and Peretti et al (2102) are utilising PET and kinetic analysis to examine and explore other metabolic differences in the brain due to the disease by using a different tracer.

QQ-plots could be generated to assess the level of variance, or histograms to get a visual representation of the data. This can help us to better understand the data and the level of centrality.

A key tool in this process is AMIDE, the tool is an excellent piece of software but exploring another available software could yield different or better results. PMOD imaging software is another tool used for PET imaging processing (About us – PMOD Technologies LLC, 2021). It would be interesting to see if the three considerations below. Unlike AMIDE, PMOD software is not free and requires a license.

1. Selecting ROI's is very important, PMOD, or another software available may be able to isolate ROI with more accuracy
2. The Gaussian filter used, as part of the free software to remove noise can lead to "holes" appearing in the images. Holes can occur due to the filter removing too much detail. PMOD or another software may have better filters.
3. Exploring other options to conduct the assessment could lead to different results. It would be worth exploring other softwares for example, PMOD imaging software to see if there is a difference in performance of the softwares and results.

2.5 Conclusion

This study set out to quantify biological parameters for different regions in a PET data set. Exploratory analysis proved useful in understanding the data set and its make-up. Through the employment of a 1-compartmental model, biological parameters are determined for different regions of the brain.

Alzheimer's appears to affect blood flow in the brain. Blood flow was estimated using two techniques, a parametric and non-parametric analysis. The levels of activity of an area susceptible to the disease yielded lower levels compared to the healthy region. This is consistent with other studies.

Correlations between parametric and non-parametric analysis implied the kinetic analysis is an appropriate method to measure functional parameters when quantifying PET data. The brain is a complex and complicated organ, identifying areas with a very high degree of precession is quite difficult. For this reason, it appears heterogeneity distorted the results of the parametric estimates, with the non-parametric results yielding better results.

Another aim of this study was to calculate kinetic estimates using free software. Using R and AIMDE, estimates for flow, fractional blood volume, and distribution volume were calculated.

2.6 References

Alzheimer's Society Ireland: http://www.alzheimer.ie/eng/Media-Centre/Facts-About-Dementia

Cummings J., (2004). Alzheimer's disease. New England Journal of Medicine, 351, 56-57

Deo, A., Borson, S., Link, J., Domino, K., Eary, J., Ke, B., Richards, T., Mankoff, D., Minoshima, S., O'Sullivan, F., Eyal, S., Hsiao, P., Maravilla, K. and Unadkat, J., 2014. Activity of P-Glycoprotein, a β-Amyloid Transporter at the Blood–Brain Barrier, Is Compromised in Patients with Mild Alzheimer Disease. Journal of Nuclear Medicine, 55(7), pp.1106-1111.

Haense, C.,Herholz,K.,Jagust,W.J.,&Heiss, W.,(2009). Performance of FDG PET for detection of Alzheimer's disease in two independent multicentre samples (NEST-DD and ADNI). Dementia Geriatric Cognitive Disorders, 28, 259-266.

Hair, Black, Babin, Anderson, Tathan. (2006). Multivariate Data Analysis, Sixth Edition.

Huisman, M., van Golen, L., Hoetjes, N., Greuter, H., Schober, P., Ijzerman, R., Diamant, M. and Lammertsma, A., 2012. Cerebral blood flow and glucose metabolism in healthy volunteers measured using a high-resolution PET scanner. EJNMMI Research, 2(1), p.63.

Jobst,K. A, Smith, A.D., Barker, C. S., Wear, A., King, E.M.,Smith,AAnslow, A.P., Molyneux, A.j., Shepstone, B. J., & Soper, N., (1992). Association of atrophy of the medical temporal lobe with reduced blood flow in the posterior parietotemporal cortex in patients with a clinical and pathological disgnosis of Alzheimer's disease. Journal of Neurology Neurosurgery & Psychiartry, 55(3), 190-194.

Kety, S. Schmidt, C., (1948). The Nitrous Oxide Method for the Quantitative Determination of Cerebral Blood Flow in Man: Theory, Procedure and Normal Values ", Journal of Clinical Investigation 27(4), 476-483.

Koopman, T., Yaqub, M., Heijtel, D., Nederveen, A., van Berckel, B., Lammertsma, A. and Boellaard, R., 2017. Semi-quantitative cerebral blood flow parameters derived from non-invasive [15O]H2O PET studies. Journal of Cerebral Blood Flow & Metabolism, 39(1), pp.163-172.

Loening, A.M., Gambhir, S.S., (2003). AMIDE: A Free Software Tool for Multimodality Medical Image Analysis. Molecular Imaging, 2(3), 133-137.

Lorr M. (1983). Clustering Analysis for the Social Scientists. Jossey-Bass.

Milestones. Alzheimer's Disease and Dementia. https://www.alz.org/alzheimers-dementia/research_progress/milestones

Nishimura, T., Hashikawa, K., Fukuyama, H., Kubota, T.,Kitamura, S., Matsuda, H., Hanyu, H., Nabatame, H.,Oku, N., Tanabe, H., Kuwabara, Y., Jinnouchi, S., & Kubo, A. (2007). Decreased cerebral blood and prognosis of Alzheimer's disease: A multicentre HMPAO-SPECT study. Annals of Nuclear Medicine, 21(1), 15-23.O'Sullivan . F, Muzi, M., Spence, A., Mankoff, D.,O'Sullivan, J., Fitzgerald, N., Newman, G., Krohn, A., (2008). Non-parametric Residue Analysis of Dynamic PET Data with Application to Cerebral FDG Studies in Normals. J Am Stat Assoc, 104(486), 556-571

O'Sullivan, F. (2006). Locally constrained mixture representation of dynamic imaging data from PET and MR studies. Journal of Biostatistics, 7(2), 318-338.

Pavlopoulos, S., Thireou, T., Kontaxakis, G., Santos, A. (2007). Analysis and interpretation of dynamic FDG PET oncological studies using data reduction techniques. Biomedical Engineering Online, 6(36).

Peretti, D., Vállez García, D., Reesink, F., van der Goot, T., De Deyn, P., de Jong, B., Dierckx, R. and Boellaard, R., 2019. Correction: Relative cerebral flow from dynamic PIB scans as an alternative for FDG scans in Alzheimer's disease PET studies. PLOS ONE, 14(3), p.e0214187.

Phelps, M., (2000). Positron Emission Tomography provides molecular imaging of biological process. PNAS, 97(16), 9226-9233.

Pmod.com 2021. About us – PMOD Technologies LLC. [Online] Available at <https://www.pmod.com/web/?page_id=1006>

Society of Necular Medicine (2009). PET PROS: Alzheimer's disease and molecular Imaging: get the facts. Retrieved from http://www.snm.org/docs/PET PROS/Alzheimer's.pdf

Terry, R., Katzman, R., Bick, K., & Sisodia, S. (1999). Alzheimer's Disease, 2nd Edition, Lippincott Williams and Wilkins.

The Whole Brain Atlas http://www.med.harvard.edu/AANLIB/cases/caseNA/pb9.htm

U.S. National Institute of Health http://www.nia.nih.gov/Alzheimers/AlzheimersInformation/GeneralInfo/)

Watabe H., Ikoma Y., Kirmura Y., Naganawa M., and Shidahara M. (2006). PET Kinetic Analysis-Compartmental Model. Annals of Nuclear Medicine, 20(9), 583-588.

Weisberg S. (2005). Applied Linear Regression, 3rd Ed. Wiley.

Appendix 2.A

Time activity curves of the whole Brain

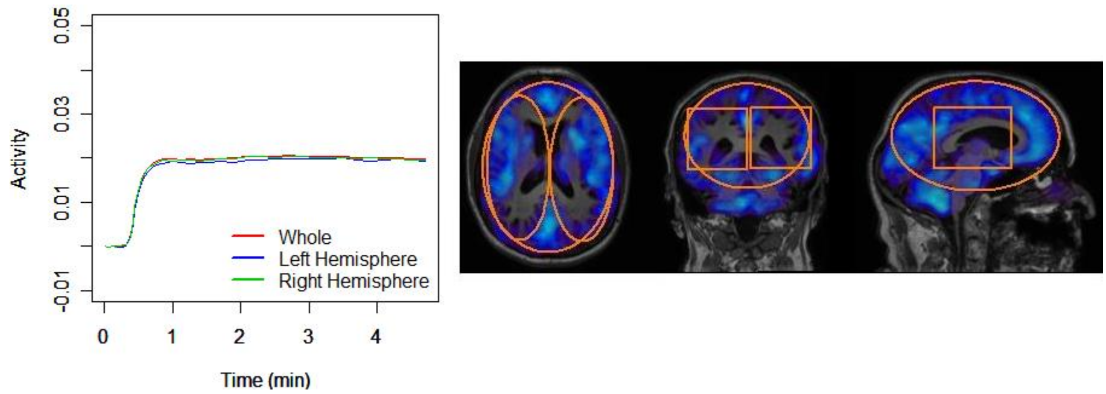

Figure 2. 25: Time Activity Curves of the whole brain along with the left and right hemisphere. PET water dynamic image with MR scan superimposed showing the whole brain along with the left and right hemisphere.

Table2.13: Parametric & Non-Parametric estimates of the whole brain with the left and right hemisphere.

Region	**Flow**	**VB**	**VD**
Whole Brain	1.1042±0.0197 2.1265±0.0031	0.1284±0.0186 0.1313±0.0001	4.3524±3.689 1.3018±0.002
Left Hemisphere	1.0529±0.0195 2.1578±0.0042	0.1242±0.0186 0.1204±0.0002	4.3028±2.896 1.3267±0.0028
Right Hemisphere	1.0877±0.0186 2.1492±0.0045	0.1267±0.0178 0.1206±0.0002	4.3028±1.921 1.3125±0.0030

Appendix 2.B

Estimates of all the regions of interest were generated using non-linear least squares. Voxel sizes of all regions are given.

Table 2.14: Non-Parametric Functional Estimates and Standard Errors for ROI's.

Region	δ ± s.e	VB ± s.e.	K1 ± s.e.	k2± s.e.
Whole Brain 77,842*	0.0102±0.01301	0.1284±0.0186	1.1042±0.0197	0.2537±0.0119
Left Hemisphere 29,934*	0.0104±0.0135	0.1242±0.0186	1.0529±0.0195	0.2447±0.0123
Right Hemisphere 29,105*	0.0082±0.0121	0.1267±0.0178	1.0877±0.0186	0.2494±0.0114
Small White Matter 1,362*	-0.0127±0.0252	0.0822±0.0242	1.0208±0.0192	0.2232±0.0118
Medium White Matter 2,291*	0.0126±0.0131	0.1000±0.0147	0.9761±0.0140	0.2125±0.0093
Large White Matter 3,486*	0.0063±0.0144	0.0877±0.0161	0.9729±0.0139	0.2121±0.0090
Small Grey Matter 1,294*	-0.0186±0.0247	0.0896±0.0322	1.2500±0.0246	0.2832±0.0125
Medium Grey Matter 2,326*	0.0105±0.0171	0.1209±0.0237	1.1724±0.0247	0.2794±0.0143
Large Grey	0.0100±0.0165	0.1265±0.0239	1.2200±0.0249	0.2843±0.0140

Region	δ ± s.e	VB ± s.e.	K1 ± s.e.	k2± s.e.
Matter 3736*				
Hippocampus (Left) 108*	0.0291±0.0254	0.1944±0.0368	1.0653±0.0455	0.1891±0.0281
Hippocampus (Right) 117*	-0.0163±0.0239	0.1997±0.0494	1.0995±0.0509	0.1995±0.0292
Cerebellum Grey 1,109*	0.0102±0.0226	0.1031±0.0331	1.5576±0.0275	0.3172±0.0119
Cerebellum White Matter 998*	-0.0246±0.0502	0.0427±0.0659	1.3829±0.0270	0.2675±0.0095
Blood Region 1 36*	0.0254±0.0358	0.2438±0.0612	1.8068±0.0842	0.3705±0.0359
Blood Region 2 40*	-0.0183±0.1646	0.0633±0.2809	1.8844±0.1284	0.3074±0.0349
Blood Region 3 33*	-0.0446±0.0034	0.9766±0.0372	0.6117±0.0562	0.1111±0.0600
White Matter Near BR1 112*	0.0163±0.0330	0.1523±0.0468	1.1452±0.0547	0.2927±0.0333
White Matter Near BR2 119*	-0.0117±0.0196	0.0922±0.0536	1.3599±0.0373	0.2177±0.0171
White Matter Near BR 3	-0.0163±0.0241	0.1464±0.0388	1.1124±0.0365	0.1922±0.0204

Region	δ ± s.e	VB ± s.e.	K1 ± s.e.	k2± s.e.
113*				
Grey Matter Near BR1 98*	0.0074±0.0587	0.1057±0.0746	1.2475±0.0663	0.2706±0.0349
Grey Matter Near BR2 105*	0.0042±0.0521	0.1377±0.0979	1.5267±0.1018	0.3446±0.0469
Grey Matter Near BR3 93*	-0.0257±0.0159	0.2476±0.0487	1.1718±0.0530	0.2476±0.0487

*denotes voxel size of that particular region.

Appendix 2.C

The mean values were used for the non-parametric residue analysis. The boxplots in **Figures 2.26**, **2.28**, and **2.30** are the activity levels for flow, VB and VD for all the regions. The data for the non-parametric analysis was viewed in AMIDE. **Figures 2.27**, **2.29**, and **2.31** represent flow, VB, and VD, respectively.

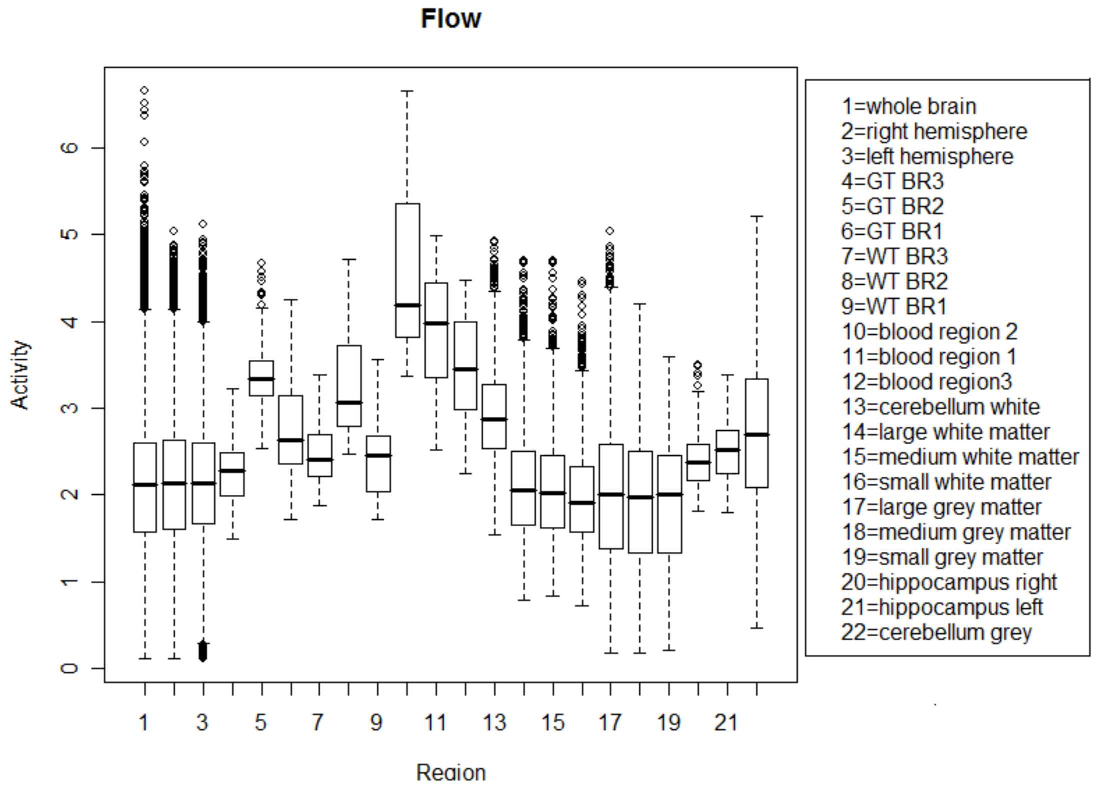

Figure 2.26: Boxplots of Activity Representing Flow.

The flow parameter for the non-parametric analysis is illustrated in **Figure 2.27**.

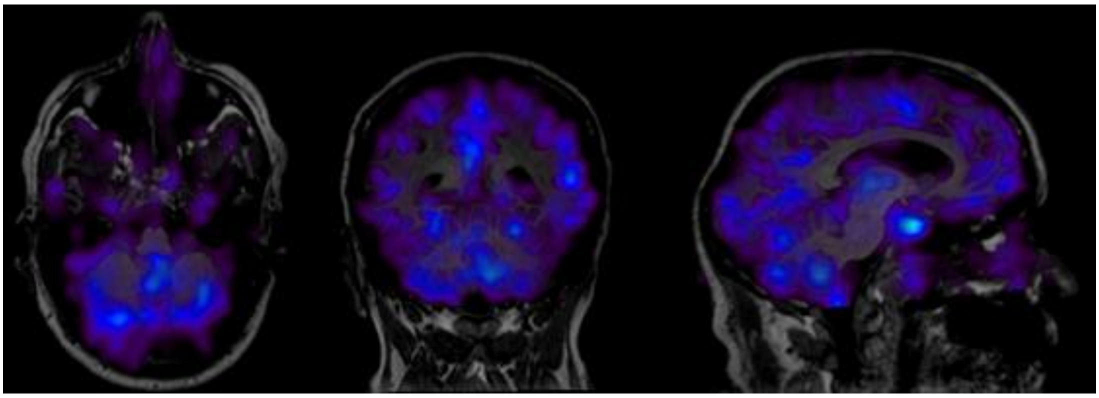

Figure 2.27: Parim generated in AMIDE Representing Flow Co-Registered with MRI Scan.

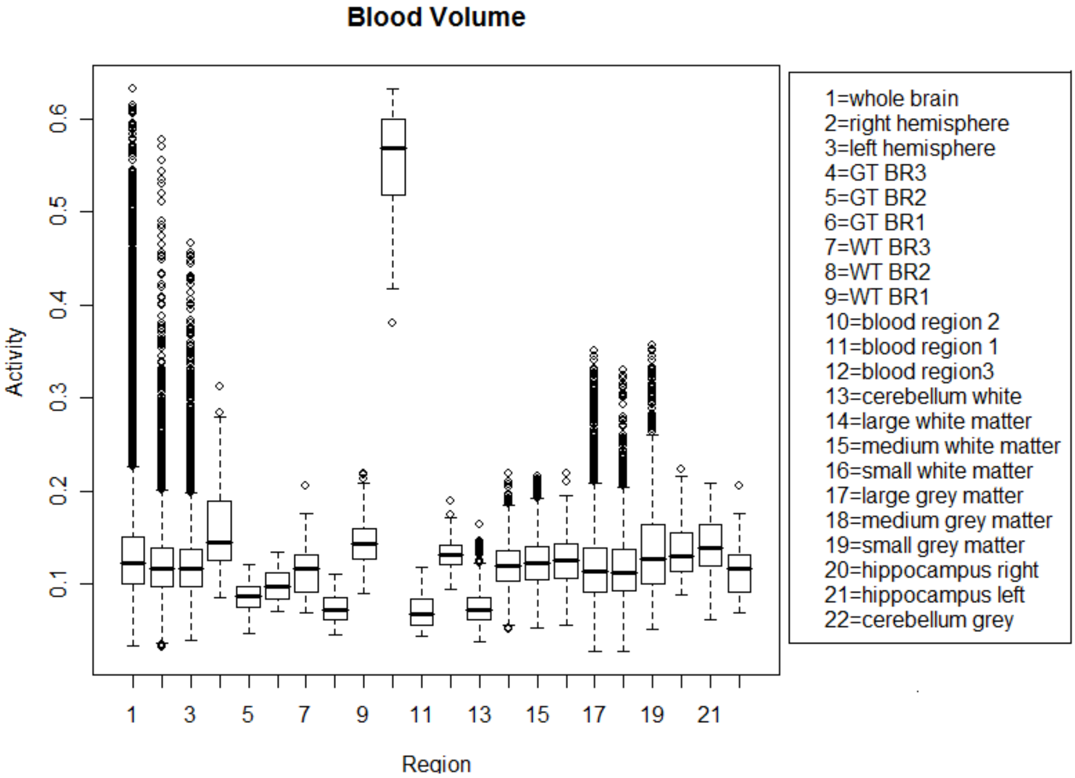

Figure 2.28: Boxplots of Activity Representing VB.

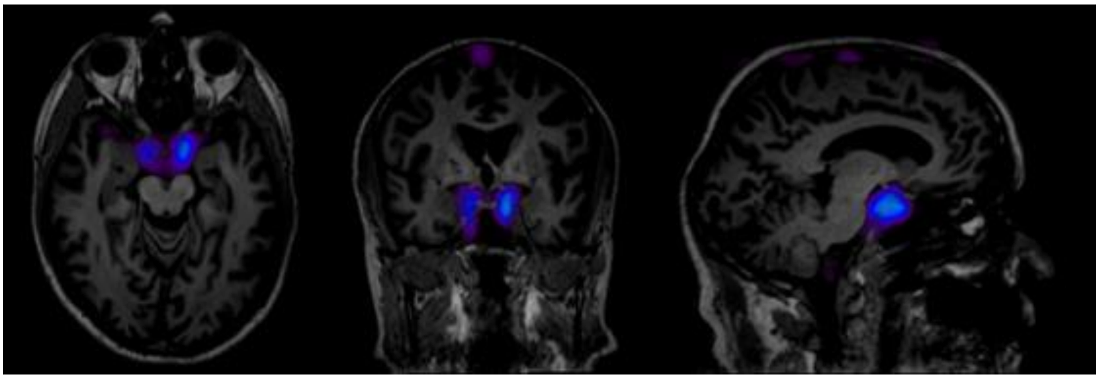

Figure 2.29: Parim Generated in AMIDE Representing VB Co-Registered with MRI Scan.

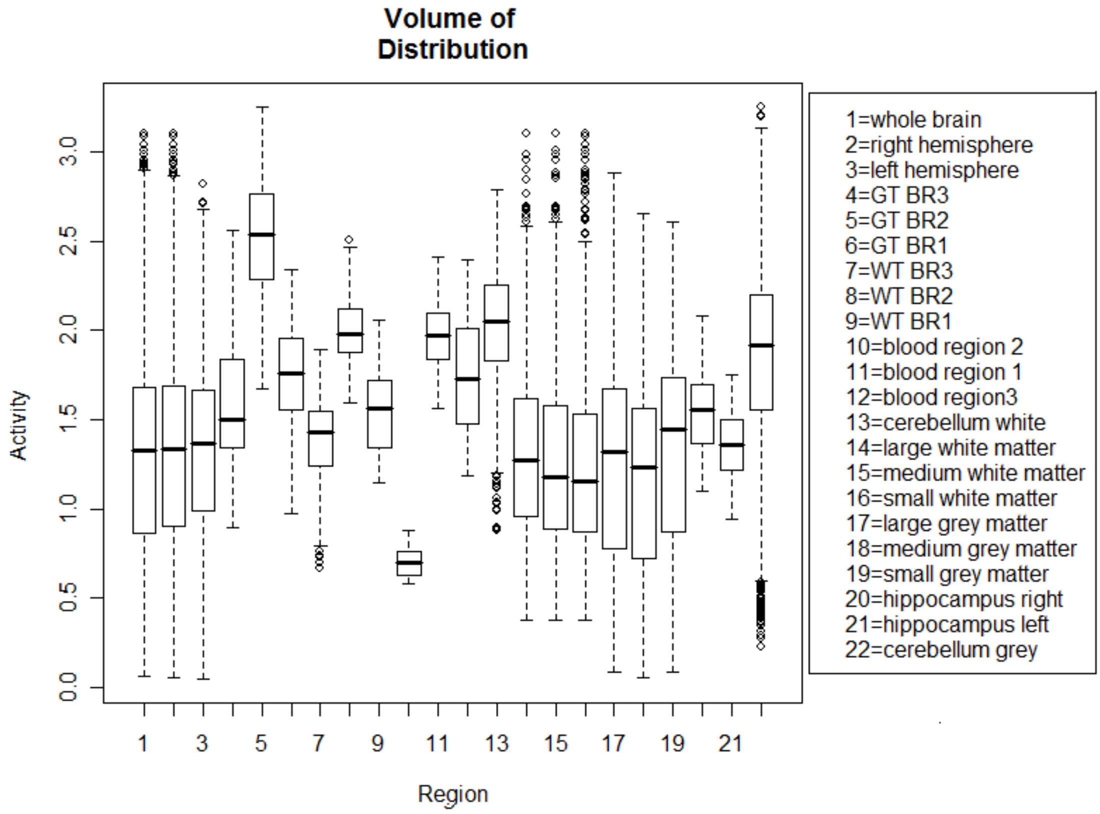

Figure 2.30: Boxplots of Activity Representing VD.

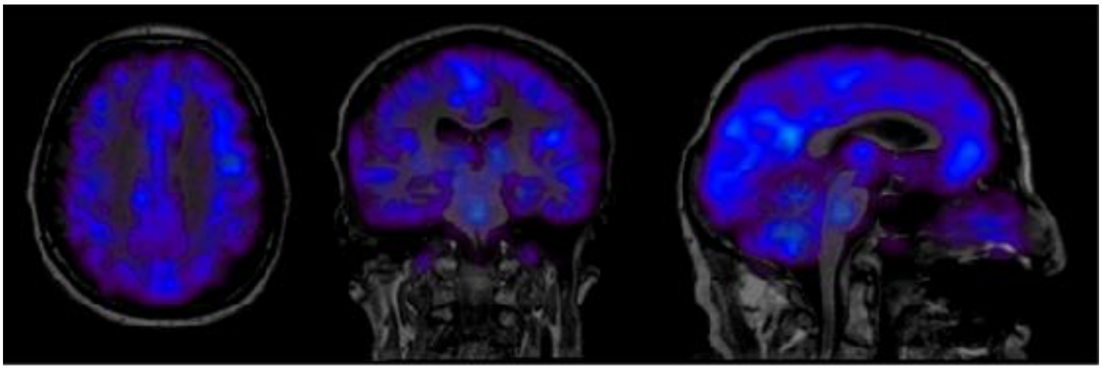

Figure 2.31: Parim Generated in AMIDE Representing VD Co-Registered with MRI Scan.

Appendix 2.D

Blood region 3 is very similar to the AIF used in our study. **Figure 32** represents the AIF used in our study.

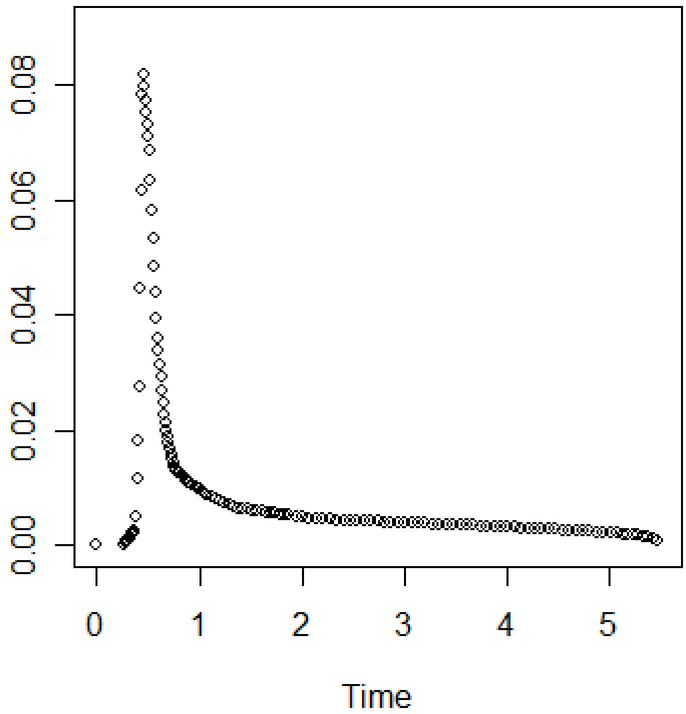

Figure 2.32: AIF used in our Study.

Chapter 3: GRADUATION OF CRUDE RATES BY STANDARD TABLE METHOD: POISSON MODEL

Supervisor: Damian Conway

3.1 Introduction

Mortality rate tables are a crucial input into all pricing and valuation exercises carried out by life insurers for all types of insurance and pension products. Insurance and pension companies perform calculations that are based on death and continued life expectancies; they are dependent on reliable and consistent life tables (Barnett, 1951). Some examples of these calculations include insurance premiums, reserve calculations, and actuarial liabilities. Different types of products provide benefits contingent on life expectancy, for example, life insurance products typically pay benefits on the death of the life insured and annuity products continue to pay benefits while the insured life is alive, annuity benefits typically cease on the death of the life insured. It is vital that the rates used for pricing such products are correct and appropriate. If rates prove to be inappropriate, they may result in losses for the company and an extreme consequence may be insolvency.

The starting point for the insurer is getting the right data. In the industry today, there are many sources of data that can be used for setting assumptions used in calculations. Some examples include: national statistics compiled by the Central Statistic Office (CSO), internal data of insurers, internal data of reinsurers, industry tables and published tables by working groups such as the Continuous Mortality Investigations. Reinsurers are large multinational companies who provide services and insurance arrangements for insurance companies (Pacific Life Re – About, 2021) (IFE SA2, 2018). Reinsurers would have data for multiple (different) insurance companies. Actuaries construct tables based on data available. There is a working group known as Continuous Mortality Investigations that publishes tables, which can be used for pricing of insurance contracts.

Insurers need to be able to adjust the data available so as to make it relevant to the cohort of lives they are trying to calculate liabilities for. Understanding the factors of risk and their mitigating risk is very important for insurers (IFE ST2, 2016). As outlined in actuarial notes for life insurance, risk management is vitally important for insurers, risk cannot be fully mitigated, however it can be reduced. Examples of risk factors that can impact life insurance are age, smoking status and health status. Mitigants that can be put in place to protect against those risks include but are not limited to underwriting questions, which are used to further quantify the risk posed by individual lives current health status or increased pricing, whereby insurers add loadings to underlying rates so as to protect against adverse experience, for example imposing higher rates for smokers compared to non-smokers. According to Forbes, smoker rates in America are higher than non-smoker rates and they vary with age, for example for a 30-year-old male, premiums are on average 2.7 times higher, while for a 40 year old male premiums are almost 3.7 times higher. In Ireland, as expected due to the lower life expectancy of smokers, smoker rates are also higher, however the difference in premiums is not as extreme is and usually just under double at age bands, and tends to not vary as drastically by age (Advisor, lion.ie 2020, bonkers.ie 2020). Age does not impact the smoker loading; instead, a generic smoking uplift is applied to the underlying rates.

The different sources of data mentioned above have been produced based on different approaches to risk assessment and indeed different cohorts of lives. Therefore, it is important to base assumptions on general tables and experience, which are most relevant to the lives currently under investigation. The best results depend on the data being used in the study.

The CSO compile data on the general population, the mortality rate for a 50-year-old male non-smoker in Ireland can differ between counties, for example a man from Cork compared to a man in Dublin. There are many socio-economic factors impacting mortality for example, how affluent the area is, education levels, urban vs rural (IFE, 2016). All these factors and assumptions are combined to determine life tables. Within datasets there are bias and trends (IFE, 2016), in general, crude mortality rates increase by age for both genders (Irish Insured Lives Mortality Investigation, 2019). The mortality rates are calculated for each age and for each gender. The assumed mortality rate for the general population is heavier than the insured population i.e. the probability of someone dying in the general population is higher than the insured population (IFE, 2020). This occurs due to the selection process, insurers will only accept lives of a particular standard, this is assessed through medical underwriting.

Other factors also influence the difference in experience, for example, economic status and psychological disposition (Houston, 1959).

A basic example, let us suppose that there are two insurers, insurer A and B. Insurer A does not have any data of their own, they use the CSO tables to price their life insurance contract. Insurer B use their own data, assume they have been selling life contracts for many years and have built up a lot of data and experience on these lives. The insurer would use their experience to price the product. If we assume everything else being equal i.e. other assumptions such as expenses and withdrawals are equal, this would mean that insurer A's contract would be more expensive for customers. How does this impact insurer A? Being more expensive could lead to poor sales and bad publicity as their price is higher than the other companies. Likewise, if they have lighter mortality rates which are incorrect, compared to B, their price would be cheaper. Being cheaper could lead to a lot of sales but would mean the actual deaths they would experience would be worse than expected, which would impact profit and solvency.

For each contract developed by an insurer, the insurer must understand the risk factors of the product and proposed target market so that all available data to price the underlying risk. There are many considerations, such as product type, profile of insured person (gender, age, smoking status, socio-economic group), and geographical location. The insurer must be sure that the table they use is applicable to the lives under consideration.

An insurer must consider what data they have available themselves or the data available in the market. Before selecting a mortality basis to be used in the calculations, a study of the mortality experience of a group of lives with similar characteristics should be considered. Tables can be constructed from the insurer's own internal data, tables can be sourced from reinsurers or existing available industry tables could be used.

This study investigates an insurer using it's own data, known as the crude rates and an available life table to create a set of useable rates. Creating the useable rate is done by a process known as a graduation process by standard table. To carry out a graduation, a set of crude rates must be collected. The crude rates are not used directly as they are not regarded as reliable. The method of calculating the crude rates depends on the model used to estimate the number of deaths. This study concentrates on the Poisson model (IFE, 2008; Benjamin & Pollard, 1993). Other models can also be used to model deaths, such as the binomial model and the two state Markov model. The crude rates differ between the Poisson and binomial model due to the different exposed-to-risk calculation used. The Poisson model uses the central exposure-to-risk when it is calculating the crude rates, while the binomial model uses the initial exposure-to-risk.

A group of lives are studied for a defined period of time, and a number of factors are recorded for each life. Some of the important pieces of information collected are age, number of lives in study, the number of deaths, and their exposure time. Age specific mortality rates of the lives in the study are calculated based on a particular model of mortality. These are known as crude rates.

Crude rates are usually rough values, with erratic jumps at the extreme ages, as there is a lack of information collected, (Irish Insured Lives Mortality Investigation, 2019). Crude rates by nature have independent sampling error, as they are estimated independently. The rates can be inconsistent, that is, the mortality rate prescribed to an individual aged 60 years old could be lower than that of an individual aged 55, which would contradict the well-known fact that as age increases, rates of mortality increase. Therefore, crude rates require development and refinement before being used by insurers to assess and price the risks covered by them.

A smoothing process known as graduation is carried out on our crude rates. London (1985) (IFE, 2008) summarises some of the varied definitions. Here, graduation is defined as "an effort to represent a physical phenomenon by a systematic revision of some observations of that phenomenon" and "the process of securing, from an irregular series of observed values of a continuous variable, a smooth regular series of values consistent in a general way with the observed series of values". The aim of graduation is to get a good balance between smoothing and adherence to the data (IFE, 2008). A study, which uses graduation for crude mortality rates states that graduation is a method or principle by which a set of observed crude rates can be adjusted to be used as a suitable basis for computations (Abid et al, 2006). However, graduation cannot remove gaps or underlying flaws present in the data set due to the collection of incorrect data.

Smoothing the data and getting the rates to adhere to the data as much as possible are conflicting goals. As such, by improving the smoothness of the rates, the rates do not adhere to the data and vice versa. This leads us to overgraduation and undergraduation.

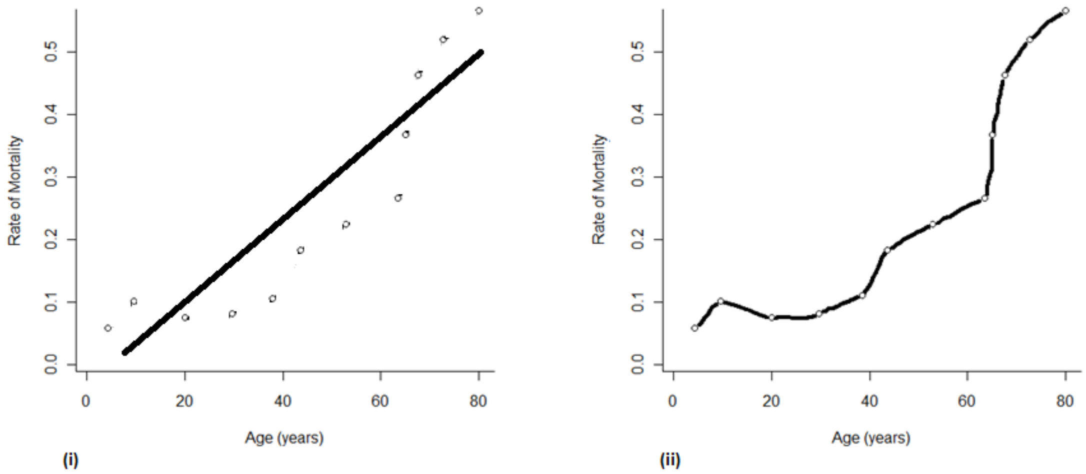

Figure 3.1: Graduated mortality rates (line) with (i) over and (ii) undergraduation. Points represent crude rates.

Overgraduation means rates are too smooth and do not adhere to the data, while on the other hand, undergraduation means they adhere to the data too much and are not smooth enough. **Figure 3.1** (i) illustrates overgraduation, and we can see from age 20 – 70 it over-estimates, while in the early life and late ages it underestimates. **Figure 3.1** (ii) shows us the effects of undergraduation; here, there is excellent adherence to the data but no smoothness. Lives aged about 15 have a higher mortality level than those aged 20 years old.

If insurers are consistently overestimating or underestimating the underlying mortality, it can have huge financial implications, resulting from more claims than expected and insufficient premium rates being earned. **Figure 3.2** shows overestimating and underestimating graduated rates.

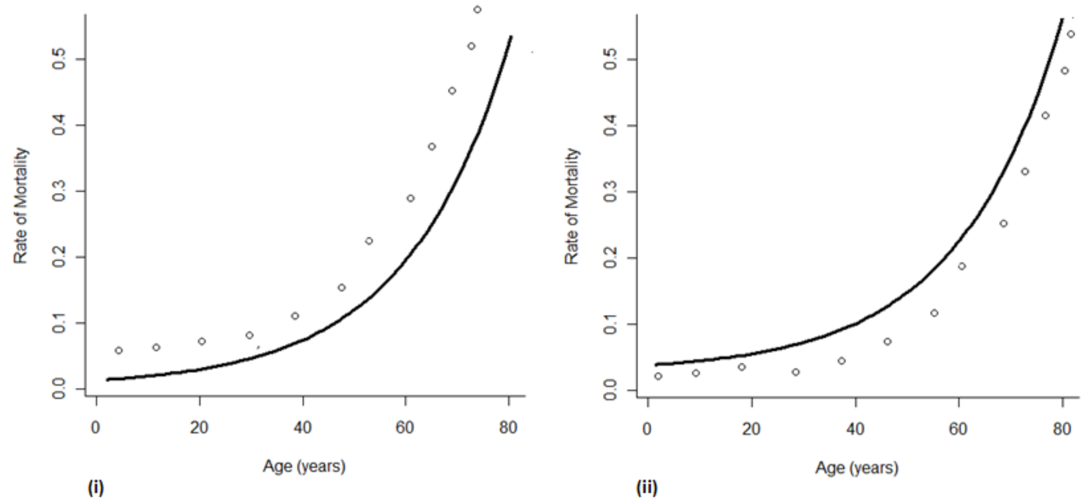

Figure 3.2: Graduated mortality (line) rates showing (i) underestimation and (ii) overestimation of the underlying crude rates (points).

As mentioned previously, the products insurance and pension companies sell are contingent on life and death. An insurance company pays out if an individual dies early. If the rates are low, then they are underestimating mortality. That is, the insurance company think individuals are going to live longer than they will. Based on these rates, they will not charge a high enough premium to cover the costs, and this will result in losses; the effect of this could be lead to very high losses and, in extreme cases, even insolvency. If the life rates are too high, then the insurer are overestimating mortality, which means life insurance companies think people will die sooner than they actually will. Mortality being assumed too high would result in the premium being too high and could make their life products uncompetitive, which will then affect several factors; in particular, profit. Lower than expected sales volumes could lead to costs not being covered. If the insurer provide pension products, for example an annuity, this product pays benefit while the life assured is alive. If the insurer have underestimated the rates, then individuals will live longer than expected and collect their pension for longer than expected. The premium being charged will not cover these costs and will lead to losses and again in extreme cases, insolvency.

A number of statistical tests are carried out to try to find a balance between smoothness and adherence to the data. The aims of this current study are to (1) calculate a set of crude rates from a given population, (2) carry out a graduation procedure, and (3) apply statistical tests to achieve the correct balance in smoothness and adherence to the data.

In 2018, there was study conducted, which explored the different statistical techniques for mortality graduation (Ramonat and Aktuar, 2018). Their study is based on the Canadian Pensioners Mortality study and utilised first principles to develop a set a graduated rates from crude rates. The study also used the Poisson model as its distribution function for the number of deaths. The shape of the rates produced by that study were similar to this study. Two aims of Ramonat and Aktuar's study (2018) is to show these statistical techniques are still valid today and to demonstrate to Actuaries in North America the power of these techniques and how they can improve their understanding of mortality of a given cohort. Ramonat and Aktuar's study conducted statistical tests to validate the results, some of those tests were Pearsons' Chi-square test, signs test and deviation test. These tests were also conducted in this study to validate the results.

In 2016 and 2018 there were two studies conducted which produced life tables for Ireland and England respectively. In both studies, crude rates were determined, which have a similar shape to the rates produced in this study. The study conducted (which focused on the English life tables by Dodd et al., (2018)), used the simple linear link function and Possion model. The application of these techniques can be used with any data set. Another study conducted in 2013 by Richards, Kaufhold and Rosenbusch, showed how these techniques can be used by companies to create bespoke rates and pricing for cohorts of business. The shape of the crude and graduated rates produced in these two studies for life tables were similar to the shape produced by this study. Another study conducted in the Slovak Republic, investigated the mortality experience of insured lives over an 11-year period by data insurers (Pacáková, Jindrová. and Seinerová, 2013) The aim of their study was to produce a set of graduated mortality rates to be used for insurance pricing. Pacáková et al, (2013) also demonstrate the impact on insurance premiums based on different mortality experience. The shape of the crude and graduated rates produced by Pacáková et al, (2013) are similar to this study. Pacáková et al, (2013) used another method of graduation as part of their study and the results are similar; the rates were smoothed using a time series technique.

Pacáková et al, (2013) results were validated using similar statistical tests as this study. These techniques are still powerful tools for understanding and deriving mortality rates.

3.2 Methods

This section describes the methodology used for the 3 steps in the process, (1) describe the model used to estimate the force of mortality for the given dataset. (2) Examine the method of graduation by reference to standard table methods is examined and (3) examine the statistical tests conducted on the data.

3.2.1 Different Notation

In the following sections, we will use the following notation.

μ_x refers to the true rate of mortality at age *x.*

$\hat{\mu}_x$ refers to the numerical value of the crude estimate at age *x.*

$\tilde{\mu}_x$ refers to the random variable representing the crude estimator at age *x.*

$\overset{\circ}{\mu}_x$ refers to the graduated rate at age *x.*

$\overset{s}{\mu}_x$ refers to the standard table rate at age *x.*

q_x *refers to the probability a life aged x will die before they are x+1 (provided from a life table)*

$\dot{q}_x$ *refers to the probability a life aged x will die before they are x+1, (provided from an observed dataset).*

θ_x refers to the actual number of deaths.

3.2.2 The Poisson Distribution

The Poisson distribution is used to model rare events that happen over a specific timeframe. Suppose we observe N individuals and we assume their deaths follow a Poisson distribution. It is a discrete probability distribution where random variable (RV) can only take positive values (Baines, 2008). A RV, X is said to follow a Poisson distribution with mean λ, (λ > 0), given the probability function of X is

$$P(X = x) = \frac{e^{\lambda}\lambda^{x}}{x!} \tag{1}$$

Where the mean, E(X) = λ and the Variance of X, Var (x) = λ.

3.2.3 The Poisson Model

Using the Poisson distribution and theory on exposed-to-risk, we can derive information about the underlying force of mortality (IFE, 2008).

Force of Mortality

The force of mortality at age x is denoted μ_x and is an important quantity in survival models (IFE, 2008). μ_x is an instantaneous measure of mortality at age x. μ_x is the continuous equivalent of the discrete quantity q_x which is the mortality rate used in the binomial model. The quantity μ_x is defined as

$$\mu_x = \lim_{h->0^+} \frac{1}{h} \times P[T \leq x + h \mid T > x] \tag{2}$$

The future life of an individual is a random variable, denoted by T. T is continuously distributed on the interval [0,ω], where 0 < ω <∞. ω is the maximum age a life can take. The AM92 tables have this limiting age at 120 years.

The force of mortality at age x+t denoted μ_{x+t} can be defined as

$$\mu_{x+t} = \lim_{h->0^+} \frac{1}{h} \times P[T \leq x + t + h \mid T > x + t] \tag{3}$$

given $x \geq 0$, $t > 0$.

Exposure-to-Risk

Consider the time period of (t, t+1) and all lives at time 0 are aged x. These lives are observed until the end of the investigation period or death. The number of lives at the beginning of the study are the number of lives at risk, which is referred to as the exposed-to-risk or initial exposed-to-risk, E_x. The number of lives aged x last birthday, which die during the period, is Θ_x. The mortality rate $\dot{q}_x$ is calculated

$$\dot{q}_x = \frac{\Theta_x}{E_x} \tag{4}$$

In reality, not all lives are the exact age at the beginning of the investigation period or lives may leave the investigation period due to circumstance other than death (e.g. withdrawal from an insurance contract due to affordability). Lives could join the investigation during the investigation period (e.g. purchase of an insurance contract).

Given these factors, it can be difficult to correctly allocate the time spent in an investigation and so reasonable assumptions must be made. Suppose the number of lives aged between exact age x and x + dx at time t is P(x,t)dx. In the time interval (t, t + dt), the number of expected deaths are

$$\int_t \int_x \mathrm{P}(x,t)\,\mathrm{dx}.\mu_x dt \tag{5}$$

Central exposure-to-risk at age x, E_X^C, is the amount of time lived for a given age range in an investigation period. The number of observed years lived in the investigation is given by

$$\int_t \int_x \mathrm{P}(x,t)\,\mathrm{dx}dt \tag{6}$$

Where

$$\int_0^1 \int_x^{x+1} \mathrm{P}(x,t)\,\mathrm{dx}dt \tag{7}$$

represents the number of years of life lived for lives aged between [x , x+1] during the investigation period.

A discrete model for N lives takes the following form:

$$E_x^c = \sum_{i=1}^{N}(b_i - a_i)(1-d_i) + \sum_{i=1}^{N}(t_i - a_i)d_i \tag{8}$$

Only lives in the age range x to x+1 and during the period (a , b) are used in the calculation. Each life contributes only the time lived in that age range during the investigation period. If a life dies, the amount of time contributed is $(t_i - a_i)$, where t_i is the time of death of an individual life. Otherwise, the length of the investigation period was applicable.

The central exposed to risk calculation is not affected by complicated multiple state or multiple decrement models. The central exposure-to-risk calculation can be carried out even with incomplete observations.

When using the Poisson model, the total observed waiting time is denoted as E_X^C. Suppose there are N individuals observed and it is assumed that their deaths follow a Poisson distribution, and the force of mortality is a constant μ. Then, the RV for death (D) is said to follow the Poisson distribution with parameter μE_X^C.

$$D_x \sim Poisson\ (\mu E_X^C) \tag{9}$$

The number of deaths recorded in this interval is denoted d and thus d is a sample value of the RV D. The probability function of D becomes

$$P[D = d] = \frac{e^{-\mu E_x^c} (\mu E_x^c)^d}{d!} \qquad 10)$$

Estimating the Underlying Force of Mortality

The true value of μ_x cannot be established, so using the information collected regarding the number of deaths observed and the total exposed-to-risk, we estimate the true force of mortality. We get the Poisson likelihood estimator for the constant μ (IFE, 2008; Baines, 2008).

$$\tilde{\mu}_x = \frac{D}{\mu E_X^C} \tag{11}$$

The estimator $\tilde{\mu}_x$ has the following properties:

$E[\tilde{\mu}_x] = \mu$, that is, it is unbiased.

$Var[\tilde{\mu}_x] = \frac{\mu}{E_X^C}$, which is the smallest variance possible for an unbiased estimator. μ.

The maximum likelihood estimate (MLE) of μ is $\hat{\mu}$ which is expressed as,

$$\hat{\mu} = \frac{d_x}{E_x^c} \tag{12}$$

$\hat{\mu}$ represents the crude estimate of the force of mortality. The maximum likelihood estimator is the best linear unbiased predictor, using this estimator will provide the most efficient estimates of the data.

This is appropriate, as linear least squares will be used to determine the coefficients in the link function for calculating the graduated rates. In addition, as the datasets increase, the estimate will converge to the true value. If a larger dataset was utilised, this would allow the same approach to be used. The value of the crude force of mortality was calculated by dividing the number of deaths at age x in our study by the central exposure time at age x.

Definition of Age & Rate Interval

It is crucial in our study that the ages definition for the number of deaths corresponds to the age definition the central exposure is based on. The age label for this study is age last birthday; based on this age definition the rate interval can be deduced. A rate interval is a one-year time period where a life's age label remains the same throughout. Based on our age definition, the rate interval in this study concentrates on the interval (x , x+1), as per (7) (IFE, 2008). The Poisson model uses $\hat{\mu}_{x+0.5}$ to estimate the crude rate over the rate interval.

3.2.4 Describing the Data

The dataset for this study consisted of 50,000 male life insurance policyholders. The study was conducted by a reinsurer who collected data from a number of insurance companies for term insurance business and focused on a five-year period beginning on the 1st of January 2005 and ending on the 1st of January 2010. This period is referred to as the observation period.

The dataset, is a snapshot of each insurer's book of business at that time, consisting of 50,000 policyholders with existing policies on the 1st of January 2005. All relevant data should be used if it adds value. At extreme age ranges, the level of available data reduces, and this can lead to unreliable estimates of the crude rates (Irish Insured Lives Investigation, 2019). Information regarding each policyholder was recorded, namely, each policyholder's date of birth, date of entry to the policy, outcome of policy, and date of outcome of policy. There are three possible outcomes for each policyholder involved in the study.

1. Survival (I): They survived the duration of the observation period and the policy remained in force.
2. Surrender (S): They surrendered their policy during the observation period.
3. Death (D): They died during the observation period.

Using this information, the central exposure to risk was calculated. The central exposure only records the time spent under a certain observation by an individual. The central exposure of a life with age label x was calculated between two dates, A and B.

Date A is stated as the latest of:

- The date of reaching age label x,
- The start of the investigation,
- The date of entry.

Date B is defined as the earliest of:

- The date of reaching age label x+1,
- The end of the investigation,
- The date of exit.

There are two possible reasons for exit; either the policyholder surrenders or dies. If either of these occur, then the time spent alive or in the investigation period is used towards the calculation of the central exposure to risk for year of age [x, x+1]. Therefore, a policyholder who exits the policy only contributes to a fraction of that year. The same happens for new policyholders, in that they only contribute to a fraction of the year of entry. The number of policyholders who die during each year is recorded.

Crude rates of mortality were calculated using all this information. The data was prepared, and all calculations were carried out using Microsoft Excel.

3.2.5 Life Tables and Changes in Mortality

There are two types of life tables: national life tables and life tables issued by insurance companies. The life tables are known as standard tables if they are based on reliable information for particular applications (IFE, 2008; Benjamin & Pollard, 1993). Life tables in the UK are released every ten years and are compiled using census data and deaths registration data based on the population of England and Wales.

In the UK, insurance companies supply information to the Continuous Mortality Investigation Bureau (CMIB). The CMIB compile all this information and release tables that are used for different types of business. An example of standard tables still used in insurance and pension businesses today are the "92 series". The series was based on data of the number of deaths dx and the central exposure-to-risk between the years 1991 and 1994. Different tables for both male and female were released, with applications to different aspects of insurance and pensions business.

An important process for each insurance company is that their rates are consistent with these standard tables. The shape of mortality does not generally change, however mortality levels have been falling due to better healthcare and living standards. Therefore, it is appropriate that the overall shape of a particular group is consistent with the standard tables. For this reason, the "92 series" is one of the most widely used standard tables and the table is based on past mortality experience.

3.2.6 Graduation of Crude Rates with Reference to Life Table

Crude force of mortality rates, $\hat{\mu}_x$, were calculated for each age using the sample data.

These rates are crude estimates, which were based on a relatively small sample size when compared to the overall population; before being used more widely they must be adjusted so as to improve their credibility. By smoothing these rates using a graduation method involving standard tables, it allows the rates to become more credible and reliable. The rates are then more appropriate for use in actuarial calculations. This process is known as graduation.

There are a number of reasons why crude rates are not appropriate for general use. For example, they can contain errors from sampling, for example the recording of the date of death being misclassified or the sample size being too low, as well as different age profiles and demographic regions. In this instance, the rates determined for one region would not be appropriate to use in another because some regions may have a lower level of health than others, which in turn would create higher levels of mortality.

In this study, we graduate our results with the use of standard tables. The standard table rates introduce extra information into the graduation by expressing prior opinion about the pattern or shape of the true rates we are trying to estimate (London, 1985). It is a smoothing process that makes our estimates more reliable, gives us the ability to extract information from adjacent ages, and removes random sampling errors (McCormack, 2016). For example, smoothing removes large irregularities between two age groups. Although the aim of graduation is to obtain a good balance between smoothing and adherence in the data, graduation cannot remove errors from the data set due to data quality for example incorrect data collection.

The first step in graduation with reference to a standard table is to choose a suitable standard table. A suitable table should be well defined and have a similar shape to the population used in this study; for example, have similar demographic, region, age profile, and smoking status of the sample dataset. This study uses the AM92 tables. The AM92 tables were collated by the Faculty and Institute of Actuaries for male lives, who have a long history of providing this information to the market for use by general insurance companies and pension funds. The AM92 tables are based on the UK population for insured lives and the sample data mortality is similar to this mortality class. AM92 was selected as its shape and characteristics are similar to the population used in this study, i.e. it represents a mortality study, in general as age increases mortality and deaths increase. AM92 rates are used to price term assurance products, the data in this study is from lives who purchased term assurance products. A plot of the crude rates in this study and AM92 have a similar shape. The similar shape can be seen in **Figure 3.4** and the plot in **Appendix 3.B**.

There are other tables, which could be considered, for example 00 series or S1 series (CMI Mortality and Morbidity Tables | Institute and Faculty of Actuaries, 2021). Each table is created based on specific characteristics, for example, the S1 series are mortality rates based on experience from a UK self-administered pension scheme. Choosing the S1 series would be appropriate if the underlying characteristics of that experience were similar to this study's data. The mortality experience for pension products can differ to the mortality experienced for other products, for example, the mortality experienced for individuals in a pension product can differ to a term assurance product.

The standard tables force of mortality at age x+0.5 is defined as, $\overset{s}{\mu}_{x+0.5}$, the graduated force of mortality at age x is $\overset{o}{\mu}_{x+0.5}$. The relationship between the crude mortality rates and the rates from the standard table is represented using a simple link function f(.) with unknown parameters.

It is assumed that the crude force of mortality rates $\hat{\mu}_{x+0.5}$ will be referenced to a standard table using the linear relationship,

$$\overset{0}{\mu}_{x+0.5} = a + b\,\overset{s}{\mu}_{x+0.5} \tag{13}$$

A linear relationship is assumed between the crude and standard table rates based on the shape of the rates. Another similar study conducted in 2018 by Ramonat and Aktuar used linear link function when performing a graduation of mortality rates. Using a linear link function is a simple approach to graduation but provides reliable results. The use of a linear function is appropriate based on the linear shape in **Figure 3.6**.

Other link functions could be considered, for example logit or probit. Another study was conducted by McCormack in 2016, which applied the method of graduation to census data using a B spline method. This study chose a linear function due to the similar relationship between the underlying data of the study and AM92 tables. The unknown parameters a and b were calculated using weighted linear least squares (Weisberg, 1985) between the crude rates and the standard table rates.

The parameters estimated are chosen to minimise the weighted sum of squared residuals.

$$\sum_{\text{All ages x}} w_x(\hat{\mu}_{x+0.5} - \overset{o}{\mu}_{x+0.5})^2 = \sum_{All \text{ ages x}} w_x[\hat{\mu}_{x+0.5} - (a + b\overset{s}{\mu}_{x+0.5})] \tag{14}$$

The weights, w_x, should be proportional to the reciprocal of the variance of the crude weights, thus for our study we have the following:

$$w_x = \frac{E_x^c}{\hat{\mu}_x} \tag{15}$$

Weighted linear least squares were used to estimate the parameters as it is an efficient method when using small data sets. Weighted regression assigns less weight to values with high variance. When estimating a and b, the regression model focuses on the case with little variance. The weights use the reciprocal of the variance because this will give more weighting to the ages where the estimated variance is low. More weight is assigned to areas where we are more confident of the true rates. Less weight is assigned to areas of high variance and where we are uncertain about the true rates.

The graduated rates can be calculated by putting in the estimated parameters into the formula:

$$\overset{0}{\mu}_{x+0.5} = a + b\overset{s}{\mu}_{x+0.5} \tag{16}$$

Statistical tests of goodness-of-fit and smoothing will be carried out on the graduated rates versus the underlying crude rates to see if our graduation was successful.

3.2.7 Statistical Tests of a Mortality Experience

Five different statistical tests to our data set, (1) the Chi-square test, (2) individual standardised deviations test, (3) cumulative deviation test, (4) signs test, and (5) grouping of signs test. These tests are described in other literature (IFE, 2008; Benjamin and Pollard, 1993) and are prescribed in guidelines by the actuarial profession (Barnett, 1951). The tests are used to determine if the newly created graduated rates are sufficiently smooth to the crude rates and within an acceptable level. A study by Ramount and Kaufhold, (2018) also use these methods as part of the validation.

Test (1) the Chi-square test, is the most widely used statistical test, however, the Chi-square test has many limitations. Two examples of limitations are; a large number of small deviations balanced out by a few large deviations or a large number of deviations of the same sign. Limitations occur because the Chi-square test assumes observations are independent, this can lead to unreliable results. Other tests are required to highlight these limitations. Tests 2 to 5 are performed to provide information on what the Chi-square test would overlook. Tests 2 and 3 examine the standardised deviations for normality, tests 4 and 5 test whether the deviations are randomly distributed, with respect to age (Benjamin and Pollard, 1993).

(Test 1) Chi-Square Test

The purpose of a Chi-square test is to detect discrepancies between observed and expected data. In this study, it examines the expected number of deaths based on our new graduated rates versus the actual deaths (Benjamin and Pollard, 1993). The observed values should be consistent with the model generating the expected values. The goodness of fit of a particular statistical model describes how well it fits a set of observations.

The Chi-square test statistic was defined by Karl Pearson as

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{17}$$

where O_i represents the observed values and E_i the expected values.

The hypotheses for the test are given as:

H_o: There is no difference between the graduated rates and the true underlying crude rates.

H_a: There is a difference between the graduated rates and the true underlying crude rates.

The null hypothesis is rejected if the test statistic is greater than a critical value. If the Chi-square test statistic is very high, then we can say there are discrepancies between the observed number of deaths and expected number of deaths. There are two main assumptions with the Chi-square test: the lives in the study must be independent, and there is no heterogeneity of mortality within a particular age group. The expected number of deaths must be high enough for a particular age group for the test to be valid. A rule of thumb is to use at least five cases in each age group (IFE, 2008).

The first step is to examine the data with some summary statistics, for example, check for any negative values in the death column, this cannot be a valid value as there cannot be negative deaths in a year. If a negative value exists, it should either be adjusted or excluded, a reasonable assumption should be documented if adjusted (IFE 2020). Examples of additional data checks are missing data or outliers. Data checks were performed using Excel to generate some simple summary statistics and through a quick visual inspection. The critical value for this test is calculated by noting the degrees of freedom in the test and using a five percent threshold, by finding the value in the Chi-square tables. The null hypothesis is rejected if,

$$\chi = \chi^2_{m-p,0.95} \tag{18}$$

Where m denotes the degrees of freedom and p the unknown parameters in the link function. The number m is the number of age groups used in the study. The degrees of freedom calculation are based on (66 -2 = 64). The test statistic is outlined in (19). Thus, the null hypothesis is rejected if,

$$\chi = \chi^2_{66-2,0.95} \tag{19}$$

Conclusions of the Chi-Square Test

If the null hypothesis is rejected, this implies that the observed data differs from the expected. This could suggest a poor fit or overgraduation. If there are large deviations at a given age group this can be used to identify ages where the fit is bad, test (2) will examine this. On the other hand, if the test statistic is very low then this could be a result of undergraduation.

Limitations of the Chi-Square Test

The Chi-square test fails to detect a number of defects, these are;

a. The Chi-square test could be satisfied but the underlying distribution assumptions would not be satisfied. This can occur if there is large deviation balanced out by a large number of small deviations.

b. The Chi-square test can fail to detect consistent bias if it is small, this can occur if there are cumulative deviations over age groups. This bias limitation could lead to the graduation being either biased above or below the data by a small amount.

c. If there are several positive or negative deviations grouped together, this would lead to bias either up or down over those particular age groups. This bias limitation is known as runs or clumps in our test.

d. As the Chi-square test is based on squared deviations, we cannot tell the direction of the bias or the nature of the lack of adherence.

If defects a and b are undetected, they could have serious financial implications (IFE, 2008). If the mortality curve is higher than the standard tables, then an insurance company issuing pensions and annuity contracts would be at financial risk. Insurance companies are susceptible to loss if lives live longer than expected, so they must not overestimate mortality. The opposite applies to life assurance contracts. Insurance companies are vulnerable to making a loss if the mortality curve is underestimated. Insurance companies make a loss from early death, as the premiums charged would be insufficient to cover the cost of the benefit to be paid.

Test (2) Individual Standardised Deviations Test.

Test (2) concentrates on the distribution of the standard deviation values, z_x, and compares the actual and expected z_x values, it tests for normality, and values should resemble a normal distribution. The test is conducted by using another Chi-square test, this time using the z_x values. The test assumes that the deaths are normally distributed and deaths at successive ages are independent.

Test (2) addresses defect (a) of the limitations of the Chi-square test. Test (2) detects the presence of large deviations, which could be offset by a lot of small deviations. Test 2 can detect overall goodness of fit. This problem could be due to over/under graduation of the rates.

The hypothesis for the test is given as:

- H_0: There is no difference between the graduated rates and the true underlying crude rates.
- H_a: There is a difference between the graduated rates and the true underlying crude rates.

The assumption of the individual standardised deviations test is that the normal approximation provides a good approximation at all ages.

$$D_x \sim Normal(E_x^c \overset{\circ}{\mu}_{x+0.5}, E_x^c \overset{\circ}{\mu}_{x+0.5}) \tag{20}$$

And the standard deviations, z_x are calculated as before,

$$z_x = \frac{d_x - E_x^c \overset{\circ}{\mu}_{x+0.5}}{\sqrt{E_x^c \overset{\circ}{\mu}_{x+0.5}}} \tag{21}$$

The z_x values were calculated for the Chi-square test in Excel. The initial check on the data is that no more than five percent of the absolute values are greater than two in absolute size.

The standard normal distribution, is divided into subintervals of (-∞,-3), (-3,-2), (-2,-1), (-1,0), (0,1), (1,2), (2,3) and (3, ∞), the expected distribution of observations in each subinterval are as follows; 0%, 2%, 14%, 34%, 34%, 14%, 2% and 0% (see **Figure 3.3**).

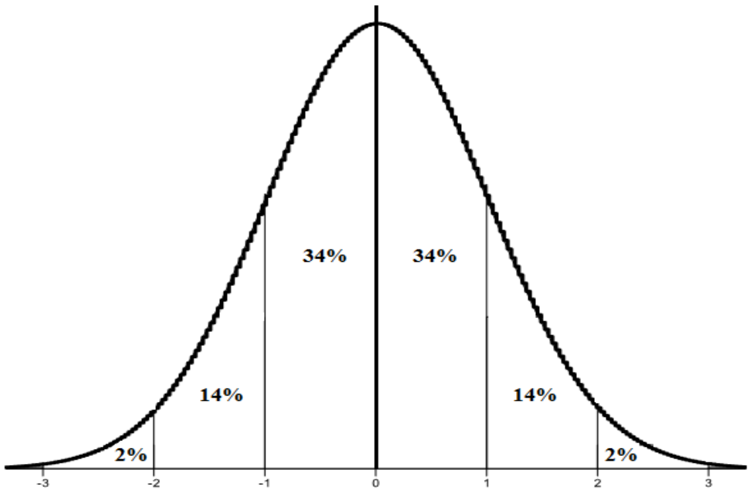

Figure 3.3: Expected number of observations in each Interval based on a normal distribution.

The expected values in each interval are produced by multiplying the degree of freedom for the test by the distribution value for that interval.

The number of expected z_x values and the number of observed z_x values are compared. The z_x are grouped on the intervals described above. A Chi-square test is performed on the distribution of expected and observed values in each interval with a critical value, χ as per (17).

The degrees of freedom for this test is the number of groups (n, in this case equal to 8) minus one, the degrees of freedom is equal to seven. The subintervals -2 and 2 can be combined with the subintervals next to them so that the expectations are not less than one (Cochran, 1954).

The null hypothesis is rejected if,

$$\chi > \chi^2_{n-1,0.95} \tag{22}$$

Visually inspecting the features of the distribution of standard deviations is another method of examining the z_x values. In particular, if the number of age groups are very low, then this method is quite effective as it is easy to review. If there was a large number of groups, then it would take a considerable amount of time and increase risk of error. There are four particular aspects of the normal distribution to look at.

- The observed values and the expected values should be very similar for each interval.

- The presence of outliers, i.e. z_x values lying outside the range (-3, 3) should not be present in our test. Ideally, all z_x values should be in (-3, 3) to coincide with the normal distribution.

- In this test, we want at least half the z_x values to lie in the interval $\left(\frac{-2}{3}, \frac{2}{3}\right)$. This is because the probability of a unit normal random variable falling in this range is almost a half. It follows that approximately half of the standardised deviations are less than $\left(\frac{2}{3}\right)$ and half greater. Thus

  ○ $$Approx \frac{1}{2}\left|\frac{d_x - E_x^c \overset{\circ}{\mu}_{x+0.5}}{\sqrt{E_x^c \overset{\circ}{\mu}_{x+0.5}}}\right| < \frac{2}{3} \tag{23}$$

- The presence of values in the tail of the distribution is due to large values of the z_x's. If this is the case, there is a problem of overgraduation or the existence of duplicates. On the other hand, if there is a lot of z_x values in the interval $\left(\frac{-2}{3}; \frac{2}{3}\right)$ this could indicate undergraduation or heterogeneity.

In this case, we want there to be an equal number of positive standard deviations as there is negative to coincide with the symmetric nature of the normal distribution. A positive bias is seen in the graduation if there is a higher number of positive z_x. values compared to negative. Thus, we say the graduated values are too low. On the other hand, if we have a lot of negative z_x values, the graduation will have a negative bias and thus the graduated rates are too high.

If the z_x are not consistent with the normal distribution, we can conclude that the observed mortality rates do not coincide with the model the rates assumed in the graduation. Inspection of the four features of the normal distribution will locate and identify the discrepancy. A strength of the individual standardised deviations test is being able to look at the distribution of the standard deviation and making an immediate assessment of the graduation.

Test 3 Cumulative Deviation Test

The purpose of the cumulative deviation test is to detect defect C of the limitations of the Chi-square test.

H_o: There is no difference between the graduated rates and the true underlying crude rates.

H_a: There is a difference between the graduated rates and the true underlying crude rates.

This test can be applied to the whole age range or just part of the age range. In this instance, age range is the overall number of age groups, so 25 to 90 is the age range while an age group refers to a number of observations for a given age. Age ranges of financial importance should be tested for discrepancies, this approach could be taken to investigate ranges which pose the greatest risk to insurers. For example, an insurer may have a particular large number of policies for certain age ranges, (e.g. people aged between 30 and 45 buying mortgage protection insurance) so more analysis could be conducted to ensure the rates are accurate, pricing reflects the true rate and reduces risk for insurers. The test also tests for overall goodness of fit.

The cumulative deviation test checks to see that the total number of deaths is consistent to the model with mortality rates assumed in the graduation.

The deviations are calculated by

$$z_x = d_x - E_x^c \overset{\circ}{\mu}_{x+0.5} \tag{24}$$

And they have the approximate distribution,

$$z_x \sim Normal(0, E_x^c \overset{\circ}{\mu}_{x+0.5}) \tag{25}$$

Thus, if we were to consider all age ranges, the deviations and their approximate distribution will be as follows

$$\sum_{All\ \text{Ages}} \left(d_x - E_x^c \overset{\circ}{\mu}_{x+0.5} \right) \sim Normal \left(0, \sum_{All\ \text{Ages}} E_x^c \overset{\circ}{\mu}_{x+0.5} \right) \tag{26}$$

Once we have our approximate normal distribution, we standardise it, and we get,

$$\frac{\sum_{All\ Ages} \left(d_x - E_x^c \overset{\circ}{\mu}_{x+0.5} \right)}{\sqrt{\sum_{All\ Ages} E_x^c \overset{\circ}{\mu}_{x+0.5}}} \sim Normal(0,1) \tag{27}$$

To examine the number of positive and negative runs, a two-tailed test is completed. The assumption of the cumulative deviation test is that the normal distribution is a good approximation of all ages.

Using our prepared data, we must decide what age groups we want to analyse. In this study, the deviations over all the age groups are examined. The sum of the deviations summed over all age groups is calculated and divided by the sum of the expected number of deaths summed over all age groups. The test statistic, T, is outlined below

$$T = \frac{\sum_{\substack{All \\ Ages}} \left(d_x - E_x^c \overset{\circ}{\mu}_{x+0.5} \right)}{\sqrt{\sum_{\substack{All \\ Ages}} E_x^c \overset{\circ}{\mu}_{x+0.5}}} \tag{28}$$

Using the test statistic, T, and standard normal tables, the probability value is calculated at the 5% level.

Ideally, the magnitude of the test statistic will to be close to zero. If the magnitude of the test statistic is very high, then the graduated values will be biased. The direction of the bias can be concluded by the sign of the test statistic. Thus, the graduated rates will be too low if the test statistic is positive, however the rates will be too high if the test statistic is negative.

Test (4) Signs Test

The purpose of the signs test is to detect defect B of the limitations of the Chi-square test. Test 4 states, the Chi-square test can fail to detect consistent bias if it is small. This could lead to the graduation being either biased above or below the data by a small amount.

H_o: There is no significant difference between the graduated rates and the underlying crude rates at each age range.
H_a: There is a significant difference between the graduated rates and the underlying crude rates at each age range.

Test 4 is a simple test of overall bias and will identify if there is an excess of positive or negative deviation values, which would cause the graduation rates to be too low or too high, respectively (IFE, 2008; Benjamin & Pollard, 1993). If the graduated rates are not biased, then we would expect roughly half to be above the crude rates and the other half to be below. Under the null hypothesis, there is no difference in the mortality rates in the standard tables and graduated rates. The difference in the expected and observed deaths are independent normal random variables. Thus, the signs of each individual deviation are independent and are therefore equally likely to be either positive or negative. Due to this fact, we apply a two-tailed test, and we say they follow a binomial distribution.

Suppose we have m age groups, the number of values above the crude rates should have $B\left(m, \frac{1}{2}\right)$ distribution. A high or low number of positive or negative deviations will indicate a bias in the rates.

Using our prepared data, the first step is to examine the signs of the z_x values. The number of positive and negative values are counted, this is the same as counting the number of graduated rates that are above/below the crude rates.

The number of positive or negative deviations is calculated. P is the test statistic where P is equal to the number of positive z_x values; likewise, it could be set equal to the negative z_x values. Thus

$$P \sim \left(m, \frac{1}{2} \right) \tag{29}$$

with a probability function,

$$P(P = x) = \binom{m}{x} \left(\frac{1}{2} \right)^m, x = 0,1,...,m \tag{30}$$

This is a two-tailed test, the test will be rejected if there is an excess of either positive or negative values. The binomial distribution is a discrete distribution, so we must choose a value k*, which is the lowest k value that satisfies the following test T, where T is equal to

$$\sum_{j=0}^{k} \binom{m}{x} \left(\frac{1}{2} \right)^m \geq 0.025 \tag{31}$$

This test will be accepted at a 5% significance level if k*≤P≤m-k*. Use of p-values can also be used. The p-values for the binomial distribution are tabulated in the actuarial tables. This table is only useful if the number of groups in our study are from one to 12 or 20 otherwise, we must use a normal approximation.

$$P \sim Normal \left(\frac{m}{2}, \frac{m}{4} \right) \tag{32}$$

Where m is the number of age groups in our study. Since the binomial distribution is a discrete distribution and we want to approximate it using the normal distribution, we must apply a continuity correction to our calculation. This is because the normal distribution is a continuous distribution.

The null hypothesis is rejected if the test statistic falls inside the lower 0.025% or upper 0.975% regions of the unit normal distribution.

If there is an excess of positive or negative z_x values, this leads us to believe that the graduated rates are too low or too high, respectively. Investigating the position of the signs is a good indication of what ages have the worst effect on the graduation.

Test (5) Grouping of Signs Test

The purpose of this test is to detect overgraduation, looking for excessive runs or "clumping" of deviations of the same sign (IFE, 2008; Benjamin & Pollard, 1993).

Test (5) looks at the actual number of same signs groups experienced in the age groups and compares to the number of expected, if they were randomly generated. If the graduation process has been successful then there should be neither too many nor too few runs of the same sign. If the effect positive runs (i.e. overgraduation) is only examined, this test is a one sided test.

H_o: There is no difference between the graduated rates and the true underlying crude rates.

H_a: There is a difference between the graduated rates and the true underlying crude rates.

Let G = Number of groups of positive standardised deviations

Suppose there are l age groups, this means there are l deviations. The n_1 deviations are positive and n_2 deviations are negative $n_1 + n_2 = l$. positive runs versus n1 positive deviations and n2 negative deviations randomly placed. Thus, we want to compute the probability of positive groups is at least G given n1 and n2.

To do so let $t \leq G$.

The number of ways of arranging $t \leq G$ positive groups among n_2 negative signs is

$$\binom{n_2+1}{t}$$

To arrange the n_1 positive signs into t positive groups is

$$\binom{n_1-1}{t-1}$$

Finally, we have $\binom{m}{n_1}$ ways of placing n1positive and n2 negative signs. Thus, the probability of exactly t positive groups is calculated by

$$\frac{\binom{n_1-1}{t-1}\binom{n_2+1}{t}}{\binom{m}{n_1}}$$

When performing this test on the dataset, we must first look at the sign of the deviations at each age. We then count the number of groups of positive runs. This value is equal to G. Every pair of positive groups must be interrupted by a negative value\group. Thus, the number of positive or negative groups must be small or large. A one-tailed test is sufficient in this case. We want the smallest k such that

$$\sum_{t=1}^{k}\frac{\binom{n_2+1}{t}\binom{n_1-1}{t-1}}{\binom{n_1+n_2}{n_1}} \geq 0.05$$

The null hypothesis is rejected at a 5% significance level, if G < k.

To approximate the Stevens test, the mean is can be calculated

$$n_1 \cdot \left(\frac{n_1+1}{n_1-n_2}\right)$$

The variance can be calculated

$$\frac{(n_1 n_2)^2}{(n_1+n_2)^3}$$

An approximation of the Steven's test can be approximated by

$$\frac{G\ -\ mean}{variance}$$

The test is accepted at the 5% significance level.

If m (total number of deviations) is large, i.e. m>20, then we can use a normal approximation, based on the results by Pollard (1977).

$$G \sim N\left(\frac{n_1(n_2+1)}{n_1+n_2}, \frac{(n_1 n_2)^2}{(n_1+n_2)^3}\right)$$

We must apply a continuity correction to the above. The continuity correction is applied as a normal approximation is being applied to a binomial distribution. It is tested at a 5% significance level using the standard normal tables.

If the test is rejected, this indicates that there are too few runs, and the rates are overgraduated. The rates do not adhere closely enough to the crude data and may be consistently too high or too low over certain parts of the table.

3.2.8 Smoothness

Smoothness can be defined in terms of differentiability, and it can be approximated using the following equations (IFE, 2008).

$$\text{The first difference: } \Delta \overset{o}{\mu}_x = \overset{o}{\mu}_{x+1} - \overset{o}{\mu}_x$$

$$\text{The second difference: } \Delta^2 \overset{o}{\mu}_x = \Delta\overset{o}{\mu}_{x+1} - \Delta\overset{o}{\mu}_x$$

$$\text{The third difference: } \Delta^3 \overset{o}{\mu}_x = \Delta^2\overset{o}{\mu}_{x+1} - \Delta^2 \overset{o}{\mu}_x$$

The implied smoothness criterion in the case of mortality graduation is that the third differences of the graduated quantities are:

- Small in magnitude compared with the quantities themselves; and
- Should progress regularly.

A check of the impact of smoothness was performed between the three different rates. These can be seen in the Results section.

3.3 Results

The results in this section were carried out using Microsoft Excel and R (a statistical computer package). We begin by looking at how we calculated our crude rates and then how, using a simple link function, we produced a set of graduated rates. Then, we look at the five statistical tests.

3.3.1 Crude Mortality Rates

Using the data available, a set of crude mortality rates estimates. These crude rates were calculated in Microsoft Excel, with the following formula, $\hat{\mu} = \frac{d_x}{E_x^c}$.

A sample of these rates can be seen in **Table 3.1.** and a visual plot of the crude rates are provided in **Figure 3.4.**

Table 3.1 : Crude Mortality Rates for Ages 25 to 29 from Sample Data.

Age x	Dx	E_x^c	$\hat{\mu}_x$
25	0	1,906	0
26	0	2,480	0
27	1	7,625	0.00013
28	2	7,980	0.00025
29	1	7,995	0.00013

The full set of crude rates for our study can be seen in **Appendix 3.A**. The lives in our study span from 25 to 90 years old.

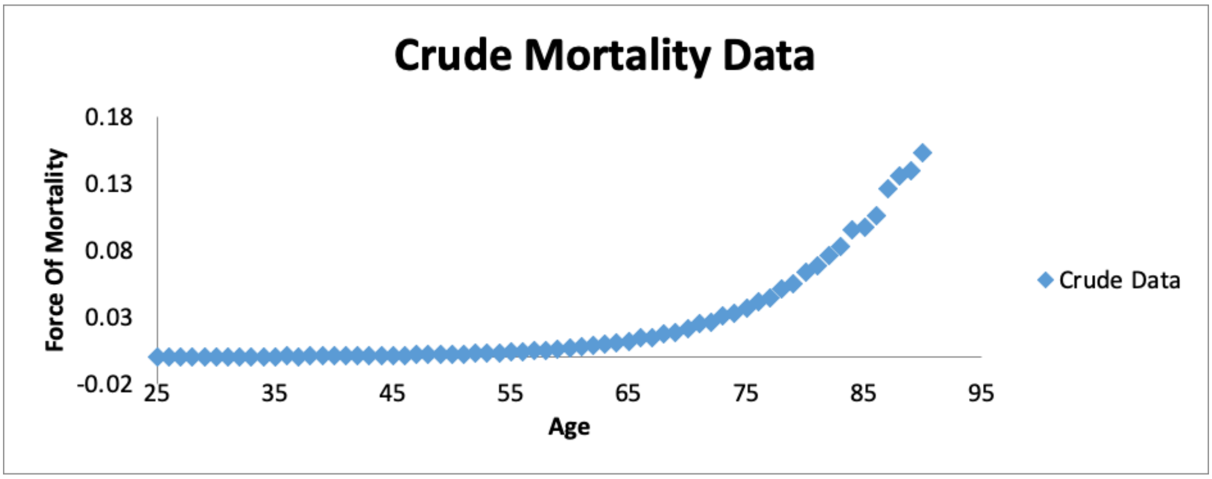

Figure 3.4: Crude Mortality Rates for the Population in this Study.

The mortality pattern of the crude rates demonstrated an exponential increase with age. In general, this is expected as mortality increases with age. However, the overall rates can be deemed rough, for example at some age ranges the value is 0, this is because there were no deaths observed in that age category. This can be seen for ages 25 and 26. The shape of the crude rates is very similar to the shape of a stand table. A plot of the shape of the standard table is provided is **Appendix 3.C**.

The logarithmic view of the crude rates in **Figure 3.5** shows a strong linear relationship for the age structures. This allowed a simple linear link function to be used.

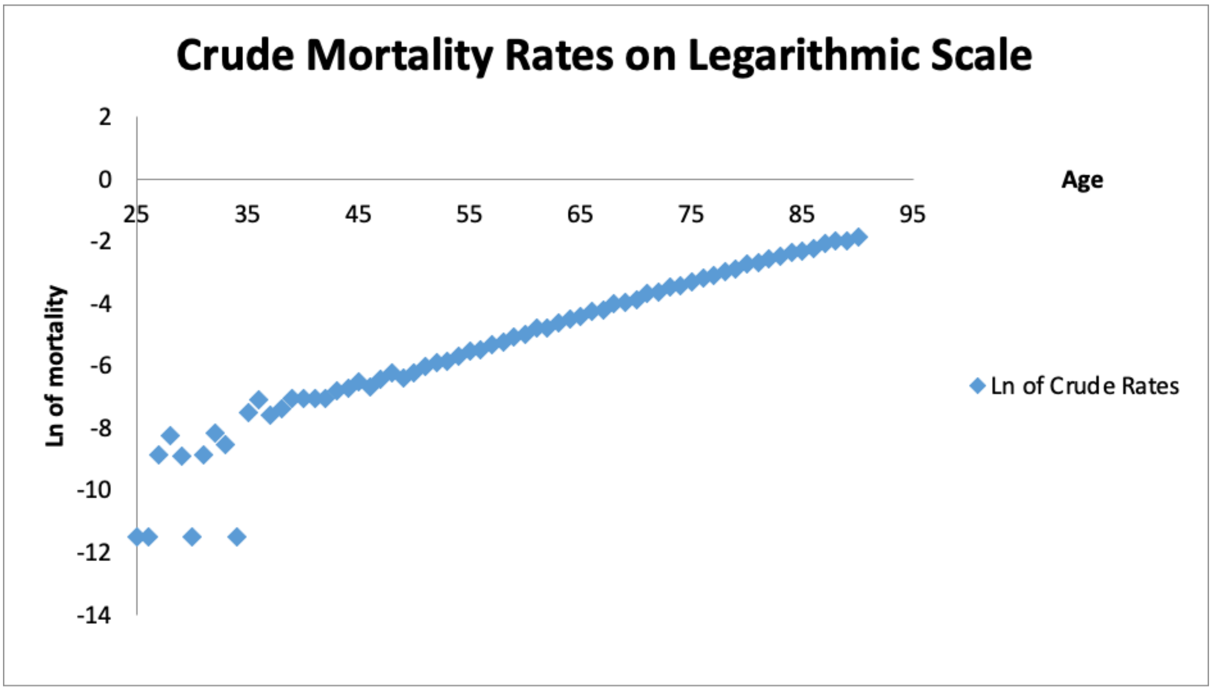

Figure 3.5: Crude Mortality Rates for the Population in this Study on Logarithmic Scale.

The link function used is

$$a+b\mu_{x+0.5}$$

The unknown parameters a and b were calculated in Microsoft Excel based on the crude rates and the standard table rates. Using Excel, the optimal value of a and b were calculated. Estimates for a and b are in **Table 3.2**.

Table 3.2 : Unknown Parameter Estimates Based on Least Squares.

Parameters	**a**	**b**
Estimates	-6.7E-05	0.86

Substituting these values into our simple link function, a set of graduated rates are created.

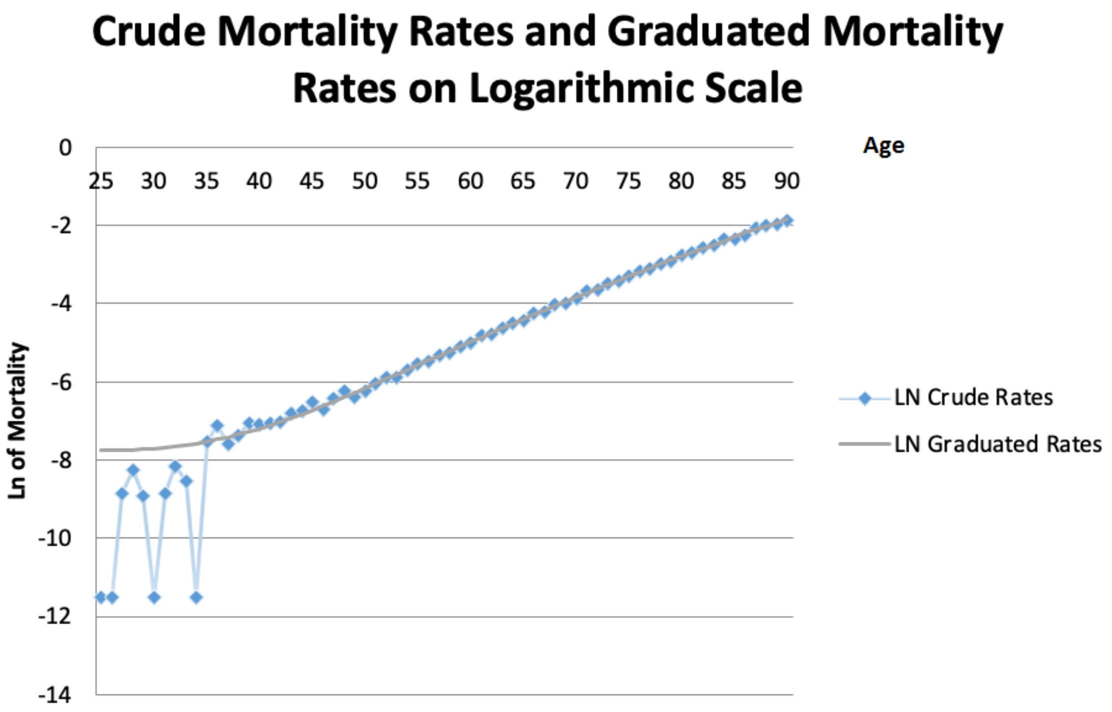

Figure 3.6: Crude Mortality Rates and Graduated Rates on Logarithmic Scale for ages 25 to 90.

Table 3.3 shows the crude rates, graduated and standard table morality rates for ages 25 to 29 from data.

Table 3.3 : Crude, Graduated and Standard Table Rates for Mortality 25 to 29.

Age x	dx	E_x^c	$\hat{\mu}_x$	$\mathring{\mu}_x$	$\overset{s}{\mu}_x$
25	0	1,906	0	0.00061	0.00063
26	0	2,480	0	0.00063	0.00065
27	1	7,625	0.00013	0.00065	0.00067
28	2	7,980	0.00025	0.00067	0.00071
29	1	7,995	0.00013	0.00071	0.00074

A snapshot of the values for each of the crude rates, graduated and standardised rates from 25 to 29. The graduated and standard rates are non-zero and increasing year on year. The crude rates have value 0, this reflects no deaths in that year. Having a 0 value is unacceptable, as the probability of death is always above 0. Mortality rates can decrease but this is seen in younger lives; from 25 upwards there should be a year-on-year increase. The full set of rates from 25 to 90 can be seen in **Appendix 3.A**. A graph of all the values of the mortality can be seen in **Figure 3.7**.

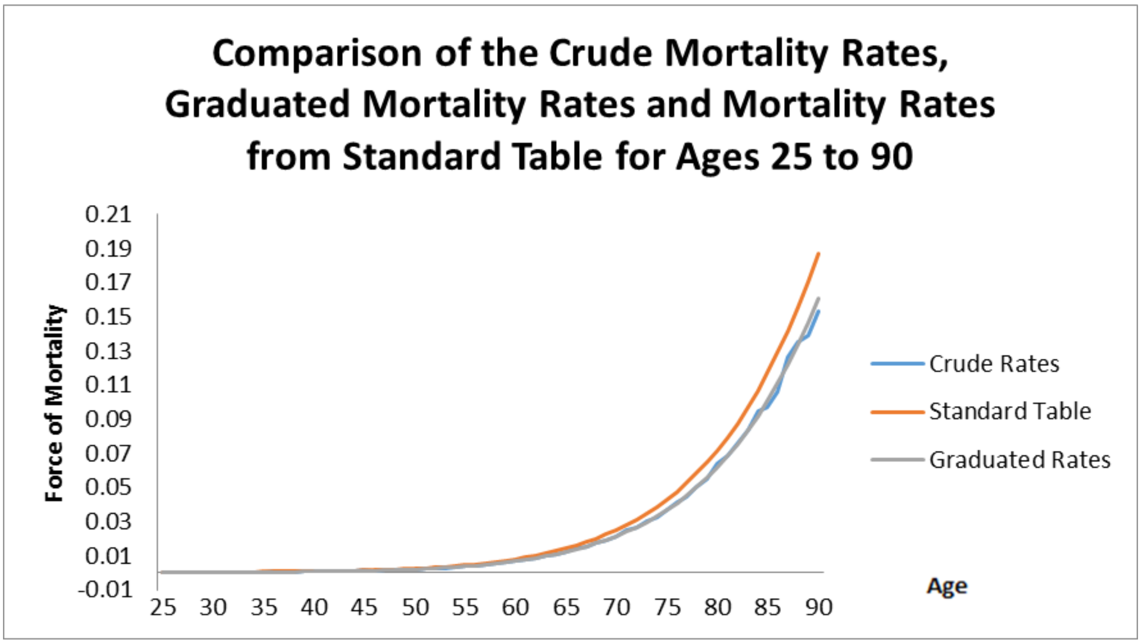

Figure 3.7: Comparison of crude rates, standard table rates and graduated rates for the force of mortality for ages 25 to 90.

Figure 3.7 shows that the curve generally increases as age increases. It is hard to see the benefits of the graduated rates. As seen in **Table 3.3**, the crude rates increase and decrease as age increases, a further example of this can be seen for a life aged 36 and then 37, the crude rates decrease from 0.00081 to 0.00049, while both the standard and graduated both increase, given the age increases. **Figure 3.8** gives a visual representation of the differences in the rates from age 25 to 50.

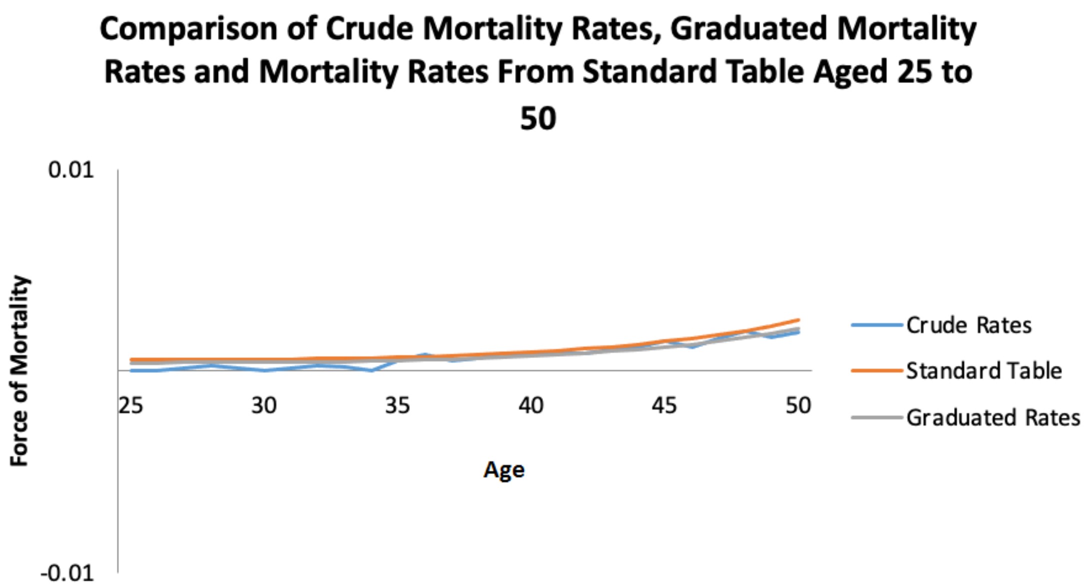

Figure 3.8: Comparison of crude rates, standard table rates and graduated rates for the force of mortality for ages 25 to 50.

To examine the difference further, the log normal rates were calculated, this will be referred to the "Ln rates" for the remainder of the chapter. A sample list of the values can be seen in **Table 3.4**.

Table 3.4: The ln of Crude, Graduated and Standardised Force of Mortality Rates from ages 25 to 29.

Age x	dx	E_x^c	Ln $\hat{\mu}_x$	Ln $\overset{\circ}{\mu}_x$	Ln $\overset{s}{\mu}_x$
25	0	1906	-11.5129	-7.77526	-7.47692
26	0	2480	-11.5129	-7.77526	-7.47515
27	1	7625	-8.87387	-7.77288	-7.46987
28	2	7980	-8.25483	-7.76578	-7.46114
29	1	7995	-8.87387	-7.75873	-7.45076

Table 3.5 Ln of Crude, standardised and graduated morality rates for ages 25 to 29 from data.

Given the Ln 0 does not exist for any age with 0, the value 0.00001 was added to 0 and the Ln of 0.00001 was used. A full list of rates can be found in **Appendix 3B**. **Figure 3.9** shows a comparison of the Ln of the different force of mortality rates.

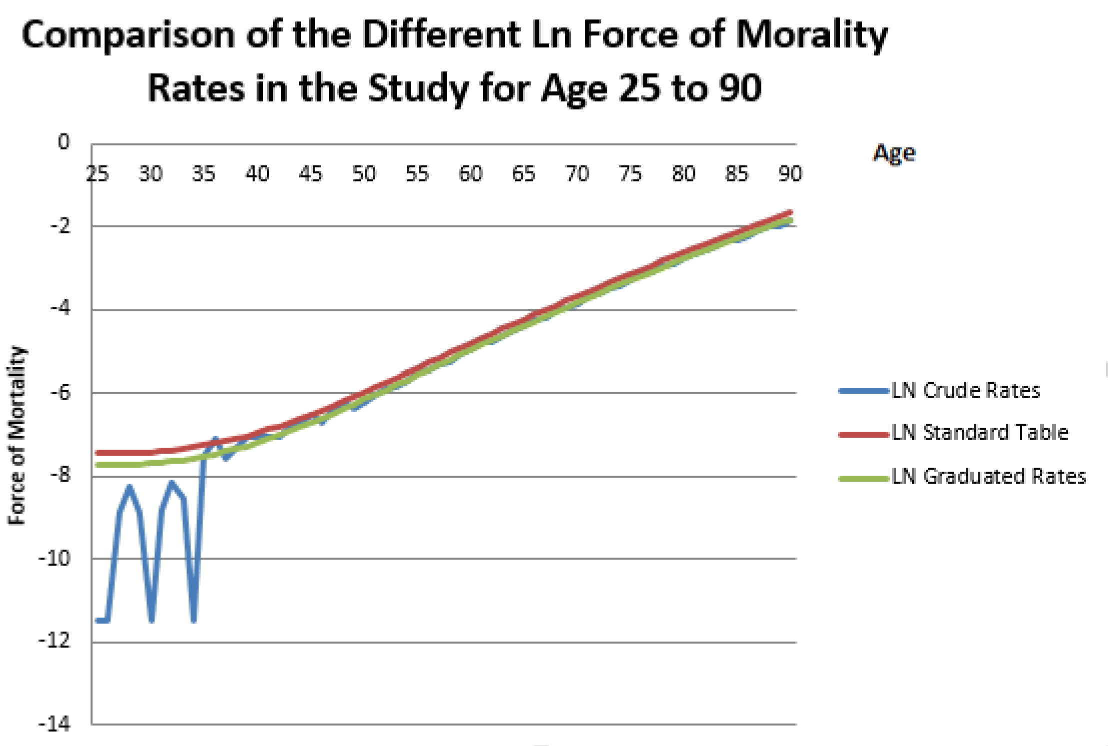

Figure 3.9: Comparison of the Ln of crude rates, Ln of the standard table rates and Ln of the graduated rates for the force of mortality for ages 25 to 90.

Figure 3.9 shows the graphs of each set of rates. The crude rates are much more volatile than the other values, in particular at the younger ages as no deaths can be recorded for an age range. This occurs as the probability of death is low for younger individuals; in our study we have ages where no deaths were recorded. With the general population or bigger data sets, death will occur in each age range. All three graphs increase with age, as would be expected, but the crude rates jump. This is due to the data not having any deaths.

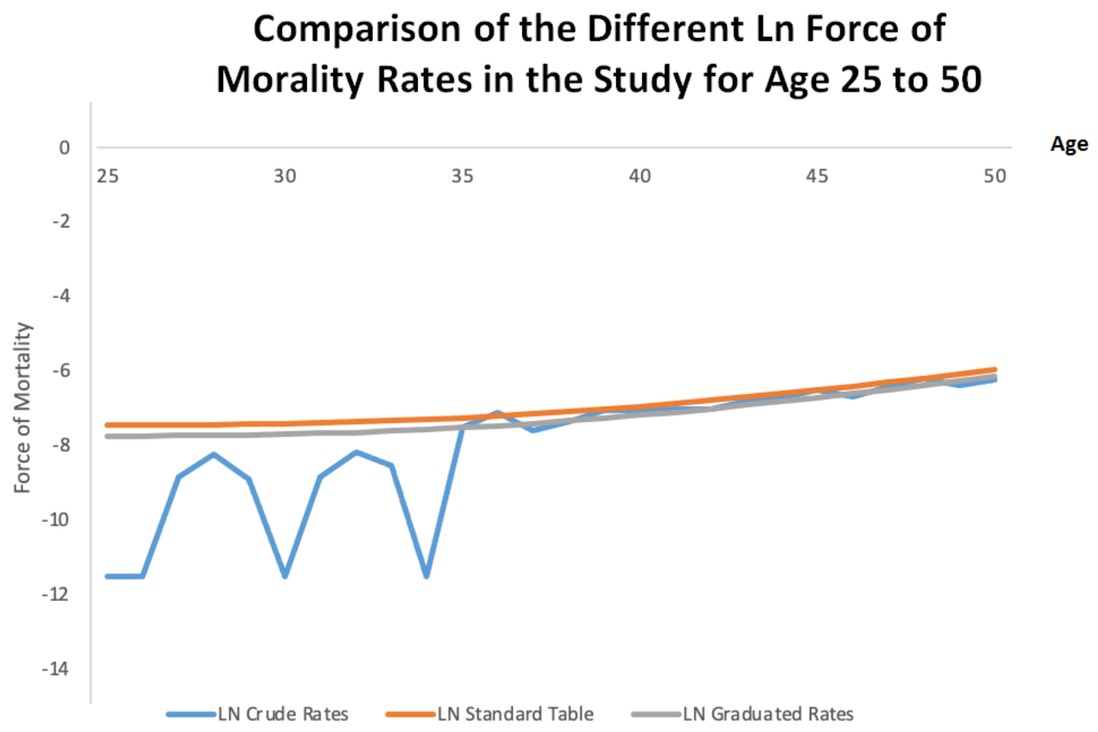

Figure 3.10: Comparison of the Ln of crude rates, Ln of the standard table rates and Ln of the graduated rates for the force of mortality for ages 25 to 50.

The graduated rates are similar to the standard table rates and are less volatile. The standard and graduated graphs are much smoother than the crude graph. A test of smoothness was conducted by looking at the third difference in rates. This is where:

$$\text{The third difference: } \Delta^3 \overset{o}{\mu}_x = \Delta^2 \overset{o}{\mu}_{x+1} - \Delta^2 \overset{o}{\mu}_x$$

For smoothness, these values should be very low, compared to the values themselves.

Table 3.5 gives the 3rd difference for the age range of 25 to 29; a full list of values can be seen in **Appendix C**. ∆25 represents the change from 25 to 26, ∆26 represents the changes from 26 to 27 and so on for each age range. A value is deemed smoother based on the magnitude of the value, lower and negative are more desirable.

Table 3.5: Difference in Crude, Graduated and Standardised Force of Mortality 25 to 29.

Δ **x**	$\hat{\mu}_{x+1} - \hat{\mu}_x$	$\overset{\circ}{\mu}_{x+1} - \overset{\circ}{\mu}_x$	$\overset{s}{\mu}_{x+1} - \overset{s}{\mu}_x$
Δ **25**	-0.00014	-4.3E-07	-5E-07
Δ **26**	-0.00023	4.3E-07	5E-07
Δ **27**	0.00025	8.6E-07	1E-06
Δ **28**	0.00026	-4.3E-07	-5E-07
Δ **29**	-0.00025	4.3E-07	5E-07

Table 3.5: Third order difference values for force of mortality for crude, graduated and standard rates. The smaller the value compared to the actual value the smoother it is.

From **Table 3.5**, the values for the crude rates are much larger than the standard and graduated values. The smaller the third difference values is compared to the actual value, the smoother it is. This is further illustrated in **Figure 3.10**. A full list of values for the ages 25 to 90 are in **Appendix 3.D**.

3.3.2 Statistical Tests

In this section, we concentrate on the results of the five statistical tests. We look at the Chi-square test, cumulative deviations test, the signs test, the grouping of signs test, and the individual standardised deviations test.

Test (1) Chi-Square Test

The results of test (1), the Chi-square are reported in **Table 3.6**. Test (1) was performed to assess overall goodness of fitness.

Table 3.6: Results of Test (1) the Chi-Square Test.

Groups	**66**
Parameters	2
DF	64
Critical value	83.68
Test statistic	72.89

Test (1) is accepted at 5% level of significance which would suggested there was an overall goodness of fit. Further tests were conducted to examine any limitations which the Chi-square test does not pick up.

Test (2) Individual Standard Deviations Test

This individual standard deviations test detects the presence of large deviations which could be offset by a lot of small deviations. The critical value is based on the normal tables with m degrees of freedom at a 5% significance level.

Table 3.7: Actual and Expected Values from Visual Inspection.

	(-inf,-3)	(-3,-2)	(-2,-1)	(-1,0)	(0,1)	(1,2)	(2,3)	(3,inf)
Actual z_x	0	1	9	16	24	15	1	0
expected	0	1.32	9.24	22.44	22.44	9.24	1.32	0
(A-E)^2/E	0	0.078	0.006	1.848	0.108	3.591	0.078	0

The degrees of freedom in this case is 7. There are 8 intervals/groups in this test. The test statistic, χ = 5.71. the test statistic is the summation of the final row of **table 3.7**.

The critical value is equal to 14.07. The test passes at the 5% significance level. This would suggest an overall goodness of fit

Overall Shape and Visual Inspection

Visually inspecting the features of the distribution of standardised deviations can give us a good indication about the z_x values. Using **Table 3.7**, we can compare our actual distribution of z_x values versus the expected ones. Beginning at the extremes, $\pm(inf, 3)$ it is good to see no values in either, as the expected is zero. Looking at the other cells, the values are relatively close. The largest deviations in the actual and expected occur at (-1, 0) and (1, 2). With the large difference in the (1, 2) range, this would suggest higher number of positive values in this range. This could occur die to higher level of deaths occurring at different age ranges.

Outliers

There are no values falling in the extremes, $\pm(inf, 3)$. Ideally, values in the extremes are not desired, as they are highly unlikely to occur and will skew the data. The data is contained in the intervals contained in (-3,3), this is the desired distribution. This is presented in **Table 3.7**.

Absolute Deviation

Absolute deviation examines the number of z_x values, found in the interval $(-2/3, 2/3)$. Ideally, approximately half the values will fall into this interval, this would be 33 in this study.

Table 3.8: The number of z_x Values Inside and Outside (-2/3,2/3).

Age Groups	**66**
z_x values within the interval (-2/3,2/3)	20
z_x values outside the interval (-2/3,2/3)	46

There are 46 values outside the interval, with 20 values inside the interval as per **Table 3.8**. This can occur due to large values of z_x values in the data. Large number of z_x values in an interval can be due to rates being overgraduated or the presence of duplicates.

Symmetry

The desired result is to have an equal number of positive and negative values. In this study, there are 40 positive values compared to 26 negative values. Ideally, we would want these values to be closer to the distribution of results, which is similar in shape to the normal distribution. There are more positive results that would suggest that the graduated rates are a bit low. **Figure 3.10** also shows the graduated results to be lower than the standard tables.

Test (3) Cumulative Deviation Test

Test (3) is to verify the total number of deaths is consistent with the model given the mortality rates. Test (3) wants to verify the fit of the graduated rates to the original crude rates.

Table 3.9: Results of Cumulative Deviations Test.

N(0,1), 95%	**1.96**
Test Statistic mux	0.002

The desired outcome is to be as close to zero as possible, which is the case here. The test statistic is less than the critical value, which suggests that the test does not provide evidence that the graduated rates are biased. That is the number of deaths experienced is consistent with the model and we cannot reject the null hypothesis.

Test (4) Signs Test

This test was carried out to determine the overall bias in the rates. The Chi-square test again does not detect this. The number of groups in our study is larger than 20, so we used the normal approximation of the binomial. The test statistic in this case was, T = -1. 60 with an associated p-value of 3.22%, which is greater than the two tailed 5% level of significance of the standard normal distribution. We have no reason to reject the null hypothesis, there is no significant difference between graduated rates and underlying crude rates.

Test (5) Grouping of Signs Test

The presence of overgraduation in the rates was tested using the grouping of signs test. We had over 20 age groups, so a normal approximation and continuity correction were used to determine the test statistic. The test statistic G = 1.10, which is inside our 1.96 threshold and a p-value of 13.57%, the null hypothesis is accepted. The graduated rates are not overgraduated and there is a sufficient amount of runs in the deviations.

3.4 Discussion

In this section, the results, observations, and inferences are discussed. The limitations of the study and future recommendations are also addressed.

3.4.1 Crude and Graduated Rates

The crude rates produced were positive sloping with age but had gaps at certain ages. The gaps occurred due to no deaths being observed at those ages. Gaps in the rate due to missing or lack of data is one of the reasons for graduating the rates. As a general principle, mortality rates increase year-on-year. This principle does not hold for mortality rates around age range 0 to 3 and 20 to 25, this is due to high infancy mortality rates and higher mortality rates for males in ages between 20 and 25, this phenomenon is due to higher mortality rates usually linked to accidents (McCormack, 2016). These irregularities are not seen in this study as our data set is from ages 25 to 90. The increases in mortality can be difficult to see year on year. Viewing the logarithmic values of the crude rates provides a better visualisation of the relationship of mortality and age structures. **Figure 3.6** shows a very definite linear relationship between mortality and age.

Figure 3.7 and **3.9** plots of the crude rates, the standard and logarithmic view have a similar shape to the crude rate plots produced in other studies. When both plots are compared to the plots constructed by Pacáková et al, (2013) or Ramonat and Kaufhold (2018), the shape and direction are similar. This indicates that mortality increases as age increases. On visual inspection of the plot of logarithmic crude rates in this study compared to the plots in the study by McCormack, the slopes look very similar, which would suggest the mortality experience is similar. The slope for their study was not recorded but the slope for this study was 0.86. However, the study by Ramonat and Kaufold (2018), which investigated lives from 50 to 100 had a slope of 0.11. The slopes are not comparable given the different age ranges they are based on. A slope of 0.11 would indicate a steeper incline which would suggest a much higher level of mortality rates and an increase in mortality. This is to be expected at higher age ranges. The shape and direction were similar but the level of mortality for their study was much higher given the age range. The difference in shape of the mortality is not just down to age, other factors are also impacting, for example the product and lives are different, the analysis Ramonat and Kaufold (2018) was conducted on Canadian lives as part of a pension study, this study used data for lives with a term assurance contract.

The link function used in this study was a linear link function. This function was used due to the relationship between mortality and age. As seen in **Figure 3.6**, there is a clear linear relationship. The shape of the rates also had a linear relationship with the standard tables. This function is a simple function to apply to these studies. A linear link function was also used in another study by Ramonat and Kaufold (2018). Ramonat and Kaufold (2018) were promoting the use of this function given its ease to implement and the reliable results.

The graduated results produced were smoother than the crude rates but less smooth than the standard table. There were no sudden drops or spikes in the rates, which were present in the crude rates due to ages with no recorded deaths. This adheres to the principle of increasing mortality rates with age, besides the exceptions stated earlier. There is a large difference between the crude rates and standard table, especially between the age range 25 to 35 years, **Figure 3.10**. If those crude rates were used to price a life insurance contract, the mortality cost would be too low for the mortality risk the company would be taking on. Low premiums due to low mortality assumptions could lead to large volumes of business being sold and miss-priced, if the product is not priced correctly it could have negative financial implications for the company.

Between the ages of 35 and 40 the crude rates bounce from above the standard rate to below the graduated rate. At age 35 the crude rates are higher than both the graduated and standard rates. The crude rates being higher could reflect a particularly bad year of deaths for that age group. It may not be appropriate to apply the full crude rate at this point as this would be too high a mortality assumption based on one set of data. If this rate was applied, the cost of mortality used in pricing an insurance contract could be too high. Instead, further investigation should be conducted, for example ensuring the number of deaths recorded is correct. The graduated rate may be a more appropriate mortality assumption.

The rates begin to converge from 40 onwards, with the shape of the rates very similar. The crude rates are still rough and bumpy, however they are not as volatile as the earlier age ranges. A reduction in the volatility could be due to the number of observed deaths increasing in quantity with the ages. As mortality is low for younger lives the number of deaths is more sensitive to change, however the number of observed deaths increase at higher ages. The magnitude in difference between the rates is reduced. The graduated rates should be appropriate.

Using other tables such as the Irish Life Tables or the English Life Tables would provide different estimates. Life tables are a method of summarising different characteristics, which impact mortality by age and gender. When selecting a life table, it should reflect the current business and considerations should also be made for future improvements.

3.4.2 Statistical Tests

Statistical tests were carried out to validate if the graduated results created are a good fit. The initial test performed was the Chi-square test to look at overall goodness of fit and then additional statistical tests to back up the Chi-square results, as the former is known to have a number of limitations.

The Chi-square test passed, this would which suggest that either the graduated rates produced are a good fit or there could be the presence of overgraduation. Overall, the goodness of fit is good. Tests 3, 4 and 5 all passed outright. Test 2 has a few parts to it, the individual standard deviation test passed, but there was some ambiguity with the normal distribution checks. The visual inspection passed, but the absolute deviation test failed. Failing the absolute deviation test could suggest overgraduation within the rates. The overall view of the Individual Standardised Deviation test and the features of normality is that the rates coincide with the normal model. The studies conducted by Pacakova, Jindrova and Seinrova (2018) and Romonat and Kaufhold, (2018) used the same test to qualify the graduation process and the graduated results.

3.4.3 Limitations and Recommendations

This study used the Poisson model to estimate the number of deaths in the study. There are other distributions models that could be used to model deaths, for example, the binomial or Weibull distribution. Other studies which have used the binomial model as a model for death (Pacakova, Jindrova and Seinrova, 2018).

Complete and accurate data is vital for the analysis and development of reliable crude rates. This data set had age groups that had no recorded deaths. Either this was correct and no deaths were observed, or the data contained errors. As much data as possible should be used, as larger data sets will provide more reliable results. The data ranged for 5 years, which is an appropriate time frame for mortality studies. Future studies should use a similar timeframe; longer time periods could lead to the risk of heterogeneity in the rates. In most societies mortality continues to improve with medical advancements and society becoming more health conscious. The improvement in mortality in Ireland over the last 30 years has increased dramatically with the rate of improvement beginning to slow down as lives reap the benefit of healthier living and better health care (IFA, 2020).

The graduated rates were developed by a simple linear link function and the coefficients calculated with linear least squares. The simple link function used in this study assumes mortality increases at an exponential rate with age. Some age ranges may actually behave differently and a different model could be more appropriate for example, the Makeham model could be used as a link function. The Makeham model can capture mid-mortality ranges better (Romonat and Kaufhold, 2018). Other studies have used time series or B spline process as part of their graduation process (McCormack, 2016).

This study used AM92 tables as the standard table, there are a number of different tables available. The English Life Tables ELT17 could be used, these tables also have male mortality tables. ELT17 tables could have a better relationship with the existing data and provide better graduated results. ELt17 tables were updated in 2018, these tables are up to date and could include allowance for new mortality trends. The final use of the graduated rates should be considered, if the rates are used to price term assurance business then tables that reflect that risk should only be used.

3.5 Conclusion

There are many factors that can impact mortality assumptions when pricing life insurance products, one very important factor is age, however an example of some other factors are product type, smoker status or gender. There are a number of sources of life tables and each has a different purpose. This study created set crude mortality rates based on observations from a term assurance product. A set of crude mortality rates were produced. The shape of the rates were in line with other studies and rates showed mortality increasing with age. The rates were bumpy and year-on-year had erratic jumps. The erratic jumps were due to the size and quality of the data, for example for some ages there were no deaths observed, which we know is not true as deaths occur at all ages.

On inspection of the logarithmic crude rates, there is a clear linear relationship. A simple link function was used for the graduation of the crude rates using a standard table. The linear link function is a simple yet powerful tool when graduating crude rates and is still widely used today in the industry. The shape of mortality can differ between age ranges, with mortality increasing and decreasing in the following age ranges 0 to 5 years and 20 to 25 years, this study did not have that problem.

The graduated rates produced were smooth and removed the erratic behaviour of the crude rates. The graduated rates were a similar shape to rates in other studies. The biggest divergence between all the rates occurs between the ages of 25 to 40 with the rates beginning to converge. One reason for this divergence is the missing death observations at certain ages, however the graduated rates smooth out the blips. This is more appropriate for mortality rates and rates to be used pricing life insurance.

Statistical tests were conducted to assess the goodness of fit of the graduation. The tests conducted were also used in other studies to determine the appropriateness of graduation. All 5 tests passed, which would indicate that there was an overall goodness of fit of the graduation.

3.6 References

Advisor, N. and Advisor, N., 2021. How Smoking Affects Life Insurance Premiums | Lion.ie. [online] Lion.ie. Available at: <https://lion.ie/mortgage-protection-insurance/smokers-life-insurance/>

Barnett, H. A. R. (1951). Graduation Tests and Experiments. Journal of the Institute of Actuaries, 11, 15-74.

Benjamin, B., & Pollard, J. H. (1993). The Analysis of Mortality and Other Actuarial Statistics. London UK: The Faculty and Institute of Actuaries.

Bonkier.ie https://www.bonkers.ie/compare-mortgage-protection-insurance/

Broffit, J. D. (1984). Maximum Likelihood Alternatives to Actuarial Estimators of Mortality Rates. Transactions of Society of Actuaries 36.

Forbes.com. 2021. [online] Available at: <https://www.forbes.com/advisor/life-insurance/smokers/> [Accessed 11 July 2021].

Houston, D. (1959). A Comparison of Mortality Rates of Insured Lives with Those of the General Population. The Journal of Insurance, 26(3), 13-23. doi:10.2307/250334

IFE (2008). Acted Study Materials: 2008 Examinations. Subject CT4.Oxford, UK: Actuarial Education Company.

IFE (2020). Acted Study Materials: 2020 Examinations. Subject ST2.Oxford, UK: Actuarial Education Company.

IFE (2016). Acted Study Materials: 2016 Examinations. Subject SA2.Oxford, UK: Actuarial Education Company.

IFOA https://www.actuaries.org.uk/learn-and-develop/continuous-mortality-investigation

Irish Insured Lives Mortality Investigation https://web.actuaries.ie/sites/default/files/2019-03/190306%20-%20Irish%20Insured%20Lives%20Mortality%20Investigation%20Report%20Final.pdf

London, D., (1985). Graduation: The Revision of Estimates. Abington, CT: ACTEX Publications.

McCormack, K, 2016, Methodological Note Producing the Irish Life Table, No. 16 2010-12

Pacáková, V., Jindrová, P., Seinerová, K., 2013, Mortality models of insured population in the Slovak Republic

Richards, S., Kaufhold, K. and Rosenbusch, S., 2013. Creating portfolio-specific mortality tables: a case study. European Actuarial Journal, 3(2), pp.295-319.

Weisberg, S. (1985). Applied Linear Regression. New York: Wiley.

Appendix 3.A

A list of values for each age interval from 25 to 90.

Table 3.11: Crude, Graduated and Standard Table Rates for Mortality 25 to 90

Age x	**dx**	E_x^c	μ_x	$\overset{\circ}{\mu}_x$	$\overset{s}{\mu}_x$
25	0	1,906	0	0.00042	0.00057
26	0	2,480	0	0.00042	0.00057
27	1	7,625	0.00013	0.00042	0.00057
28	2	7,980	0.00025	0.00043	0.00058
29	1	7,995	0.00013	0.00043	0.00058
30	0	7,065	0	0.00044	0.00059
31	1	7,522	0.00013	0.00045	0.0006
32	3	11,000	0.00027	0.00047	0.00062
33	2	10,770	0.00019	0.00048	0.00064
34	0	10,841	0	0.0005	0.00066
35	7	13,165	0.00053	0.00053	0.00069
36	19	23,499	0.00081	0.00056	0.00073
37	12	24,290	0.00049	0.00059	0.00077
38	20	32,097	0.00062	0.00063	0.00081
39	31	36,372	0.00085	0.00068	0.00087
40	32	38,173	0.00084	0.00074	0.00094
41	38	43,885	0.00087	0.00081	0.00102
42	41	47,059	0.00087	0.00089	0.00111
43	59	54,083	0.00109	0.00098	0.00121
44	65	55,442	0.00117	0.00108	0.00133
45	88	60,042	0.00147	0.0012	0.00147
46	89	73,070	0.00122	0.00133	0.00163
47	121	74,530	0.00162	0.00149	0.00181
48	156	79,224	0.00197	0.00167	0.00201
49	131	78,423	0.00167	0.00187	0.00225
50	156	80,021	0.00195	0.0021	0.00251
51	210	87,307	0.00241	0.00236	0.00282
52	255	92,638	0.00275	0.00266	0.00316
53	231	82,185	0.00281	0.00299	0.00355
54	305	89,784	0.0034	0.00337	0.00399
55	368	92,429	0.00398	0.0038	0.00448

Age x	dx	E_x^c	μ_x	$\mathring{\mu}_x$	$\overset{s}{\mu}_x$
56	411	98,867	0.00416	0.00428	0.00504
57	585	120,000	0.00488	0.00482	0.00567
58	593	113,103	0.00542	0.00543	0.00638
59	648	105,636	0.00613	0.00611	0.00717
60	811	119,930	0.00676	0.00688	0.00806
61	1052	129,034	0.00815	0.00774	0.00906
62	1023	121,768	0.0084	0.0087	0.01018
63	1022	104,471	0.00978	0.00977	0.01142
64	1152	103,124	0.01117	0.01097	0.01281
65	1098	91,905	0.01195	0.0123	0.01436
66	1254	87,543	0.01432	0.01379	0.01609
67	1228	81,877	0.015	0.01544	0.018
68	1395	77,961	0.01789	0.01728	0.02013
69	1475	78,735	0.01873	0.01932	0.0225
70	1654	78,976	0.02094	0.02157	0.02512
71	1585	63,005	0.02516	0.02407	0.02802
72	1744	66,415	0.02626	0.02684	0.03123
73	1893	62,151	0.03046	0.02989	0.03477
74	1854	57,000	0.03253	0.03326	0.03868
75	1779	48,472	0.0367	0.03697	0.04299
76	1896	45,253	0.0419	0.04106	0.04774
77	1774	39,937	0.04442	0.04555	0.05296
78	1587	31,172	0.05091	0.05049	0.05869
79	2245	41,026	0.05472	0.05591	0.06498
80	1995	31,208	0.06393	0.06185	0.07188
81	2285	33,306	0.06861	0.06836	0.07943
82	2152	28,188	0.07635	0.07547	0.08769
83	1998	24,000	0.08325	0.08325	0.09671
84	1547	16,261	0.09514	0.09173	0.10655
85	1358	14,000	0.097	0.10098	0.11729
86	866	8,160	0.10613	0.11105	0.12898
87	858	6,806	0.12607	0.122	0.14169
88	355	2,621	0.13544	0.1339	0.1555
89	162	1,163	0.1393	0.14681	0.17049
90	255	1,660	0.15361	0.16081	0.18674

Appendix 3.B

A list of values for each age interval from 25 to 90. These values are the log of the values in **Appendix 3.A**.

Table 3.12: The ln of Crude, Graduated and Standardised Force of Mortality 25 to 90

Age x	dx	E_X^C	Ln $\hat{\mu}_x$	Ln $\overset{\circ}{\mu}_x$	Ln $\overset{s}{\mu}_x$
25	0	1,906	-11.5126	-7.77526	-7.47692
26	0	2,480	-11.5129	-7.77526	-7.47515
27	1	7,625	-8.87387	-7.77288	-7.46987
28	2	7,980	-8.25483	-7.76578	-7.46114
29	1	7,995	-8.87387	-7.75873	-7.45076
30	0	7,065	-11.5129	-7.74477	-7.43369
31	1	7,522	-8.87387	-7.72647	-7.41359
32	3	11,000	-8.18072	-7.70183	-7.38902
33	2	10,770	-8.51719	-7.67347	-7.35874
34	0	10,841	-11.5129	-7.63964	-7.32176
35	7	13,165	-7.52394	-7.59691	-7.27882
36	19	23,499	-7.10621	-7.54831	-7.22934
37	12	24,290	-7.6009	-7.49295	-7.17433
38	20	32,097	-7.36979	-7.432	-7.11355
39	31	36,372	-7.05858	-7.36346	-7.04587
40	32	38,173	-7.07027	-7.28902	-6.97176
41	38	43,885	-7.03559	-7.20751	-6.89188
42	41	47,059	-7.03559	-7.12219	-6.80701
43	59	54,083	-6.81245	-7.02992	-6.71713
44	65	55,442	-6.74224	-6.93307	-6.62258
45	88	60,042	-6.51571	-6.83265	-6.52385
46	89	73,070	-6.70074	-6.72794	-6.42225
47	121	74,530	-6.41918	-6.62107	-6.31664
48	156	79,224	-6.22466	-6.51032	-6.20863
49	131	78,423	-6.38896	-6.39733	-6.09816
50	156	80,021	-6.23481	-6.28289	-5.98588

Age x	dx	E_X^C	Ln $\hat{\mu}_x$	Ln $\overset{\circ}{\mu}_x$	Ln $\overset{s}{\mu}_x$
51	210	87,307	-6.02399	-6.16677	-5.87208
52	255	92,638	-5.89252	-6.04952	-5.75687
53	231	82,185	-5.87102	-5.93131	-5.64081
54	305	89,784	-5.68104	-5.81248	-5.52421
55	368	92,429	-5.52396	-5.69314	-5.40724
56	411	98,867	-5.47984	-5.57381	-5.28956
57	585	120,000	-5.32056	-5.45427	-5.17204
58	593	113,103	-5.24953	-5.33519	-5.05459
59	648	119,930	-5.09228	-5.09734	-4.93729
60	811	129,034	-4.99496	-4.97928	-4.82047
61	1,052	121,768	-4.80814	-4.86174	-4.70389
62	1,023	104,471	-4.77821	-4.74466	-4.58782
63	1,022	103,124	-4.62609	-4.62834	-4.4723
64	1,152	91,905	-4.49354	-4.51268	-4.35745
65	1,098	87,543	-4.42644	-4.39783	-4.24331
66	1,254	81,877	-4.24512	-4.28388	-4.12987
67	1,228	77,961	-4.19917	-4.17066	-4.01722
68	1,395	78,735	-4.02273	-4.05832	-3.90535
69	1,475	78,976	-3.97688	-3.94687	-3.79433
70	1,654	63,005	-3.86547	-3.83636	-3.68413
71	1,585	66,415	-3.68222	-3.72675	-3.57484
72	1,744	62,151	-3.63937	-3.61805	-3.46644
73	1,893	57,000	-3.49108	-3.5103	-3.35894
74	1,854	48,472	-3.42541	-3.40349	-3.25238
75	1,779	45,253	-3.30465	-3.29765	-3.14674
76	1,896	39,937	-3.17228	-3.19277	-3.04205
77	1,774	31,172	-3.11384	-3.08886	-2.93829
78	1,587	41,026	-2.97748	-2.98592	-2.8355
79	2,245	31,208	-2.90533	-2.88396	-2.73368
80	1,995	33,306	-2.74987	-2.78298	-2.63281
81	2,285	28,188	-2.67923	-2.68298	-2.53292
82	2,152	24,000	-2.57236	-2.58397	-2.43399

Age x	dx	E_X^C	Ln $\hat{\mu}_x$	Ln $\overset{\circ}{\mu}_x$	Ln $\overset{s}{\mu}_x$
83	1,998	16,261	-2.48579	-2.48594	-2.33606
84	1,547	14,000	-2.35234	-2.38892	-2.2391
85	1,358	8,160	-2.33294	-2.29287	-2.14312
86	866	6,806	-2.24302	-2.19782	-2.04814
87	858	2,621	-2.07088	-2.10376	-1.95413
88	355	1,163	-1.99912	-2.01069	-1.86111
89	162	1,660	-1.97109	-1.91862	-1.76908
90	255	0	-1.87325	-1.82754	-1.67804

Appendix 3.C

Plot of the AM92 standard table.

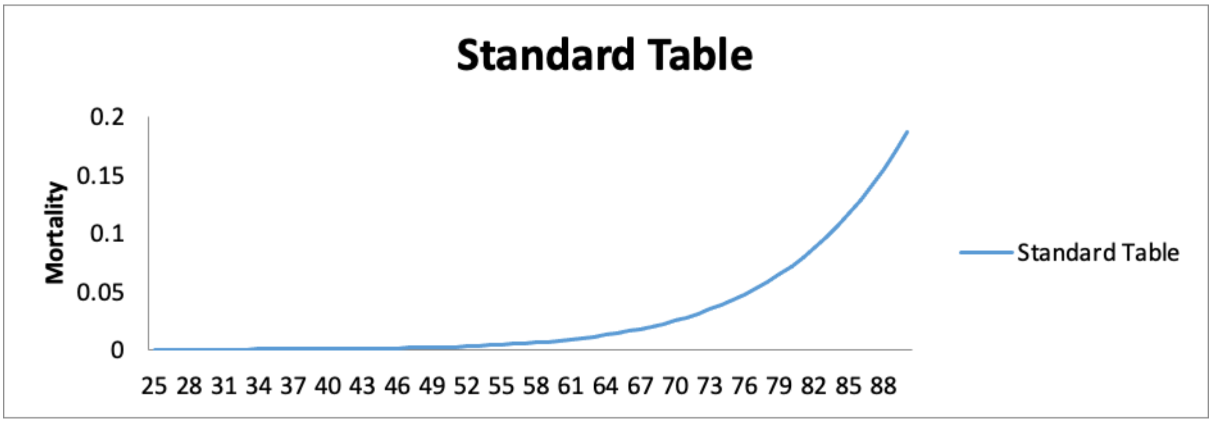

Figure 3.11: A Plot of the standard Mortality Rates of AM92.

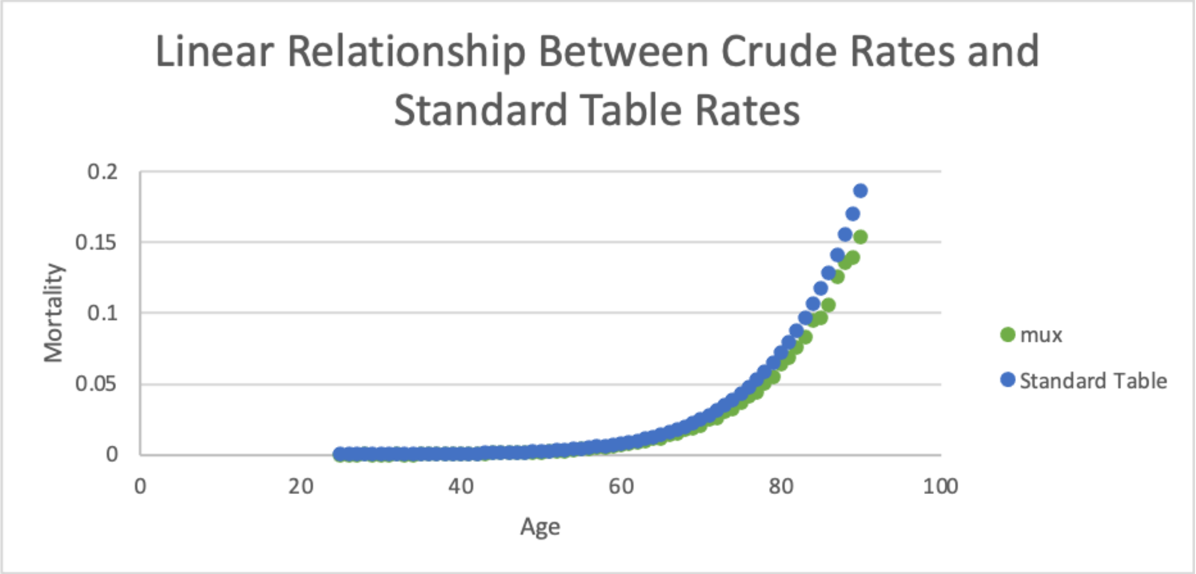

Figure 3.12: A Plot of the crude rates and standard Mortality Rates of AM92.

A plot of the residuals of the AM92 mortality rates and crude rates.

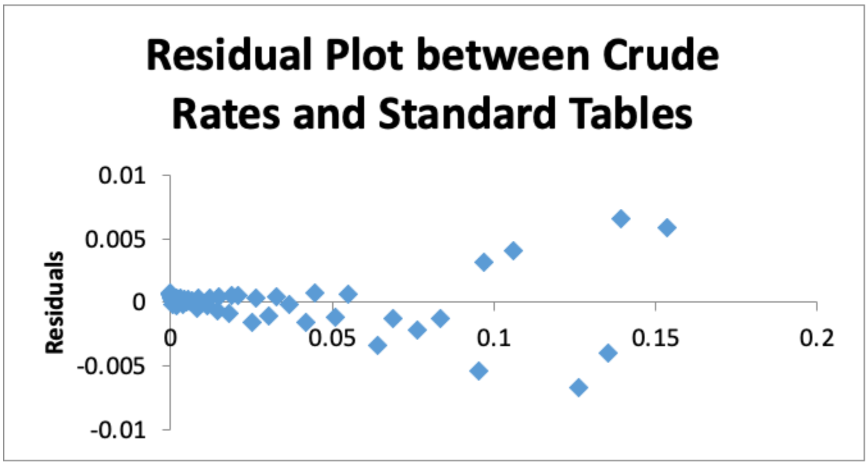

Figure 3.13: A Plot of the Residuals of AM92.

Appendix 3.D

The table below has the third differential at each age. Δ25 represents the changes from 25 to 26, Δ26 represents the changes from 26 to 27 and so on for each age range. A value is deemed smoother based on the magnitude of the value, lower and negative are more desirable.

Table 3.13: Difference in Crude, Graduated and Standardised Force of Mortality 25 to 90

Δ **x**	$\hat{\mu}_{x+1}-\hat{\mu}_x$	$\overset{\circ}{\mu}_{x+1}-\overset{\circ}{\mu}_x$	$\overset{s}{\mu}_{x+1}-\overset{s}{\mu}_x$	Δ **x**	$\hat{\mu}_{x+1}-\hat{\mu}_x$	$\overset{\circ}{\mu}_{x+1}-\overset{\circ}{\mu}_x$	$\overset{s}{\mu}_{x+1}-\overset{s}{\mu}_x$
Δ **25**	-0.0001	-4.3E-07	-5E-07	Δ 58	0.00103	1E-05	1.1E-05
Δ **26**	-0.0002	4.31E-07	5E-07	Δ 59	-0.0019	1E-05	1.1E-05
Δ **27**	0.00025	8.62E-07	1E-06	Δ 60	0.00228	1E-05	1.3E-05
Δ **28**	0.00026	-4.3E-07	-5E-07	Δ 61	-0.0011	1E-05	1.3E-05
Δ **29**	-0.0003	4.31E-07	5E-07	Δ 62	-0.0006	1E-05	1.4E-05
Δ **30**	-0.0002	8.62E-07	1E-06	Δ 63	0.00221	1E-05	1.7E-05
Δ **31**	0.00013	8.62E-07	1E-06	Δ 64	-0.0033	2E-05	1.8E-05
Δ **32**	0.00082	4.31E-07	5E-07	Δ 65	0.00393	2E-05	1.9E-05
Δ **33**	-0.001	-3.3E-19	-1.1E-19	Δ 66	-0.0043	2E-05	2.1E-05
Δ **34**	-0.0003	4.31E-07	5E-07	Δ 67	0.00342	2E-05	2.2E-05
Δ **35**	0.00103	8.62E-07	1E-06	Δ 68	0.00064	2E-05	2.4E-05
Δ **36**	-0.0003	1.72E-06	2E-06	Δ 69	-0.0051	2E-05	2.7E-05
Δ **37**	-0.0003	8.62E-07	1E-06	Δ 70	0.00621	2E-05	2.9E-05
Δ **38**	0.00028	4.31E-07	5E-07	Δ 71	-0.0052	3E-05	3E-05
Δ **39**	-6E-05	1.72E-06	2E-06	Δ 72	0.00424	3E-05	3.4E-05
Δ **40**	0.00024	1.29E-06	1.5E-06	Δ 73	-0.0011	3E-05	3.5E-05
Δ **41**	-0.0004	1.72E-06	2E-06	Δ 74	-0.0037	3E-05	3.8E-05
Δ **42**	0.00035	1.72E-06	2E-06	Δ 75	0.00664	3E-05	4E-05
Δ **43**	-0.0008	1.29E-06	1.5E-06	Δ 76	-0.0066	4E-05	4.4E-05
Δ **44**	0.00119	3.02E-06	3.5E-06	Δ 77	0.00807	4E-05	4.7E-05
Δ **45**	-0.0007	2.58E-06	3E-06	Δ 78	-0.0099	4E-05	5E-05
Δ **46**	-0.0006	2.15E-06	2.5E-06	Δ 79	0.00758	5E-05	5.4E-05
Δ **47**	0.00122	3.02E-06	3.5E-06	Δ 80	-0.0039	5E-05	5.5E-05
Δ **48**	-0.0004	3.45E-06	4E-06	Δ 81	0.00581	5E-05	6E-05
Δ **49**	-0.0003	3.88E-06	4.5E-06	Δ 82	-0.015	6E-05	6.4E-05
Δ **50**	-0.0002	4.31E-06	5E-06	Δ 83	0.01728	6E-05	6.7E-05
Δ **51**	0.00082	3.88E-06	4.5E-06	Δ 84	0.00355	6E-05	7.1E-05

Δ x	$\hat{\mu}_{x+1}-\hat{\mu}_x$	$\overset{\circ}{\mu}_{x+1}-\overset{\circ}{\mu}_x$	$\overset{s}{\mu}_{x+1}-\overset{s}{\mu}_x$	Δ x	$\hat{\mu}_{x+1}-\hat{\mu}_x$	$\overset{\circ}{\mu}_{x+1}-\overset{\circ}{\mu}_x$	$\overset{s}{\mu}_{x+1}-\overset{s}{\mu}_x$
Δ**52**	-0.0005	5.6E-06	6.5E-06	Δ 85	-0.0214	6E-05	7.5E-05
Δ**53**	-0.0004	6.46E-06	7.5E-06	Δ 86	0.00503	7E-05	7.7E-05
Δ**54**	0.00095	4.74E-06	5.5E-06	Δ 87	0.016	7E-05	8.2E-05
Δ**55**	-0.0009	7.32E-06	8.5E-06	Δ 88	-0.1784	0.1759	-0.2042
Δ**56**	0.00087	7.75E-06	9E-06	Δ 89	0.32155	0.3356	0.38973
Δ**57**	-0.0008	7.32E-06	8.5E-06				

Chapter 4: A STUDY OF STRUCTURAL EMPOWERMENT IN NURSING IN IRELAND: FORMAL POWER

Supervisor: Kathleen O'Sullivan

Disclaimer: Use of this data was granted by Sheila Creedon. The use of this data outside this report requires explicit permission on behalf of the aforementioned person from Kathleen O'Sullivan.

4.1 Introduction

A recent paper published by the Economic and Social Research Institute (ERSI) stated that the total public health expenditure in Ireland in 2009 was close to €15.4 billion (ESRI, 2010). This relates to an increase of around 100 per cent in real-term costs since 2000. In 1997, the figure was close to €3.6 billion (Irish Department of Health and Children, 2007). This increase in expenditure shows that it is an important aspect in Irish society. In the most recent CSO figures available, this figure was €22.5bn in 2018 (Health – CSO – Central Statistics Office, 2021). This figure is set to be even larger today, 2021, however this will be due to the natural trend upwards and the Covid-19 pandemic. The number one goal of the healthcare system is to treat and care for the patients, with nurses being the cornerstone of this goal, and with continued supervision of patients at all times being their primary function. As nurses have such an important role in the healing process, it is essential that the most experienced nurses be retained and the best calibre of individuals are attracted to the profession, as what nurses do and don't do has a direct impact on a variety of patient outcomes, including in-hospital deaths (Benner, 1984). Therefore, a high-quality nursing sector is essential.

Recent reports state that there is a shortage of nurses, in developed countries, in particular (Chaguturu & Vallabhaneni, 2005). The Economic Co-operation and Development (OECD) published a report in 2008 reporting fears of an increase in the supply and demand gap for nurses. Nurses are believed to be leaving the profession and it has become a less attractive option, due perhaps to burnout and high levels of illness-related absenteeism (Laschinger, Finnegan, Shamian & Wilk, 2004; Cho, Laschinger & Wong, 2006; Greco, Laschinger & Wong, 2006). This problem needs to be addressed so that the nursing community can grow stronger.

Studies focused on methods of retaining and attracting nurses were carried out in America (McClure, Poulin, Sovie & Wandlet, 1983; Kramer & Schamalenberg, 1988; Kramer, 1990). The outcome of studies on nurse empowerment has given five common characteristics. These characteristics are known as the forces of magnetism. The first characteristic describes the importance and status of nurses within the organisation, the second was the ability to make independent clinical decisions. The third looked at their control over patients, and the fourth looked at the nurses' clinical responsibilities at the unit level to promote accountability and continuity of care. The final characteristic is that there is an established culture outlining nurses' importance within the organisation (Aiken, Smith & Lake, 1994).

These key points mention in above are the reasons for attracting nurses to hospitals and these hospitals are known as magnet hospitals. Magnet status is awarded to hospitals by The American Nurses' Credentialing Centre (ANCC), that match a particular set of criteria (http://www.nursingadvocacy.org/faq/magnet.html). The ANCC look at hospitals where nurses deliver excellent patient outcomes, where nurses feel they have a high sense of job satisfaction, that there is a low staff nurse turnout, and that there are suitable mechanisms in place for solving problems.

The Institute of Medicine (IOM) described nursing as essential to patient care, with poor management practice and negative working conditions having adverse effects on patient safety (Armstrong, Laschinger, 2005). A study of magnet hospitals showed that hospitals that encourage unit rate decision-making were more likely to provide superior decision making.

An exploratory test was done to see if there was a link between structural empowerment, magnet hospital characteristics and patient safety (Armstrong & Laschinger, 2005; Laschinger, Almost & Tuer-Hodes, 2003). The exploratory test found that there was a link between structural empowerment, magnet hospitals, and patient safety, and the results of the exploratory test reported that nurses practising in magnet hospitals felt more empowered (Upenicks, 2003). In light of the findings of the exploratory test, structural empowerment is an important feature of magnet hospitals.

Kanter (1977, 2003) describes empowerment through six conditions in the workplace, 1) information, 2) support, 3) resources, 4) opportunities, 5) formal power, and 6) informal power. Kanter believes these factors have a big influence on employees' attitudes and behaviour than their own personal characteristics.

Empowerment is an important part of the nursing profession. Structurally empowered environments have been linked to increased patient safety (Kohn et al., 2000; Bernholtz et al., 2004). Many studies have identified a positive relationship between structural empowerment and nursing outcomes such as work effectiveness (Laschinger & Wong, 1999), job satisfaction (Laschinger & Havens, 1996), and organisational commitment (McDermott et al., 1996). As mentioned earlier, a lot of nurses are suffering from burnout, however Hatcher and Laschinger (1996) found that structural empowerment has a relieving effect on levels of burnout amongst nurses. Thus, it is fair to say that empowerment creates a better environment for nurses, which in return creates a better environment for patients.

Since this study was conducted, research and study into nurse empowerment continues. An area which is still of interest today is retaining and attracting nurses into the profession. Nurse empowerment has a positive impact on job satisfaction. In 2013, a systematic literature review was conducted between the relationship of workplace empowerment and nurse job satisfaction (Cicolini, Comparachi and Simonetti, 2013). The approach in their review was different to the approach employed in this study. Instead of using a questionnaire and engaging with nurses, Cicolini et al, (2013) read close to 600 articles that examined the relationship between nurse job satisfaction, psychological empowerment and structural empowerment. Twelve papers were used for their analysis. Cicolini et al, (2013) determined that there was a positive relationship between nurse empowerment and job satisfaction, but that structural empowerment and psychological empowerment effected nurse empowerment differently. Job dissatisfaction being a main cause for staff turnover. Cicolini et al, (2013) found that a satisfying work environment for nurses related to structural empowerment in the workplace. This positive relationship leads to positive job satisfaction. However, another systematics reviewed in 2018, showed a positive relationship between psychological empowerment and job satisfaction. This review used 1,572 articles relating to the relationship between job satisfaction and psychological empowerment of nurses. A study conducted in 2011 explored the relationship between structural empowerment and anticipated turnover amongst critical care nurses. (Hauck, Quinn Griffin and Fitzpatrick, 2011), the method used in that their was similar to this study, and they collected data from nurses using a questionnaire. The study by Hauck et al (2013) found there was a positive relationship between structural empowerment and anticipated staff turnover.

Education is a factor that has a positive influence on nurse empowerment, a study in 2018 showed that empowerment programmes improved patient welfare (Amiri, Khademian and Nikandish, 2018), in particular, the effect of an innovative educational empowerment programme on patient safety. A report in 2019 by the Irish Nurses and Midwives Organisation (INMO), called for an end to the recruitment freeze in Ireland and also for an increase in the funding levels to combat the current nursing level shortage and retention levels. The INMO want to combat this by increasing the number of nurses educated at undergraduate level. The INMO's aim is to provide safe staffing levels and a working environment that does not contribute to adverse outcomes for staff and patients (Oireachtas.ie, 2019). Adequate staffing levels will contribute to a positive work environment as it will allow staff to have more flexibility, this is one of the key attributes of informal power. Removing the recruitment freeze and increasing educated nurses will supply more nurses to institutions. Nurses at university are becoming more rounded individuals now as universities promote additional transferable skills, work experience and training (Student Experience, 2021). These additional skill sets could make nurses more adaptable, which is linked to formal power. These initiatives should have a positive impact towards empowerment for new and existing staff.

This study investigates how demographics and job-related and work environment characteristics impact nurses' perceived level of formal power.

4.2 Methods

This section examines the questionnaire from which the data was gathered. This section also describes the survey design and the participants used in the study. We explain the data manipulation procedure and other areas of the analysis. To begin this section, the target population for the study and the instrument that was used to collect the information is described.

4.2.1 Participants

The questionnaire was sent out to the homes of 4,000 nurses on behalf of the researcher by An Bord Altrnais; An Bord Altranais maintain the live registry of nurses in Ireland and holds the only national database of nurses in Ireland. Out of the 4000 sample of nurses, there were 885 questionnaires returned. This gave a response rate of approximately 22%.

4.2.2 Instrument-Questionnaire

The data for this study was collected using a self-completed questionnaire. The questionnaire comprises of 51 questions in total. The survey is divided into three sections:

1. Section one is focused on demographic questions and is made up of 10 questions varying between numeric and categorical single responses questions.
2. Section two is based on perceptions of support in the work environment, known as structural empowerment.
3. The final section is based on hospital culture.

The questions comprised both categorical and quantitative variables.

Table 4.1: Breakdown of Questionnaire.

Section	Question Number	Description of Questions
Demographics (pp.1-3)	1-10	Gender, Type of Job, Race, Hospital Name, Type of Shift
Support in Work Environment (pp.4-5)	11-31	Empowerment Instrument
Hospital Culture (pp.6-8)	32-51	Culture Instrument

Section 1: Demographic Section

The demographic section was based on two types of questions, both quantitative and categorical questions. There were six categorical questions and these looked at gender, ethnicity, job description, whether the respondent worked full or part time, type of life, and place of employment. The final four questions of the section were numerical and addressed age, year of first qualification, length of time in current job, and the respondent's current hospital. Sections two and three contained instruments, respectively, the structural empowerment instrument and the hospital culture instrument, these two instruments were used to examine the respondents' perceptions of empowerment and hospital culture.

Section 2: Structural Empowerment

The Conditions of Work Effectiveness Questionnaire (CWEQ 1) is the instrument used to measure health workers' perceptions of access to elements of structural empowerment (Laschinger, Finegan, Shamian & Wilk, 2001). The instrument is comprised of 31 items, which are based on work done by Kanter (1977, 1993) based on the theory of structural empowerment. As stated in the introduction, Kanter believes that health workers' behaviour arises due to the work environment and not their own personal characteristics.

In our study, we used a shorter modified version of the (CWEQ1), namely, the Conditions of Work Effectiveness Questionnaire (CWEQ 2). This instrument uses 19 items to measure health workers' perceptions of empowerment. Empowerment is measured by six subscales, 1) information, 2) support, 3) resources, 4) opportunities,

5) formal power, and 6) informal power. The difference between the CWEQ 1 and 2 instruments is that it only uses three items from each of Kanter's empowerment structures for opportunity, information, support and resources, along with three items from the Job Activities Scale (Laschinger, 1996) and a final four from the Organisational Relationship Scale to measure formal and informal power, respectively.

The 19 items of the CWEQ 2 are made up of six subscales, five of the six subscales, information, support, resources, opportunity, and formal have three items in each subscale. The final subscale, informal power, is comprised of four items. The response to each item is measured on a 5-point Likert scale. The score for each subscale is determined by taking the average of the responses for each item in the subscale. For each subscale containing three items, a score is only valid if respondents answered at least two items in the subscale, otherwise a score of zero is applied. The four-item subscale score is only valid if at least three items are answered, otherwise a score of zero is applied. When a zero is applied, the respondent's score is assigned as a missing value (Laschinger, 2001). Scores for the subscales ranged from 1-5 when valid.

Information

Information means having access to information in relation to organisational goals and policy changes. A high score on this scale translates to an individual who perceives a high level of access to information.

Support

Support is a result of feedback and guidance from both peers and superiors. Support refers to an individuals' access to risk taking and independent decision making (Laschinger, 2001). A high score on this scale translates to an individual who perceives a high level of access to support.

Resources

Resources involve having money, equipment, supplies and the time needed to accomplish organisational goals (Laschinger, 2001). A high score on this scale translates to an individual who perceives a high level of access to resources.

Opportunity

Opportunity is a result of a combination of a willingness to grow, a sense of change and independence within the organisation that will increase knowledge and skills (Laschinger, 2001). A high score on this scale translates to an individual perceiving that they have high levels of access to opportunity.

Formal Power

This study is focusing on formal power. Formal power is described as being flexible, adaptable, visible and being central to an organisational purpose and goals (Laschinger, 2009). A high score on this scale translates to an individual who perceives a high level of formal power.

Informal Power

Informal power directly results from positive relationships with superiors, peers and subordinates. This results in a positive effective work alliance. A high score on this scale translates to an individual who perceives a high level of informal power.

The scores of each of the subscales are summed to give an overall structural empowerment score (6-30). Scores ranging from 6 to 13 are described as low levels of empowerment, 14 to 22 as moderate levels of empowerment, and 23 to 30 as high levels of empowerment. Thus, higher scores represent higher perceptions of structural empowerment. The items included in each scale are included for reference in **Appendix 4.A**.

Section 3: Hospital Culture

The final section of the questionnaire is based on the hospital culture scales. The hospital culture scare is taken from Shortell's Quality Improvement Survey (QIIS) (Shortell et al., 1995). This scale examines the perceived culture in a respondent's hospital. Culture is based on the values, norms, and beliefs of an organisation that shapes it behaviour (Shortell et al., 1995).

The hospital culture scale was developed by Zammuto and Krakower (1991). The scale consists of 20 items and is divided into five subscales that examine a different hospital characteristic. These are (1) hospital character, (2) hospital managers, (3) hospital cohesions, (4) hospital emphases, and (5) hospital rewards. Each of these subscales are made up of four items. Using these items, we can create four hospital culture types: (1) group culture, (2) developmental culture, (3) hierarchical culture and (4) rational culture.

The first item in each of the five subscales relates to the group culture scale. Group culture is defined as the extent to which the values, norms and beliefs which are strongly linked to affiliation, teamwork and participation (Shortell et al., 1995). The second item in each scale makes up the developmental culture scale. Developmental culture is believed by individuals to be culture based on risk taking, innovation, and change. The third and fourth items in each of the five subscales make up the hierarchal and rational culture scale respectfully. Hierarchal culture is defined as an individual's beliefs that culture should reflect values and norms linked to bureaucracy. The final culture, rational culture, is defined as an individual's perception of culture through emphasis and achievements.

The culture instrument is measured using a modified version the "Quality Improvement Implementation Survey 2 (QIIS-2)" (Shortell et al., 2000). The items in the QIIS-2 instrument are scored out of 100. In this study, a simplified approach is used, the instrument uses a 5-point Likert scale, and each item can score between 1-5. A respondent providing an answer of 1 represented a strong agreement with the item, 3 represented a neutral reply and 5 stated they had a strong disagreement with the statement. For a score to be valid, respondents must answer at least three questions for that culture type. A score for each culture type is the average of the responses to the respondents' particular score for that culture type. Respondents who did not answer three or more questions were assigned as missing. The scores for each item were reversed, this ensured that a higher score represented more of that culture type. The dominating culture type is the culture which has the highest score. The items in this instrument are shown in **Appendix 4.B**.

4.2.3 Study Design

The design of this study is categorised as a quantitative descriptive design (Heppner, Kivlighan & Wampold, 1998). Specifically, a complex passive research design was used, and the data was collected through the use of a survey tool. As previously described in chapter one, a quantitative descriptive design is a process in which data is utilised to gather information about a particular population. Passive research designs have the aim of exploring the relationship between variables, without the researcher manipulating some aspect of the study. Such designs are one of the major forms of quantitative descriptive research. Simple passive designs explore relationships between just two variables, while more complex designs aim to explain the relationship among multiple variables.

To ensure anonymity of respondents, the researcher involved was not allowed to provide the questionnaire used in this study in a supplementary appendix. Respondents were assigned a confidential identification number to ensure anonymity. The confidentiality of all data was maintained and was stored on a password protected secure drive on the researcher's laptop.

4.2.4 Data Preparation

Data analysis was carried out using SAS 9.2 software. The data was received in Microsoft Excel format, with each respondent assigned a confidential ID number. Prior to statistical analysis, it was necessary to adjust certain variables. The groupings for demographic are provided in the Results section 4.3. Numerical values were grouped for example, the age of respondents was grouped into suitable age ranges, he length of time in current job and length of time in current workplace. For the question relating to the respondents' job description, some grouping contained quite low numbers and the researcher combined the original groupings into the following classes: nurse, nurse manager, clinical nurse specialist, public health nurse, midwife, and education.

Four new variables were created based on the respondents' current place of employment, which was determined in section one of the questionnaire. These variables examined the employers' management type, HSE location, type of care supplied, and whether the employer had an associated religious affiliation. Management type was grouped into three categories: HSE, voluntary, and not publicly funded. HSE location was divided into the following categories: Dublin North East, Dublin Mid-Leinster, South, and West. Type of care was classified into six groups: university teaching hospital, regional hospital, general hospital, public health, private, or specialist care. Religious affiliation was classified as yes or no.

4.2.5 Reliability Analysis

A reliability analysis is carried out on the items that make up the formal power scale and the four culture types. The reliability analysis determined the internal consistency of the scales and it established whether a set of items measured a unidimensional construct (Hair, Black, Babin, Anderson & Tatham, 2006). Cronbach's alpha is calculated for each scale. A high Cronbach's alpha represents a high inter correlation, thus showing a high internal consistency.

DeVellis (1991) outlines the recommended range for Cronbach's alpha to lie somewhere between 0.7 to 0.9. A value between 0.6 and 0.65 is considered undesirable, between 0.65 and 0.7 as minimally desirable while a value greater than 0.9 is said to be too strong. If the Cronbach's alpha for a subscale is between 0.7 and 0.9, then a summated score can be calculated.

4.2.6 Univariate Analysis

In this study, the relationship between perception of formal power and each of the demographics, job-related and work environment characteristics is investigated. Univariate analysis is utilised to examine the effects of the explanatory variable under study on each individual response variable. Univariate analysis is used in determining associations between a respondent's perceived empowerment and each of the 10 demographic variables separately. A two-sample t-test is employed when there is a single factor and the factor has two factors. When there are three or more factor levels (groups) for a variable, the single-factor between-subjects analysis of variance procedure is conducted.

Comparing Two Means

Two Sample T-Test

The independent two sample t-test is used to establish if two independent samples are representative of populations with different mean values (Rice, 1995; Sheskin, 2004). In essence, the sample t-test determines if the samples represent populations with different mean values. This test is a hypothesis-based procedure, whereby two sample means are calculated to represent the population means. A test statistic is calculated as follows:

$$\frac{\bar{X}_1 - \bar{X}_2}{SE(\bar{X}_1 - \bar{X}_2)}$$

Where $\bar{X}_1$ is the sample mean of population 1, $\bar{X}_2$ is the sample mean of population 2 and $SE(\bar{X}_1 - \bar{X}_2)$ is the standard error of the difference between the sample means of the two populations.

A hypothesis test is put in place to determine if there is a significant difference between the two mean values. The hypothesis test for the two-sample t-test is given as follows:

$$H_0 = \mu_1 = \mu_2$$

$$H_a = \mu_1 \neq \mu_2$$

where the null hypothesis is a hypothesis of zero difference and the alternative is non-directional and is therefore evaluated using a two-tailed test.

The following assumptions are the foundations on which the two-sample t-test is applied: (1) that each sample is randomly selected; (2) that the underlying populations of the samples are normally distributed; and (3) that there is homogeneity of variance.

The single-factor between-subjects analysis of variance (one-way ANOVA) is used to evaluate the mean values at different factor levels (groups) (Rice, 1995; Sheskin, 2004). The one-way ANOVA is employed to evaluate whether there is a difference between at least two means in a set of k independent samples. The ANOVA model is given as follows:

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

Y_{ij} is the response variable for the j^{th} observation for the i^{th} level, μ_i is the mean response values at the i^{th} factor level and ε_{ij} is the error term, which is independent $N(0, \sigma^2)$.

The hypothesis test examines the ANOVA model to determine if the mean response values at any factor level are significantly different from at least one other level. The hypothesis statement is as follows:

$$H_0 : \mu_1 = \mu_2 = ... = \mu_r$$
$$H_a : \text{ not all } \mu_i \text{ are equal}$$

The alternative hypothesis indicates that there is a difference between at least two of population variances. The total variability is explained by the variability between and within the groups. Dividing the between-groups variability by the within-groups variability calculates the F-ratio. The null hypothesis of zero difference is tested using the F-statistic.

There are numerous assumptions, which underlie the ANOVA procedure. Firstly, that the samples are independent. Secondly, the data is normally distributed and, finally, that there is equality of variance between each group. The normality assumption is examined through the use of a normal probability plot of the residuals, where the points should form an approximate straight line. Formal testing for normality is completed through use of the Shapiro-Wilk and Kolomogorov-Smirnov tests. The null hypothesis for both tests states that the groups come from a normal population. The homogeneity of variance is determined by plotting the residuals against the predicted values. Non-random behaviour of the residuals indicates that this assumption is invalid. Levene’s test is conducted to formally test for homogeneity; the null hypothesis states that the variance is equal across groups.

Tukey's Test

If the one-way ANOVA procedure produces significant results, then post-hoc analysis is carried out to establish where the differences lie. Tukey's Honest Significant Difference (HSD) test is the most frequently used multiple comparison test, used concurrently with the ANOVA procedure (Sheskin, 2004). HSD is used when the researcher wishes to make multiple pairwise comparisons. Tukey's HSD test for each mean comparison is determined using the q statistic:

$$q = \frac{\overline{X}_i - \overline{X}_j}{\sqrt{MS_{wg}\frac{\frac{1}{n_i}+\frac{1}{n_j}}{2}}}$$

where $\overline{X}_i$ is the i^{th} factor mean, $\overline{X}_j$ is the j^{th} factor mean, MS_{wg} is the within groups mean square, n_i is the number of observations in the i^{th} category and n_j is the number of observations in the j^{th} category.

Sheskin (2004) states that the alternative hypothesis used in Tukey's HSD testing procedure should be non-directional. Therefore, Tukey's HSD hypothesis statement is as follows:

$$H_0 : \overline{X_i} = \overline{X}_j$$

$$H_a : \overline{X_i} \neq \overline{X}_j$$

The underlying assumptions of Tukey's HSD testing procedure are that: (1) the samples are independent; (2) the means are from normally distributed populations; and (3) there is homogeneity of variance.

4.2.7 Multivariate Analysis

Multivariate analysis describes the association between the response variable and the explanatory variables. In this study, multivariate analysis was performed to determine what combination of demographic, job-related and work environment characteristics affect a respondent's perception of formal power. Regression analysis, a statistical tool, was used to investigate the relationship between variables (Hair et al., 2006; Kutner, Neter, Nachtsheim & Wasserman, 2004). Hair et al. (2006) state that "regression analysis is by far the most widely used and versatile dependence technique".

Multiple linear regression is an extension of bivariate linear regression, in that the response (perception of formal power) variable is explained by more than one explanatory (demographic characteristics, job-related or work environment characteristics) variable.

The multiple linear regression model is as follows:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i$$

where Y_i denotes the response in the i^{th} case, $\overline{X}_{ip}$ is the value of the p^{th} independent variable for the i^{th} case. β_0 is the Y intercept, β_p is the partial regression coefficient and ε_i is the random error term where $\varepsilon_i \sim N(0, \sigma^2)$. In the regression model, a case is only included if, that case must have observations across the response variable and all p independent variables.

Regression analysis is a means to explore all types of dependent relationships and can be used for predictive and explanatory purposes. Explanation with multivariate regression allows the researcher to examine the degree and character of the relationship between the response and explanatory variables. The following steps outline the model building process.

Candidate Variables

Multivariate regression begins with the selection of variables to be included in a model (Hair et al., 2006). The logic behind minimising the number of independent variables to include in the model is the desire to determine the most parsimonious model that sufficiently describes the data. The more independent variables that are contained in the model, the greater the chance of specification error. Specification error is caused by inclusion of irrelevant variables, and the exclusion of relevant variables. Therefore, it is fundamental to determine a strategy for the selection of candidate variables.

Firstly, univariate analysis is performed for each independent variable. Any variable whose univariate testing procedure obtained a p-value less than or equal to 0.25 is regarded as a candidate for the regression model (Hosmer & Lemeshow, 2000). Once the variables have been selected, the model building process can be continued.

Categorical Variables

It is a critical component of regression analysis that the data is metric in nature before fitting the regression model. However, through the use of dummy variable coding (Draper & Smith, 1998; Hair et al., 2006) it is possible to include non-metric data. A dummy variable is created when non-metric variables account for different factor levels. If there are k levels in a non-metric independent variable, k-1 dummy variables need to be created. It is critical to choose a reference category when using dummy variable coding (a reference category is the excluded level of the non-metric independent variable when the dummy variable is formed). The two methods for creating dummy variables are reference cell coding, and effects coding.

The more common of the two methods of dummy variable coding is reference cell coding. In reference cell coding, each category in a variable is represented by either 0 or 1 and indicates the absence or presence of a categorical effect. The regression coefficients are used to signify the factor level differences in the dependent variable from the reference category, where the reference category receives all zero values. The second method of dummy variable coding is effects coding. This method differs from reference cell coding in that the reference category is given a value of -1. The regression coefficient indicates the difference between any groups from the mean of all other groups pooled.

The method of coding used in this study is reference cell coding. The number of dummy variables is always the total number of categories minus one. For example, if there are three categories A, B and C, there are two dummy variables required (see

Table 4.2 below). Category C is chosen as the reference category, two dummy variables are created D1 and D2.

Table 4.2: **Example of Reference Cell Coding.**

Category	**Dummy Variable**	
	D1	D2
A	1	0
B	0	1
C	0	0

In the regression model, the D1 coefficient compares A to C, this receives a code of 1. D2 compares B to C and receives a code of 1. Category C receives a code of 0 for D1 and D2 as it is the reference category. The regression coefficient for D1 is the difference between the average response to A and the average response to C. A similar interpretation applies to the D2 dummy variable.

Estimation Technique

Candidate variables must be chosen for the overall model. The first step of the process is to conduct a univariate analysis that examines the relationship between the response and each explanatory variable separately. Any variable which received a p-value of 0.25 or less was regarded as a candidate for inclusion in the regression model (Hosmer & Lemeshow, 2000).

The independent variables are now in the appropriate form to be entered into the regression equation. There are three basic techniques for selecting variables to be included in the model: forward; backward; and stepwise regression.

In the forward addition regression method, variables are sequentially added for inclusion based on the variable's statistical significance. The initial model contains an intercept. This is the current model. For each independent variable not in the current model, the increase in R^2 and associated p-value are calculated for its inclusion in the model. The variable with the largest increase to R^2 and has a p-value less than 0.05 is added to the current model. The model now contains the intercept and an independent variable. The process repeats itself from this point to include the next independent variable that has the largest increase in R^2 and a p-value less than 0.05. This procedure is repeated until all independent variables that produce the largest increase in R^2 and are statistically significant are included in the model. Once a variable is added to the model, it is not removed.

Backward multiple linear regression modelling begins with the inclusion of all the variables included in the model. This is the current model. For each independent variable in the current model, the reduction in R^2 and associated p-value are calculated for its removal from the model. The variable with the smallest reduction on R^2 and has a p-value greater than 0.05 is removed from the current model. The model now contains the intercept and all independent variables less the one removed. The process is repeated, each time it removes a variable that has the smallest reduction on R^2 and has a p-value greater than 0.05. This procedure is repeated until all independent variables in the current model produce reductions R^2 that are statistically significant. Once a variable is removed from the model, it cannot be considered for inclusion.

There are two versions of the stepwise regression method: forward addition with a test for backward elimination or backward elimination with a test for forward addition. The most common approach is the modified version of the forward selection method. In this study, forward addition with a test for backward elimination was conducted, it will be referred to as stepwise multiple linear regression modelling from this point. The process begins with the method outlined in the forward addition procedure for the first two steps. The variable that significantly increase R^2 is added to the model. This is the model. From the remaining variables not in the model the process is repeated, and another variable is added. The model has two variables and an intercept. Backward elimination is performed at this point. The variables in the model are examined for removal as outlined in the backward elimination process. Each time a variable is added to the model, backward elimination is applied. This process continues until none of the variables outside the model meet the criteria to be added to the model. The variables identified by the stepwise process and variables that were significant ($0.05 \leq p \leq 0.01$) were evaluated by the backward multiple linear regression model. The variables that were retained within the model were refitted to a simultaneous multiple regression model.

Interactions

Once a model that contains all of the significant variables is created, it is then necessary to examine the independent variables in question more closely and consider the requirement of including an interaction term. In a model, the presence of an interaction between two variables implies that one variable varies as a function of another variable. The inclusion of such an interaction term describes a situation where the influence of two independent variables on the dependent variable is not additive.

To determine the appropriateness of including an interaction term in a model, the interaction terms must first be created. To determine if the interactions significantly contribute to the model, an F test was conducted. Two-way interactions were considered to identify those that made clinical sense. These relevant interactions were added to the final model, only interactions that attained statistical significance were retained. The inclusion of an interaction term must be based on statistical analysis, as well as practical considerations. For the final model, no interactions attained a statistical significance.

Interpreting the Regression Model

Once a final model is decided upon, a simultaneous regression model is fitted to the data. All the independent variables are entered into the regression equation, and this method affords no model selection processes. To estimate the unknown parameters of the model, an ordinary, least squares estimation procedure is conducted. The best model fit minimises the sum of squared residuals.

After the model is fitted to the data, the regression equation can be interpreted by evaluating the estimated regression coefficients. These coefficients represent the extent of the effect that the independent variable has on the dependent variable, and the sign of the coefficient gives the direction of that effect. In multiple regression, the coefficients express the change in the dependant variable when there is a change to the independent variables. It is fundamental to hold all other variables constant when interpreting the effect of an independent variable.

Assumptions

It is essential to examine the underlying assumptions of regression analysis after the relationship has been fitted to the data. Firstly, it is vital to check that the data is normally distributed. This is explored using a normal probability plot where the plot should approximately form a straight line. Departures from this line indicate departures from normality. Secondly, it has to be ensured that there is homogeneity of variance. This is seen when the residuals are plotted against the predicted values. This plot should exhibit a horizontal band of random scatter with no trends. The linearity of the phenomenon measured does not have to be validated when using non-metric data. When applying a regression model, these assumptions must be verified.

The assumptions for all statistical procedures were validated through visually interpreting normal probability plots and plots of residuals against predicted values. The plots for the final model only are presented in **Appendix 4.C**.

4.3 Results

4.3.1 Demographics

The demographic section describes details of gender, type of job, and information on work environment collected from the questionnaire. We examine the demographics based on the personal characteristics, details of current employment, and finally details of work environment. The frequency and percentage responses for the demographic characteristics are listed in **Table 4. 3**.

Table 4.3: Demographic Characteristics.

Characteristic	Category	Frequency	%
Age (Years) (n=875)1	20-29	94	10.7
	30-39	311	35.5
	40-49	260	29.7
	50-59	169	19.3
	60 +	41	4.7
Gender (n=880)	Female	853	96.9
	Male	27	3.1
Year of First Qualification (n=777)	1966-1998	616	79.3
	1999-2005	130	16.7
	2006-2008	31	4.0
Ethnicity (n=882)	Caucasian	804	91.2
	Other	78	8.8

n = number who responded to this characteristic.

Out of the 875 respondents to the survey, over one third (35.5%) were aged between 30 and 39 years, around 29.7% were aged between 40 and 49 years old, with 19.3% aged between 50 and 59 years. The youngest and oldest intervals have the fewest responses, with nearly 10.7% and 4.7% responding, respectively. The majority of the respondents were female (96.9%), caucasian (91.2%), and had qualified in or before 1998 (79.3%). Around 16.7% of those who responded qualified between 1996 and 2005. While the remaining 4% qualified from 2006 to 2008.

The frequency and percentage responses for the job-related characteristics are listed in **Table 4.4**.

Table 4.4: Job-Related Characteristics.

Characteristic	Category	Frequency	%
Job Description (n=859)1	Nurse	495	57.6
	Nurse Managers	178	20.7
	CNS	57	6.6
	Midwife	50	5.8
	PHN	46	5.4
	Education	33	3.8
Length of Time in Current Job (years) (n=865)	0-<1	62	7.2
	1-<3	148	17.1
	3-<10	393	45.4
	10-<20	178	20.6
	20+	84	9.7
Full/Part Time (n=868)	Full Time	545	62.8
	Part Time	323	37.2
Shift Schedule (n=871)	Day Shifts Only	444	51.0
	Shift Rotation	344	39.5
	Night Shifts Only	55	6.2
	Other	28	3.2
Length of Time in Place of Employment (years) (n=828)	0-<1	51	6.2
	1-<3	124	15.0
	3-<10	348	42.0
	10-<20	193	23.3
	20+	112	13.5

[1]Number who responded to this characteristic.

Of the respondents, over half were nurses (57.6%), while nurse managers accounted for 20.7%. The other categories pooled together resulted in around 21.6% of respondents. These comprised CNS (6.6%), public health nurses (5.4%), midwives (5.8%), and individuals based in education (3.8%).

Around 24.3% of respondents have been in their current job less than 3 years, with 6.2% accounting for individuals who are in their current job less than a year and 17.1% of people in their current job between 1 and 3 years. Around 45.4% of respondents, just under half, have been working in their current job between 3 and 10 years. Nearly 20.6% of respondents have been in their current job between 10 and 20 years, with a further 9.7% in their current job for over 20 years.

There are 545 individuals who work full time, these individuals represent 62.8% of the people surveyed. Doing day shifts only, accounts for the highest shift chosen by the respondents with 51% choosing this option, while only 6.2% did night shift only; 39.5% of respondents usually do shift rotation, while the final 3.2% chose other.

Forty-two per cent of the respondents have worked in their respective employment between 3 and 10 years. Around 23.3% have worked in their current job between 10 and 20 years. Under 3 years and above 20 years combine to represent 28.5% of the respondents, with 15% and 13.5%, respectively.

The frequency and percentage responses for the work environment characteristics are listed in **Table 4. 5**.

Table 4.5: Work Environment Characteristics.

Characteristic	Category	Frequency	%
HSE Location (n=752)1	Dublin Mid Leinster	223	29.7
	South	219	29.1
	West	168	22.3
	Dublin North East	142	18.9
Management Type (n=787)	HSE	467	59.3
	Voluntary	205	26.0
	Not Publicly Funded	115	14.6
Type of Care (n=772)	General	189	24.5
	Specialist	156	20.2
	University Teaching	153	19.8
	Regional	103	13.3
	Public Health	90	11.7
	Private	81	10.5
Religious Affiliation (n=787)	No	664	84.4
	Yes	123	15.6

[1]Number who responded to this characteristic.

Table 4.5 shows us the figures for work environment and begins with the HSE location. There are four regions to choose from and 29.1% of respondents are located in the South, with a further 22.3% based in the West. The next two locations include Dublin, with 29.7% working in Dublin Mid Leinster, while the final portion are based in Dublin North East (18.9%).

The majority of the respondents (59.3%) work in a HSE managed hospital, while 26% worked in a voluntary hospital. The final 14.6% work in a not publicly funded institution. The vast majority (84.4%) said their place of employment had no religious affiliation.

Type of care was split into six categories, with general being the most popular, at 24.5%. Specialist accounted for 20.2% of responses, with 19.8% of respondents involved in university teaching. Regional care was at 13.3%, with 11.7% involved in public health. The final amount of (10.5%) worked in private clinics.

4.3.2 Analysis of Scales

To determine if the items that make up the formal subscale are internally consistent, a Cronbach's alpha is computed. A Cronbach's alpha value of 0.764 was produced for the formal power subscale. This value is within the recommended interval outlined by Devellis (1991). The formal power scale comprises three questions. Of the 885 individuals who were surveyed, there were 849 valid responses and 36 missing values. The mean value for the scale was 2.82, with a standard deviation of 0.92, which would indicate a wide variety of responses.

The internal consistency of each culture had to be examined by calculating their Cronbach's alpha, having examined these values we could calculate the score for each hospital culture as outlined in the methods section. The Cronbach's alpha for each of the scales can be seen in **Table 4.6**.

Table 4.6: Cronbach's Alpha for Culture Type.

Culture Type	Cronbach's Alpha Value
Group	0.805
Developmental	0.764
Hierarchical	0.495
Rational	0.484

As we can see from **Table 4.6**, both the group and developmental culture types have Cronbach's alpha values that fall inside the recommended level, as described by Devellis (1991). Thus, we can formulate scores for these culture types. **Table 4.6** also shows us that the Cronbach's alpha for the hierarchical and rational scales are not within the recommended level. Because of this fact, these culture types were not explored further. The scores of the culture type for each respondent were explored to see which culture was dominant. Based on this score, respondents were classified into a specific hospital culture type.

Table 4.7 shows the frequency and percentage of the respondents who perceive their work environment to be either developmental or group culture type.

Table 4.7: Culture Type Work Environment Characteristic.

Scale	Category	Frequency	%
Culture	Developmental	482	70.4
(n=685)1	Group	203	29.6

[1]Number who responded to this characteristic.

Over seventy percent (70.4%) thought their work environment to have a developmental culture type.

4.3.3 Relationship between each Demographic, Job-related and Work Environment Characteristic and Perception of Formal Power

For each of the categories contained in the demographic, job-related or work environment characteristics, the average perception of formal power is compared. In order to do this, a t-test or ANOVA was conducted. If the ANOVA p-value obtained is statistically significant then Tukey's test is employed to see where the differences were. The mean and standard deviation for each category for the demographic characteristics is presented in **Table 4.8**, together with the p-value for the test statistic employed.

Table 4.8: Mean, Standard Deviations (SD) and P-Value for Demographic Characteristics.

Characteristic	Category	Mean (S.D.)	P-Value
Ethnicity (n=882)	Non-Caucasian	2.75 (0.91)	<0.01
	Caucasian	3.32 (0.76)	
Gender (n=880)	Male	3.07 (0.82)	0.017
	Female	2.75 (0.85)	
Age (n=875)	20-29	2.67 (0.92)	0.453
	30-39	2.77 (0.88)	
	40-49	2.82 (0.95)	
	50-59	2.81 (0.89)	
	60 +	3.02 (0.83)	
Year of First Qualification (n=777)	1966-1998	2.82 (0.93)	0.662
	1999-2005	2.76 (0.86)	
	2006-2008	2.72 (0.82)	

There are significant differences in the average perception of formal power in relation to gender (p = 0.017) and ethnicity (p <0.01). No significant differences were found for age (p=0.453) or year of first qualification (p=0.662).

As both characteristics only had two factor levels, we did not need to apply Tukey's test. Thus, caucasians had a significantly higher perception of formal power in comparison to non-caucasians. Males had a significantly higher perception of access to formal power, in comparison to females.

For each current employment characteristic, the average perception of formal power is compared, using a t-test or ANOVA. The mean and standard deviation for each category is presented in **Table 4.9**, together with the p-value for the test statistic employed.

Table 4.9: Mean, Standard Deviations (SD) and P-Value for Job-Related Characteristics.

Characteristic	Category	Mean (S.D.)	P-Value
Job Description (n=859)1	Nurse Nurse Managers CNS Midwife PHN Education	2.71 (0.90) 2.93 (0.94) 3.17 (0.89) 2.93 (0.90) 2.55 (0.97) 3.17 (0.88)	<0.001
Shift Schedule (n=871)	Day shifts only Shift rotation Night shifts only Other	2.92 (0.90) 2.65 (0.90) 2.82 (0.94) 2.61 (0.79)	<0.001
Length of Time in Place of Employment (n=828)	0-<1 1-<3 3-<10 10-<20 20+	2.88 (0.84) 2.91 (1.01) 2.80 (0.91) 2.67 (0.84) 2.83 (0.94)	0.076
Length of Time in Current Job (n=865)	0-<1 1-<3 3-<10 10-<20 20+	2.89 (0.86) 2.85 (0.95) 2.77 (0.92) 2.72 (0.87) 2.84 (0.92)	0.480
Full/Part Time (n=868)	Full Time Part Time	2.81 (0.92) 2.76 (0.89)	0.616

Significant differences in the average perception of formal power were obtained for job description (p<0.01) and shift schedule (p<0.01). No significant differences were found for whether the respondents worked full time or part time (p=0.616). There were also no significant differences in the length of time in their current job (p=0.48), and the length of time in current place of employment (p=0.076).

Pairwise comparisons by Tukey's test revealed a significant difference in the average perceptions of formal power between respondents who described their job as nurse and those who described their job as clinical nurse specialist (p=0.004). In addition, there is a significant difference in the average perception of formal power between respondents who described their job as clinical nurse specialist and public health nurse (p=0.004). Those who described their job as clinical nurse specialist had significantly higher average perception of formal power compared to those who described their job as nurse and those who described their job as public health nurse. The remaining pairwise comparisons showed no significant differences in average perception of formal power and are, therefore, not reported.

Respondents who work day shifts had a significant difference in the average perception of formal power when compared to those who worked on shift rotation.

T-tests and ANOVA were conducted to compare the average perception of formal power for each work environment characteristic. The mean and standard deviation for each category is presented in **Table 4.10**, together with the p-value for the test statistic employed.

Table 4.10: Mean, Standard Deviations (SD) and P-Value for Work Environment Characteristics.

Characteristic	Category	Mean (S.D.)	P-Value
Management Type (n=787) 1	HSE	2.67 (0.89)	<0.001
	Voluntary	2.93 (0.85)	
	Not Publicly Funded	3.16 (1.01)	
Religious Affiliation (n=787)	Yes	3.10 (0.88)	<0.001
	No	2.73 (0.90)	
Culture Type (n=685)	Developmental	2.84 (0.90)	<0.001
	Group	2.67 (0.91)	
HSE Location (n=752)	Dublin North East	2.89 (0.84)	0.007
	Dublin Mid Leinster	3.00 (0.96)	
	South	2.68 (0.86)	
	West	2.61 (0.90)	
Type of Care (n=772)	University Teaching	2.84 (0.89)	0.803
	Regional	2.95 (0.90)	
	General	2.79 (0.87)	
	Public Health	2.71 (0.97)	
	Specialist	2.71 (0.84)	
	Private	2.74 (0.98)	

Significant differences in the average perception of access of formal power were obtained for HSE location (p=0.007), management type (p<0.001), religious affiliation (p<0.001), and culture type (p<0.001). Perceptions of formal power did not differ for type of care (p=0.803).

Respondents who worked in a religiously affiliated place of employment had a higher perception of formal power, compared to those whose workplace did not have a religious affiliation.

Respondents who perceived their place of employment to have a developmental culture type had a higher average perception of access to formal power, when compared to those who perceived their workplace to have a group culture type.

The significant differences in HSE location were further explored by pairwise comparison, as seen in **Table 4.11**. For each pairwise comparison, the p-value associated with the Tukey's test is given.

Table 4.11: Tukey's Test for HSE Location.

Comparison	p-value
(†) Dublin Mid Leinster vs. Dublin North East	0.806
South vs. Dublin North East (†)	0.385
West vs. Dublin North East (†)	0.284
West vs. Dublin Mid Leinster (†)	0.026
South vs. Dublin Mid Leinster (†)	0.018
West vs. South (†)	0.990

(†) Indicates group with greater average perception of formal power.

Tukey's test revealed significant differences in respondents' perceptions of formal power for Dublin Mid Leinster compared to West (p=0.026) and Dublin Mid Leinster compared to South (p=0.018). No significant differences were found for West compared to South (p=0.990), Dublin North East compared to South (p=0.385), and Dublin North East compared to West (p=0.284).

Respondents who worked in Dublin Mid Leinster had significantly greater average perception of formal power compared to those who worked in West and those who worked in South.

The significant differences in management type were further explored by pairwise comparison, as seen in **Table 4.12**. For each pairwise comparison, the p-value associated with the Tukey's test is given.

Table 4.12: Tukey's Test for Management Type.

Comparison	P-value
(†) Voluntary vs. HSE	<0.001
(†) Not Publicly Funded vs. HSE	<0.001
(†) Not Publicly Funded vs. Voluntary	0.010

(†) Indicates group with greater average perception of formal power.

Pairwise comparisons by Tukey's test revealed a significant difference in the average perception of formal power between respondents who work in a voluntary managed place of employment compared to those who worked in a HSE managed place of employment ($p<0.001$). Respondents' perception of formal power was significantly different for those who worked in a not publicly funded managed place of employment compared to those who worked in a HSE ($p<0.001$) and voluntary ($p=0.010$) managed place of employment. Those who worked in a voluntary managed place of employment had a greater average perception of formal power compared to those who worked in a HSE managed setting. Respondents who worked in a not publicly funded managed setting had a greater average perception of formal power, compared to those who worked in a voluntary managed setting and those who worked in a HSE managed setting.

4.3.4 Multiple Regression

To determine which combination of the demographic, current employment, and work environment characteristics affected a respondent's perception of formal power, a multiple regression analysis was applied to the data.

In order to develop an appropriate multiple regression model, we must identify particular candidate variables from the 14 questions that the demographics, current employment, and work environment characteristics comprise of. For a variable to be a successful candidate, it must have a p-value less than 0.25 (Hosmer & Lemeshow, 2000), the results are in **Table 4.13**.

Table 4.13: Univariate Analysis P-Values.

Characteristic	P-Value
Ethnicity	<0.001
Job Description	<0.001
Shift Schedule	<0.001
Management Type	<0.001
Religious Affiliation	<0.001
Culture Type	<0.001
HSE Location	0.007
Gender	0.017
Current Hospital of Employment	0.076
Age	0.453
Current Job	0.480
Full/Part Time	0.616
Year of Qualification	0.622
Type of Care	0.846

Using the criteria outlined by Hosmer & Lemeshow (2000), ethnicity, job description, shift schedule, management type, religious affiliation and culture will all be used in our model as the have an associated p-value of <0.001. HSE location, gender, and current hospital employment also match our criteria, as they have p-values of 0.007, 0.017, and 0.076 respectively.

With the candidate variables selected, a stepwise regression model was performed. The variables retained in the model and the change in R^2 are shown in **Table 4.14**.

Table 4.14: Summary of the Stepwise Multiple Regression Analysis (n = 572).

Step	Variable Entered	Change	Model	P-Value
1	Management Type	0.045	0.045	<0.001
2	Job Description	0.038	0.083	0.0002
3	Ethnicity	0.022	0.105	0.0002
4	Culture Type	0.007	0.112	0.0383

Stepwise multiple linear regression modelling identified four variables to include in the model. The process examined the change on R^2 by each candidate variable, and if it was statistically significant. Management type was the first variable added to the model, it accounted for the highest increase in R^2 and was statistically significant. Management type produced an increase in R^2 of 4.5%, the overall R^2 was 4.5%. The second variable entered into the model was job description, it increased the R^2 by 3.8%, total R^2 was 8.3%. Ethnicity was the third variable entered into the model, it increased R^2 by 2.2%, the overall R^2 was 10.5%. The last variable entered into the model was culture type, it increased R^2 by 0.7%, this produced a total model R^2 of 11.2%.

The variables that were not included in the model were examined, none of them were statistically significant. A backward multiple linear regression model was conducted on management type, job description, ethnicity and culture type. The test statistic and associated p-values are presented in **Table 4.15**.

Table 4.15: Final Multiple Linear Regression Model.

Variable	F-Statistic	P-Value
Management Type	30.14	<0.001
Job Description	25.68	0.0002
Ethnicity	22.02	0.0002
Culture Type	15.98	0.0383

The final model contained management type, job description, ethnicity and culture type. R^2 for the model is 11.2%. **Table 4.16** shows the partial regression coefficients of the final model with associated standard errors and p-values. The reference categories for each variable are as follows: for management type, the reference category is HSE. For job description the reference category is nurse, for ethnicity it is caucasian and for culture type it's developmental. Pairwise comparisons were run holding all other variables in the model constant. The regression coefficient is the difference between the averages of the groups in the comparison.

Table 4.16: Regression Coefficients Estimates for Final Model.

Characteristic	Pairwise Comparison	Regression Coefficient (S.E.)	P-Value
Management Type	Voluntary vs. HSE	0.23 (0.07)	0.0059*
	Not Publicly Funded vs. HSE	0.61 (0.09)	<0.001*
Job Description	Nurse Manager vs. Nurse	0.2262 (0.058)	0.0099*
	CNS vs. Nurse	0.4291 (0.145)	0.0032*
	PHN vs. Nurse	0.1996 (0.209)	0.3395
	Midwife vs. Nurse	-0.0080 (0.148)	0.9571
	Education vs. Nurse	0.8494 (0.180)	<0.001*
Ethnicity	Non-Caucasian vs. Caucasian	0.55 (0.13)	<0.001*
Culture Type	Group vs. Developmental	-0.20 (0.08)	0.0083*

*Significant at 5% level.

For management type, the perceived level of formal power of respondents who work voluntarily or in an institution that is not publicly funded, is significantly higher than that of candidates who work for the HSE.

For job description, the perceived level of formal power of respondents who work as a nurse manager, a CSN or work in education, is significantly higher than respondents who are nurses.

For ethnicity, the perceived level of formal power for non-caucasians is significantly higher than caucasians.

For culture type, the perceived level of formal power of respondents' part of developmental culture is higher than respondent's part of groups.

The reference variable was changed and pairwise comparison was conducted. There was only one further comparison for management type which is seen in **Table 4.17**.

Table 417: Management Type Pairwise Comparison.

Characteristic	Pairwise Comparison	Regression Coefficient (S.E.)	P-Value
Management Type	Not Publicly Funded vs. Voluntary	0.38 (0.10)	0.0014*

The perceived level of formal power of respondents who worked in institutions that are not publicly funded is significantly higher than respondents who work in voluntary institutions.

The reference category for job description was changed. Ten pairwise comparisons are presented below.

Table 4.18: Job Description Pairwise Comparison.

Characteristic	Pairwise Comparison	Regression Coefficient (S.E.)	P-Value
Job Description	CNS vs. Nurse Manager	0.203 (0.155)	0.1911
	PHN vs. Nurse Manager	-0.267 (0.217)	0.9026
	Midwife vs. Nurse Manager	-0.234(0.158)	0.1401
	Education vs. Nurse Manager	0.623 (0.198)	0.0017*
Job Description	PHN vs. CNS	-0.230 (0.246)	0.3516
	Midwife vs. CNS	-0.437 (0.196)	0.0264*
	Education vs. CNS	0.420 (0.229)	0.0668
Job Description	Midwife vs. PHN	-0.208 (0.247)	0.2471
	Education vs. PHN	0.650 (0.275)	0.2750
Job Description	Education vs. Midwife	0.857 (0.231)	0.0002*

*Significant at 5% level.

The estimated average perception of formal power for respondents who work in education are significantly higher than respondents who were nurse managers. With those working in education also perceived to have a significantly higher level of formal power compared to midwives.

The estimated average perception of formal power for respondents who work as clinical nurse specialists are significantly greater than those who work as midwives.

4.4 Discussion

In this section, we discuss the results, observations, and inferences based on our results. The limitations of the study and future recommendations will also be addressed.

4.4.1 Formal Power Subscale

Cronbach's alpha was calculated for the formal power scale and based on DeVellis (1991) threshold, our value was in the recommended interval. Thus, the formal power scale was internally consistent. A mean value of 2.8 was achieved, the respondents' perceived perception of formal power was, on average, good. The mean score and standard deviation produced by our analysis is very similar to other studies conducted (Armstrong & Laschinger, 2006; Laschinger & Finegan, 2005; Khatun et al, 2020). The findings of this study were inline on an international level.

4.4.2 What Combination of Demographic, Job-Related and Work Environment Characteristics have a Significant Impact on a Respondents' Perception of Formal Power?

In the univariate analysis, there were fourteen characteristics considered. Nine of the demographic, job-related and work environment characteristics had significant differences for the average perceived formal power. Multiple linear regression analysis was conducted on these nine characteristics. These are (1) ethnicity; (2) gender; (3) job description; (4) shift schedule; (5) length of time in current place of employment; (6) management type; (7) religious affiliation; (8) culture type and (9) HSE location. The other five characteristics were not included in the multiple linear regression analysis. The multiple linear regression identified four characteristics as candidate variables. These variables are significant on respondents' perception of formal power. The four variables explained a significant amount of the variance. They were management type, ethnicity, job description and culture type.

The percentage of variance explained by these four variables is 11.2%. This value is low compared to a similar study where the model explained 42% of the variance (Laschinger, Wong, McMahon and Kaufmann, 1999). The study by Laschinger et al, (1999) had few respondents with 527 compared 885 in this study, however their respondents were all of the same job type and were all staff nurses, while the respondents in this study were from varying sectors of the nursing profession. The respondents from this study were from 45 different hospitals compared to only two hospitals in the other study. Having respondents from multiple different hospitals could be a reason why the model explains less variance. The experience they experienced could differ greatly between the hospitals and the different type of nurses. The low level of variance could suggest that there are other characteristics not included in this study, which explain a respondents' perception of formal power.

Formal power in the context of this study is (1) innovation on the job, (2) the amount of flexibility, and (3) the amount of visibility of work-related activities within the institution. The differences in the perception of formal power between management types shows that candidates from not publicly funded organisations have a higher perception of formal power compared to both the HSE and voluntary organisations. Additionally, respondents in voluntary organisations had a higher perception of formal power than candidates in the HSE. In a study in 2020 by Khatun et al, they found that work settings had a significant relationship to structural empowerment. Formal power scored the lowest out of all six structural empowerment characteristics in their study, with a mean score of 2.89, which it is similar to the mean score of this study. Work setting can impact formal power, for example it can dictate the level of flexibility or innovation of nurses. Management can change work settings by providing flexibility to nurses or encouraging innovation, this could increase the perception of formal power for nurses. A study in 2018 of the nurse empowerment environment between public and private hospitals showed higher levels of empowerment in private hospitals when compared to public hospitals (Pires et al. 2018). The study by Pires at al, (2018) highlighted characteristics similar to formal power in the context of this study. Key favourable areas were control over their environment, which could relate to the amount of flexibility within the workplace. Positive relationships among nurses and physicians in private institutions compared to public setting. The positive relationships could promote a nurse being central and visible to an organisational purpose and goals. Another study conducted by Scot et al. (2003) identified a number of similar characteristics to inhibit empowerment. The study by Scot et al. (2003) found that education before commencing in a health care institution was necessary to be empowered but not essential in their study. Respondents who experienced professional respect, personal control and power, inhibited empowerment. These characteristics are similar to characteristics in this study, having autonomy or flexibly, control over their own role. Professional respect would be important for positive relationships with management, which are positive for empowerment. Respondents from not publicly funded institutes have a higher level of empowerment when compared to those in publicly funded organisations. Having a higher level of empowerment could be due to relationships between management and nurses or down to flexibility or own autonomy in the workplace in the private sector. Positive experiences in these areas would result in a positive impact on respondents' level of empowerment. Given the characteristics central to inhibit empowerment are the similar for formal power, this may result in a higher level of formal power.

Respondents who are with voluntary organisations may have a higher level of flexibility when compared to publicly funded organisations, given they are volunteering. At the time of the study there was a recruitment freeze in the HSE and there were fresh calls to remove the freeze, improve funding, and increase and improve education and staffing levels in 2019 (Oireachtas.ie, 2019). Inappropriate staffing levels will impact staff flexibility in the work environment and could also impact internal relationships. This negative effect could lead to staff burnout and overall retention in the workplace, as seen in another study by Lachinger 2009, which is a problem for the HSE (Oireachtas.ie, 2019). A lack of flexibility and poor relationships within the organisation could contribute to a lower level of perceived empowerment for respondents. Furthermore, if there are staff shortages, this can lead to existing staff being overwhelmed with work and could reduce the level of care for patients. Staff shortages can reduce flexibility, which can have a negative impact on relationships between nurses and patients, and nurses and management. Reduced flexibility could lead to a reduced level empowerment for respondents and in particular respondents who are part of a publicly funded institute (Pires et al. 2018, Amiri, Khademian and Nikandish, 2018). Greater emphasis on adequate nursing staff levels to properly complete all job requirements should be monitored, it is clear that nurses play a crucial role in patient care (WHO, 2010). With more balance in the staffing levels, organisational goals and purpose become more transparent, along with greater flexibility for staff.

The difference in the perception of formal power between job descriptions shows that respondents in positions of higher authority had a higher perception of formal power, compared to those who were not. This is expected given one of the key ideas of formal power is being central to an organisation's goals and priorities (Laschinger, 2009). Two groups, nurses in education and clinical health nurses showed particularly high levels of formal power compared to other groups. A similar study, which examined empowerment of a group of nurses from different departments from a hospital in Brazil showed the difference in formal power scores between the departments (Moura et al., 2020). The formal power mean score was lower than this study at 2.59. The activities in each department differed for the nurses, this would suggest a nurse job description impacts formal power and likewise structural empowerment for nurses. The level of education and training of the nurses in the study varied, close to 59% of the nurses were specialised and all had third level education. In a study by Scott et al, they identify factors contributing to empowerment. Education is the highest factor on that list, this could be one of the main factors nurses involved in education and clinical health experience higher levels of formal power. Education nurses who had the highest level of perceived formal power may feel they have more innovation in their work as they can dictate what they are teaching, they may have increased flexibility in their work or working hours, however also fully understand their position within the organisation's goals and purposes, which are all central to formal power. Nurses and midwives have the lowest levels of perceived formal power, these respondents are also educated and skilled, but their training or education may not have continued. A study conducted in 2016 by Kukkanen et al, showed graduate nurses had high levels of empowerment at the beginning of their career but need support and career opportunities to contribute to higher levels of empowerment. Nurses are the largest subgroup in the study, they may have limited ability for innovation in their work or flexibility given the challenges in staffing levels

The difference in the perception of formal power between ethnicity types indicated that non-caucasians had a significantly higher perception of formal power than caucasians. A study conducted by the Royal College of Surgeons in Ireland indicated that Ireland depends highly on migrant nurses (Humphries, Brugha & McGee, 2009). The majority of these nurses come from non-EU countries such as India and the Philippines. Ethnicity or where a respondent is from could imply that perception of formal power is cultural. In Indian business, they are often very hierarchally structured, with decision makers at the top feeding down, this could be the case in a hospital setting (Business Culture in India – UK India Business Council, 2021). Thus, respondents from these countries could, in general, give greater respect to authority compared to their Irish counterparts. Respondents from these countries could be more likely to not speak negatively of their work situation. The caucasian group were the dominant group in the study, a qualitative study could be conducted to see why they have a lower perception of formal power. Training and education promote empowerment as a whole and formal power. Perhaps respondents from these areas receive better training or an understanding of how they fit or are central into an organisation's goals and purpose, which promote empowerment and perceived formal power (Scott et at 2003, Laschinger 2009).

The difference in the perception of formal power between culture types, indicated that developmental culture had a significantly higher perception of formal power than group culture. This is to be expected, developmental culture focuses on innovation and change, these form one of the pillars of formal power. To ensure individuals feel they have formal power, organisations will need to change to help their staff. Due to budgetary constraints, increasing nursing levels may not be feasible, however they could adapt a new culture type, in this case developmental culture. Here, through new guidance and direction, each individual may feel more connected with the institute and get the recognition they deserve.

4.4.3 Limitations and Recommendations

Self-reporting questionnaires have many advantages and disadvantages. Respondents can answer them quickly and they can be designed to ensure anonymity and privacy of the respondent. Unfortunately, for postal surveys and online surveys, respondents can answer questions to make them shine in a better light or in a way they feel the author of the study wants them to (Heppner et al., 1988). Collecting the data from a post survey is more tedious compared to an online survey. Survey Monkey now allows the user to view their data and responses in an easy manner and run analysis. The author of the study has to bear the cost and ultimately has a risk with the returned information containing errors (Ilieva, Baron & Healey, 2002).

This study was completed through a postal survey, this method of collection can lead to a drop off in response rates. When this study was conducted, it could not use an online survey as there were no e-mail address available. The response rate for this study was 22%. A similar study conducted online had a 30% response rate (Pires et al, 2016). A future study should consider using a larger nurse register such as the INMO and be completed online. Future studies should also consider the use of incentives for completion, which could increase the response rate. There are pros and cons to using incentives to increase response rates, they increase the cost to the researcher, however some incentives can increase response rates while others are still unproven (Survey Monkey, 2021). Cash incentives are seen to increase responses rates but have a direct cost to the researcher. Another study suggested using observational data collection methods or potentially viewpoints from other parties for example patients (Amiri, Khademian and Nikandish, 2021). In the context of this study, the satisfaction levels of the patients could be compared between the different nurse types and their level of empowerment. A follow up communication and survey should have been sent to candidates to try to boost the number of respondents. Heppner et al. (1988) suggests that response rates can be increased by 10-15% by a follow up questionnaire. Follow up communications should be incorporated into any further study. Online providers such as Survey Monkey, allow users to set up automatic reminders.

Only 11% of variance was explained using the four variables in the multiple linear regression model, therefore there must be additional variables not captured in the study, which explains the perception of formal power. Other studies explored other factors, which may have explained empowerment Scott et al (2003), for example, in that study the highest level of education, years in education and job satisfaction were asked. In future studies, improvements to the questionnaire could be made with the aim of increasing the level of variance explained. To increase the level of variance explained additional questions could be asked, which include the factors that are outlined in Scott et al (2003). Qualitative questions could be included to examine the context of formal power for example, what elements of the definition of formal power are significant when it comes to ethnicity, formal power in this study is described as (1) innovation on the job, (2) the amount of flexibility, and (3) the amount of visibility of work-related activities within the institution. Where and how do these differ between the caucasians and non-caucasians?

When constructing the instrument for future surveys, the ranges responses are grouped in should be examined for example, the question for years of qualifications groups the responses into 3 different groups. The first group spans 30 years, the second 6 years and the third 2 years. Further studies could consider more suitable options or there should be information provided about why these groupings are appropriate. Another study, which asked a similar question about work experience, split the responses into 5-year ranges (Khatun, 2001). An assessment with equal categories should be considered. Different options for age ranges for years since receiving a qualification may provide a better insight into the difference in respondent's perception of formal power. More work experience has a positive relationship with structural empowerment (Khatun, 2001), examining the differences in perceived formal power across length of time working and job type could be interesting.

4.5 Conclusion

The aim of this study was to identify which demographics, job-related and working environment characteristics influence nurses' perception of formal power. The study found that respondents perceived on average to have a moderate level of formal power. Four of fourteen characteristics were found to be significantly related to the perception of formal power. The characteristics were ethnicity, type of job, the type of management they worked for, and the culture perceived in the workplace.

The mean score for formal power was in-line with other studies. The level of variance explained by these four characteristics is quite low when compared to other studies. One study explained nearly three times the variance with their model, this would suggest that there are other variables, which could be included in the model. Another study identified other significant characteristics that positively contributed to empowerment (Scot et al, 2003). Some of the characteristics in the study by Scot et al, (2013) include job satisfaction and the number of years in education. This study used respondents from a large number of hospitals and across different nurse types. Perhaps concentrating on one nurse type or fewer hospitals could provide different results.

Nurses in education had the strongest level of perceived formal power when compared with other respondents. Respondents who worked in public hospitals had lower levels of perceived power. Non-Caucasians had a higher level of perceived formal power when compared to Caucasians. Respondents in developmental culture type had a higher level of perceived formal power compared to groups.

It is recommended that a further study be conducted, a study that explores qualitative and quantitative questions. Further studies should consider other characteristics, other studies found significant to empowerment. The future study could look for qualitative responses around the difference in level of formal power between Caucasians and non-Caucasians. The survey could be conducted online to reach a wider audience if possible. Other studies have explored the relationship between empowerment and job satisfaction. Understanding respondents' level of job satisfaction could provide some further insight into the levels of perceived formal power.

There are a wide range of factors impacting structural empowerment, this study has examined formal power. It would be interesting to see where formal power ranks in the 6 scales that make up structural empowerment. Improving formal power can occur through initiatives in university and in the workplace. It is in the interest of the hospital management to increase levels of empowerment as whole with a focus on formal power, an increase in formal power could have wider positive effects as it can lead to an increase job satisfaction and patient care.

4.6: References

Aiken, L., Smith, H., &Lake, E. (1994). Lower medicine mortality among a set of hospitals known for good nursing care. *Medical Care*, 32(8), 771-787.

Amiri, M., Khademian, Z. and Nikandish, R., 2021. *The effect of nurse empowerment educational program on patient safety culture: a randomized controlled trial*.

Benner, P. (1984). *From novice to expert*. Menlo Park, CA: Addison-Wesley Publishing.

Chaguturu, S., & Vallbhaneni, S. (2005). Aiding and abetting – nursing crises at home and abroad. *The New England Journal of Medicine*, 353(17), 1761-1763.

Cho, J., Laschinger, H., & Wong, C. (2006). Workplace empowerment, workplace engagement and organisational commitment of new graduate nurses. *Canadian Journal of Nursing Leadership*, 19 (3), 43-60.

Cso.ie 2021 Health – CSO – Central Statistics Office. [online] Available at: https://www.cso.ie/en/releaseandpublications/ep/p-syi/statisticalyearbookofireland2020/soc/health/

Department of Health and Children. Retrieved on 25 August,2010, from http://www.dohc.ie/publications/health_in_ireland2007.html

Faulkner, J. & Laschinger, H. (2008). The effects of structural and psychological empowerment on perceived respect in acute care nurses. *Journal of Nursing Management*, 16 (2), 214-221.

Greco, P., Laschinger, H. & Wong, C. (2006). Leader empowering behaviours, staff nurse empowerment, and work engagement/burnout. *Canadian Journal of Nursing Leadership*, 19 (4), 41-56.

Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (2006). *Multivariate Data Analysis* (6th ed.), London, Pearson Prentice Hall International.

Hatcher, S., & Laschinger, H. (1996). Staff nurses' perceptions of job empowerment and level of burnout: A test of Kanters' theory of structural power in organisations. *Canadian Journal of Nursing Administration,* 9(2), 74-94.

Heppner, P.P., Kivlighan, D.M., & Wampold, B.E. (1998). *Research design in counselling* (2nd ed.). California: Wadsworth.

Humphries, N., Brugha, R. & McGee, H. (2009). A profile of migrant nurses in Ireland: Nurse migration project policy brief 4. Dublin Royal College of Surgeons, Ireland.

Ilieva, J., Baron, S. & Healey, N. (2002). Online survey in marketing research: Pros and cons. *International Journal of Market Research*, 44(3), 361-376.

Irish Department of Health and Children. (2007). *Health in Ireland: Key trends*. Dublin: Irish.

Kanter, RM. (1977). *Men and Women of the Corporation*. New York: Basic Books.

Kanter, RM. (2003). *Men and Women of the Corporation.* 2nd ed. New York: Basic Books.

Kramer, M. (1990). The magnet hospitals-excellence revisited. *Journal of Nursing Administration,* 20 (9), 35-44.

Kramer, M., & Schmalenberg, C. (1988). Magnet hospitals: part 1Institutions of excellence. *Journal of Nursing Administration*, 18 (1), 13-24.

Laschinger, H. (2008). Effect of empowerment on professional practice environments, work environments, and patient care quality: Further testing the worklife model. *Journal of Nursing Care Quality*, 23 (4), 322-330.

Laschinger, H., & Wong C. (1999). Staff nurse empowerment and collective accountability: Effect on perceived productivity and self-rated work effectiveness. *Nursing Economics*, 17 (6), 308-316.

Laschinger, H., Finnegan, J., Shamian J. & Wilk, P. (2004). A longitudinal analysis of the impact of workplace empowerment on work satisfaction. *Journal of Organisational Behaviour*, 25 (4) 527-545.

Laschinger, H.K., Wong, C., McMahon, L., & Kaufmann, C. (1999). Leader behaviour impact on staff nurse empowerment, job tension and work effectiveness. *Journal of Nursing Administration,* 29 (5), 28-39.

Laschinger, H., Shamian, J & Thomson, D. (2001a). Impact of magnet workplace characteristics on nurses perceptions of trust, burnout, quality of care, and work satisfaction. *Nursing Economics*, 19(5), 209-219.

Laschinger H., & Havens D. (1996). Staff nurse work empowerment and perceived control over nursing practice. Conditions for work effectiveness. *Journal of Nursing Administration,* 26(9), 27-35.

Laschinger, H., Almost, J., & Tuer-Hodes, D. (2003). Workplace empowerment and Magnet workplace characteristics: Making the link. *Journal of Nursing Administration*, 33(7/8), 410-422.

Laschinger, H. (n.d.). The University of Western Ontario: Publish Web Server. Retrieved from http://publish.uwo.ca/~hkl/instrumentCWEQ.html

Moura, L., Camponogara, S., Santos, J., Gasparino, R., Silva, R. and Freitas, E., 2020. Structural empowerment of nurses in the hospital setting. Revista Latino-Americana de Enfermagem, 28.

OECD (2008). The Looming Crises in the Health Workforce: How Can OECD Countries Respond?

Data.oireachtas.ie. 2021. [online] Available at: <https://data.oireachtas.ie/ie/oireachtas/committee/dail/32/joint_committee_on_health/submissions/2019/2019-11-13_opening-statement-phil-ni-sheaghdha-general-secretary-irish-nurses-and-midwives-organisation-inmo_en.pdf> [Accessed 6 July 2021].

Patrick, A. & Laschinger, H. (2006). The of structural empowerment and perceived organization support on middle 'level nurse managers' role satisfaction. *Journal of Nursing Management*, 14, 13-22.

Pires, B., Oliveira, L., Siqueira, C., Feldman, L., Oliveira, R. and Gasparino, R., 2018. Nurse work environment: comparison between private and public hospitals. Einstein (São Paulo)*, 16(4).*

Scott, A., Matthews, A. & Corbally M. (2003). *Nurses' and midwifes understanding and experiences of empowerment in Ireland* (ISBN 1-873820-24-0). Dublin, Irish Department of Health and Children.

Sheskin, D.J. (2004). *Handbook of parametric and non-parametric statistical procedures* (3rd *ed.)*. United States of America: Chapman & Hall/CRC.

Shortell, S. M., O'Brien, J. L., Carman, J. M., Foster, R. W., O'Connor, E. J., Hughes, E., &

Shortell S.M., Jones R.H., Rademaker A.W., Gillies R.R., Dranove D.S., Hughes E.F., Budetti P.P.,

Tukey, J. W. (1953). *The problem of multiple comparisons*. Unpublished paper, Princeton University, Princeton, New Jersey.

Upenicks VV. (2003). Nurse perceptions of job satisfaction and empowerment: Is there a difference between nurses employed at Magnet hospitals versus non-magnet Hospitals? *Nursing Management*, 34(2), 43-44.

World Health Organisation. (2010). Retrieved November 2, 2010 from http://www.who.int/hrh/nursing_midwifery/leaflet.pdf

Zammuto, R. F., &. Krakower J. Y. (1991). Quantitative and qualitative studies of organizational culture. *Research in Organizational Change and Development*, 5, 83-114.

Appendix 4.A

Questions that make up CWEQ-II

(A)-Opportunity

Challenging work,

The chance to gain new skills and knowledge on the job,

Tasks that use all your skills and knowledge.

(B)-Information

The current state of the hospital,

The values of top management,

The goals of top management.

(C)-Support

Specific information about things you do well,

Specific comments about things you could improve,

Helpful hints or problem-solving advice.

(D)-Resources

Time available to do necessary paperwork,

Time available to accomplish job requirements,

Acquiring temporary help when needed.

(E)-Formal

The rewards for innovation on the job are,

The amount of flexibility in my job is,

The amount of visibility of my work-related activities within the institution is.

(F)-Informal

Collaborating on patient care with doctors,

Being sought out by peers for help with problems,

Being sought out by managers for help with problems,

Seeking out ideas from professionals other than doctors e.g. Physiotherapists Occ. Therapists, Dieticians.

Appendix 4.B

Questions that make up Hospital Culture

Hospital Character

This hospital is a very personal place. It is a lot like an extended family. People seem to share a lot of their private lives.

This hospital is a very dynamic and entrepreneurial place. People are willing to stick their necks out and take risks.

This hospital is a very formalized and structured place. Bureaucratic procedures generally govern what people do.

This hospital is very production oriented. Getting the job done is a major concern. People are not very personally involved.

Hospital Managers

Managers in this hospital are warm and caring. They seek to develop employees' full potential and act as their mentors and guides.

Managers in this hospital are risk-takers. They encourage employees to take risks and be innovative.

Managers in this hospital are rule-enforcers. They expect employees to follow established rules, policies, and procedures.

Managers in this hospital are coordinators and coaches. They help employees meet the hospital's goals and objectives.

Hospital Cohesion

The glue that holds this hospital together is loyalty and tradition. Commitment to this hospital runs high.

The glue that holds this hospital together is commitment to innovation and development.

The glue that holds this hospital together is formal rules and policies. Maintaining a smooth-running operation is important here.

The glue that holds this hospital together is the emphasis on tasks and goal accomplishment. A production orientation is commonly shared.

Hospital Emphases

This hospital emphasises human resources i.e. values its staff. High cohesion and morale in the organisation are important.

This hospital emphasises growth and acquiring new resources. Readiness to meet new challenges is important.

This hospital emphasises permanence and stability. Efficient, smooth operations are important.

This hospital emphasises competitive actions and achievement. Measurable goals are important.

Hospital Rewards

This hospital distributes its rewards equally among its members. It is important that everyone from top to bottom be treated as equally as possible.

This hospital distributes its rewards based on individual initiative. Those with innovative ideas and actions are most rewarded.

This hospital distributes rewards based on rank. The higher you are, the more you get.

This hospital distributes rewards based on the achievement of objectives. Individuals who provide leadership and contribute to attaining the hospital's goals are rewarded.

Appendix 4.C

An assumption in regression analysis is the normality of the residuals. **Figure 4.1** shows the normal probability plot.

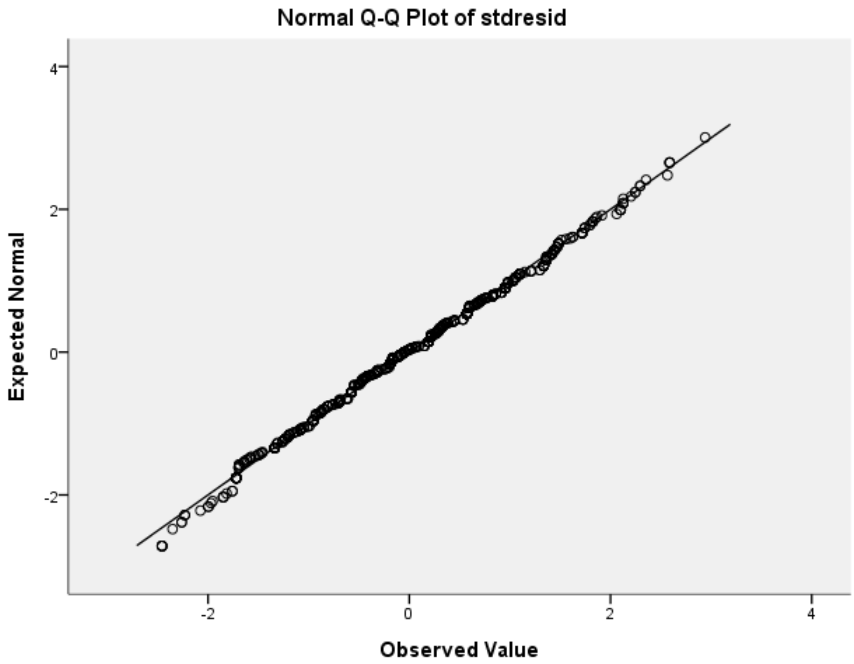

Figure 4.1 Normal Probability Plot for the Final Multiple Linear Regression Model.

There are no major deviations from normality of the residuals in the final multiple regression model.

Figure 4.2 shows the plot of the standardised residuals against the predicated values. This validates the constant variance assumption.

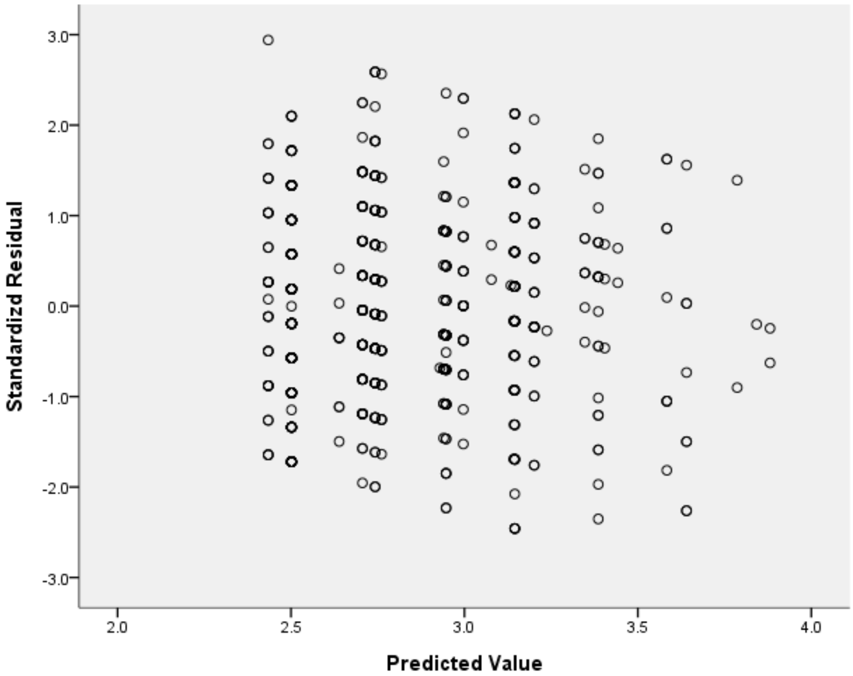

Figure 4.2 Plot of Standardised Residuals versus the Predicated Values.