Statistical modelling to determine the influence of vaccine dose and prior *Mycobacterium tuberculosis* exposure on antigen-specific T cell responses

Author: Kelly A. Williams

Supervisors: Prof. Francesca Little
Associate Prof. Elisa Nemes

Co-supervisor: Dr Anele Gela

Dissertation submitted to the Department of Statistical Sciences at the University of Cape Town in partial fulfillment of the requirements for the degree of Master of Science in Biostatistics.

April 1, 2024

In collaboration with the South African Tuberculosis Vaccine Initiative.

Declaration

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.
2. I have used the IEEE convention for citation and referencing. Each contribution to, and quotation in, this report from the work(s) of other people has been attributed, and has been cited and referenced.
3. This report is my own work.
4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as their own work or part thereof.

Signature: .....
K. Williams

Date: April 1, 2024

Abstract

This dissertation investigates the effects of two subunit vaccines H1:IC31 and H56:IC31 as well as prior *M.tb* sensitization on the immune responses of three cohorts of South African adolescents and adults. The primary outcomes are frequencies of antigen-specific CD4 T cells expressing different combinations of immunological markers over three time points. Two *M.tb* antigens are investigated: Ag85B and ESAT-6. The dissertation compares the results produced by the standard procedures that would typically be employed in the immunology research community to investigate these aims with the results produced by employing a mixed effect modelling approach. Not only is it of interest to investigate whether the results agree, but also to investigate the difference in inference that one can make and whether the mixed effect modelling approach is able to provide greater insight into the data. Methods typically employed by the immunology community that are used in this thesis are non-parametric pair-wise tests and the data analysis pipelines mixture models for single-cell assays (MIMOSA) and combinatorial polyfunctionality analysis of single cells (COMPASS). For the mixed effect modelling approach, generalized linear mixed effect models with various hierarchical structures as well as latent variable models are employed. Results suggest that 5 µg of the vaccine induces the strongest immune response. The mixed effect modelling approach showed good potential in terms of depth of analysis and ease of interpretation, however many model assumptions were violated making inference difficult. The standard approaches where much more cumbersome to implement and interpret and resulted in significant multiple testing concerns.

Contents

List of Figures

List of Tables

List of Abbreviations

Ag85B	Antigen 85B
AIC	Akaike information criterion
ANOVA	Analysis of variance
AUC	Area under the curve
BCG	Bacille Calmette-Guérin
BIC	Bayesian information criterion
COMPASS	Combinatorial polyfunctionality analysis of single cells
CCR7	CC-chemokine receptor 7
EM	Expectation-maximization
ESAT-6	Early secretory antigen target-6
FDR	False discovery rate
FS	Functionality score
GAIC	Generalised Akaike information criterion
GAM	Generalised additive model
GAMLSS	Generalised additive models for location, scale and shape
GLM	Generalised linear model
GLMEM	Generalised linear mixed effect model
IFN-γ	Interferon-gamma
IGRA	Interferon-gamma release assay
IL-2	Interleukin-2
IL-17	Interleukin-17
LM	Linear regression model
LMEM	Linear mixed effect model
LVM	Latent variable model
MAD	Median Absolute Deviation
MCMC	Markov chain Monte Carlo
MEER	Maximum experiment-wise error rate
MIMOSA	Mixture models for single-cell assays
ML	Maximum likelihood
M.tb	*Mycobacterium tuberculosis*
PFS	Polyfunctionality score
PHA	Phytohaemagglutinin
PID	Participant identification number
QFT	QuantiFERON-TB Gold
qq-plots	Quantile-quantile plots
REML	Restricted maximum likelihood
SATVI	South African Tuberculosis Vaccine Initiative
SSI	Statens Serum Institute

TB	Tuberculosis
T_{CM}	Central memory T cells
T_{EM}	Effector memory T cells
Th1	T helper 1
T_N	Naive T cells
TNF	Tumour-necrosis factor
TRF	Total response frequency
Trial A	H1-THYB04
Trial B	H56-032
Trial C	H56-03
T_{TE}	Terminal effector T cells
TVR	Total vaccine response
UN	United Nations
WHO	World Health Organisation
ZIP	Zero-inflated Poisson
ZINB	Zero-inflated negative binomial

Chapter 1

Introduction

1.1 Introduction

Tuberculosis (TB) is currently the second leading cause of death from a single pathogen globally, with The World Health Organisation (WHO) estimating a death toll of 1.6 million during 2021 [1]. The WHO estimated that 10.6 million people fell ill with TB in 2021, the vast majority of whom were adolescents and adults (approximately 89%). Both the WHO and the United Nations (UN) have set goals of ending TB by 2030 and 2035 respectively. Interrupting the transmission of *M.tb* is key to eliminating TB disease and meeting these targets. This can be achieved through the development of novel drugs, better diagnostic tests and more effective vaccines for TB [2].

TB is an airborne disease caused by the bacteria *Mycobacterium tuberculosis* (*M.tb*) [1]. When an individual has contracted *M.tb* infection, they can remain healthy and non-infectious, or progress to TB disease, which may be symptomatic and infectious. *M.tb* can also disseminate to any part of the body, and when it causes disease in the lungs it is referred to as pulmonary TB and is considered infectious [3]. Extra-pulmonary TB (i.e. *M.tb* that has disseminated to areas other than the lungs) is considered non-infectious. About 70% of symptomatic TB cases are pulmonary and thus infectious [4]. Transmission can occur when an individual with pulmonary TB expels *M.tb* by coughing, sneezing or spitting, making pulmonary *M.tb*-infected individuals the reservoir for *M.tb*. The WHO estimates that an individual with pulmonary TB can infect 5-15 susceptible individuals per year through close contact [1]. Adolescents and adults with pulmonary TB are primarily responsible for transmission of *M.tb* and thus are the demographic that should be targeted when developing interventions to stop the transmission of TB [5].

TB disease is both treatable and curable, although it is estimated that without proper treatment, 45% of HIV-negative individuals with active TB and nearly all HIV-positive individuals with active TB would die [1]. The standard treatment of drug-susceptible TB disease is a 6 month course of four antimicrobial drugs. However, there are often delays in treatment-seeking due to symptoms frequently being mild for the first few months. This can result in transmission from the unknowingly infected individual to others, furthering the spread of TB despite eventual treatment of the disease. Once an individual has recovered, they are not necessarily immune and they may develop TB again via exogenous *M.tb* infection (progression upon re-infection) or via endogenous *M.tb* infection (progression of the existing infection, also known as reactivation or relapse) [6]. Thus treatment of the disease alone is not enough to interrupt the transmission of TB disease and preventative steps such as vaccines are necessary as well.

Bacille Calmette-Guérin (BCG) is the currently licensed vaccine against TB and has been administered to infants at birth for over 100 years. BCG primarily protects children against TB meningitis and miliary disease [7]. In South Africa, BCG is administered at birth with a

coverage of 89% among 1-year-olds as of 2021 [8]. However, there is evidence that by adolescence BCG provides inconsistent protection against pulmonary TB [9]. There is also limited evidence that the BCG (re-)vaccination of adolescents and adults is effective in preventing TB [10]. This is significant because adolescents and adults with pulmonary TB are the primary transmitters of TB. Thus, a vaccine that is effective in both adolescents and adults is key to interrupting the transmission of TB disease [11, 12].

The apparent ineffectiveness of BCG re-vaccination has been theorised as being attributed to prior *M.tb* sensitisation and has lead to the masking and blocking hypotheses. The masking hypothesis suggests that environmental sensitisation to *M.tb*, or closely-related non-tuberculous mycobateria, provides enough protection against TB such that BCG re-vaccination barely offers any additional protection. The blocking hypothesis suggests that environmental sensitisation to *M.tb* prior to vaccination may trigger an immune response capable of blocking the replication of the BCG vaccine or accelerating its clearance before an effective immune response can be stimulated [13]. In TB endemic countries such as South Africa, it is estimated that more than three-fourths of the adult population are *M.tb*-infected [14]. Thus both hypotheses imply that in the context of South Africa, a vaccine targeted at adolescents and adults would need to be effective regardless of *M.tb* sensitisation.

A key component of an individual's immune response to *M.tb* is the recognition of *M.tb* antigens by the individual's immune system. *M.tb* antigens are molecular structures expressed by the pathogen and are the key ingredient in *M.tb* vaccine development [15]. Two immuno-dominant *M.tb* antigens included in the candidate vaccines evaluated in this project are antigen 85B (Ag85B) and early secretory antigen target-6 (ESAT-6). Ag85B is expressed by both BCG and M.tb, and therefore the immune recognition of this antigen is expected in both *M.tb* unsensitized as well as sensitized individuals before vaccination. On the other hand, ESAT-6 is expressed by *M.tb* and not BCG, therefore the immune recognition of this antigen is only expected in *M.tb* sensitized individuals before vaccination.

This project aims to investigate the effects of two subunit vaccines H1 and H56, administered with IC31 adjuvant, on the immune responses of three cohorts of adolescents and adults, the target demographic for preventing TB transmission. It will also investigate the effect of prior *M.tb* sensitisation on the immune responses of participants, thus investigating whether the vaccines would be immunogenic in an TB endemic country such as South Africa. Since vaccine regimen is a key factor for determining vaccine effectiveness, it is also of interest to investigate the effect of vaccine regimen, specifically vaccine concentration and number of administrations, on the immune response. The immune response shall be quantified by analysing the antigen-specific CD4 T cell magnitude, functional quality and memory profiles of the participants, where the antigens of interest are Ag85B and ESAT-6.

1.2 Aims and objectives

There are three specific, immunological aims for this study:

1. To investigate the effects of vaccine regimen, *M.tb* sensitisation and antigen specificity on CD4 T cell *magnitude*.

2. To investigate the effects of vaccine regimen, *M.tb* sensitisation and antigen specificity on CD4 T cell *functional quality*.

3. To investigate the effects of vaccine regimen, *M.tb* sensitisation and antigen specificity on CD4 T cell *differentiation*.

In the context of this project, vaccine regimen is determined by the concentration and the number of administrations of the vaccine. QuantiFERON-TB Gold (QFT) status shall be used as a proxy for *M.tb*-sensitisation, where QFT negative (QFT-) individuals are assumed to be *M.tb*-unsensitised and QFT positive (QFT+) individuals are assumed to be *M.tb*-sensitised. Furthermore, the antigens of interest in this project are Ag85B and ESAT-6.

With regards to the first aim, the magnitude of the T cell response shall be quantified by the number of *cytokine positive* cells observed. In the context of this project, cytokine positive cells are CD4 T cells which express at least one of the following cytokines (also known as functional markers): interleukin-2 (IL-2), interleukin-17 (IL-17), tumour-necrosis factor (TNF) and interferon-γ (IFN-γ). CD4 T cells generally only express cytokines after antigen recognition, thus cytokine positive cells can be classified as antigen-specific cells. Consequently, measuring the number of cytokine positive cells after in vitro stimulation is roughly equivalent to measuring the number of antigen-specific cells, which is why it can be used to quantify the immune response. It is hypothesised that *M.tb*-sensitised individuals will require a lower vaccine concentration to induce the greatest magnitude of cytokine positive CD4 T cells compared to *M.tb*-unsensitised individuals. This hypothesis shall be investigated by analysing the relationship between QFT status, vaccine concentration, number of vaccine administrations and the frequencies of cytokine positive CD4 T cells.

For the second aim, T cell functional quality refers to the pattern of cell functional marker expression. Thus, the frequency of cells expressing specific combinations of functional markers is of interest rather than simply the frequency of cytokine positive cells. It has been shown that better quality T cell responses are associated with disease non-progression and protection, where cells with better quality responses are those with greater degrees of functionality (i.e. express more functional markers) [16]. It is hypothesised that T cell functional quality is influenced by vaccine concentration and *M.tb* sensitisation and that high frequencies of T cells expressing IL-2 and TNF together and T cells expressing IL-2, TNF and IFN-γ together are associated with lower concentrations of the vaccine in *M.tb*-sensitised individuals compared to unsensitised individuals. This will be investigated by analysing the relationship between QFT status, vaccine regimen and the frequencies of antigen-specific T cells producing all functional markers measured on the dataset. It is also hypothesised that the functional profiles of T cells recognising Ag85B are distinct from those of T cells recognising ESAT-6 and that after vaccination, the functional profiles of T cells recognising Ag85B or ESAT-6 differ between *M.tb*-sensitised and *M.tb*-unsensitised individuals. This shall be investigated by comparing the

functional profiles of Ag85B- and ESAT-6-specific T cells induced by different vaccine regimens and analysing their relationship with QFT status.

With regards to the third aim, it is hypothesised that a high vaccine concentration in *M.tb*-sensitised individuals drives T cells towards a more differentiated phenotype compared to unsensitised individuals. A cell has a more differentiated phenotype compared to another cell depending on the specific combination of memory markers expressed by that cell. In the context of this project, the memory markers that will be measured are CCR7 and CD45RA. This hypothesis shall be investigated by measuring the frequency of antigen-specific (i.e. cytokine positive) T cells expressing different combinations of memory markers and analysing the relationship between this and vaccine dose as well as QFT status. It is also hypothesised that the differentiation profiles of T cells recognising Ag85B are distinct from those of T cells recognising ESAT-6 and that after vaccination, the differentiation profiles of T cells recognising Ag85B or ESAT-6 differ between *M.tb*-sensitised and *M.tb*-unsensitised individuals. This shall be investigated by comparing the differentiation profiles of Ag85B- and ESAT-6-specific T cells induced by different vaccine regimens and analysing their relationship with QFT status.

The statistical aim is to compare the results produced by the standard procedures that would typically be employed in the immunology research community to investigate these aims with the results produced by employing a mixed effect modelling approach. Not only is it of interest to investigate whether the results agree, but also to investigate the difference in inference that one can make and whether the mixed effect modelling approach is able to provide greater insight into the data.

Chapter 2

Literature review

2.1 Progression of *M.tb* infection

After inhalation of *M.tb* bacteria, the bacteria can either by eliminated by the body's immune system or, in the case of pulmonary TB, it can disseminate to the lungs and infect alveolar macrophages, a type of white blood cell located in the pulmonary alveoli that encapsulate and kill microorganisms, remove dead cells and enhance the action of other immune cells [17]. Later, it may disseminate to other myeloid cells, establishing infection [18, 19]. The *M.tb* bacteria employs several tactics to allow sufficient time for it to establish a chronic infection in the lungs before an effective T cell response can be mounted. One way that it does this is by delaying the transportation of the *M.tb* bacteria and its products to the regional lymph nodes, thus delaying the priming and migration of T cells to the infection site, consequently delaying the body's immune response [20]. In order to launch an effective immune response, T cells need to traffic to the infected tissues timeously where they must be able to recognise and respond to the *M.tb* antigens. Thus, in order for a TB vaccine to be effective, antigen-specific T cells induced by vaccination should be able to migrate to the lungs and mount an effective immune response against *M.tb* [21].

Once T cells have reached the infection site, they play a key role in granuloma formation [22]. Granulomas are organised and localised masses of immune cells containing infected macrophages, stimulated macrophages, lymphocytes and other immune cells [23]. The inability of granulomas to contain the *M.tb* bacteria can lead to *M.tb* dissemination and thus onset of active disease [24]. Thus, granulomas are important for controlling and containing *M.tb* infection [23]. CD4 T cells expressing TNF have been shown to play a critical role in granuloma formation, while CD4 T cells expressing IFN-γ have been shown to play an important role in containing the *M.tb* bacteria within the granuloma and infected macrophages [25, 26, 27, 28].

2.2 TB vaccines

The two subunit vaccines used in this project, H1 and H56, were both developed as pre- and post-exposure vaccines specifically targeted at TB endemic populations. Both vaccines are formulated with a Th1-priming adjuvant IC31. The role of an adjuvant is to enhance the immunogenicity of a vaccine and hence achieve a more durable immune response [29]. Adjuvant IC31 strengthens the induced immune response by promoting differentiation of naive CD4 T cells into T helper 1 (Th1) cells (i.e. CD4 T cells expressing IFN-γ, TNF and/or IL-2) which have been shown to play an essential role in preventing the rapid progression of *M.tb* infection [30, 20].

Both vaccines contain protein antigens ESAT-6 and Ag85B, however H56 contains an extra

antigen, Rv2660c. Antigens are molecular structures that are recognized by the immune system, and a vaccine includes antigens that are specific for the target pathogen [31]. Ag85B is present in both BCG and *M.tb*, however both ESAT-6 and Rv2660c are only present in *M.tb* [32]. Ag85B has been shown to be highly expressed during the acute phase of *M.tb* infection as well as frequently recognized by Th1 cells, thus making it highly immunogenic [33]. ESAT-6, on the other hand, is expressed throughout *M.tb* infection and has been shown to be highly immunogenic in human and animal models [34, 35]. Rv2660c is a hypothetical *M.tb* antigen whose transcripts are upregulated during nutrient starvation, however its exact function is unknown [36]. For this project, both vaccines H1 and H56 will be treated as equivalent since only the Ag85B and ESAT-6 stimulated response, the two antigens present in both vaccines, will be investigated and not the Rv2660c stimulated response.

While both functional antigen-specific CD4 T cells and CD8 T cells are induced by *M.tb* infection, the majority of TB vaccines, including those studied in this project (H1 and H56), primarily induce CD4 responses [37]. As previously mentioned, CD4 cells producing Th1 cytokines are key immune players against *M.tb* infection. Specifically, Th1 cells assist in controlling *M.tb* bacterial growth by activating infected macrophages to kill intracellular *M.tb* bacteria [27].

2.3 T cell immune responses to *M.tb*

The functional and memory markers expressed by a T cell define the cell's differentiation state, its functionality as well as its memory capabilities, hence why measuring the frequencies of these marker combinations is so important for quantifying the immune response against TB. The differentiation state of a T cell is determined by the combination of memory and functional markers expressed by that cell. Naive T cells (T_N) are the least differentiated cells, followed by central memory T cells (T_{CM}), then effector memory cells (T_{EM}) and finally the most differentiated T cells, terminal effector T cells (T_{TE}). These CD4 T cell subsets are defined by the cell's expression of memory markers CCR7 and CD45RA. Naive CD4 T cells express both CCR7 and CD45RA, T_{CM} CD4 T cells express CCR7 but not CD45RA, T_{EM} CD4 T do not express either marker and terminal CD4 T cells express CD45RA but not CCR7. Within these four CD4 T cell subsets, the expression of cytokines further determines the cell differentiation. Figure 2.1 summarises the differentiation of CD4 T cell according to the cells memory marker and Th1 cytokine expression. Continued antigen stimulation can result in the progressive cell differentiation, resulting in terminally differentiated CD4 T cells that only produce IFN-γ and have relatively short life spans. Thus, antigen exposure or innate immune factors affect the extent of T cell differentiation [16].

The memory capabilities of a CD4 T cell depend on the differentiation state of the cell (i.e. both the memory and functional markers expressed). Immunological memory is defined as the ability of the immune system to have an enhanced response upon re-challenge with the same pathogen [38]. As a cell becomes more differentiated, its long-term memory potential decreases, as seen in Figure 2.1. Naive CD4 T cells as well as T_{CM} cells expressing less than two functional markers

have the highest long-term memory potential while T_{TE} cells have the least memory potential since they are very short-lived. The induction of CD4 T cells with longer-term memory potential is an essential component of an efficacious vaccine in order to confer long-term immunity [39].

While CD4 T cell differentiation is defined based on which memory markers CCR7 and CD45RA are expressed, T cell functionality depends on the number of functional markers expressed. The more functional markers a cell expresses, the higher its degree of functionality, with polyfunctional cells expressing two or more functional markers. Figure 2.1 shows how cell functionality changes depending on the number of functional markers expressed. Continued antigen stimulation can also result in the progressive loss of a cell's functional marker production (i.e. loss of functionality) . Cell functionality is of particular interest in this project as it has been shown that greater degrees of polyfunctionality are associated disease non-progression and protection in other infectious disease models [16]. Consequently, most TB vaccines focus on inducing polyfunctional Th1 CD4 cells, although it is not known whether these cells provide sufficient protection [37].

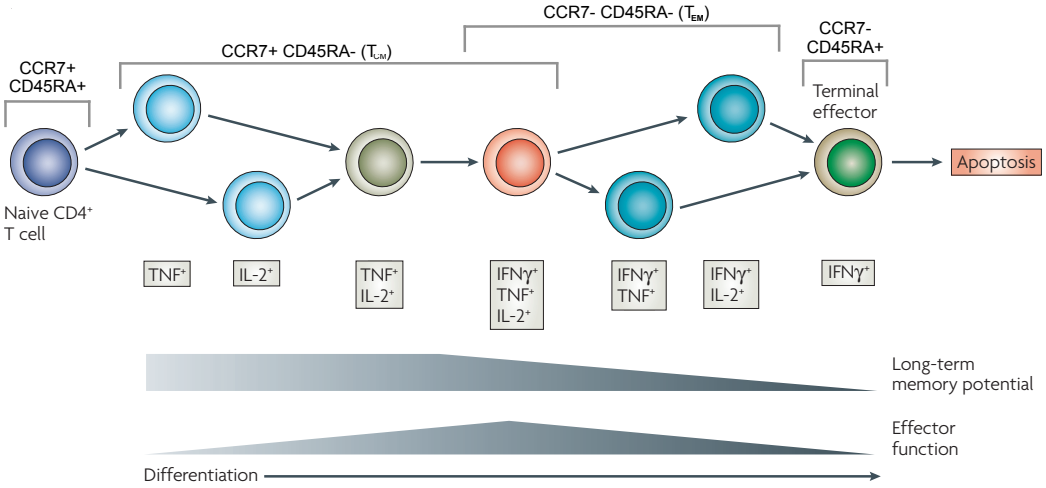


Figure 2.1: Diagram of CD4 T cell differentiation. Credit: Seder *et al.* (2008) [16].

T_{CM} cells result from antigen-naive T cells that have been stimulated by an antigen and survive after the antigen has been eliminated. T_{CM} cells are relatively long-lived and display similar behaviour to stem cells in terms of their regenerative abilities. As previously mentioned, a key characteristic of T_{CM} cells is their long-term memory potential, thus making induction of T_{CM} cells a key characteristic of an efficacious vaccine. T_{CM} cells are also noted for their ability to trigger anamnestic responses upon re-challenge with *M.tb* as well as maintaining protective T cell immune responses during prolonged antigenic stimulation. Despite these qualities, T_{CM} cells are only considered a second line of defense due to their inability to enter infected tissues [40].

T_{EM} cells are considered the first line of defense against *M.tb* as they rapidly traffic to the site of infection and display immediate effector function on antigenic stimulation [40, 41]. However, T_{EM} cells are relatively short-lived and fail to endure under continuous antigenic stimulation during prolonged *M.tb* infection [42].

IFN-γ production has been shown to be an essential mechanism for CD4 T cell-mediated protection, playing an important role in controlling *M.tb* replication and in containing the *M.tb* bacteria within the granuloma [27, 28]. There is also evidence that humans with abnormal IFN-γ receptors are more susceptible to disease [43]. Expression of IL-2 induces proliferation, aids the survival of antigen-specific T cells and is involved in the development of memory T cells during the early infection phase [44]. Although early research suggested that IL-17 was not necessary for protective immunity against *M.tb*, more recent research suggests that IL-17 is in fact required in order to drive the induction of IL-12 [45]. There is also evidence that IL-17 plays a role in mediating vaccine-induced protection against *M.tb* [46].

2.4 QuantiFERON-TB Gold (QFT)

One of the key aims of this project is to investigate the effects of *M.tb*-sensitisation on immune response to vaccination. In order to do this, QFT status shall be used as a proxy for *M.tb*-sensitisation. QFT is an *in vitro* test used to measure *M.tb* infection. Specifically, it is an Interferon-Gamma Release Assay (IGRA) that measures the amount of IFN-γ released from an individual's T cells in whole blood when stimulated with antigens ESAT-6 and culture filtrate protein-10 (CFP-10). The significance of using ESAT-6 and CFP-10 is that these antigens are expressed by *M.tb* alone and not BCG or most environmental mycobacteria, thus increasing the specificity of the test [47]. The IFN-γ response of the negative control is subtracted from the IFN-γ response after antigen stimulation. If an individual has a background subtracted response greater than the cut-off of 0.35 international units per mL (IU/mL), then they are classified as QFT+ while those below the cut-off are classified as QFT- [48]. However, this cut-off is a practical interpretation of the test results and there is evidence that results within the board-line range of 0.20 to 0.99IU/ml should be interpreted with caution [49], implying that classification of borderline cases may be inaccurate. Despite this, QFT status rather than quantitative values shall be used in this project since it is still common to do so and because it simplifies the analysis.

2.5 Statistical methods traditionally used

Statistical methods traditionally used by the immunology community consist mostly of simplistic and low dimensional techniques such as ad hoc rules based on fold-changes, Hotelling's T2 statistics, exact tests of 2×2 contingency tables and graphical displays of summary statistics [50]. Such methods put a large emphasis on statistical hypothesis testing and p-values and thus typically result in multiple comparisons across participants as well as across cell subsets which need to be adjusted for, resulting in a loss of power. Another disadvantage of these methods is their low dimensionality whereby subjects and/or cell subsets are tested separately without allowing for the sharing of information across observations [51].

Recently, to combat these shortcomings two data analysis pipelines have become increasingly

popular in the immunology community: mixture models for single-cell assays (MIMOSA) [51] and combinatorial polyfunctionality analysis of single cells (COMPASS) [50]. Both methods have the advantage of sharing information across participants and COMPASS has the additional advantage of being compatible with multivariate data. These two data analysis pipelines, along with other non-parametric techniques such as graphical displays and non-parametric statistical tests, shall be discussed in detail in Chapter 4. Chapter 4 will also introduce and discuss the more sophisticated statistical modelling used in this dissertation.

Chapter 3

Study Design and Data Manipulations

3.1 Study design

This study is a retrospective study which uses data and samples collected from three clinical trials conducted by the South African Tuberculosis Vaccine Initiative (SATVI) in collaboration with Statens Serum Institute (SSI). The three trials used are trials H1-THYB04 [52], H56-032 [53] and H56-035 [36]. Throughout this project, these three trials shall be referred to as trials A, B and C respectively.

Trial A was a Phase II clinical trial consisting of a total of 240 adolescents. Figure 3.1 shows the experimental design used for this trial. Each participant was randomly assigned to one of four groups, with randomisation accounting for the fact that half of the participants in each group should be QFT- and the other half should be QFT+. Depending on their group allocation, each participant was administered either 15 µg or 50 µg of the H1 vaccine either once off or twice with an interval of 56 days between doses.

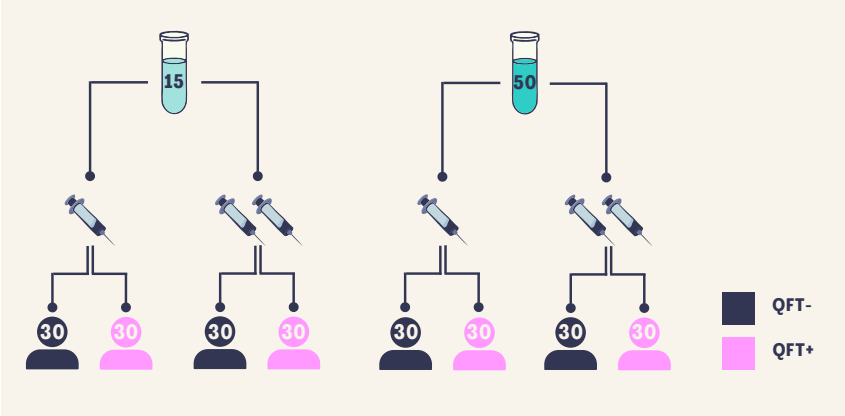


Figure 3.1: Diagram of the Trial A cohort. Each figure represents a group, with the number of participants in that group indicated by the number. The number in the vial represents the concentration in µg of the vaccine used. The needles represent the administration schedule of the vaccine: one needle for one administration and two for two administrations. The groups are colour coded by QFT status.

Trial B was a first-in-human, open label Phase I clinical trial with a very small cohort of only 25 adults, as seen in Figure 3.2. QFT+ participants were randomly assigned to one of two groups and QFT- participants formed the third group. Each participant received three administrations of either 15 µg or 50 µg of vaccine H56 with 56 day intervals between doses.

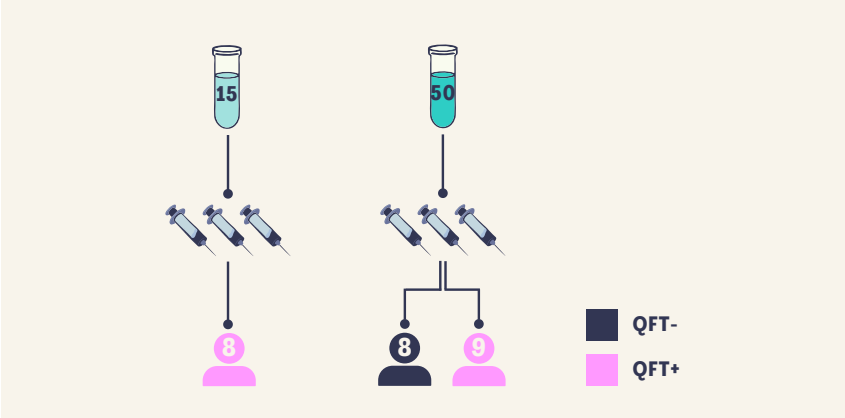


Figure 3.2: Diagram of the Trial B cohort. Each figure represents a group, with the number of participants in that group indicated by the number. The number in the vial represents the concentration in µg of the vaccine used. The 3 needles represent the 3 administrations of the vaccine that each participant received.The groups are colour coded by QFT status.

Trial C was a Phase I/IIa clinical trial enrolling 81 healthy adults, as seen in Figure 3.2. The first phase of the trial was the dose selection stage which enrolled 45 QFT- participants and randomly assigned them to one of three groups. Each participant was administered two doses of either 5 µg, 15 µg or 50 µg of H56. Based on the results of the first phase of the trial, a dose of 5 µg of the vaccine was then chosen for the schedule selection phase of the trial. This phase enrolled 12 QFT- and 24 QFT+ participants and randomly assigned the QFT+ participants to one of two groups and the QFT- participants formed the third group. Participants were then administered either two or three doses of 5 µg.

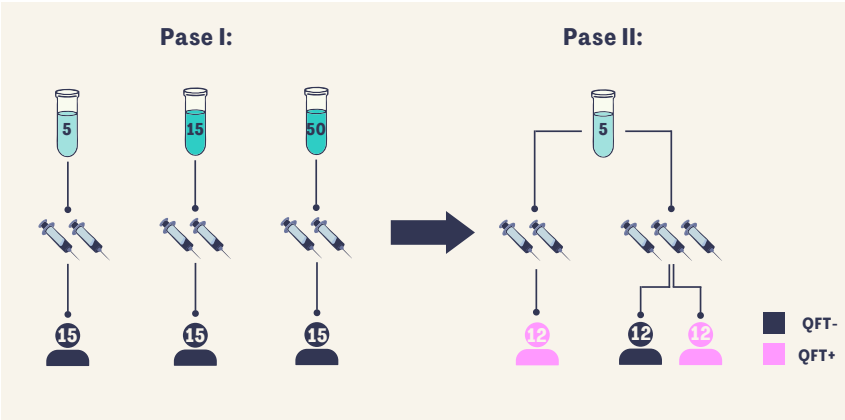


Figure 3.3: Diagram of the Trial C cohort. Each figure represents a group, with the number of participants in that group indicated by the number. The number in the vial represents the concentration in µg of the vaccine used. The needles represent the administration schedule of the vaccine: 2 needles for 2 administrations and 3 for 3 administrations. The groups are colour coded by QFT status.

3.2 Data integration

Data generated from the three trials was combined for this study in order to span a wider range of vaccine doses and number of administrations compared to each individual study, as

seen in Figure 3.4. Data from all participants in trial B was included in the combined cohort, however only data from a subset of participants from trials A and C were used due to time constraints on the data pre-processing steps, specifically gating, that needed to be performed by the immunologists prior to the analysis of the data.

Gating refers to the process of selecting successive subpopulations of cells where only cell data from the final subpopulation will be used for statistical analysis [54]. Each trial's investigator had employed a different gating strategy, thus a new unified gating strategy was applied across all three trials in FlowJo™ (a software package used to analyse flow cytometry data) in order to minimize between trial bias. Gating was performed manually to each whole blood sample one at a time, as is common, which is a very time-consuming process.

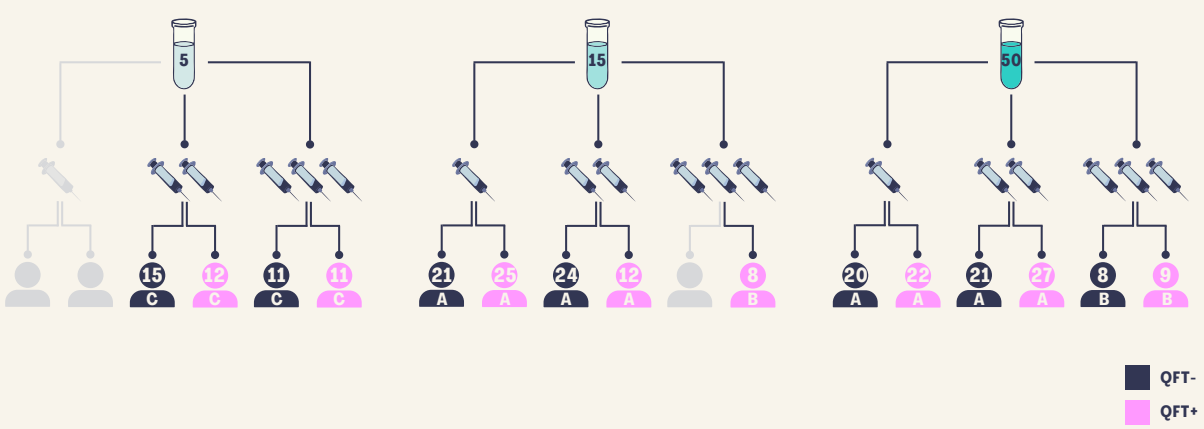


Figure 3.4: Diagram of the full cohort that will be used for this project. Each figure represents a group, with the number of participants in that group indicated by the number and the trial from which those participants came from indicated by the letter. The number in the vial represents the concentration in µg of the vaccine used. The needles represent the administration schedule of the vaccine: one needle for one administrations, two for two administrations, etc. The groups are colour coded by QFT status. No data is available for the groups in grey.

The key differences between the three trials are the type of vaccine administered and the age group of the cohorts. Trial A administered vaccine H1 to a cohort of adolescents while trials B and C administered vaccine H56 to two cohorts of adults. A key assumption of the study design is that the two vaccines are essentially equivalent. As previously discussed, both the H1 and H56 vaccines contain M.tb antigens Ag86B and ESAT-6, with the only difference being that H56 contains a third antigen, Rv2660c, while H1 does not. The rationale for treating H1 and H56 as equivalent is that the function of antigen Rv2660c is unknown [36] and that only stimulation with Ag85B and ESAT-6 will be analysed, and immune responses to Rv2660c were ignored. With regards to the confounding factor age, this can be accounted for and investigated in the statistical modelling approach of the analysis.

The data collected was longitudinal with whole blood samples collected at different time points for different participants depending on which trial they were enrolled in and what dosage schedule they were on. In order to combine the data collected in these three trials, it was decided to harmonise these time points by choosing three universal time points: a baseline (pre-exposure), peak and memory time point. Figure 3.5 outlines how this was done for each dose schedule. The peak time point needed to be two weeks after the last vaccine dose and the memory time

point needed to be 6 to 10 months after the last vaccine dose, hence why the time points fall on different days depending on the number of administrations received of the vaccine. Because the sample collection days were not the same across trials, samples collected from participants receiving two vaccine administrations on day 224 in trial A and day 292 in trial C were classified as memory time point data. Similarly, for participants receiving three vaccine administrations, samples collected on day 210 in trial B and day 292 in trial C were classified as memory time point data.

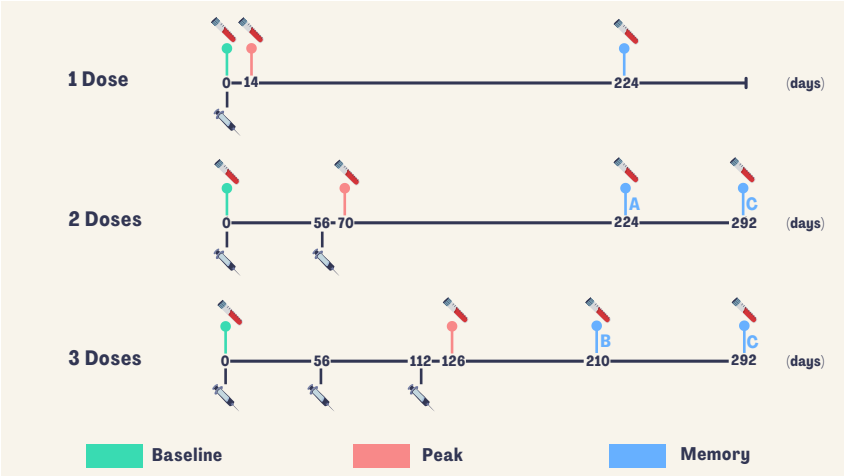


Figure 3.5: Diagram of the unified sample selection schedule for the combined datasets. A separate timeline is given for each vaccine schedule and the syringes indicate the days that the vaccine was administered. The red vials indicate the days of interest where samples were collected, each day colour coded according to whether it is a baseline, peak or memory response.

The diagram summarising the data integration process and noting which observations were missing or removed can be found in the Appendix, Figure B.1.

3.3 Importing data

After the data was re-gated in FlowJo™, it needed to be imported into the statistical software environment `R` [55] in order for the counts and frequencies of cells expressing different marker combinations to be computed. Usually these counts are computed in FlowJo™, but for larger datasets where many more markers have been measured this would not be possible. It was decided to compute the counts in `R` because it was desirable to establish a new data analysis pipeline that is compatible with larger datasets. This is particularly necessary in the context of this study because while the counts for whole blood dataset could have been computed using FlowJo™, a related peripheral blood mononuclear cell (i.e. blood cells with a round nucleus [56]) dataset which will be analysed in a follow-up project would be far too large. Packages `flowCore` [57], `flowWorkspaceData` [58] and `CytoML` [59] were used to import the FlowJo™ work spaces into `R` and create a function that could compute the counts and frequencies of marker combinations.

3.4 Data description

Whole blood samples were collected from each participant at each time point (baseline, peak and memory). Each participant has 4 whole blood samples for a given day, one stimulated with Ag85B, one stimulated with ESAT-6, one stimulated with phytohaemagglutinin (PHA) and one unstimulated. For each of these samples, the frequency of each marker combination is recorded. A specific participant and time point combination shall be referred to as a *participant visit* (e.g. PID 1010 at baseline) and a specific participant, time point and stimulus combination shall be referred to as *one observation* (e.g. PID 1010 at baseline under stimulation with ESAT-6). Thus, provided that no observations are missing, each participant will have three participant visits and a total of $3 \times 3 = 9$ observations. Note that the PHA observations are only relevant for the data cleaning process and will not be used in the statistical analysis, hence why there are three stimuli rather than four.

The explanatory variables of interest are participant-specific variables QFT status, administration and concentration as well as observation-specific variables stimulus and time point. QFT status will be a categorical variable used as a proxy for *M.tb*-sensitisation. Administration refers to the number of administrations of the vaccine that the participant received (either 1, 2 or 3), while concentration refers to the concentration of the vaccine that the participant received (either 5 µg, 15 µg or 50 µg). These two variables will be included in the models as continuous variables in order to limit the complexity of the model structure. Stimulus will be a categorical variable with categories Ag85B, ESAT-6 and unstimulated. Demographic variables sex, ethnicity and age are also available for inclusion in the models, although they are not of primary interest.

The response of interest are the frequencies of cells expressing different combinations of the markers seen in Table 3.1. The markers with rationale "Function" in the table are functional markers, also known as cytokines, specifically induced upon antigen stimulation, where the specific combination of functional markers expressed by a cell describe the functional quality of that cell. The markers with rationale "Phenotype" are phenotypic markers, also known as memory markers, and describe the memory state of the cell at the steady state (i.e. not significantly modulated by stimulation). The different possible combinations of markers expressed by cells shall be referred to as *marker combinations*, the different combinations of functional markers as *functional combinations* and the different combinations of memory markers as *memory combinations*. For the whole blood dataset there are $2^6 = 64$ possible marker combinations. Not all combinations are relevant or observable, so this number will be reduced in the dimension reduction step of the analysis.

Table 3.1: Flow cytometry panel showing the markers measured for the dataset.

Fluorochrome	**Marker**	**Rationale**
APC-H7	CD3	
BV786	CD4	T cells
BUV805	CD8	
PE	IL-2	
PE-Cy7	TNF	Functions
AF700	IFN-γ	
BV605	IL-17	
PE-CF594	CCR7	Phenotype
BV570	CD45RA	

There are a variety of outcomes of interest, all of which are frequencies of either CD4 T cells or cytokine positive CD4 T cells. The frequency may be background subtracted, meaning that the unstimulated frequency is subtracted from the stimulated frequency. The response of interest will depend on the statistical method employed, and shall be discussed in further detail in Chapter 4.

3.5 Data cleaning

Data cleaning was done as per the SATVI Immunology Lab Data Management for Clinical Vaccine Trials standard operating procedure number S137-02 with some amendments. Below are the detailed amended steps:

(a) Exclude all participants visits for which the unstimulated control is missing, since all observations will require their corresponding unstimulated observation in order to calculate the response of interest.

(b) Plot the frequency of cytokine positive CD4 T cells for all unstimulated observations and use this plot to identify any potential outliers. To determine if the outliers should be removed, look at previous notes made during the re-gating process.

(c) Calculate the frequency of cytokine positive CD4 T cells for each observation (stimulated and unstimulated). Next, calculate the median value of this frequency for all unstimulated observations only. Using this value, calculate the Median Absolute Deviation (MAD) by taking the absolute value of the difference between the median and the actual frequency for each unstimulated observation, and then taking the median of these resulting differences. If a participant visit has a PHA-induced frequency less than or equal to the unstimulated median frequency + 3MAD, then check the Ag85B and ESAT-6 frequencies and see if any of those pass the cut-off. If none of the frequencies pass the cut-off, then exclude the whole participant visit. In other words, if the frequency for at least one stimulus passes

the cutoff, then the participant visit will not be excluded. If the PHA responses are not available for the participant visit, use Ag85B responses instead, and if that is not available, use ESAT-6 instead. This is done to check that the assay was performed correctly. If no response is detected to any of the stimulations, a technical error may have occurred.

(d) Exclude all participant visits for which the PHA-, Ag85B- and ESAT-6-induced cytokine positive CD4 T cells are less than or equal to the corresponding frequency for the unstimulated control.

Because the goal is to create a unified dataset, the above points were applied to the dataset after combing the data from all three trials, instead of applying the steps to each trial's data separately and then combining the data to form the cleaned dataset. Overall, only one participant visit was removed, the details of which can be found in Figure B.2 in Appendix B.

Chapter 4

Statistical Methods

This chapter shall outline the statistical methods used to investigate the three aims of this project. For each aim, a mixed effect modeling approach shall be employed as well as a "standard" approach using statistical procedures typically employed by the immunologist community. The specific responses analysed using each statistical method is dependent not only on the statistical method used but also on the aim being investigated. While some methods such as combinatorial polyfunctionality analysis of single cells (COMPASS) and to some extent mixture models for single-cell assays (MIMOSA) shall be used to investigate only one aim, many of the methods discussed here shall be applied to more than one aim. This means that for many of the statistical methods used, there is more than one corresponding response variable. As such, for many of the statistical methods discussed the exact calculation of the response variable will not be given in this chapter but rather in subsequent chapters which are separated by aim and which apply the methods outlined here.

The layout of this chapter is as follows: Section 4.1 introduces notation that shall be used throughout this chapter and in subsequent chapters. Section 4.2 will describe the statistical methods typically employed by the immunologist community, that being non-parametric methods, MIMOSA and COMPASS. The mixed effect modelling approach will be broken up into two sections: Section 4.3.1 shall discuss generalised linear mixed effect models (GLMEM's) and Section 4.3.2 shall discuss latent variable models (LVM's).

4.1 Notation

The following sections require the introduction of some important notation. Let V be the total number of visits across all participants, $i \in \{1, ..., I\}$ be the specific participant ID and $j \in \{1, ..., J\}$ be the specific visitation day (either baseline, peak or memory). Thus, in the case of no missing values, $V = I \times J$. Let c be a specific cytokine combination (for example TNF+IL2-IFNγ+IL17-) and c^+ be a specific cytokine positive combination. Thus, $c \in \{1, ..., C\}$ where $C = 2^4$ and $c^+ \in \{1, ..., C-1\}$ since there is only one cytokine combination which is not cytokine positive (TNF-IL2-IFNγ-IL17-). Let s stand for stimulated (either Ag85B or ESAT-6) and u stand for unstimulated.

Let $n_{ijc}^{(s)}$ be the number of CD4 T cells expressing cytokine combination c in the sample collected from participant i at time point j after stimulation with s. Similarly, let $n_{ijc}^{(u)}$ be the number of CD4 T cells expressing cytokine combination c in the sample collected from participant i at time point j when unstimulated.

Let $n_{ijc^+}^{(s)}$ be the number of CD4 T cells expressing *cytokine positive* combination c^+ in the sample collected from participant i at time point j after stimulation with s.

Let

$$N_{ij}^{(s)} = n_{ij.}^{(s)} = \sum_{c=1}^{C} n_{ijc}^{(s)} \qquad \text{and} \qquad N_{ij}^{(u)} = n_{ij.}^{(u)} = \sum_{c=1}^{C} n_{ijc}^{(u)}.$$

In other words, $N_{ij}^{(s)}$ is the total number of CD4 T cells observed in the sample collected from participant i at time point j after stimulation with s. Similarly, let

$$N_{ij+}^{(s)} = \sum_{c^+=1}^{C-1} n_{ijc^+}^{(s)} \qquad \text{and} \qquad N_{ij+}^{(u)} = \sum_{c^+=1}^{C-1} n_{ijc^+}^{(u)}.$$

In other words, $N_{ij+}^{(s)}$ is the total number of *cytokine positive* CD4 T cells observed in the sample collected from participant i at time point j after stimulation with s.

Finally, define $n_{ijm}^{(s)}$ as the number of CD4 T cells expressing *memory combination* m in the sample collected from participant i at time point j after stimulation with s. Here, m is a specific combination of two memory markers CCR7 and CD45RA (for example CCR7+CD45RA-) and hence $m \in \{1, ..., M\}$ where $M = 2^2$. Define $n_{ijm}^{(u)}$ similarly. Let $n_{ijm^+}^{(s)}$ be the number of *cytokine positive* CD4 T cells expressing memory combination m^+ in the sample collected from participant i at time point j after stimulation with s. Here, $m^+ \in \{1, ..., M^+\}$ where $M^+ = 2^2$ still but now m^+ refers to cells that express memory combination m *and* are also cytokine positive (i.e. express at least one cytokine). Define $n_{ijm}^{(s)}$ similarly. Note that this means that

$$N_{ij}^{(s)} = \sum_{m=1}^{M} n_{ijm}^{(s)} \qquad \text{and} \qquad N_{ij}^{(u)} = \sum_{m=1}^{M} n_{ijm}^{(u)},$$

and

$$N_{ij+}^{(s)} = \sum_{m^+=1}^{M^+} n_{ijm^+}^{(s)} \qquad \text{and} \qquad N_{ij+}^{(u)} = \sum_{m^+=1}^{M^+} n_{ijm^+}^{(u)}.$$

4.2 Standard approach

The traditional approach to data analysis employed by the immunologist community predominately consists of data visualisation techniques, such as box-plots and longitudinal plots, coupled with non-parametric significance tests [50]. Since the data used is usually very skew and zero-inflated and thus non-normal, these significance tests need to be non-parametric. The two significance tests that shall be used in this project are the Mann–Whitney U test for comparisons between independent samples and the Wilcoxon signed-rank test for comparisons between dependent samples. Both the Mann–Whitney and Wilcoxon tests test whether the observations

from two samples are drawn from the same distribution. However, analysing immunological data where there are many different data subsets and many hypotheses of interest, as in this project, necessitates multiple comparisons and thus presents the challenge of adjusting for multiple testing.

Two data analysis pipelines that have recently become popular in the immunologist community are COMPASS and MIMOSA. These data analysis pipelines, together with data visualisation techniques and significance tests shall constitute the statistical analysis approach employed in this project in order to be representative of the standard data analysis approach employed by the immunologist community. Section 4.2.1 will discuss statistical power, significance levels and multiple testing in relation to the significance tests that shall be conducted in this project. Sections 4.2.2 and 4.2.3 shall describe the data analysis pipelines COMPASS and MIMOSA and how they will be used in this project.

4.2.1 Statistical significance, power and multiple testing

As discussed by Murphy *et al.* (2010) [60], concluding that something is *statistically significant* usually says more about the study design rather than the actual results. This is because significance depends upon the pre-specified significance level α as well as the sample size of the experiment. The significance level α is the probability of making a type I error, which is defined as the probability of rejecting the null hypothesis when it is in fact true, or more specifically in the case of the significance tests used in this project, the probability of concluding that there is a difference when in fact there is not. A related concept is *statistical power*, which is defined as the probability of rejecting the null hypothesis when it is in fact wrong [60], or in other words, the probability of finding a difference if it exists.

The problem with conducting multiple significance tests, is that with each additional significance test the maximum experiment-wise error rate (MEER), the probability of an investigator concluding that there is at least one statistically significant difference regardless of whether this is true or not, increases [61]. Thus, if a researcher wishes to be able to conclude statistical significance, not adjusting for multiple testing would result in the MEER being much higher than expected, and increases the probability of the investigator committing a Type I error.

According to Bender *et al.* (2001) [62], adjusting for multiple testings consists of three main steps

1. identifying which significance tests belong to a single experiment,
2. choosing which error rate to control, and
3. choosing the appropriate multiple testing adjustment.

In order to adjust for multiple testing, one first needs to determine how many tests are being performed, a task around which there is much debate and little consensus. Bender *et al.* (2001)

[62] suggest that significance tests belong to one single experiment if all of the tests and only these tests are used to make one final conclusion or decision. Since there are many aims and hypotheses that shall be investigated in this project, it was decided to consider each plot as a single experiment, as each plot was created to answer a slightly different question. The plots used in this project were either collections of box-plots or longitudinal profiles, where the aim was to compare the distribution of the data between the different box-plots or time points. Thus, adjusting for multiple testing shall be done within each plot separately.

Controlling the MEER is generally seen as the best protection against type I errors and thus produces the strongest statistical inference [62]. Hence, controlling the MEER is generally recommended for confirmatory analyses [62, 63]. The Bonferroni correction is one of the most simple and popular methods of controlling MEER, however it comes at the cost of significantly reduced power and is generally not recommended for when the number of tests exceeds five [61, 62].

Another, more recent method of adjusting for multiple testing is by controlling the expected proportion of false positives, called the false discovery rate (FDR), rather than the MEER. Controlling the FDR rather than the MEER is less conservative with respects to Type I error but has the advantage of being more statistically powerful [61]. As such, controlling of the FDR is recommended for exploratory analyses as oppose to confirmatory analyses [63]. Gelman *et al.* (2012) [61] note that controlling the FDR is particularly appropriate in fields such as genetics where it is expected to observe a couple of real effects amidst a large quantity of zero effects, for example when examining the effect of a treatment on differential expression of genes. The Benjamini-Hochberg procedure is a popular method for controlling the FDR, and has been shown to be more powerful than the Bonferroni correction [64].

For this project, it was decided to adjust for multiple testing by controlling the FDR using the Benjamini-Hochberg procedure for two main reasons. Firstly, because of the many aims and hypotheses specified, this project should be considered more of an exploratory analysis rather than a confirmatory analysis, and thus controlling the FDR is more appropriate as more power is desired. Secondly, this project aims to examine the effect of antigen stimulation on differential biomarker expression, and due to the large number of marker combinations considered, it is expected that many of these marker combinations will not be differentially expressed and thus are biologically irrelevant.

Although this project will adjust for multiple testing, it is important to understand this is not a perfect solution and it does not imply that the probability of a Type I is now, say $\alpha = 5\%$, and thus that any observed p-values below $\alpha = 5\%$ are necessarily statistically significant. Adjusting for multiple testing is still useful as it does reduce the probability of a Type I error but it also runs the danger producing an over-confident investigator. As such, this project will avoid concluding *statistical significance* but rather comment on the strength of observed results, in order to emphasise that the epidemiological conclusions made in this project are exploratory only.

4.2.2 MIMOSA

Mixture models for single-cell assays (MIMOSA) is a Bayesian hierarchical framework developed by Finak *et al.* (2014) [51] aimed at testing differential biomarker expression between two conditions using single-cell assays. In the context of this thesis, MIMOSA was used to identify participant visits for which the proportion of cytokine positive cells was significantly different for the stimulated sample taken on that visit compared to the unstimulated sample also taken on that visit for the same participant. Such participant visits are known as responders. Since there are two stimuli, Ag85B and ESAT-6, MIMOSA was applied separately for each stimulus, resulting in two datasets, one for each stimulus. The theory described below is a summary of that presented by Finak *et al.* (2014) [51] and Finak *et al.* (2020) [65], adapted to be specific to how it will be used in this project.

For Ag85B there are $V = 732$ participant visits (732 Ag85B-stimulated observations and 732 unstimulated observations) and for ESAT-6 there are $V = 734$ participant visits. The variables of interest are total number of cytokine positive cells $N_{ij+}^{(s)}$ and $N_{ij+}^{(u)}$ as well as the total number of CD4 T cells $N_{ij}^{(s)}$ and $N_{ij}^{(u)}$. Note that $N_{ij}^{(s)}$ is equal to the total number of cytokine positive CD4 T cells ($N_{ij+}^{(s)}$) plus the total number of CD4 T cells expressing no cytokines, and similarly so for $N_{ij}^{(u)}$ and $N_{ij+}^{(u)}$. The vector of observed counts for participant i on visit j can be denoted as $\boldsymbol{y}_{ij} = (N_{ij+}^{(s)}, N_{ij+}^{(u)})$. In other words, observations are paired, containing a stimulated and an unstimulated component.

For each participant visit, $N_{ij+}^{(s)}$ and $N_{ij+}^{(u)}$ can be modelled using binomial distributions as follows

$$\left(N_{ij+}^{(s)} \mid p_{ij}^{(s)}\right) \sim \text{Bin}\left(N_{ij}^{(s)}, p_{ij}^{(s)}\right)$$

and

$$\left(N_{ij+}^{(u)} \mid p_{ij}^{(u)}\right) \sim \text{Bin}\left(N_{ij}^{(u)}, p_{ij}^{(u)}\right),$$

where $p_{ij}^{(s)}$ is the unknown proportion of s-stimulated CD4 T cells which are cytokine positive, and similarly so for $p_{ij}^{(u)}$.

MIMOSA tests for differential expression by fitting a two-component mixture model where the two components represent two competing hypotheses. The first component models the null hypothesis of no response, while the second component models the alternative hypothesis of response. Specifically, under the null hypothesis the proportion of cytokine positive cells is the same in the stimulated and unstimulated samples (although counts may differ), while under the alternative hypothesis the proportion of cytokine positive cells is greater in the stimulated sample compared to the unstimulated sample. This can be denoted as follows

$$H_0 : p_{ij}^{(s)} = p_{ij}^{(u)}$$
$$H_1 : p_{ij}^{(s)} > p_{ij}^{(u)}.$$

For non-responders, both $N_{ij+}^{(s)}$ and $N_{ij+}^{(u)}$ would be generated by the same distribution, $\text{Bin}(N_{ij}^{(u)}, p_{ij}^{(u)})$, while for responders $N_{ij+}^{(u)}$ would be generated by $\text{Bin}(N_{ij}^{(u)}, p_{ij}^{(u)})$ and $N_{ij+}^{(s)}$ would be generated by $\text{Bin}(N_{ij}^{(s)}, p_{ij}^{(s)})$.

Although the mixture model is participant visit specific, the models share information across all participant visits through the use of exchangeable Beta priors on the unknown proportions. For non-responders,

$$H_0 : (p_{ij}^{(u)} \mid z_{ij} = 0) \sim \text{Beta}(\alpha^{(u)}, \beta^{(u)}),$$

while for responders

$$H_1 : (p_{ij}^{(u)} \mid z_{ij}^{(s)} = 1) \sim \text{Beta}(\alpha^{(u)}, \beta^{(u)}) \quad \text{and} \quad (p_{ij}^{(s)} \mid z_{ij}^{(s)} = 1) \sim \text{Beta}(\alpha^{(s)}, \beta^{(s)}).$$

$z_{ij}^{(s)}$ is an indicator variable that equals 1 if participant i on day j is a responder for stimulus s and zero otherwise, and it's assumed that $z_{ij}^{(s)} \sim \text{Be}(w^{(s)})$ are independent draws from a Bernoulli distribution with probability $w^{(s)}$, where $w^{(s)}$ is the unknown proportion of responders for stimulus s. $\alpha^{(u)}$, $\beta^{(u)}$, $\alpha^{(s)}$ and $\beta^{(s)}$ are the unknown hyper-parameters that allow information to be shared across all participants. The parameter vector $\boldsymbol{\theta} = (\alpha^{(u)}, \beta^{(u)}, \alpha^{(s)}, \beta^{(s)}, w^{(s)})$ can then be estimated either in an empirical-Bayes fashion using an expectation-maximization (EM) algorithm or using a fully Bayesian approach by use of a Markov chain Monte Carlo (MCMC) algorithm. For this project, the MCMC algorithm was used as it is generally more stable than the EM algorithm.

Using the estimated posterior probabilities, the false discovery rates (FDR) are obtained by using "q-values" as the estimated FDR. These q-values are calculated using the methods introduced by Storey and Tibshirani (2003) [66].

Filtering out non-responders

Applying MIMOSA to the ESAT-6 dataset and the Ag85B dataset separately produced FDR for each stimulated observation. All observations with an estimated FDR greater than 0.01 were classified as non-responders. In other words, a certain participant visit could be classified as a responder for, say, Ag85B but also classified as a non-responder for ESAT-6. A further data filtering step was then applied: all stimulated observations for which the frequency of

cytokine positive CD4 T cells was less than 3 times that when unstimulated were also classified as non-responders. These filtering steps have been used to select responding observations in previous SATVI studies [67, 68]. In Chapter 5, the response rate shall be investigated while in Chapter 7, a new response variable shall be defined which can only be calculated using responding observations.

It should also be noted that the act of filtering out observations using MIMOSA is in a sense determining statistical significance. One is identifying responders as those whose proportion of cytokine positive cells was significantly different for the stimulated sample compared to the unstimulated sample.

4.2.3 COMPASS

Combinatorial polyfunctionality analysis of single cells (COMPASS) is a Bayesian hierarchical framework developed by Lin *et al.* (2015) [50] aimed at modelling all observed cell subsets and identifying those that are most likely to have antigen-specific responses. In the context of this thesis, COMPASS was used to identify cytokine combinations which are expressed at a significantly higher frequency after stimulation with a certain antigen and thus are antigen-specific. COMPASS has the advantage of providing posterior probabilities for each participant visit *for each cytokine combination* rather than just one posterior probability for each participant visit as MIMOSA did. As with MIMOSA, COMPASS also had to be applied separately for each stimulus, Ag85B and ESAT-6. The theory described will be a summary of that presented by Lin *et al.* (2015) [50], adapted to be specific to how it will be used in this project.

Assume that the set $1, ..., C$ of all possible cytokine combinations is arranged in ascending order according to their degree of functionality, except for the last subset C which is set to be the null with zero degree of functionality. Let $\boldsymbol{n}_{ij}^{(s)} = \left(n_{ij1}^{(s)}, ..., n_{ijC}^{(s)}\right)$ and $\boldsymbol{n}_{ij}^{(u)} = \left(n_{ij1}^{(u)}, ..., n_{ijC}^{(u)}\right)$, where $n_{ijc}^{(s)}$ and $n_{ijc}^{(u)}$ are as defined in Section 4.1. $\boldsymbol{n}_{ij}^{(s)}$ and $\boldsymbol{n}_{ij}^{(u)}$ can the be jointly modelled by the following multinomial distributions

$$\left(\boldsymbol{n}_{ij}^{(s)} | \boldsymbol{p}_{ij}^{(s)}\right) \sim MN\left(N_{ij}^{(s)}, \boldsymbol{p}_{ij}^{(s)}\right), \quad \text{and} \tag{4.1}$$

$$\left(\boldsymbol{n}_{ij}^{(u)} | \boldsymbol{p}_{ij}^{(u)}\right) \sim MN\left(N_{ij}^{(u)}, \boldsymbol{p}_{ij}^{(u)}\right), \tag{4.2}$$

where $\boldsymbol{p}_{ij}^{(s)} = \left(p_{ij1}^{(s)}, ..., p_{ijC}^{(s)}\right)$ and $\boldsymbol{p}_{ij}^{(u)} = \left(p_{ij1}^{(u)}, ..., p_{ijC}^{(u)}\right)$ are two unknown proportion vectors, where $p_{ijc}^{(s)}$ is the proportion of CD4 T cells expressing cytokine combination c for participant i on visit j when stimulated.

The following two competing hypotheses can be considered in order to detect the responding participant visits

$$H_0 : \boldsymbol{p}_{ij}^{(u)} = \boldsymbol{p}_{ij}^{(s)},$$
$$H_1 : \exists c \in \{1, .., C-1\} \quad \text{such that} \quad \boldsymbol{p}_{ij}^{(u)} \neq \boldsymbol{p}_{ij}^{(s)}.$$

In other words, under the null hypothesis, there is no difference in the proportion of cytokine producing cells between stimulated and unstimulated samples, and the proportion parameter is shared across the two multinomial models. Under the alternative hypothesis, there is at least one cytokine combination c which is considered antigen-specific.

A binary indicator $\boldsymbol{\gamma}_{ij} = (\gamma_{ij1}, ..., \gamma_{ijC})'$ can be introduced so that $\gamma_{ijc} = 1$ if $p_{ijc}^{(s)} = p_{ijc}^{(u)}$ and zero otherwise. Consequently, $\boldsymbol{p}_{ij}^{(s)} = \boldsymbol{p}_{ij}^{(u)}$ if and only if $\boldsymbol{\gamma}_{ij}$ is a vector of zeros and $\boldsymbol{p}_{ij}^{(s)} \neq \boldsymbol{p}_{ij}^{(u)}$ otherwise. Thus, $\boldsymbol{\gamma}_{ij}$ introduces dependencies between the two sample conditions and Equations 4.1 and 4.2 can be rewritten as follows

$$\left(\boldsymbol{n}_{ij}^{(s)} | \boldsymbol{p}_{ij}^{(s)}, \boldsymbol{\gamma}_{ij}\right) \sim MN\left(N_{ij}^{(s)}, \boldsymbol{p}_{ij}^{(s)}\boldsymbol{\gamma}_{ij} + \boldsymbol{p}_{ij}^{(u)}(\mathbf{1} - \boldsymbol{\gamma}_{ij})\right), \quad \text{and}$$
$$\left(\boldsymbol{n}_{ij}^{(u)} | \boldsymbol{p}_{ij}^{(u)}, \boldsymbol{\gamma}_{ij}\right) \sim MN\left(N_{ij}^{(u)}, \boldsymbol{p}_{ij}^{(u)}\right).$$

Parameters $\boldsymbol{p}_{ij}^{(s)}$, $\boldsymbol{p}_{ij}^{(u)}$ and $\boldsymbol{\gamma}_{ij}$ are the parameters of the model and are modelled in terms of unknown hyperparameters that share information across participant visits. Again, the MCMC algorithm is used to estimate these hyperparameters.

For this project, there are three COMPASS outcomes of interest: posterior probabilities, functional scores (FS) and polyfunctional scores (PFS). The posterior probabilities are available for each cytokine combination for each participant visit and are the posterior means of the binary indicators $\boldsymbol{\gamma}_{ij}$. These posterior probabilities can be interpreted as the probability that a cytokine combination is antigen-specific for a given participant visit and thus can be used to quantify the cell subset responses.

COMPASS also provides a FS and a PFS for each participant visit. Both scores summarise the response across the different cell subsets for each participant visit. The advantage of these scores over the posterior probabilities is that they provide a single qualitative value for each participant visit, making it easier for statistical analysis, comparisons across treatment groups and correlation with outcomes of interest.

The FS is defined as the posterior mean of the average number of antigen-specific cell subsets, irrespective of their degree of functionality, denoted as follows

$$FS_{ij} = \sum_{c=1}^{C-1} \frac{\hat{\gamma}_{ijc}}{C-1},$$

where $\hat{\gamma}_{ijc}$ is the posterior mean of γ_{ijc} estimated using the MCMC algorithm. FS can be interpreted as the proportion of antigen-specific cell subsets for a given participant visit among all possible subsets.

The PFS is the mean of the posterior probabilities, weighted by the cell subset's degree of functionality and normalised by the total number of possible cell subsets that could be observed, given the number of markers considered, denoted as follows

$$PFS_{ij} = \frac{\sum_{c=1}^{C-1} \hat{\gamma}_{ijc} \times d(c) / \binom{M}{d(c)}}{(M \times (M+1)/2)},$$

where M is the number of cytokines measured, and $d(c)$ is the degree of functionality of cell subset c and is simply equal to the number of cytokines expressed by the cell.

4.3 Mixed effect modelling approach

The immune response data used in this project presents many challenges, including the following

1. the response variable, which will be some type of frequency specific to the aim being addressed, is highly skewed and zero inflated,
2. the data is longitudinal with multiple nested classification factors (time point nested in marker combination nested in stimulus nested in participant),
3. there are many explanatory variables and in some cases many response variables, and
4. the response is not linear over time.

In order to address these challenges, generalised linear mixed effect models (GLMEM's) as well as latent variable models (LVM's) were used to model the data. In the subsequent sections, the theory underlying these models shall be presented in order to show how these models account for the above challenges and to justify why they were chosen for this dataset.

4.3.1 GLMEMs using GAMLSS

This section shall define GLMEMs using the generalised additive models for location, scale and shape (GAMLSS) framework. GAMLSS refers to a type of distributional regression model where the response can have any distribution and the parameters of that distribution can depend on other explanatory variables [69]. This section begins by describing the evolution of GAMLSS from the humble linear regression model. The theory described will be a summary of that presented by Stasinopoulos *et al.* (2017) [69].

A linear regression model (LM) with response Y, r covariates $x_1, ..., x_r$ and sample size n is defined as follows

$$Y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_r x_{ir} + \epsilon_i$$

where

$$\epsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2), \qquad \text{and} \qquad i = 1, 2, ..., n,$$

where ϵ_i are independently and identically normally distributed with mean zero and variance σ^2. Note that this specification is equivalent to $Y_i \sim N(\mu_i, \sigma^2)$, where $\mu_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_r x_{ir}$ and Y_i is independently and identically normally distributed. This can be written in vector form as follows

$$\begin{aligned} \boldsymbol{Y} &\overset{\text{iid}}{\sim} N(\boldsymbol{\mu}, \sigma^2 \mathbf{1}) \\ \boldsymbol{\mu} &= \boldsymbol{X}\boldsymbol{\beta}, \end{aligned}$$

where Y is a $n \times 1$ response vector, μ is a $n \times 1$ vector of means, $\mathbf{1}$ is a $n \times 1$ vector of ones, $\boldsymbol{X}$ is a $n \times (r+1)$ design matrix and $\boldsymbol{\beta}$ is a $(r+1) \times 1$ vector of regression coefficients. Note that the design matrix will have a column of ones to account for the intercept term, hence why there are $(r+1)$ columns instead of r. When fitting a LM, the parameters that need to be estimated are σ^2 and $\boldsymbol{\beta}$, a total of $r+2$ parameters.

Generalised linear models (GLMs) build on LMs by making two key changes to the model formula

1. the response variable Y can have any distribution from the exponential family, not just the normal distribution, and
2. a monotonic link function $g(.)$ can be used to model the relationship between μ and $\boldsymbol{X}\boldsymbol{\beta}$.

A GLM can be written in vector form as follows

$$g(\mu_i) = \beta_0 + \beta_1 x_{i1} + ... + \beta_r x_{ir}$$

where

$$Y_i \overset{\text{iid}}{\sim} \xi(\mu_i, \phi), \qquad \text{and} \qquad i = 1, 2, ..., n,$$

where ξ represents a distribution from the exponential family with dispersion parameter ϕ. The above can be written in vector form as follows

$$\begin{aligned} \boldsymbol{Y} &\overset{\text{iid}}{\sim} \xi(\boldsymbol{\mu}, \phi \mathbf{1}) \\ \boldsymbol{\eta} &= g(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta}, \end{aligned}$$

where $\boldsymbol{\eta}$ is called the linear predictor. The great advantage of using a GLM over a LM is that it allows for modelling of non-normal responses, such as the data used in this project, with the exponential family of distributions. The exponential family of distributions includes a wide variety of distributions such as the normal, Poisson, gamma, inverse Gaussian and Tweedie distributions. In order to model a response that may not be normal and thus may have a range that is subset of the real line, a link function is needed to ensure that the parameter estimates remain within their appropriate range.

Generalised additive models (GAM's) build further on GLM's by allowing more flexible modelling between $\eta = g(\mu)$ and the continuous explanatory variables through the addition of smoothing functions. A GAM can be written as follows

$$\begin{aligned}\boldsymbol{Y} &\overset{\text{iid}}{\sim} \xi(\boldsymbol{\mu}, \phi\mathbf{1}) \\ \boldsymbol{\eta} &= g(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta} + s_1(x_1) + ... + s_J(x_J),\end{aligned}$$

where $s_j(x_j)$ is a chosen non-parametric smoothing function applied to continuous explanatory variable x_j, where $j = 1, 2, .., J$. Note that the explanatory variables x_j do not have to be included in the design matrix $\boldsymbol{X}$, hence the use of different subscripts. Thus the relationship between $\eta = g(\mu)$ and the explanatory variables no longer has to be linear as it was before. Since smoothing functions will not be used in models fit in this project, no further details shall be given.

GAMLSS then builds on GAM's by making three key changes to the model formula

1. The distribution of Y no longer has to belong to the exponential family,
2. Y is allowed to have a distribution with up to four different parameters (mean, scale and two shape parameters) instead of just two, and
3. All four of these parameters can be modelled as functions of the explanatory variables, not just μ.

GAMLSS can be defined as follows

$$\begin{aligned}\boldsymbol{Y} &\overset{\text{iid}}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}) \\ \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \boldsymbol{X}_1\boldsymbol{\beta}_1 + s_{11}(x_{11}) + ... + s_{1J_1}(x_{1J_1}) \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \boldsymbol{X}_2\boldsymbol{\beta}_2 + s_{21}(x_{21}) + ... + s_{2J_2}(x_{2J_2}) \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \boldsymbol{X}_3\boldsymbol{\beta}_3 + s_{31}(x_{31}) + ... + s_{3J_3}(x_{3J_3}) \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \boldsymbol{X}_4\boldsymbol{\beta}_4 + s_{41}(x_{41}) + ... + s_{4J_4}(x_{4J_4})\end{aligned}$$

where $D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$ is a four-parameter distribution, σ is the scale parameter, related to the variance by $V(Y) = \sigma^2\mu^2$, and $\boldsymbol{\nu}$ and $\boldsymbol{\tau}$ are shape parameters often related to skewness and kurtosis. The distribution $D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$ can be continuous, discrete or mixed with approximately 100 different distributions compatible with the `gamlss` [70] package in `R` and the option to specify your own custom distribution. There are also many zero-inflated distributions that are compatible with the `gamlss` package, which makes GAMLSS even more appropriate for the data used in this project.

As shown, GAMLSS models allow for linear or non-linear parametric functions of the explanatory variables, or non-linear smoothing functions of explanatory variables. However, they also allow for the inclusion of various additive terms, including random effects, which is what makes them relevant to this project.

GAMLSS allows for the addition of random effects because smoothing can be represented as a random effect model. Random effects are useful for modelling longitudinal data as they are able to capture the within-participant correlation between observations. This is achieved by assuming that there exist participant-level random effects variables $\boldsymbol{b}_i$ which provide an additional source of the variation within the data and are independently normally distributed with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\psi}$. Random effects can either be included as intercept random effects where they are simply added to the model or as slope random effects where they are allowed to interact with explanatory variables in the model. Although it is possible to include random effects on the mean, scale parameters and two shape parameters, for the purposes of this project random effects shall only be included on the mean. Thus, GAMLSS with the addition of a random effect on the mean can be defined as follows

$$
\begin{aligned}
\boldsymbol{Y} &\overset{\text{iid}}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}) \\
\boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{Z}_1\boldsymbol{b}_1 \\
\boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \boldsymbol{X}_2\boldsymbol{\beta}_2 \\
\boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \boldsymbol{X}_3\boldsymbol{\beta}_3 \\
\boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \boldsymbol{X}_4\boldsymbol{\beta}_4.
\end{aligned}
$$

For the purposes of this project, only the mean and scale parameters shall be modelled in terms of explanatory variables. If a distribution with one or two shape parameters is chosen, then these parameters shall be modelled as constants.

Model Components

There are three aspects of the model construction that need to considered prior to model specification and the model building procedure. These include

1. the link function,
2. the parameterisation of the time component, and
3. the covariates of interest.

These aspects of the model construction will be mostly the same across all models fit in this project. Other aspects of model construction which will not necessarily be the same across all models, such as selection of the appropriate conditional distribution and determination of the structure of the random effects, shall be determined during the model building procedure which is discussed in section 4.3.1.

Link-function

In order to fit a GLMEM to the data, an appropriate link function needs to be specified. The responses of interest for all three aims are frequencies and thus have range $(0, 100)$, meaning that $\mu \in (0, 100)$. The link function g is needed to convert μ from the range $(0, 100)$ to the same range as $X\beta$, which is $(-\infty, \infty)$. The first potential candidate is the log link function, which maps from the positive reals to the reals. Using the log link,

$$\begin{aligned} \boldsymbol{X\beta} &= g(\mu) = \log(\mu) \in (-\infty, \infty), \\ \mu &= g^{-1}(\boldsymbol{X\beta}) = e^{\boldsymbol{X\beta}} \in (0, \infty). \end{aligned}$$

Technically this is not ideal because μ should have an upper bound of 100, but since all the frequencies are extremely low and nowhere near this upper bound, this is likely to not be a problem. However, if one did want to ensure that the upper bound was preserved, one could convert the response variable Y to a proportion rather than a frequency and then use a logit link function which maps from the $(0, 1)$ to the reals. Converting the response to a proportion (i.e. dividing it by 100) would result in $\mu \in (0, 1)$. Using a logit link,

$$\begin{aligned} \boldsymbol{X\beta} &= g(\mu) = \log\left(\frac{\mu}{1-\mu}\right) \in (-\infty, \infty), \\ \mu &= g^{-1}(\boldsymbol{X\beta}) = \frac{1}{1+e^{-\boldsymbol{X\beta}}} \in (0, 1), \end{aligned}$$

and thus the upper bound is preserved. However, using a logit link may exacerbate the existing problem of numerical underflow (because of dividing Y by 100) as well as make interpretation of variable coefficients more difficult. Thus, it was decided to rather use the log link function for all our models.

Parameterisation of the time component

An important consideration when fitting a GLMEM is that parameterisation of the time com-

ponent. It is expected that for all responses, the vaccine-induced immune response will increase from baseline to peak time point and then decrease from peak to memory. Thus, the cell frequencies are not linear over time but are instead triangular in shape, suggesting that perhaps a spline would be appropriate. However, another issue is the choice of using categorical time points baseline, peak or memory or using the actual sample collection days which are continuous. The problem with using days is that the longitudinal profiles of the response will differ significantly depending on the participant and the number of vaccine administrations they received. Referring back to Figure 3.5, participants who received 3 administrations of the vaccine would have a linear increase in response between days 0 and 128, but the profile of someone who received only one administration of the vaccine will be triangular on that same range, peaking at day 14 and then decreasing to day 128. This would make fitting a spline impossible as the break point would need to be different for different participants, depending on the number of vaccine administrations that they received. Thus, in order to ensure consistent response profiles, it was decided to use the categorical time points since all participants should experience a linear increase from baseline to peak, and then a linear decrease from peak to memory. Since a categorical time variable shall be used, it is not necessary to apply any transformations to the time component, such as modelling it as a spline, since it essentially already acts as a spline with a break point at the peak time point. Thus, all of the GLMEMs that will be fit will include the following time component

$$\beta_1 I_j^{Peak} + \beta_2 I_j^{Mem},$$

where I_j^{Peak} is an indicator variable that equals 1 if time point j = peak and zero otherwise, and similarly so for I_j^{Mem}, thus using baseline as the reference time point.

If the effect of time point differs depending on the actual day that the time point occurred, an interaction between time point and number of administrations can also be included in order to capture this variation. This is because the day that the time point occurred is mainly dependent on the number of administrations of the vaccine that the participant received. The potential addition of this interaction term to the model will be investigated during the model building process as part of the selection of interaction terms.

The reason for considering an interaction between time point and number of administrations rather than time point and days is that the former is of more interest with regards to the aims of this project compared to the latter. Furthermore, it would be redundant to include a day and time point main effect in the model (which would be necessary if including an interaction between day and time point) because time point is completely determined by day. Thus, day will not be included as a main effect either.

Covariates of interest

The key covariates of interest are categorical variables time point and QFT status as well as continuous variables vaccine concentration and number of administrations. Since these variables are present in all three aims of this project, they shall be included in all fitted models. For

certain response variables, categorical variables stimulus and marker combination may also be included. Other variables that will be considered during the fixed effect selection process are ethnicity, age and sex. These variables will only be included if they improve the fit of the overall model. The choice of interaction variables shall be determined in the model building process which is discussed in Section 4.3.1.

Single level of grouping

Multilevel models are useful when data has multiple nested classification factors, such as the data used in this project where time point is nested within marker combination nested in stimulus nested in participant [71]. Either, one would have to subset by these classification factors into smaller datasets and fit multiple mixed effect models, one to each dataset, or one can fit a multi-level mixed effect model to the whole dataset. This project shall explore both approaches by fitting mixed effect models with single, two and three levels of grouping.

The first set of models that shall be fit are single-level GLMEMs grouped at the participant level for each marker-stimulus combination, as depicted in Figure 4.1. In other words, observations are nested within participants.

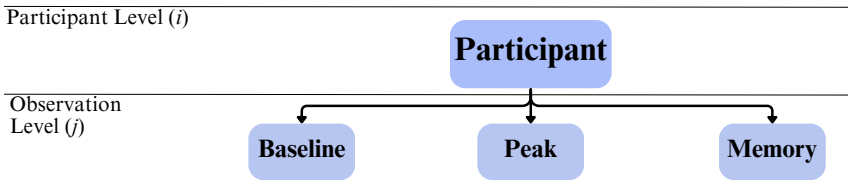


Figure 4.1: Illustration of the multivariate multilevel nature of the data used to fit the single-level GLMEM, where time point is nested within participant.

The general formula for a single-level GLMEM can be written as

$$\begin{aligned} \log(\mu_{ij}) &= \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i, \\ \log(\sigma_{ij}) &= \boldsymbol{w}_{ij}^T\boldsymbol{\alpha}, \\ i &= 1, ..., I, \qquad j = 1, 2, 3, \end{aligned}$$

where

$$\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{\psi}) \qquad \text{and} \qquad \epsilon_{ij} \sim N\left(0, \sigma_e^2\right).$$

μ_{ij} is the average CD4 T cell frequency for participant i at time j, while $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ are vectors of observation- and participant-level covariates with dimensions $p \times 1$ and $q \times 1$ respectively and $\boldsymbol{\beta}$ is a $p \times 1$ vector of variable coefficients. I is the total number of subjects in the dataset, $\boldsymbol{b}_i$ is a $q \times 1$ vector of participant-specific random effects and ϵ_{ij} is the observation-level random error (i.e. the within-group error). $\boldsymbol{b}_i$ is assumed to be independent for different i, ϵ_{ij} is assumed to

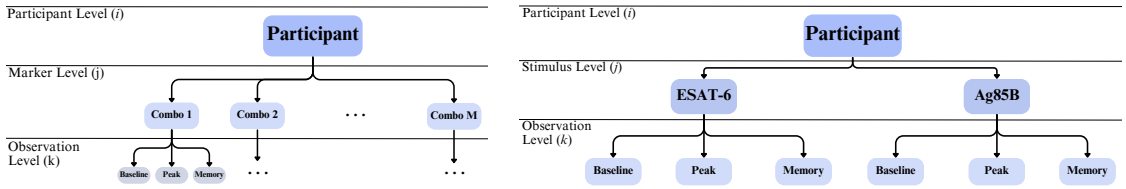
be independent for different observations and the $\boldsymbol{b}_i$ is assumed to be independent of ϵ_{ij}.

For simplicity, only μ_{ij} has been modelled with a random effect, not σ_{ij}. The models used in this project will at most model σ_{ij} as well as μ_{ij} as a function of explanatory variables, and any potential shape parameters shall be kept constant. The purpose of modelling σ_{ij} is to try cope with the within-participant heterogeneity since the `gamlss` package does not support specification of a variance function. Usually, specifying a variance and/or correlation function allows one to relax the assumptions of independent within-participant errors with constant variance. The reason for not modelling the shape parameters was that the residuals did not exhibit excess skewness or kurtosis.

It was decided to restrict the random effects to the intercept and time effects in the models. Thus, the only covariate that $\boldsymbol{z}_{ij}$ could contain is time point, along with a column of ones in order to fit a participant-specific intercept random effect.

Two levels of grouping

The second set of models that shall be fit are two-level GLMEMs which combine either markers or stimuli into one model. Thus, the two-level GLMEMs are grouped at the stimulus or marker combination level (second level) and the participant level (first level), as depicted in Figure 4.2. In other words, observations are nested in stimulus (or marker combination) which is nested in participant. Whether the model is grouped by stimulus or marker combination depends on the aim being addressed and the response variable of interest.



(a) (b)

Figure 4.2: Illustration of the multivariate multilevel nature of the data used to fit the two-level GLMEM's, where either (a) time point is nested within marker combination which is nested within participant, or (b) time point is nested within stimulus which is nested within participant.

The general formula for a two-level GLMEM can be written as

$$\begin{aligned} \log(\mu_{ijk}) &= \boldsymbol{x}_{ijk}^T\boldsymbol{\beta} + \boldsymbol{z}_{ijk}^T\boldsymbol{b}_i, \\ \log(\sigma_{ijk}) &= \boldsymbol{w}_{ijk}^T\boldsymbol{\alpha}, \\ i = 1, ..., I, \qquad & j = 1, ..., J, \qquad k = 1, 2, 3 \end{aligned}$$

where

$$\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{\psi}_1) \qquad \text{and} \qquad \epsilon_{ijk} \sim N\left(0, \sigma^2\right).$$

μ_{ijk} is the average CD4 T cell frequency for participant i, the first level of grouping, under stimulus j (or expressing marker combination j), the second level of grouping, at time point k. $\boldsymbol{x}_{ijk}$ is a vector of observation-level covariates with dimensions $p \times 1$. $\boldsymbol{b}_i$ is the first-level random effect with dimensions $q_1 \times 1$ and $\boldsymbol{z}_{ijk}^T$ is its associated vector of covariates. $\boldsymbol{b}_i$ is assumed to be independent for different i and ϵ_{ijk} is assumed to be independent for different observations. Again, the random effect is assumed to be independent of the within-group errors.

In the case where stimulus is the second level of grouping, there was the option to either include stimulus as a fixed effect or as a random effect. Since there are only two levels of stimulus, ESAT-6 and Ag85B, and since a key aim of this project is how antigen specificity effects vaccine-induced immune response and because the difference between stimuli is expected to be systematic and not random, it was decided to include stimulus as a fixed effect rather than a random effect. For these reasons, a stimulus-specific (nested within participant) random effect was not included in the above model specification, and all two-level GLMEM's with stimulus as the second level of grouping fit in this project will include stimulus as a fixed effect. Similarly, in the case where marker combination is the second level of grouping, marker combination will be included as a fixed effect rather than a random effect.

Again, it was decided to restrict the random effects to the intercept and time effects in the models. Thus, the only covariates that $\boldsymbol{z}_{ijk}$ could contain is time point and stimulus or marker combination, along with a column of ones in order to fit a participant-specific intercept random effect.

Three levels of grouping

The final set of models that shall be fit are three-level GLMEMs grouped at the marker combination level (third level), stimulus level (second level) and participant level (first level), as depicted in Figure 4.3. The purpose of this model is to be able to fit a single model to all marker combinations for both stimuli.

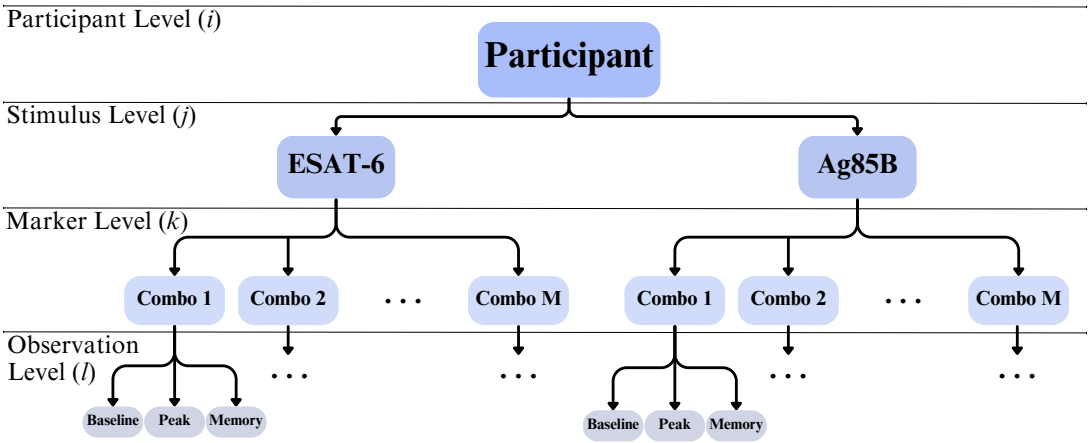


Figure 4.3: Illustration of the multivariate multilevel nature of the data used to fit the three-level GLMEM, where time point is nested within marker combination nested within stimulus nested within participant.

The general formula for a three-level GLMEM can be written as

$$\begin{aligned}
&\log(\mu_{ijkl}) = \boldsymbol{x}_{ijkl}^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{b}_{ijkl},\\
&\log(\sigma_{ijkl}) = \boldsymbol{w}_{ijkl}^T\boldsymbol{\alpha},\\
&i = 1, ..., I, \qquad j = 1, 2, \qquad k = 1, ..., C, \qquad l = 1, 2, 3,
\end{aligned}$$

where

$$\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{\psi}_1) \qquad \text{and} \qquad \epsilon_{ijkl} \sim N\left(0, \sigma^2\right).$$

μ_{ijkl} is the average CD4 T cell frequency for subject i, the first level of grouping, under stimulus j, the second level of grouping, for marker combination k, the third level of grouping, and at time point l. $\boldsymbol{x}_{ijkl}$ is a vector of observation-level covariates with dimensions $p \times 1$. $\boldsymbol{b}_i$ is the first-level random effect with dimensions $q_1 \times 1$ and $\boldsymbol{z}_i^T$ is its associated vector of covariates. $\boldsymbol{b}_i$ is assumed to be independent for different i and ϵ_{ijkl} is assumed to be independent for different observations. Again, the random effect is assumed to be independent of the within-group errors.

Using the same reasoning as was used for the two-level GLMEM, stimulus and marker combination shall be included as a fixed effects rather than random effects in all three level models. Similarly to the two-level GLMEM, the only covariates that $\boldsymbol{z}_{ijkl}$ could contain is time point, stimulus and marker combination, along with a column of ones in order to fit a participant-specific intercept random effect.

Model building procedure

The model building procedure for all GLMEM's fit in this project shall follow the following steps

1. Select an appropriate conditional distribution for the response variable by considering its marginal distribution.
2. Determining the structure of the random effects by looking at all possible structures.
3. Select the main effects using step-wise selection.
4. Select the interaction terms.
5. Select the main effects for the parameterization of σ using step-wise selection.
6. Examine model diagnostic plots and check model assumptions.

The first step is to choose an appropriate distribution for the response variable. Even though the distribution specified for a GLMEM is that for the response variable conditional on the model's explanatory variables, the distribution shall be chosen by considering only the marginal

distribution of the response variable. This is because the model's explanatory variables will only be known after steps 3 and 4 have been performed, both of which require the specification of the conditional distribution of the response. Thus, an appropriate conditional distribution of the response shall be chosen based solely on the empirical marginal distribution of the response, since this is all the information that is available and since the marginal distribution of the response usually indicates which distribution is likely to work.

The structure of the random effects shall be determined by fitting models containing only the covariates of interest, as specified in Section 4.3.1, but with different random effect structures and comparing the fit of these models. Specifically, models with only intercept random effects as well as models with intercept and different slope random effects shall be compared. For accuracy, restricted maximum likelihood (REML) shall be used instead of maximum likelihood (ML) when fitting the models, since the `gamlss` package does not allow for ANOVA comparison of models. The optimal model will be the one with the lowest generalized Akaike information criterion (GAIC) with $k = \sqrt{\log(n)}$.

After the selection of random effect structure, step-wise selection shall be used to determine whether the addition of any of the demographic variables sex, ethnicity and age will improve the model fit. This shall determine the main effects of the model.

For steps 3 and 4, step-wise selection shall also be used to select the two-way interactions since all possible two-way interactions were plausible. Of course, only interactions between the main effects selected in the previous step will be considered. After the two-way interactions were selected, three-way interactions using these two-way interactions will be added to the model if they improved the model fit, which will again determined by GAIC. Note that three-way interactions will only be considered when there were already two or more two-way interactions in the model.

Before conducting inference on a fitted model, certain underlying model assumptions must be checked in order to better understand any potential weakness of the model used. As outlined by Pinheiro and Bates (2006) [71], there are two key assumptions made when fitting a mixed effect model:

1. Within-participant errors are independent and identically normally distributed with mean of zero and variance σ_e^2 and they are independent of the random effects.
2. Random effects are normally distributed with mean zero and covariance matrix $\boldsymbol{\psi}$ and are independent between subjects.

The assumption that the within-participant errors are identically distributed with the same variance (i.e. the assumption that the within-participant errors are homoscedastic) can be assessed by plotting the normalised quantile residuals against participant-specific covariates as well as against the fitted values. If the model assumption holds, then these residuals should be homoscedastic. The normality assumptions of both the within-participant errors as well as the random effects shall be assessed using qq-plots.

Only diagnostic plots shall be used to assess these assumptions rather than hypothesis tests, since the diagnostic plots and hypothesis tests are rarely not in agreement [71]. Diagnostic plots not only allow one to identify problems in the model, but they can also potentially provide information as to what is causing the problem where as hypothesis tests cannot [72].

4.3.2 LVM's

An important limitation of GLMEM's is that they can only be used to model univariate responses or multivariate responses whose structure has been changed to a long univariate response, which is what has been done thus far. For all three project aims, the response can be considered multivariate. For the first aim, the response matrix would consist of two columns, one for each stimulus. For the second aim, the response would consist of $2 \times 16 = 32$ columns, one for each stimulus and cytokine combination pair. For the third aim, the response would consist of $2 \times 4 = 8$ columns, one for each stimulus and memory combination pair. In Section 4.3.1, multilevel, hierarchical GLMEM's were used to account for the multivariate nature of the response variable, however the response was always treated as univariate. Latent variable models (LVM's) extend the univariate GLMEM's to multivariate response variables through the use of latent variables, making them a potential alternative approach to modelling the response variables associated with each aim.

Although LVM's could be used for all three aims, this project shall only use LVM's to investigate the second aim, as the dimensionality of the response is far higher compared to that of the other aims. In order to reduce the dimensionality of the response, output from COMPASS shall be used to select the cytokine combinations that are biologically relevant, the details and results of which can be found in Section 6.2. Since COMPASS had to be applied to each stimulus separately, the cytokine combinations selected in response to ESAT-6 may not be the same as those selected in response to Ag85B. Thus, two separate LVM's would need to be fit to each stimulus: a LVM fit to a $V^{Ag85B} \times C^{Ag85B}$ response matrix and another LVM fit to a $V^{ESAT6} \times C^{ESAT6}$ response matrix, where V^{Ag85B} and V^{ESAT6} are the number of participants visits with responses available for Ag85B and ESAT-6 respectively, and C^{Ag85B} and C^{ESAT6} are the number of biologically relevant cytokine combinations when stimulated by Ag85B and ESAT-6 respectively. The LVM theory described below will be a summary of that presented by Warton *et al.* (2015) [73] and Hui *et al.* (2016) [74], adapted to be applicable to these specific response variables.

Model Structure

LVM's were initially developed to analyse abundance data in ecology, where the abundance of various different taxa, such as species, are observed at various different sites. The rows of the response matrix would be the sites and columns would be the species. Site-specific variables would be considered covariates and species-specific variables would be considered traits. LVM's allow for both covariates and traits to be included in the model.

The general form of the LVM specification covariates but without traits for the cytokine combination responses is as follows

$$g(\mu_{ijk}) = \boldsymbol{x}_{ik}^T\boldsymbol{\beta}_j + \boldsymbol{z}_{ik}^T\boldsymbol{b}_i + \boldsymbol{u}_{ik}^T\boldsymbol{\theta}_j,$$
$$i = 1, ..., I, \qquad j = 1, ..., C^s, \qquad k = 1, 2, 3$$

where

$$\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{\psi}_1) \qquad \text{and} \qquad \boldsymbol{u}_{ik} \sim N(\boldsymbol{0}, \boldsymbol{I}).$$

μ_{ijk} is the average frequency of CD4 T cells from participant i at time point k expressing cytokine combination j and $g(.)$ is the associated link function. $\boldsymbol{x}_{ik}$ is a vector of observation-level covariates with dimensions $p \times 1$. $\boldsymbol{b}_i$ is the participant-specific random effect with dimensions $q \times 1$ and $\boldsymbol{z}_{ik}^T$ is its associated vector of covariates. $\boldsymbol{u}_{ik}^T$ is the latent variable and $\boldsymbol{\theta}_j$ is the factor loadings, which together induce correlations between the cytokine combinations j of participant i at time point k. Notice that the covariate coefficients $\boldsymbol{\beta}_j$ are now cytokine combination-specific, allowing for a different covariate coefficients for each cytokine combination. The latent variables $\boldsymbol{u}_{ik}^T$ are assumed to be unknown and drawn randomly and independently from a bivariate standard normal distribution.

Latent variables

As previously mentioned, latent variables and their factor loadings induce correlations between the cytokine combinations. Thus, latent variables can be interpreted as capturing any residual covariance not explained by the model covariates and hence can be thought of as missing model predictors.

The latent variable $\boldsymbol{u}_{ik}^T$ is treated as a random effect and since it is common across all cytokine combinations for participant j and time point k, $\boldsymbol{u}_{ik}^T$ is able to induce correlation across the different cytokine combinations. The degree of the correlation between the cytokine combinations is controlled for by the factor loadings $\boldsymbol{\theta}_j$ and the number of latent variables included in the model controls the model's complexity.

The above model specification assumes that latent variables are independent across participant visits, however other correlation structures are possible. The advantage of assuming independence is that it creates a more parsimonious model and substantially decreases computation time. The inclusion of participant-specific random effects accounts for the correlation between visits from the same participant, and in fact it is recommended to include row random effects rather than assume a non-independent correlation structure for the latent variables [75].

The inclusion of latent variables also allows one to produce model-based ordination plots. Ordination is a visualisation tool that attempts to capture the main features of multivariate data and represent it in two or three dimensions. The latent variables represent the main axes of

covariation of CD4 T cell subset frequencies across the different cytokine combinations, and thus the number of latent variables determines the dimension of the ordination plot. The ordination plot is created by plotting the V^{Ag85B} (or V^{ESAT6}) estimated latent variable values as a scatter plot and then overlaying a bi-plot by plotting the corresponding C^{Ag85B} (or C^{ESAT6}) factor loadings. Since one of the goals of fitting LVM's is to create these ordination plots, all LVM's shall only include two latent variables. If ordination plots are not a priority of one's analysis, then it is suggested to fit models with different numbers of latent variables and compare the fit using diagnostic plots.

Inclusion of traits

The above LVM can be extended to include traits for the responses j. This is done by modelling the variable coefficients, β, as random effects, which reduces the number of parameters that need to be estimated. Specifically, the model coefficients β are assumed to be drawn from normal distributions

$$\beta_{wj} \sim N\left(\kappa_{0,w} + \boldsymbol{T}_j'\boldsymbol{\kappa}_w, \sigma_w^2\right),$$

where β_{wj} is the w^{th} element of $\boldsymbol{\beta}_j$ and $\boldsymbol{T}_j$ is the vector of traits corresponding to cytokine combination j. $\kappa_{0,w}$ and $\boldsymbol{\kappa}_w$ are the regression parameters relating the traits to the cytokine combination-specific variable coefficients. LVM's with the inclusion of traits shall be discussed in further detail in Section 6.3.2 where these models are applied.

Comparison of GLMEM's and LVM's

The main difference between GLMEM's and LVM's is that univariate GLMEM's modelling multivariate responses whose structure has been changed to a long univariate response do not model the residual covariation across the multiple responses. In other words, while both models use covariates and shared coefficient estimates to model covariation in the responses, LVM's use latent variables and factor loadings to also model any residual covariation not explained by the model covariates, while GLMEM's do not.

Other, more minor differences include that univariate GLMEM's share coefficients across different cytokine combinations, while LVM's allows for different cytokine combination-specific coefficient's. This is attractive because it is very likely that the relationship between immune response and the covariates is not the same for different cytokine combinations. While GLMEM's can account for this through the inclusion of interactions between the covariates and the cytokine combinations, this requires more parameters to be estimated.

While response traits (i.e. cytokine combination-specific covariates) can be included in GLMEM's, they do not necessarily result in a simpler model as they do when included in LVM's. Further-

more, the GLMEM's considered in this project do not include response traits.

Model building procedure

Hui *et al.* (2016) [75] strongly advise against using information criteria to compare model fit, instead suggesting that the stochastic search variable selection (SSVS) method provided by the `boral` package should be used for model selection instead. However, the authors note that using SSVS results in the summaries of the model coefficients such as posterior medians and highest posterior density (credible) intervals possibly being problematic. An alternative approach is to use the differences in the trace of the estimated residual covariance matrix induced by the latent variables, which quantifies how well the covariates describe the covariation between cytokine combinations. This can be used to compare LVM's in terms of how much more or less the model's covariates account for the covariation between cytokine combinations. However, while this could potentially be used for selection of model covariates this cannot be used for trait and random effect selection. Thus, the LVM's fit here shall be illustrative and not focus on the model building procedure. The choice of random and fixed effects to include in the model shall be informed by those included in the GLMEM's fit previously, rather than by comparison of LVM's with different random and fixed effects. The choice of the conditional distribution of the response as well as the link function shall also be informed by the GLMEM's. Where the fit of LVM's are compared, they shall be compared using diagnostic plots only.

Chapter 5

Investigating the effects of vaccine regimen, *M.tb* sensitisation and antigen specificity on CD4 T cell *magnitude*.

In this chapter, the first aim of this project, that is the effects of vaccine regimen, *M.tb* sensitisation and antigen specificity on CD4 T cell *magnitude*, shall be investigated. Along with exploring this aim, this chapter shall also investigate the hypothesis that *M.tb*-sensitised individuals will require a lower vaccine concentration to induce the greatest magnitude of cytokine positive CD4 T cells compared to *M.tb*-unsensitised individuals.

As discussed in Section 1.2, the magnitude of the T cell response is equivalent to the frequency of antigen-specific CD4 T cells observed. Cytokine positive cells are antigen-specific, and thus the frequency of antigen-specific CD4 T cells can be measured by measuring the frequency of cytokine positive cells. Even though cytokine positive cells should generally only be observed after stimulation with either ESAT-6 or Ag85B, in practice there are low frequencies of cytokine positive cells observed when unstimulated. Thus, one of the response variables of interest in this chapter will be the TRF, which is the background subtracted frequency of cytokine positive cells.

Notice that the TRF is calculated for all participant visits, regardless of whether that participant visit actually responded to stimulation (i.e. the frequency of cytokine positive cells before stimulation was significantly greater after stimulation). Thus, another response variable of interest in this chapter is the response rate, the proportion of participant visits which can be classified as responders out of all participant visits, where responders shall be classified using MIMOSA as outlined in Section 4.2.2.

This chapter will begin by implementing a standard approach to investigating these objectives, consisting of a non-parametric approach to investigating the TRF in Section 5.1.1, followed by the investigation of the response rate obtained using MIMOSA in Section 5.1.2. Section 5.2 will then fit single- and two-level GLMEM's to TRF and compare the fit and inference made from these models. Finally, Section 5.3 will summarise the main findings made in this chapter and compare the standard approach to the the mixed effect modelling approach.

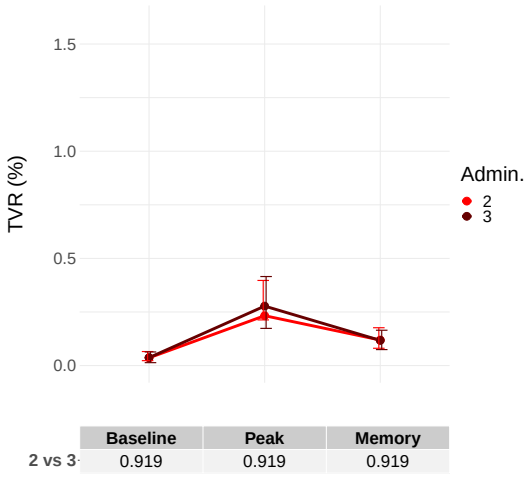
5.1 Standard approach

This section consists of a non-parametric approach exploring the TRF followed by the use of the data analysis pipeline MIMOSA. The non-parametric approach consists of graphical representations and pairwise tests of the TRF. The MIMOSA approach consists of graphical representations of the response rate obtained using MIMOSA.

5.1.1 Non-parametric approach

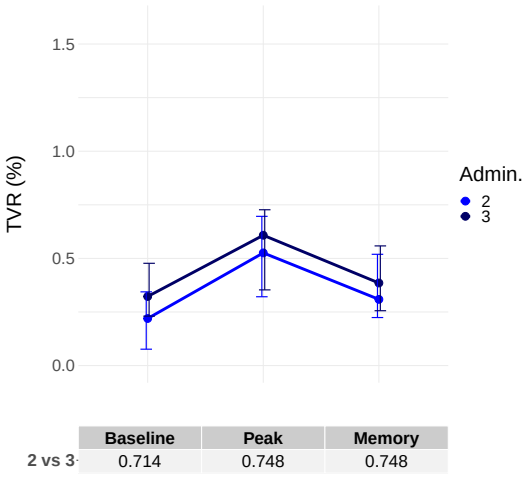
In this section, graphical representations and pairwise tests of the TRF shall be investigated. Since the TRF would need to be calculated separately for Ag85B and ESAT-6, in order to avoid unnecessary repetition, the TRF for the two stimuli shall be combined into a single response called the total vaccine response (TVR). The TVR for participant i at visit j is simply the sum of the Ag85B TRF and the ESAT-6 TRF for that participant visit. TVR is used as a measure of vaccine take, a higher TVR suggesting stronger vaccine-induced immune responses. The separate TRF plots can be found in the Appendix and will be referenced where appropriate.

In order to compare the different vaccine administrations given QFT status, time point and vaccine concentration, the longitudinal profiles of the median TVR were plot, as seen in Figure 5.1. The strongest difference in TVR is seen between one and two administrations for QFT- participants receiving either 15 µg or 50 µg of the vaccine. At the lowest concentration of the vaccine, there seems to be no difference between the number of administrations of the vaccine, in which case the least resource-intensive schedule would be preferred, specifically two administrations of 5 µg of the vaccine. For 15 µg and 50 µg of the vaccine, two administrations of the vaccine produces a much higher TVR at peak and memory for QFT- participants compared to only one administration. There is also no difference between two and three administrations. Thus, since the selected vaccine schedule would need to be administered to QFT- and QFT+ people indiscriminately, two administrations of the vaccine seems to be preferable when a concentration of 15 µg or 50 µg of the vaccine is used. Thus, these plots suggest that two administrations of the vaccine is preferable, regardless of the concentration used, with the caveat that there was no study arm with one administration of 5 µg. There was also evidence that two administrations of the vaccine was preferable when TRF was plotted for both ESAT-6 and Ag85B (see Figures C.1 and C.4).



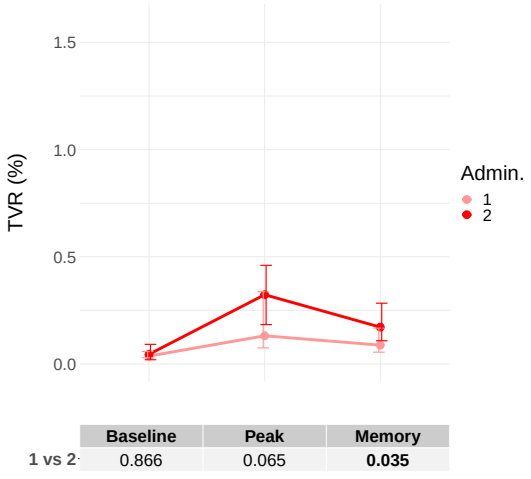
	Baseline	Peak	Memory
2 vs 3	0.919	0.919	0.919

(a) QFT- with 5 µg



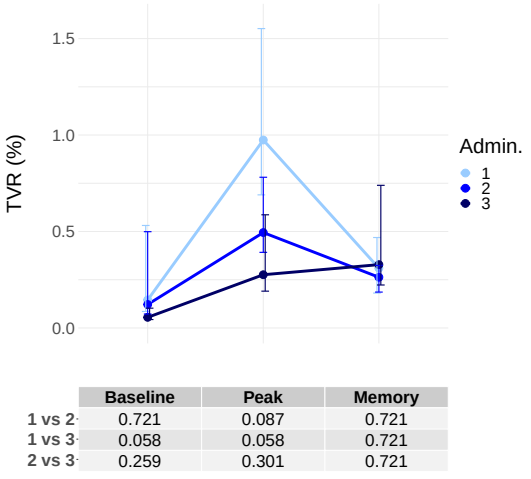
	Baseline	Peak	Memory
2 vs 3	0.714	0.748	0.748

(b) QFT+ with 5 µg



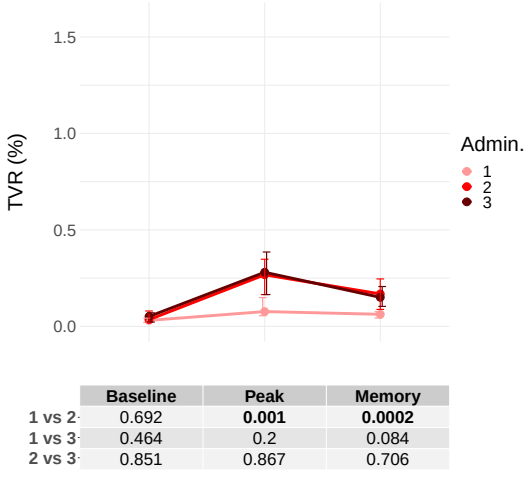
	Baseline	Peak	Memory
1 vs 2	0.866	0.065	**0.035**

(c) QFT- with 15 µg



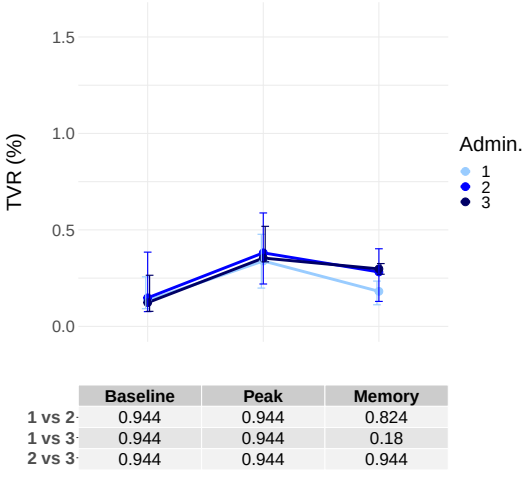
	Baseline	Peak	Memory
1 vs 2	0.721	0.087	0.721
1 vs 3	0.058	0.058	0.721
2 vs 3	0.259	0.301	0.721

(d) QFT+ with 15 µg



	Baseline	Peak	Memory
1 vs 2	0.692	**0.001**	**0.0002**
1 vs 3	0.464	0.2	0.084
2 vs 3	0.851	0.867	0.706

(e) QFT- with 50 µg

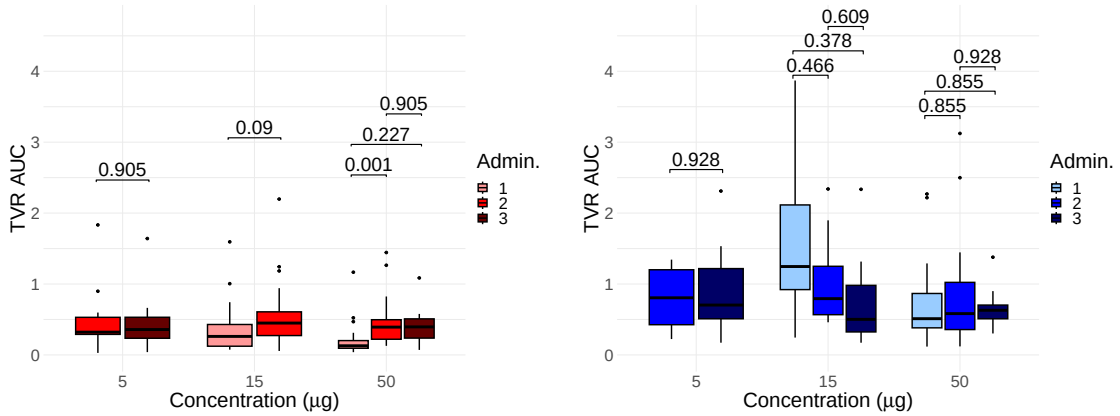


	Baseline	Peak	Memory
1 vs 2	0.944	0.944	0.824
1 vs 3	0.944	0.944	0.18
2 vs 3	0.944	0.944	0.944

(f) QFT+ with 50 µg

Figure 5.1: Longitudinal profiles of the median TVR frequencies of CD4 T cells for different administrations of the vaccine (No. of Admin.). The error bars denote the interquartile range of the TVR frequencies. The medians have been plotted for three different time points: baseline, peak and memory. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing the different numbers of administrations for the same time point are shown in the table.

To further compare the different vaccine administrations without having to subdivide by time point, the area under the curve (AUC) of the longitudinal profiles of the median TVR in Figure 5.1 were plot, as seen in Figure 5.2. The only strong difference in TVR AUC is seen between one and two administrations of 50 µg of the vaccine for QFT- participants. Since two administrations of 50 µg of the vaccine has a higher AUC TVR than one administration and is not different from 3 administrations, this suggest that two administrations is preferable when 50 µg of the vaccine is administered. There is also no difference between two and 3 administrations of 5 µg of the vaccine, suggesting that two administrations is again preferable when administrating 5 µg of the vaccine. Both of these results are in agreement with what was seen in Figure 5.1. However, while Figure 5.1 suggested that two administrations of the vaccine was also preferable when administering 15 µg of the vaccine, Figure 5.2 does not support this. A strong difference in AUC TVR is only seen between one and two administrations of 50 µg of the vaccine for QFT- participants, which is again in agreement with what was seen in Figure 5.1. Overall, Figure 5.2 also suggests that two administrations of the vaccine is preferable. These results are also in agreement with what was seen when plotting the TRF AUC for each stimulus separately, which can be seen in Figures C.2 and C.5.



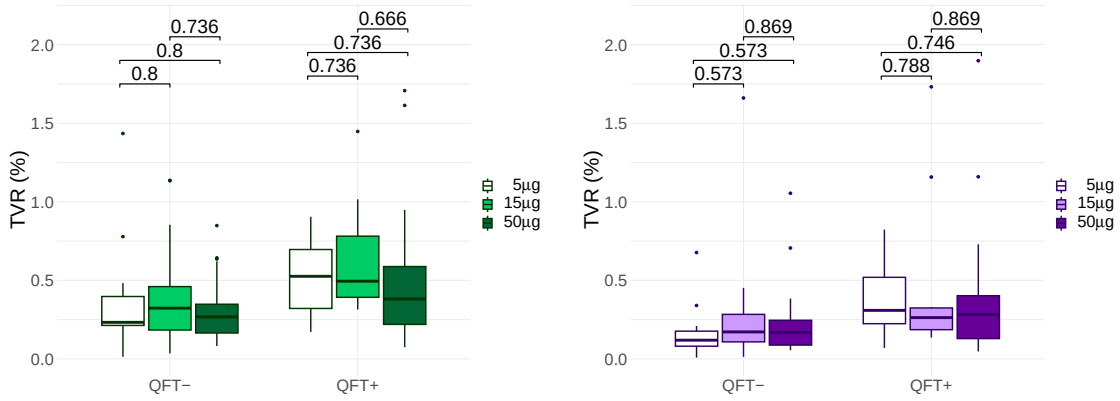
(a) QFT-

(b) QFT+

Figure 5.2: Box-plots of the TVR AUC, plotted for different QFT statuses, vaccine concentrations and number of vaccine administrations received. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing the different numbers of administrations for the same vaccine concentration are shown. For example, for the TVR AUC of QFT- participants who received 50 µg of the vaccine, the Benjamini Hochberg adjusted p-value comparing those who received one administration to those that received two administrations is 0.227.

Since the majority of the evidence seen in Figures 5.1 and 5.2 suggested that two administrations of the vaccine was preferable, it was decided that two administrations would be selected to investigate the optimal vaccine concentration. In order to determine which concentration is optimal when two administrations of the vaccine is used, the TVR at post-vaccination time points for participants who received two administrations of the vaccine was plot, as seen in Figure 5.3. At both peak and memory time points and for both QFT- and QFT+ participants, there are no differences in TVR between the different vaccine concentrations. Thus, Figure 5.3 suggests a concentration of 5 µg of the vaccine is preferable when administering two doses of the vaccine. These results are in agreement with what was seen when plotting the TRF for

each stimulus separately, which can be seen in Figures C.3 and C.6. Note that since there are no participants who received one administrations of 5 µg of the vaccine in this study, one cannot conclude that two doses of 5 µg of the vaccine is preferable over one dose, only that it is preferable over 3 doses. With regards to the hypothesis that *M.tb*-sensitised individuals will require a lower vaccine concentration to induce the greatest magnitude of cytokine positive CD4 T cells compared to *M.tb*-unsensitised individuals, the results seen here do not support this hypothesis. For two vaccine administrations, there is evidence that there is no difference in the magnitude of cytokine positive CD4 T cells between the different vaccine concentrations, for both *M.tb*-sensitised and -unsensitised individuals.



(a) Peak response

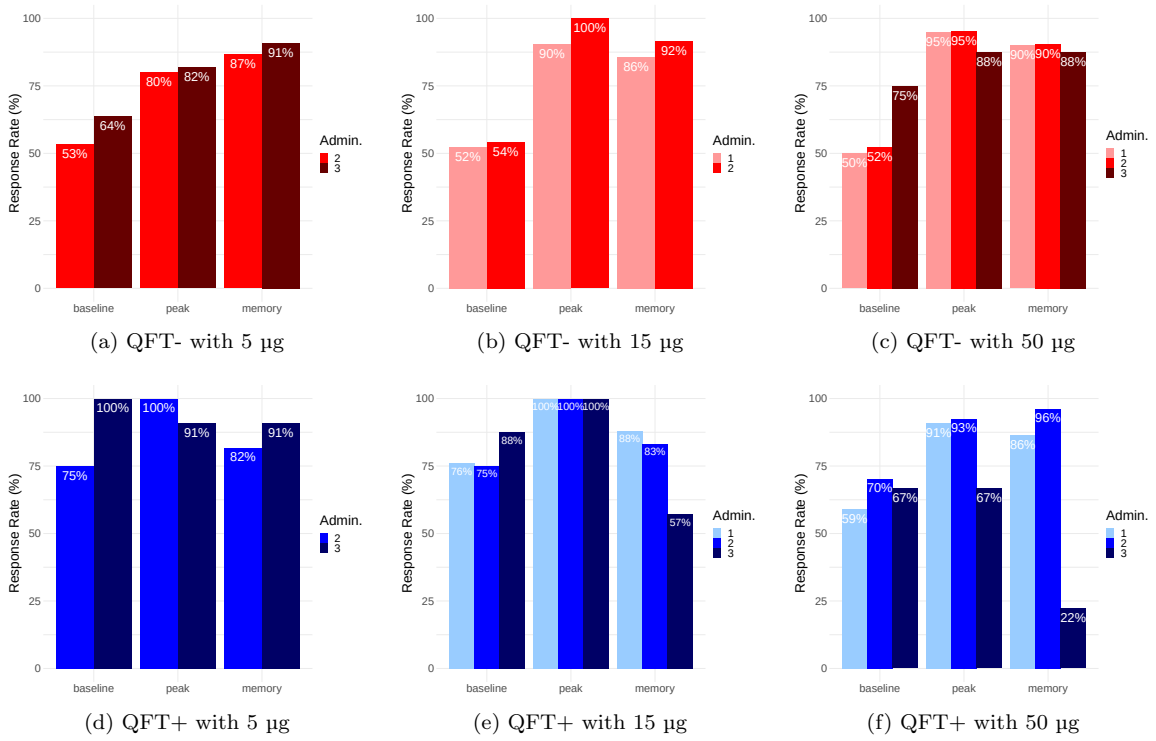
(b) Memory response

Figure 5.3: Box-plots for the peak and memory TVR frequencies for the participants who received two administrations of the vaccine. Box-plots are plotted for different QFT statuses as well as for different vaccine concentrations. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing the different vaccine concentrations for the same QFT status are shown.

5.1.2 MIMOSA

Using results of the data filtering using MIMOSA discussed in Section 4.2.2, the response rate can be used to further investigate the effects of vaccine regimen and QFT status on the magnitude of the immune response. Figures 5.4 and 5.5 show the percentage of participant visits retained as responders after stimulation (i.e. the response rate). Referring to Figure 5.4, when stimulated with Ag85B, the response rate is generally highest at peak and lowest at baseline, regardless of QFT status, vaccine concentration and number of administrations. It was expected that baseline would have the lowest response rate since it was prior to vaccination, and vaccination should increase an individual's immune response. The response rate at baseline is greater for QFT+ participants compared to QFT- participants, which is expected since QFT+ are *M.tb*-sensitised and thus should already have some level of natural immunity prior to vaccination. There does not seem to be much difference in response rate between QFT- and QFT+ participants at peak and at memory, suggesting that the level of vaccine take might not differ between QFT statuses. Although the memory response rate is generally less than the peak response rate, it is still generally higher than the baseline response rate, suggesting that the vaccine does offer some level of long-term immune responses. Two notable exceptions are the QFT+ participants receiving three administrations of either 15 µg or 50 µg of the vaccine, both

of whose response rates are very low. This could suggest that more administrations of higher doses of the vaccine do not offer long term immunity in QFT+ participants.

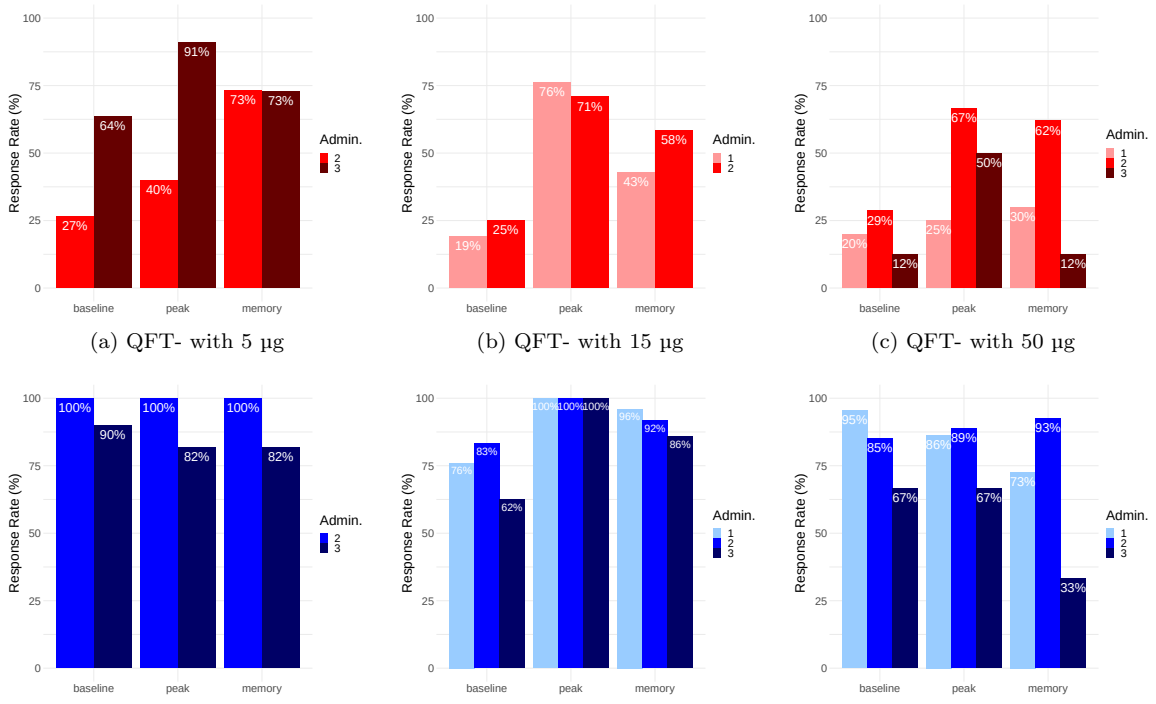


(a) QFT- with 5 µg (b) QFT- with 15 µg (c) QFT- with 50 µg

(d) QFT+ with 5 µg (e) QFT+ with 15 µg (f) QFT+ with 50 µg

Figure 5.4: Bar-plots showing the percentage of Ag85B-stimulated observations retained as Ag85B responders (i.e. the response rate) after applying the data filtering steps discussed in Section 4.2.2.

When stimulated with ESAT-6, as in Figure 5.5, the response rate is again generally highest at peak and lowest at baseline, regardless of QFT status, vaccine concentration and number of administrations. Again, the response rate at baseline is greater for QFT+ participants compared to QFT- participants, except now the difference is much larger than for Ag85B. The response rate at peak and memory generally seems to be lower for QFT- participants compared to QFT+ participants, a difference which was not seen for Ag85B. This could suggest that there is in fact a difference in the level of immunogenicity between QFT- and QFT+ participants, QFT+ participants being afforded a higher level of vaccine take. Although the memory response rate is generally less than the peak response rate, it is again still generally higher than the baseline response rate, suggesting that the vaccine does offer some level of long-term immunity. Again, participants who received three administrations of 50 µg of the vaccine had a much lower response rate, this time regardless of their QFT status. This could further suggest that more administrations of higher doses of the vaccine do not offer long term immunity.



(a) QFT- with 5 µg (b) QFT- with 15 µg (c) QFT- with 50 µg

(d) QFT+ with 5 µg (e) QFT+ with 15 µg (f) QFT+ with 50 µg

Figure 5.5: Bar-plots showing the percentage of ESAT-6-stimulated observations retained as ESAT-6 responders (i.e. the response rate) after applying the data filtering steps discussed in Section 4.2.2.

5.2 GLMEM's

In order to model the magnitude of the immune response, the TRF shall be used as the response variable modelled using three different GLMEM's. Specifically, two single-level GLMEM's shall be fit, one using the TRF under stimulation with ESAT-6 as the response variable and the other using the TRF under stimulation with Ag85B as the response variable. A two-level GLMEM grouped by stimulus nested within participant shall be fit to the TRF under stimulation with either ESAT-6 or Ag85B.

5.2.1 Candidate distributions for the response

In order to choose an appropriate conditional distribution for the response variable, the marginal distribution of the response needs to be assessed. Figures 5.6 (a) and (b) show the density plots for each of the two response variables for the single-level GLMEM's and Figure 5.6 (c) shows the density plot for the response variable of the two-level GLMEM's grouped by stimulus. All variables are continuous on the domain $[0, 100]$, highly zero-inflated and positively skewed. Possible candidate distributions are the Gamma, log-normal and Weibull, since they all at most require two parameters and are suitable for continuous right-skewed data. Note that choosing a log-normal distribution for the response would mean that essentially a linear mixed effect model (LMEM) is fit to the log-transformed response. In this case, the identity link function would

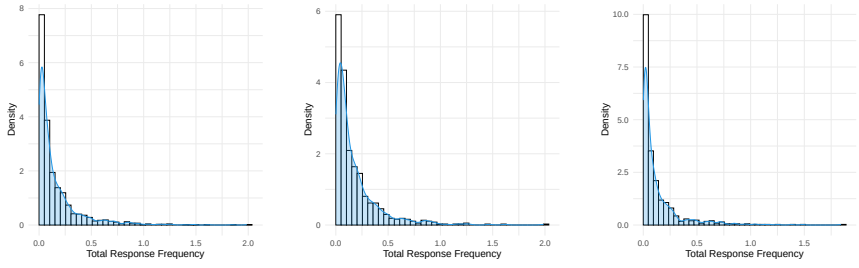
be used rather than the log link function.

For all three of these distributions, it was necessary to transform the response by adding a small non-zero term, 0.001. In the case of the Weibull and the Gamma distributions, this was necessary because a log link function was used and without the transformation, μ was usually equal to zero. In the case of the log-normal distribution, this was necessary so that the response was non-zero and could be logged.

Two more distributions of interest are the Sichel distribution and compound Poisson Tweedie distribution (which shall be referred to as a Tweedie distribution). Although these distributions are traditionally used to model count data and the data used here are frequencies, these distributions have the advantage of being well-suited to skew data. However, because the response is not only frequencies but very small frequencies close to zero, fitting a Sichel was often not possible because μ tended towards zero when it needed to be strictly greater than zero. Scaling the response by a large value would prevent μ from tending towards zero, but it would also result in a bimodal distribution (one of the peaks being at zero) which no longer has the same shape as a Sichel. The Tweedie distribution was however successful as its mean could equal zero.

One could also consider a zero-inflated distribution such as a zero-inflated Poisson (ZIP) or a zero-inflated negative binomial (ZINB). These distributions could be appropriate because they would account for the significant zero-inflation in the data. However, there are two features that make these distributions inappropriate. Firstly, the data used is frequencies, not count data which is what the ZIP and ZINB model. Secondly, all the non-zero frequencies are very small, meaning that the mean of the distribution after accounting for the zero-inflated component still tends towards zero. This is a problem because neither a ZIP or a ZINB can have a mean of zero. Again, scaling the response variable by a large number does not work.

Thus for each of the responses, we shall fit a log-normal, Gamma, Weibull and Tweedie distribution. These distributions were chosen for their ability to fit positively skewed data while still being interpretable and parsimonious, with only a maximum of three parameters necessary.



(a) ESAT-6 (b) Ag85B (c) ESAT-6 and Ag85B

Figure 5.6: Density plots of the (a) TRF after stimulation with Ag85B, (b) TRF after stimulation with ESAT-6 and (c) TRF after stimulation with either ESAT-6 or Ag85B.

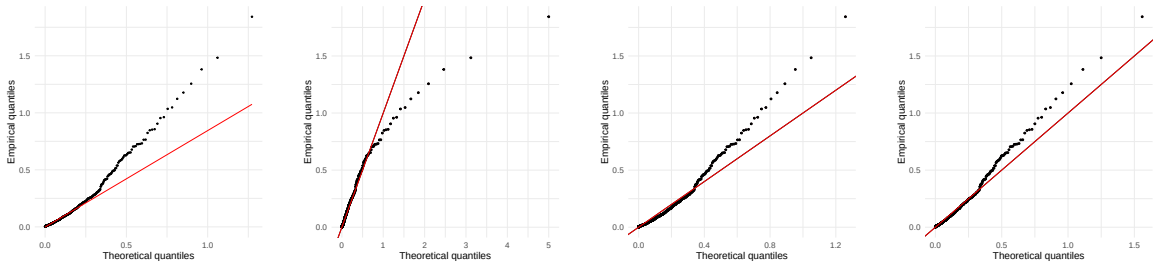
5.2.2 Single-level GLMEM: TRF after ESAT-6 stimulation

A log-normal, Gamma, Weibull and Tweedie distribution were all fit to response variable, the TRF after stimulation with ESAT-6. The log-normal, Gamma and Weibull distributions were fit using the `fitdistrplus` [76] package in `R`, while the Tweedie distribution was fit using the `cplm` [77] package. Table 5.1 summaries the resultant AIC and BIC values from the fitted log-normal, Gamma and Weibull distributions. The log-normal had the lowest AIC and BIC compared to the other distributions, suggesting that out of the three distributions the log-normal fit the data best.

Table 5.1: The AIC and BIC after fitting log-normal, Gamma and Weibull distributions to the TRF after ESAT-6 stimulation

Distribution	AIC	BIC
Log-normal	-1732.253	-1723.056
Gamma	-1608.404	-1599.207
Weibull	-1651.558	-1642.361

To further assess the fit, the empirical versus theoretical quantiles were plot, as seen in Figure 5.7. All four quantile-quantile plots (qq-plots) suggest that neither of the distributions fit the data well, however the Weibull distribution does seem to fit the data far better than the other three distributions. So, even though the Weibull did not have the lowest AIC and BIC, it was decided to assume that the conditional response follows a Weibull distribution.



(a) Tweedie (b) Log-normal (c) Gamma (d) Weibull

Figure 5.7: QQ-plots of the ESAT-6 TRF fit to (a) Tweedie, (b) log-normal, (c) Gamma and (b) Weibull distributions.

Following the model building steps laid out in Section 4.3.1, the optimal model chosen is as follows

$$\log(\mu_{ij}) = \beta_0 + \beta_1 I_j^{Peak} + \beta_2 I_j^{Mem} + \beta_3 I_i^{QFT+} + \beta_4 x_i^{Admin} + \beta_5 x_i^{Conc} + \beta_6 x_i^{Admin*QFT+} + \beta_7 x_{ij}^{Admin*Peak} + \beta_8 x_{ij}^{Admin*Mem} + b_{0i},$$

$$\log(\sigma_{ij}) = \alpha_0 + \alpha_1 x_i^{Admin},$$

$$\text{for} \quad i = 1, ..., I \quad \text{and} \quad j = 1, 2, 3,$$

where

$$b_{0i} \sim N\left(0, \sigma_0^2\right) \qquad \text{and} \qquad \epsilon_{ij} \sim N\left(0, \sigma_e^2\right).$$

Here, μ_{ij} is the average ESAT-6 TRF for participant i, the first level of grouping, at time point j. $\hat{\sigma}_0 = 0.820$ with 95% confidence interval $(0.740; 0.908)$ and $\hat{\sigma}_e = 1.081$ with 95% confidence interval $(1.015; 1.151)$. No participant-specific random effect was added on time point due to convergence issues. No three-way interactions were included in the model since they did not improve the model fit.

Model diagnostics can be found in Appendix D.1, where there was some evidence that the homogeneity assumption of the residuals and normality assumption of the random effects were violated.

Referring to the parameter estimates seen in Table 5.2, there is evidence that TRF decreases with increasing concentration and that variation in TRF increases with increasing number of administrations. There is also evidence of strong interactions between number of administrations and QFT status as well as between number of administrations and time point.

Table 5.2: Parameter estimates of the single-level GLMEM fit to the ESAT-6 TRF.

Variable	Parameter	Estimate	Std. Error	t-value	p-value
μ Coefficients:					
	β_0	-4.064	0.119	-34.117	$< 2 \times 10^{-16}$
Peak	β_1	0.992	0.130	7.640	1.09×10^{-13}
Mem.	β_2	-0.135	0.132	-1.027	0.305
QFT+	β_3	2.629	0.107	24.548	$< 2 \times 10^{-16}$
Admin.	β_4	0.209	0.055	3.819	0.0002
Conc.	β_5	-0.009	0.0009	-9.881	$< 2 \times 10^{-16}$
Admin*QFT+	β_6	-0.559	0.051	-10.854	$< 2 \times 10^{-16}$
Admin*Peak	β_7	0.001	0.062	0.022	0.983
Admin*Memory	β_8	0.315	0.063	5.006	7.66×10^{-7}
σ Coefficients:					
	α_0	0.447	0.081	5.547	4.69×10^{-8}
Admin.	α_1	0.138	0.041	3.325	0.001

Figure 5.8 was created in order to more easily interpret the interactions by plotting the expected value of logged mean $\log(\mu_{ij})$ for changing QFT status, time point and number of administrations while keeping all other variables constant using the parameter estimates in Table 5.2. QFT- participants seem to consistently have lower TRF compared to QFT+ participants. For QFT+ participants, TRF decreases with increasing vaccine administrations at all time points. However, for QFT- participants this relationship is reversed: TRF increases with increasing vaccine administrations, although this is only seen at the memory time point. Both peak and memory time points always have higher TRF compared to baseline, as expected. However, at baseline there is an unexpectedly large difference in TRF between the QFT+ participants receiving different vaccine administrations. This difference is surprising because at baseline, none of the participants have received the vaccine, and thus one would expect no difference between participants in the different vaccine arms, as seen for the QFT- participants. Perhaps this is just a result of the model trying to capture the interaction between number of administrations and QFT status at peak and memory time points. However, the addition of a three-way interaction with QFT, administrations and time point did not improve the model fit. This difference could also be due to slightly different populations enrolled in the separate trials that were merged together to form this dataset.

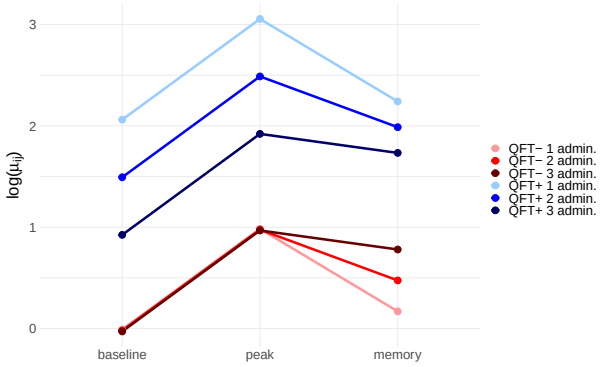


Figure 5.8: The expected value of logged mean $\log(\mu_{ij})$ for changing QFT status, time point and number of administrations while keeping all other variables constant.

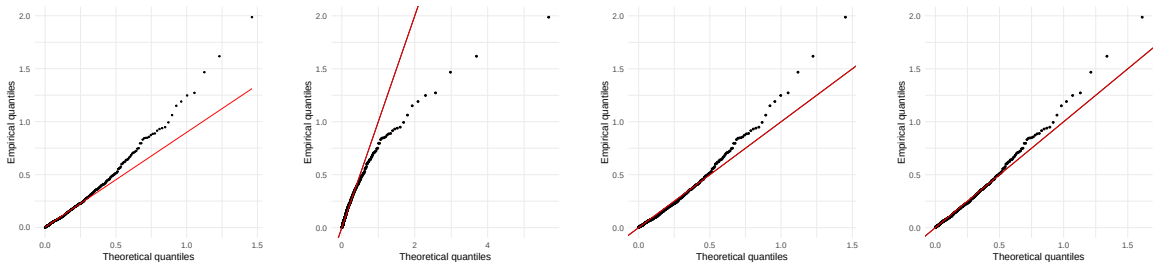
5.2.3 Single-level GLMEM: TRF after Ag85B stimulation

A log-normal, Gamma, Weibull and Tweedie distribution were all fit to response variable,TRF after stimulation with Ag85B. The log-normal, Gamma and Weibull distributions were fit using the `fitdistrplus` [76] package in `R`, while the Tweedie distribution was fit using the `cplm` [77] package. Table 5.3 summaries the resultant AIC and BIC values from the fitted log-normal, Gamma and Weibull distributions. The log-normal and Weibull had almost the same AIC and BIC values, both lower than that of the Gamma distribution. This suggests that both the log-normal and Weibull distributions are potentially good fits for the response variable.

Table 5.3: The AIC and BIC after fitting log-normal, Gamma and Weibull distributions to the TRF after Ag85B stimulation

Distribution	AIC	BIC
Log-normal	-1075.079	-1065.887
Gamma	-1064.242	-1055.051
Weibull	-1074.812	-1065.62

To further assess the fit, the empirical versus theoretical quantiles were plot, as seen in Figure 5.9. All four qq-plots suggest that neither of the distributions fit the data well, however the Weibull distribution does seem to fit the data far better than the other three distributions. Since the Weibull also had the lowest AIC and BIC, it was decided to assume that the conditional response follows a Weibull distribution.



(a) Tweedie (b) Log-normal (c) Gamma (d) Weibull

Figure 5.9: QQ-plots of the Ag85B TRF fit to (a) Tweedie, (b) log-normal, (c) Gamma and (b) Weibull distributions.

Following the model building steps laid out in Section 4.3.1, the optimal model chosen is as follows

$$\begin{aligned}\log(\mu_{ij}) = \beta_0 &+ \beta_1 I_j^{Peak} + \beta_2 I_j^{Mem} + \beta_3 I_i^{QFT+} + \beta_4 x_i^{Admin} + \beta_5 x_i^{Conc} \\ &+ \beta_6 x_{ij}^{Admin*Peak} + \beta_7 x_{ij}^{Admin*Mem} + \beta_8 I_{ij}^{Peak*QFT+} + \beta_9 I_{ij}^{Mem*QFT+} \\ &+ b_{0i},\end{aligned}$$

$$\log(\sigma_{ij}) = \alpha_0,$$

$$\text{for} \quad i = 1, ..., I \quad \text{and} \quad j = 1, 2, 3,$$

where

$$b_{0i} \sim N\left(0, \sigma_0^2\right) \qquad \text{and} \qquad \epsilon_{ij} \sim N\left(0, \sigma_e^2\right).$$

Here, μ_{ij} is the average Ag85B TRF for participant i, the first level of grouping, at time point j. $\hat{\sigma}_0 = 0.836$ with 95% confidence interval $(0.759; 0.922)$ and $\hat{\sigma}_e = 1.025$ with 95% confidence

interval $(0.964; 1.090)$. No participant-specific random effect was added on time point due to convergence issues. No three-way interactions were included in the model since they did not improve the model fit.

Model diagnostics can be found in Appendix D.2, where there was some evidence that the homogeneity assumption of the residuals and normality assumption of the residuals and random effects are violated.

Referring to the parameter estimates seen in Table 5.4, there is again evidence that TRF increases with decreasing concentration. There is also evidence of strong interactions between number of administrations and time point and QFT status and time point. Referring to Figure 5.4, there is evidence that QFT+ participants consistently have a higher TRF compared to QFT- participants. At peak and memory time points, TRF increases with increasing vaccine administrations, for both QFT+ and QFT- participants. The increase from baseline to peak appears to be much steeper for QFT- participants, suggesting a greater increase in TRF from baseline to peak compared to QFT+ participants. Difference in TRF between different vaccine administrations is more pronounced at memory time point than peak time point, suggesting that selecting the appropriate number of administrations is more important for producing long-term immune responses rather than short-term.

Table 5.4: Parameter estimates of the single-level GLMEM fit to the Ag85B TRF.

Variable	Parameter	Estimate	Std. Error	t-value	p-value
μ Coefficients:					
	β_0	-3.085	0.083	-37.379	$< 2 \times 10^{-16}$
Peak	β_1	1.355	0.114	11.884	$< 2 \times 10^{-16}$
Mem.	β_2	0.468	0.111	4.233	2.75×10^{-5}
QFT+	β_3	0.947	0.055	17.359	$< 2 \times 10^{-16}$
Admin.	β_4	-0.106	0.038	-2.781	0.006
Conc.	β_5	-0.003	0.001	-3.246	0.001
Admin*Peak	β_6	0.171	0.054	3.168	0.002
Admin*Memory	β_7	0.295	0.054	5.459	7.59×10^{-8}
Peak*QFT+	β_8	-0.546	0.077	-7.059	5.7×10^{-12}
Mem*QFT+	β_9	-0.460	0.077	-5.971	4.5×10^{-9}

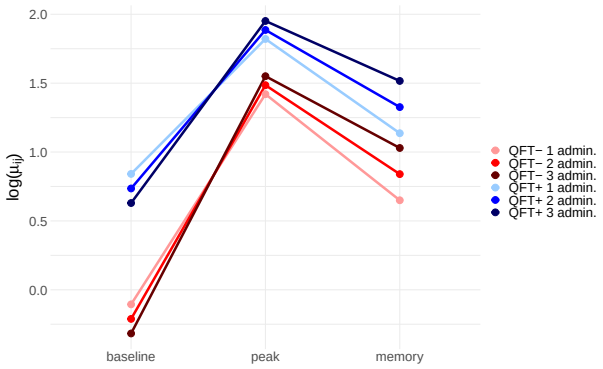


Figure 5.10: The expected value of logged mean $\log(\mu_{ij})$ for changing QFT status, time point and number of administrations while keeping all other variables constant.

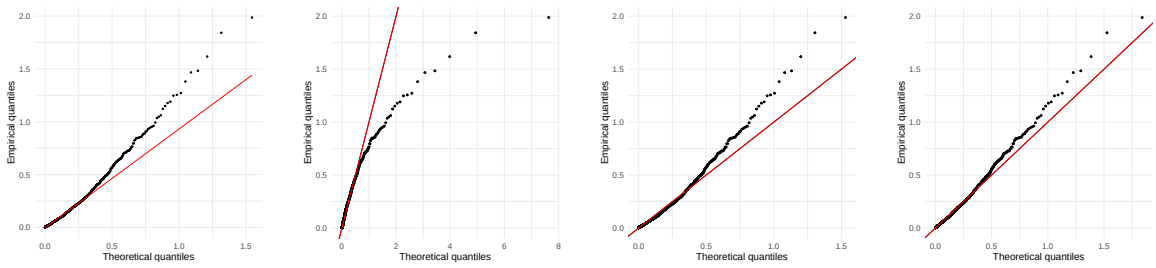
5.2.4 Two-level GLMEM: TRF after stimulation with either ESAT-6 or Ag85B

A log-normal, Gamma, Weibull and Tweedie distribution were all fit to response variable, the TRF after stimulation with either ESAT-6 or Ag85B. The log-normal, Gamma and Weibull distributions were fit using the `fitdistrplus` [76] package in `R`, while the Tweedie distribution was fit using the `cplm` [77] package. Table 5.5 summaries the resultant AIC and BIC values from the fitted log-normal, Gamma and Weibull distributions. The log-normal had the lowest AIC and BIC compared to the other distributions, suggesting that out of the three distributions the log-normal fit the data best.

Table 5.5: The AIC and BIC after fitting log-normal, Gamma and Weibull distributions to the TRF after stimulation with either ESAT-6 or Ag85B

Distribution	**AIC**	**BIC**
Log-normal	-2748.242	-2737.661
Gamma	-2635.147	-2624.567
Weibull	-2682.647	-2672.066

To further assess the fit, the empirical versus theoretical quantiles were plot, as seen in Figure 5.11. All four qq-plots suggest that neither of the distributions fit the data well, however the Weibull distribution does seem to fit the data far better than the other three distributions. So, even though the Weibull did not have the lowest AIC and BIC, it was decided to assume that the conditional response follows a Weibull distribution.



(a) Tweedie (b) Log-normal (c) Gamma (d) Weibull

Figure 5.11: QQ-plots of the TRF after stimulation with either Esat-6 or Ag85B fit to (a) Tweedie, (b) log-normal, (c) Gamma and (b) Weibull distributions.

Following the model building steps laid out in Section 4.3.1, the optimal model chosen is as follows

$$\begin{aligned}\log(\mu_{ijk}) = \beta_0 &+ \beta_1 I_k^{Peak} + \beta_2 I_k^{Mem} + \beta_3 I_i^{QFT+} + \beta_4 x_i^{Admin} + \beta_5 x_i^{Conc} + \beta_6 I_j^{ESAT6} \\ &+ \beta_7 I_{ij}^{ESAT6*QFT+} + \beta_8 I_{ik}^{Peak*QFT+} + \beta_9 I_{ik}^{Mem*QFT+} + \beta_{10} I_{ij}^{Conc*ESAT6} \\ &+ \beta_{11} x_{ik}^{Peak*Admin} + \beta_{12} x_{ik}^{Mem*Admin} + \beta_{13} I_{jk}^{Peak*ESAT6} + \beta_{14} I_{jk}^{Mem*ESAT6} \\ &+ b_{0i},\end{aligned}$$

$$\log(\sigma_{ijk}) = \alpha_0,$$

$$\text{for} \quad i = 1, ..., I, \quad j = 1, 2, \quad \text{and} \quad k = 1, 2, 3,$$

where

$$b_{0i} \sim N\left(0, \sigma_0^2\right) \qquad \text{and} \qquad \epsilon_{ijk} \sim N\left(0, \sigma_e^2\right).$$

Here, μ_{ijk} is the average TRF for participant i, the first level of grouping, after stimulation with stimulus j, the second level of grouping, at time point k. $\hat{\sigma}_0 = 0.713$ with 95% confidence interval $(0.644; 0.790)$ and $\hat{\sigma}_e = 1.083$ with 95% confidence interval $(1.041; 1.127)$. No participant-specific random effect was added on time point due to convergence issues. No three-way interactions were included in the model since they did not improve the model fit.

Model diagnostics can be found in Appendix D.3, where there was some evidence that the homogeneity assumption of the residuals and normality assumption of the random effects are violated.

Referring to the parameter estimates seen in Table 5.6, there is evidence that TRF decreases with increasing concentration for all participants, even more so if under ESAT-6 stimulation. The difference in TRF between different vaccine administrations is only strong at the memory time point, where increasing number of administrations is associated with increasing the TRF for all participants. For QFT- participants, TRF is lower under ESAT-6 stimulation compared

to Ag85B stimulation, even more so at post-vaccination time points and even more so for higher concentrations. However, for QFT+ participants the TRF is always higher under ESAT-6 stimulation compared to Ag85B stimulation, although less so at peak and memory time points and at higher concentrations. QFT+ participants are also associated with higher TRF compared to QFT- participants, regardless of stimulation, although this difference becomes less pronounced at post-vaccination time points. Due to the large number of interactions in this model, it was decided not to include a plot of the interactions as it would be difficult to interpret.

Table 5.6: Parameter estimates of the single-level GLMEM fit to the TRF after stimulation with ESAT-6 or Ag85B.

Variable	Parameter	Estimate	Std. Error	t-value	p-value
μ Coefficients:					
	β_0	-3.003	0.096	-31.273	$< 2 \times 10^{-16}$
Peak	β_1	1.449	0.128	11.316	$< 2 \times 10^{-16}$
Mem.	β_2	0.419	0.123	3.397	0.001
QFT+	β_3	0.917	0.066	13.808	$< 2 \times 10^{-16}$
Admin.	β_4	-0.068	0.040	-1.691	0.091
Conc.	β_5	-0.003	0.001	-2.483	0.013
ESAT6	β_6	-0.618	0.083	-7.414	2.27×10^{-13}
ESAT6 *QFT+	β_7	1.079	0.066	16.264	$< 2 \times 10^{-16}$
Peak*QFT+	β_8	-0.484	0.081	-5.941	3.67×10^{-9}
Mem*QFT+	β_9	-0.415	0.081	-5.105	3.83×10^{-7}
Conc*ESAT6	β_{10}	-0.008	0.002	-4.627	4.11×10^{-6}
Admin*Peak	β_{11}	0.047	0.057	0.837	0.403
Admin*Mem	β_{12}	0.279	0.057	4.873	1.24×10^{-6}
ESAT6 *Peak	β_{13}	-0.305	0.082	-3.735	0.0002
ESAT6 *Mem	β_{14}	-0.278	0.081	-3.423	0.001

5.2.5 Comment on the GLMEM's

Deviation from normality and constant variance assumptions for some models may impact on accuracy of the statistical inference. Hence, it is preferred to look at what relationships are consistently seen across all three models. As such, after assessing the results, the following inference was common amongst all three models. For all models, there was no evidence that the demographic variables age, sex and ethnicity had an effect on TRF. QFT+ participants consistently had a higher TRF compared to QFT- participants, no matter the number of administrations, concentration or stimulus. There was evidence that TRF decreases with increasing concentration, no matter the number of administrations, QFT status or stimulus. There was evidence that there exists a strong interaction between number of administrations and time point, where at the memory time point TRF increases with increasing numbers of administrations. As expected, both peak and memory time points always have a higher TRF compared to baseline. Note that a comparison between ESAT-6 and Ag85B can only be made using the two-level

GLMEM, and since there is only one of these models, it is preferred not to comment on the difference between ESAT-6 and Ag85B.

Thus, there is evidence that the lowest vaccine concentration, 5 µg, should be considered as the preferred vaccine concentration. With regards to vaccine administrations, both the one-level Ag85B GLMEM and two-level GLMEM suggested that TRF increases in order to elicit long-term immune responses, 3 administrations of the vaccine should be considered. However, while the one-level ESAT-6 GLMEM suggested that this was true for QFT- participants, it suggested that for QFT+ participants one administration produced the highest TRF at both post-vaccination time points. As such, it is inconclusive as to which number of vaccine administrations induces the highest vaccine take. With regards to the hypothesis of interest, no interaction between vaccine concentration and QFT status was observed in either model, suggesting that there is no difference in the magnitude of cytokine positive CD4 T cells between the different vaccine concentrations, for both M.tb-sensitised and -unsensitised individuals.

In terms of the interpretability of the models, the two-level GLMEM required far more interaction terms in order to account for the sharing of covariate coefficients between Ag85B and ESAT-6. Specifically, including interactions between stimulus and QFT, concentration and time point suggest that the relationship between TRF and these covariates varies between the two stimuli. The advantage of these higher-level models is that they allow for inference related to specific hypotheses, such as comparison between ESAT-6 and Ag85B in the case of the two-level GLMEM. A notable disadvantage is the increase in the number of interaction terms and model complexity (although there is also an increase in the number of observations), which can result in a model that is difficult to interpret.

5.3 Discussion

Both the standard approach as well as the GLMEM's suggested that 5 µg should be considered as the preferred vaccine concentration. However, while the standard procedure suggested that two administrations of the vaccine was sufficient, the GLMEM's produced conflicting evidence as to which number of administrations was optimal. Hence, while there is strong evidence that 5 µg of the vaccine should be considered, there is conflicting evidence as to the number of administrations that should be considered. With regards to the hypothesis of interest, both methods found no evidence that there exists a difference in the magnitude of cytokine positive CD4 T cells between the different vaccine concentrations, for both *M.tb*-sensitised and -unsensitised individuals.

A clear disadvantage of the non-parametric approach shown in Section 5.1.1 is the multiple testing which prevents one from making any strong inference from the results. However, the GLMEM's do not completely avoid the same problem. Since multiple models have been fit and used to make inference, multiple testing is still an issue and the probability of a type I error is unknown and higher than 5%. As such, the results produced by both the models and standard approach should be interpreted with caution and be used for exploratory analysis rather than

confirmatory. The purpose of the models is to illustrate how they can be used in such a setting, and in future, fitting fewer models with more of an emphasis on the model building procedure can provide stronger inference as well as including power calculations when reporting results. With respects to significance testing in the standard approach, Bender *et al.* (2001) [62] emphasise the importance of formulating a clear statistical analysis plan and pre-specifying the hypotheses and their priorities. Multiplicity becomes a significant issue when there are many different hypotheses of interest and equal priority and/or in the case of exploratory studies. Trying to investigate the effects of many different variables on the response, as done in this chapter, makes it very difficult to guard against multiple testing. The plots of the response rate created by filtering out responders using MIMOSA also suffer from multiple testing and do rely on determining statistical significance in order to categorise participant visits as responders and non-responders.

The main advantage that GLMEM's have over significance testing is that they provide effect sizes in the form of variable coefficients, where as significance tests focus more on hypothesis testing and p-values. Providing effect sizes alongside the p-values gives one greater insight into how much of an effect a given variable has on the response, which is more informative than just knowing that an effect exists, which is what p-values provide.

In conclusion, the standard approach and the mixed effect modelling approach both produced similar inference, with the main disagreement being the number of vaccine administrations that produce the greatest immune response. However, while both approaches could be used to determine which variables may have an effect on immune response, only the GLMEM's were able to provide estimates of the effect size. Both approaches suffered from issues of multiple testing, hence why all results should be interpreted as exploratory and used to guide future hypothesis formulations and experiments.

Chapter 6

Investigating the effects of vaccine regimen, *M.tb* sensitisation and antigen specificity on CD4 T cell *functional quality*

In this chapter, the second aim of this project, that is the effects of vaccine regimen, *M.tb* sensitisation and antigen specificity on CD4 T cell *functional quality*, shall be investigated. Along with exploring this aim, this chapter shall also investigate the hypothesis that T cell functional quality is influenced by vaccine concentration and *M.tb* sensitisation and that high frequencies of T cells expressing IL-2 and TNF together and T cells expressing IL-2, TNF and IFN-γ together are associated with lower concentrations of the vaccine in *M.tb*-sensitised individuals compared to unsensitised individuals. Another hypothesis of interest is that the functional profiles of T cells recognising Ag85B are distinct from those of T cells recognising ESAT-6. The last hypothesis to test was that after vaccination, the functional profiles of T cells recognising Ag85B or ESAT-6 differ between *M.tb*-sensitised and *M.tb*-unsensitised individuals.

As discussed in Section 1.2, T cell functional quality refers to the pattern of cell functional marker (i.e. cytokine) expression. Thus, the frequency of cells expressing different cytokine combinations is the response interest. Since there are four different cytokines that were measured, IL-2, TNF, IL-17 and IFN-γ, there are a total of $2^4 = 16$ different possible cytokine combinations, and thus 16 different response variables.

Due to the high-dimensionality of the response variable, COMPASS will be used both in the standard approach and the mixed effect modelling approach for variable reduction. During the standard approach, the FS and PFS produced by COMPASS shall be used as the response variable of interest, acting as two summary scores of the 16 different response variables. For the mixed effect modelling approach, the results from COMPASS shall be used to filter out cytokine combinations which are not differentially expressed.

This chapter will begin by implementing the standard approach to investigate these objectives, first analysing the FS and PFS produced by COMPASS using data visualisation and hypothesis testing in Section 6.1.1, followed by a more in-depth analysis of the frequency of the different cytokine combinations using data visualisation and hypothesis testing in Section 6.1.2. For the mixed effect modelling approach, COMPASS shall be used to first reduce the number of cytokine combinations in the response. Then, two-level GLMEM's shall be fit to the reduced selection of cytokine responses in Section 6.2 followed by LVM's in Section 6.3.

6.1 Standard approach

This section shall analyse the FS and PFS computed using COMPASS as well as the CD4 T cell frequency of the 16 different cytokine combinations. FS can be interpreted as the proportion of antigen-specific cell subsets expressed for a given participant visit among all possible subsets. The PFS can be interpreted as the FS weighted by the cell subset's degree of functionality, where higher degrees of functionality are weighted more. Thus, the PFS gives one an idea of the quality of the T cell response while FS gives one an idea of the proportion of cytokine combinations which were differentially expressed. Both FS and PFS shall be used as summary statistics of the CD4 T cell frequency of the 16 different cytokine combinations. Using FS and PFS as the response instead of the individual cytokine combinations reduces the dimensionality of the response variable, making it easier to plot and analyse.

While the FS and PFS provides a good summary measure of the functional quality of the response, it does not provide insight into the specific functional profiles making up the immune response, however the CD4 T cell frequency of the 16 different cytokine combinations does. Thus, some of the hypotheses of interest shall be investigated using both the FS and PFS response as well as the CD4 T cell frequency of the cytokine combinations response in order to make inference about not only the level of functionality but also the specific functional profiles making up the response.

Section 6.1.1 will use the FS and PFS produced by COMPASS as the response of interest. It will begin with selecting the number of vaccine administrations that induces the best quality immune responses. It will then investigate the relationship between vaccine concentration and QFT status, followed by selection of the vaccine concentration that induces the best quality immune responses. Then, the relationship between QFT status and stimulus shall be investigated, followed by an analysis of the posterior probabilities produced by COMPASS used to infer which cytokine combinations are antigen-specific.

Section 6.1.2 will then use the CD4 T cell frequency of the cytokine combinations as the response variable in order to investigate some of the hypotheses of interest in terms of specific functional profiles rather than just the overall functional quality. Firstly, the hypothesis that high frequencies of T cells expressing IL-2 and TNF together and T cells expressing IL-2, TNF and IFN-γ together are associated with lower concentrations of the vaccine in *M.tb*-sensitised individuals compared to unsensitised individuals shall be investigated. Then, the hypothesis that after vaccination, the functional profiles of T cells recognising Ag85B or ESAT-6 differ between QFT+ and QFT- individuals shall be investigated.

6.1.1 COMPASS

The below plots in Figure 6.1, 6.2, 6.3 and 6.4 were created to investigate which number of administrations of the vaccine induces the highest FS and PFS (i.e. the highest quality immune responses) with the goal of only looking at observations from participants who received

this number of administrations in future plots. Referring to Figure 6.1 which plots the FS using observations stimulated by ESAT-6, relatively large differences between administrations are seen when 5 µg or 50 µg of the vaccine was administered. For QFT+ participants who received 5 µg, the participants who received two vaccine administrations had much higher FS at peak compared to those who received three administrations. While for QFT- participants who received 50 µg, the participants who received two vaccine administrations had much higher FS at peak and memory compared to those who received one administration. For QFT+ participants who received 50 µg, the participants who received three administrations have much lower FS at baseline compared to participants who received one or two administrations. Since this was at baseline, and hence before either of the treatment arms received their vaccinations, it suggests that there was a difference between the participants chosen for each treatment arm, rather than a difference caused by the number of vaccine administrations. Thus, overall there is evidence that two vaccine administrations may have induced a higher quality immune response.

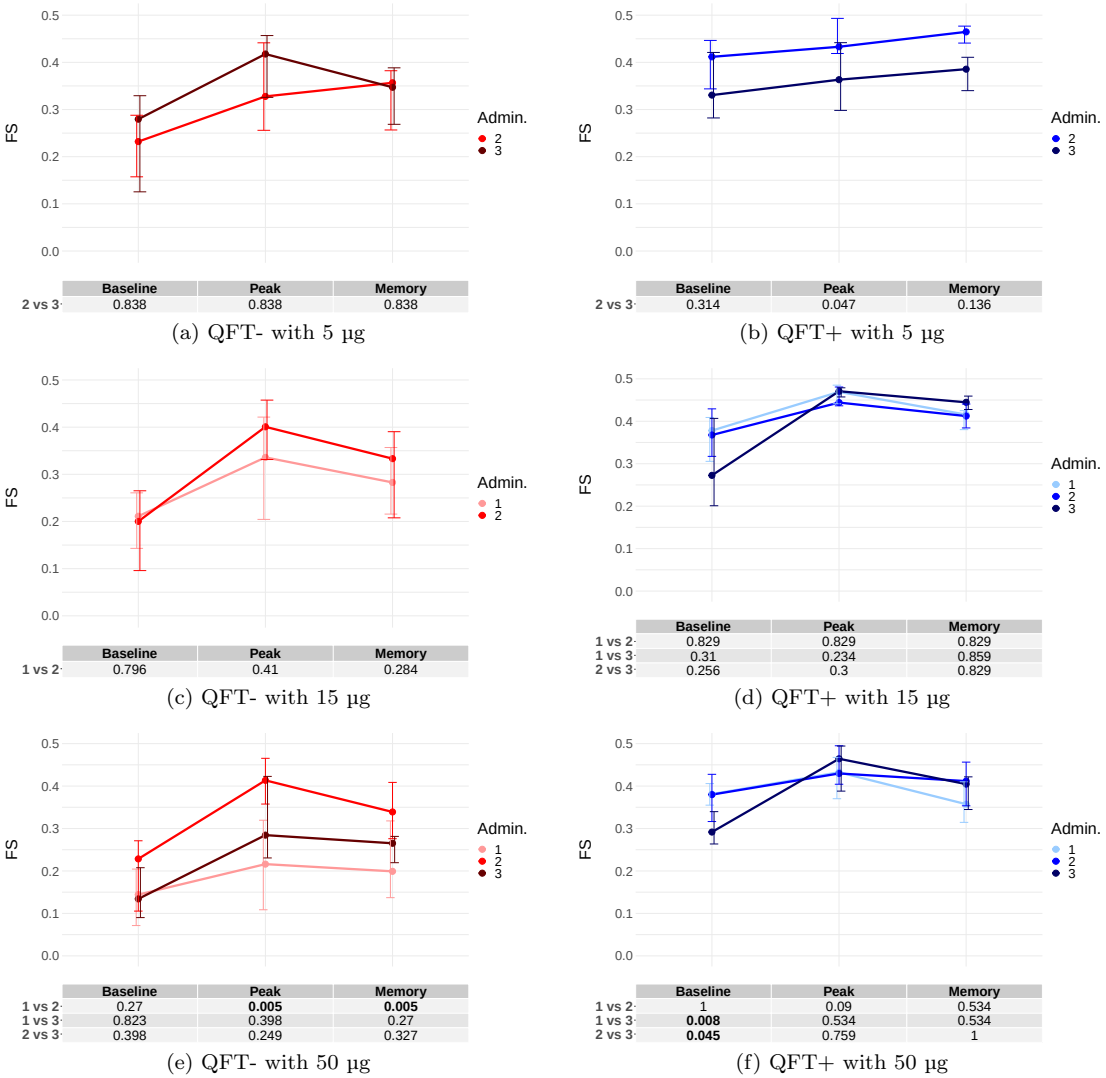
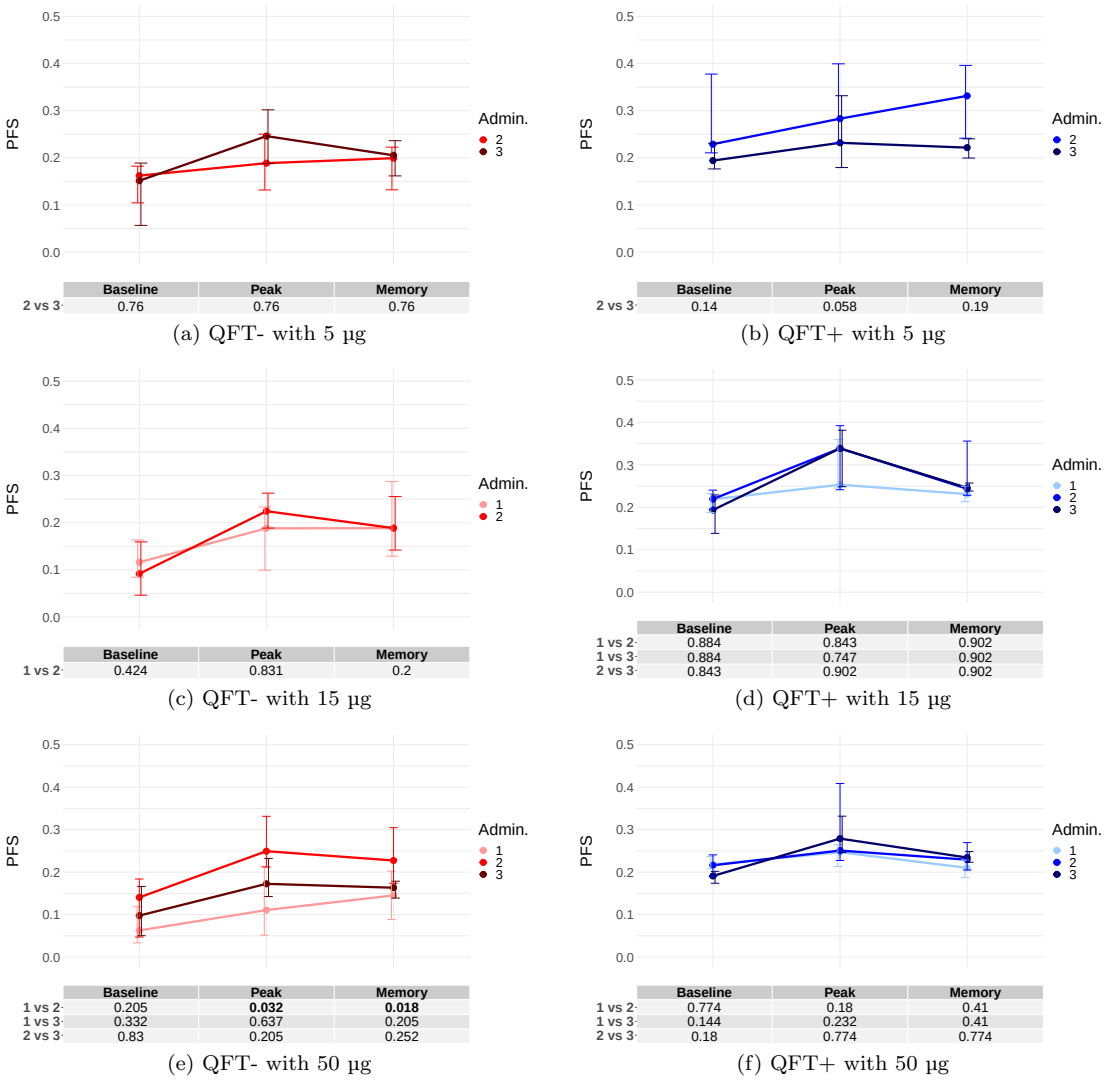


Figure 6.1: Longitudinal profiles of the median ESAT-6 FS of CD4 T cells for different vaccine administrations (Admin.). The error bars denote the interquartile range of the FS. The medians have been plotted for three different time points: baseline, peak and memory. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing the different numbers of administrations for the same time point are shown in the table.

Referring to Figure 6.2, the only potentially noticeable difference in PFS determined after ESAT-6 stimulation is seen in QFT- participants who were administered 50 µg of the vaccine. For these participants, those that received two administrations generally had a higher PFS at peak and memory compared to those who received one administration. Thus, there is evidence that two vaccine administrations may have induced a higher quality immune response.



(a) QFT- with 5 µg

(b) QFT+ with 5 µg

(c) QFT- with 15 µg

(d) QFT+ with 15 µg

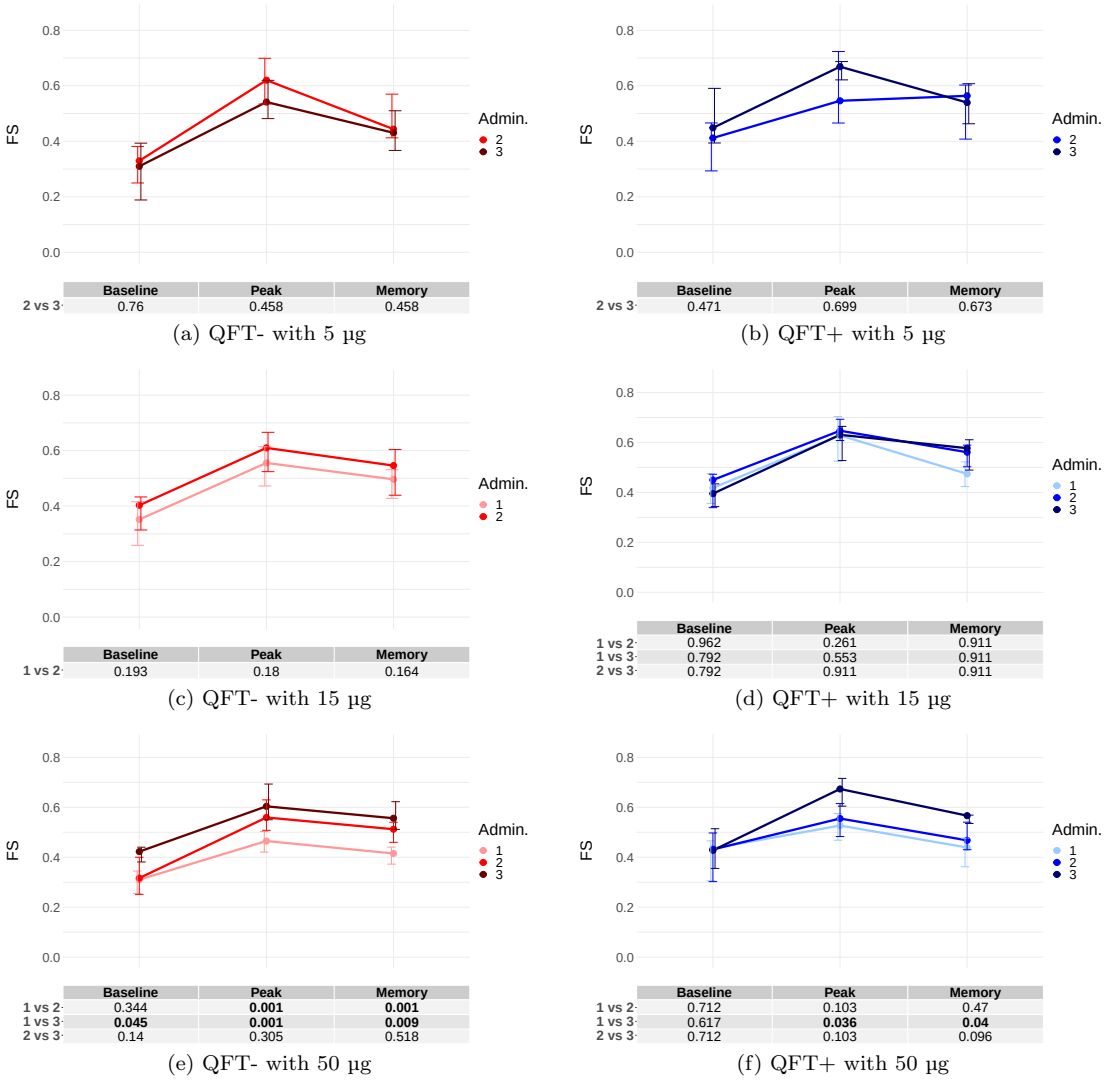
(e) QFT- with 50 µg

(f) QFT+ with 50 µg

Figure 6.2: Longitudinal profiles of the median ESAT-6 PFS of CD4 T cells for different vaccine administrations (Admin.). The error bars denote the interquartile range of the PFS. The medians have been plotted for three different time points: baseline, peak and memory. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing the different numbers of administrations for the same time point are shown in the table.

Looking at the Ag85B FS shown in Figure 6.3, the only potentially noticeable difference between administrations is seen when a concentration of 50 µg was used. For QFT- participants who received 50 µg, the participants who received one vaccine administration had lower FS at both post-vaccination time points compared to those who received two or three administrations. Since there wass seemingly no difference between two and three administrations, the least resource-intensive option, two administrations, is preferred. Note that a difference also seems to be

present at baseline between the cohort which received one administrations and the cohort which received two administrations, suggesting that there was a difference between these cohorts prior to vaccine administration. This could be a potential confounder and influence the difference seen later on between the two cohorts at peak and memory. This could mean that the difference between the two cohorts was exaggerated, however even if this was the case, the conclusion of choosing two administrations would not change because of the difference seen between one and two administrations. For QFT+ participants who received 50 µg, the participants who received three administrations generally had higher FS at both post-vaccination time points compared to those who received one administration. This indicates that three administrations would be preferred over one. However, there is no difference between two and three administrations, so again the least resource-intensive option, two administrations, is preferred.



(a) QFT- with 5 µg

(b) QFT+ with 5 µg

(c) QFT- with 15 µg

(d) QFT+ with 15 µg

(e) QFT- with 50 µg

(f) QFT+ with 50 µg

Figure 6.3: Longitudinal profiles of the median Ag85B FS of CD4 T cells for different vaccine administrations (Admin.). The error bars denote the interquartile range of the FS. The medians have been plotted for three different time points: baseline, peak and memory. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing the different numbers of administrations for the same time point are shown in the table.

Looking at the Ag85B PFS shown in Figure 6.4, again the only potentially noticeable difference between administrations is seen when a concentration of 50 µg was used. For QFT- participants who received 50 µg of the vaccine, the inference made using Figure 6.3 applies here as well when looking at PFS. As such, it shall not be repeated here. For QFT+ participants who received 50 µg, participants who received three vaccine administrations had higher PFS at post vaccination time points compared to those who received one or two administrations. This indicates that three administrations would be preferred. However, there was also a potential difference at baseline between the cohort which received two administrations and the cohort which received three administrations, suggesting that there was a difference between these cohorts prior to vaccine administration, which could be a potential confounder and influence the difference seen later on between the two cohorts. This could mean that there was actually no difference due to number of administrations between the two cohorts, in which case the conclusion would change and two administrations would be preferred. As such, it is unclear whether two or three administrations induce the highest quality immune response.

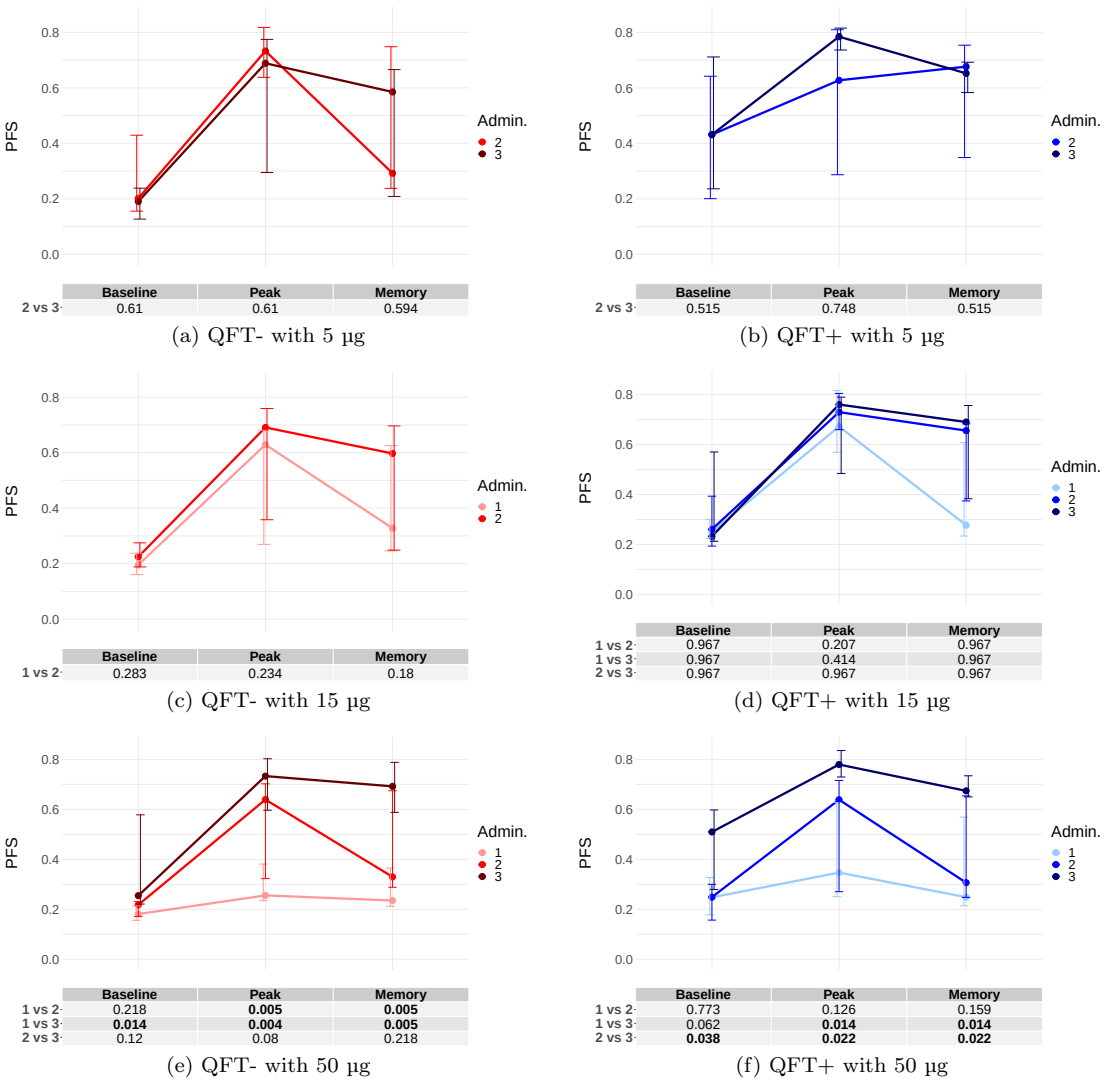


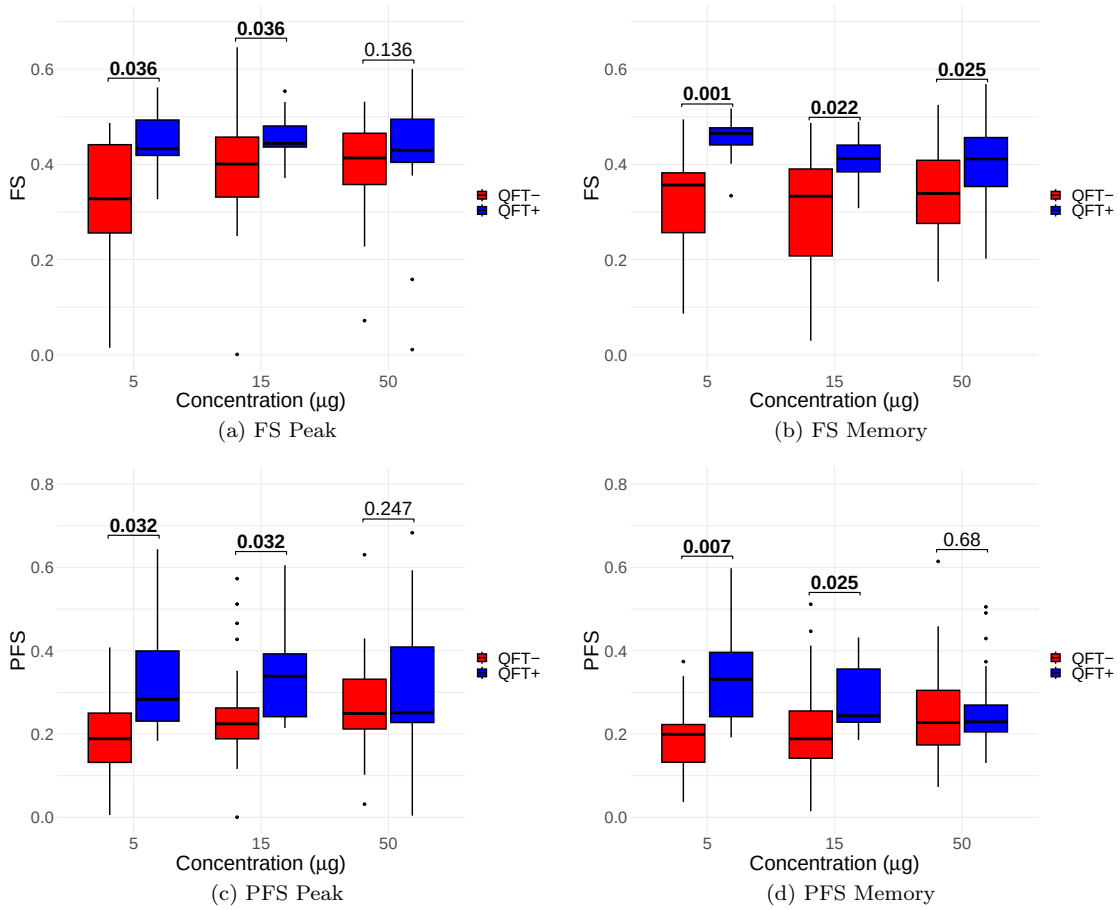
Figure 6.4: Longitudinal profiles of the median Ag85B PFS of CD4 T cells for different vaccine administrations (Admin.). The error bars denote the interquartile range of the PFS. The medians have been plotted for three different time points: baseline, peak and memory. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing the different numbers of administrations for the same time point are shown in the table.

Therefore, based on Figures 6.1, 6.2, 6.3 and 6.4, there is evidence that two administrations produced the greatest FS and PFS, and thus that two administrations may have induced the highest quality immune response. Thus, only observations from participants who received two vaccine administrations shall be used in future plots.

Before determining the corresponding concentration that shall be used to further filter observations, the hypothesis that higher quality immune responses are associated with lower concentrations of the vaccine in *M.tb*-sensitised individuals compared to unsensitised individuals shall be investigated. Thus, the below plots in Figure 6.5 and 6.6 were created to investigate the difference in functionality between QFT- and QFT+ participants for different vaccine concentrations.

Referring to Figure 6.5, the ESAT-6 FS was higher in QFT+ participants compared to QFT-

participants across all time points and vaccine concentrations, except for at the peak time point for participants who received 50 µg of the vaccine, where there was no evidence of a difference between QFT- and QFT+ participants. Similarly, the ESAT-6 PFS was seemingly higher in QFT+ participants compared to QFT- participants across all time points and vaccine concentrations, except for at both post-vaccination time points for participants who received 50 µg of the vaccine, where there was no difference between QFT- and QFT+ participants. Thus, under ESAT-6 stimulation, there is evidence that for participants who received lower vaccine concentrations, QFT+ participants had higher quality immune responses compared to QFT- participants. For 50 µg of the vaccine, there is evidence that there was no difference between QFT+ and QFT- participants.

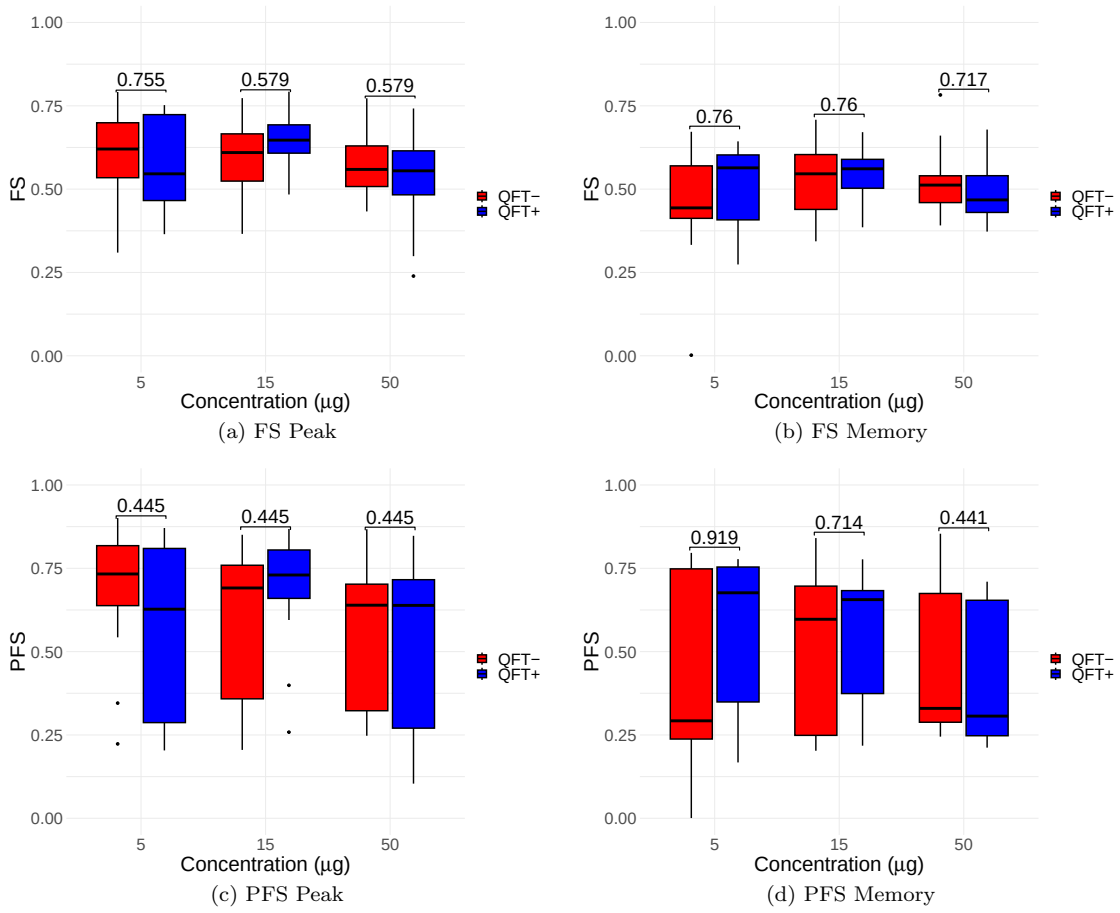


(a) FS Peak (b) FS Memory

(c) PFS Peak (d) PFS Memory

Figure 6.5: Box-plots of ESAT-6 FS and PFS for two administrations of the vaccine for different vaccine concentrations and QFT statuses. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing different QFT statuses for the same vaccine regimen are shown.

To further investigate this hypothesis, the Ag85B FS and PFS were plot, as seen in Figure 6.6. There is no evidence that there was a difference in FS or PFS between QFT+ and QFT- participants across both time points and vaccine concentrations. Thus, there is evidence that there was a difference in functionality between QFT- and QFT+ participants for different vaccine concentrations under ESAT-6 stimulation but not under Ag85B stimulation.

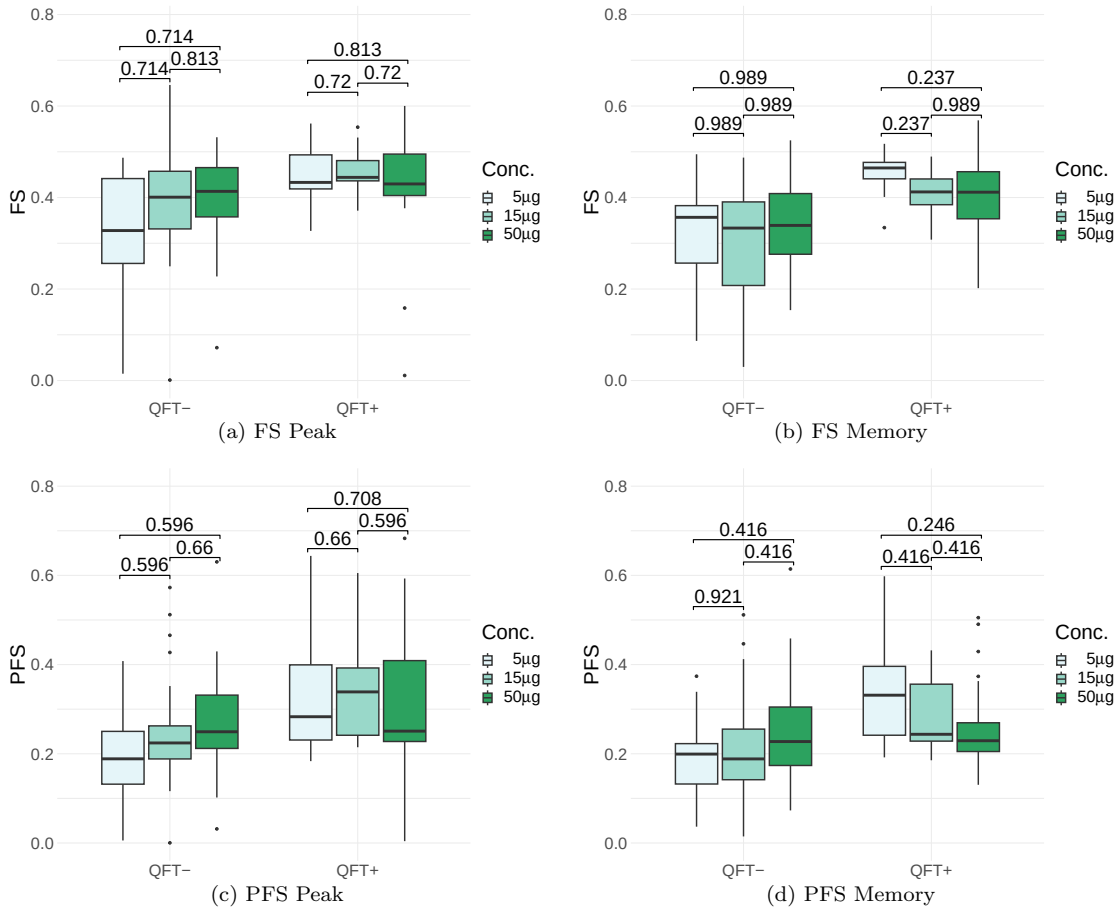


(a) FS Peak (b) FS Memory

(c) PFS Peak (d) PFS Memory

Figure 6.6: Box-plots of Ag85B FS and PFS for two administrations of the vaccine for different vaccine concentrations and QFT statuses. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing different QFT statuses for the same vaccine regimen are shown.

The below plots in Figure 6.7 and 6.8 were created to investigate which vaccine concentration induced the highest FS and PFS when considering only two administrations of the vaccine. Referring to Figure 6.7, after ESAT-6 stimulation there is no evidence that a difference in FS or PFS existed between the three concentrations for both QFT- and QFT+ participants across all post-vaccination time points.



(a) FS Peak (b) FS Memory (c) PFS Peak (d) PFS Memory

Figure 6.7: Box-plots of ESAT-6 FS and PFS for two administrations of the vaccine for different vaccine concentrations and QFT statuses. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing different vaccine concentrations for the same QFT status are shown.

Referring to Figure 6.8, after Ag85B stimulation there is again no evidence that a difference in FS or PFS existed between the three vaccine concentrations for both QFT- and QFT+ participants across all post-vaccination time points. Since there is no evidence that a difference between the vaccine concentrations existed after either stimulation, the lowest vaccine concentration is preferred as it is the least resource-intensive. Thus, there is evidence that a concentration of 5 µg of the vaccine is preferable when administering two doses of the vaccine. Hence, only observations from participants who received two administrations of 5 µg of the vaccine shall be used in future plots. Note that this is the same regimen that was chosen when analysing the magnitude of the T cell response, suggesting that this regimen was not only optimal in terms of functional quality but also in terms of magnitude of the response.

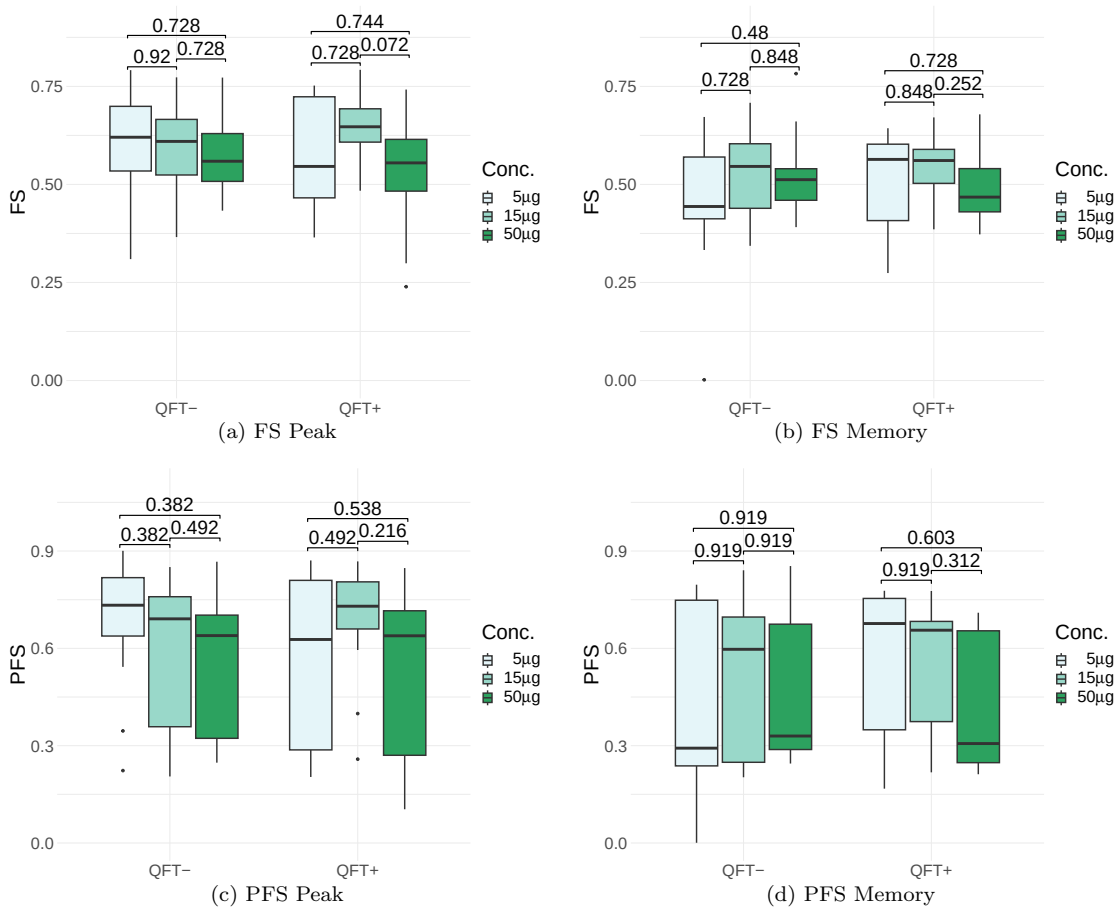


Figure 6.8: Box-plots of Ag85B FS and PFS for two administrations of the vaccine for different vaccine concentrations and QFT statuses. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing different vaccine concentrations for the same QFT status are shown.

The below plots in Figure 6.9 were created to investigate the hypothesis that after vaccination, the functional profiles of T cells recognising Ag85B or ESAT-6 differed between QFT+ and QFT- individuals. When comparing QFT status, there is strong evidence that QFT+ participants had higher FS and PFS compared to QFT- participants across both post-vaccination time points after ESAT-6 stimulation but not after Ag85B stimulation. When comparing the two stimuli, the samples after Ag85B stimulation seemed to have had higher FS and PFS, regardless of QFT status and time point. Thus, there is strong evidence that after vaccination QFT+ participants had higher quality immune responses to ESAT-6 stimulation compared to QFT- participants, however this is not observed for Ag85B. There is also evidence that Ag85B induced higher quality immune responses compared to ESAT-6.

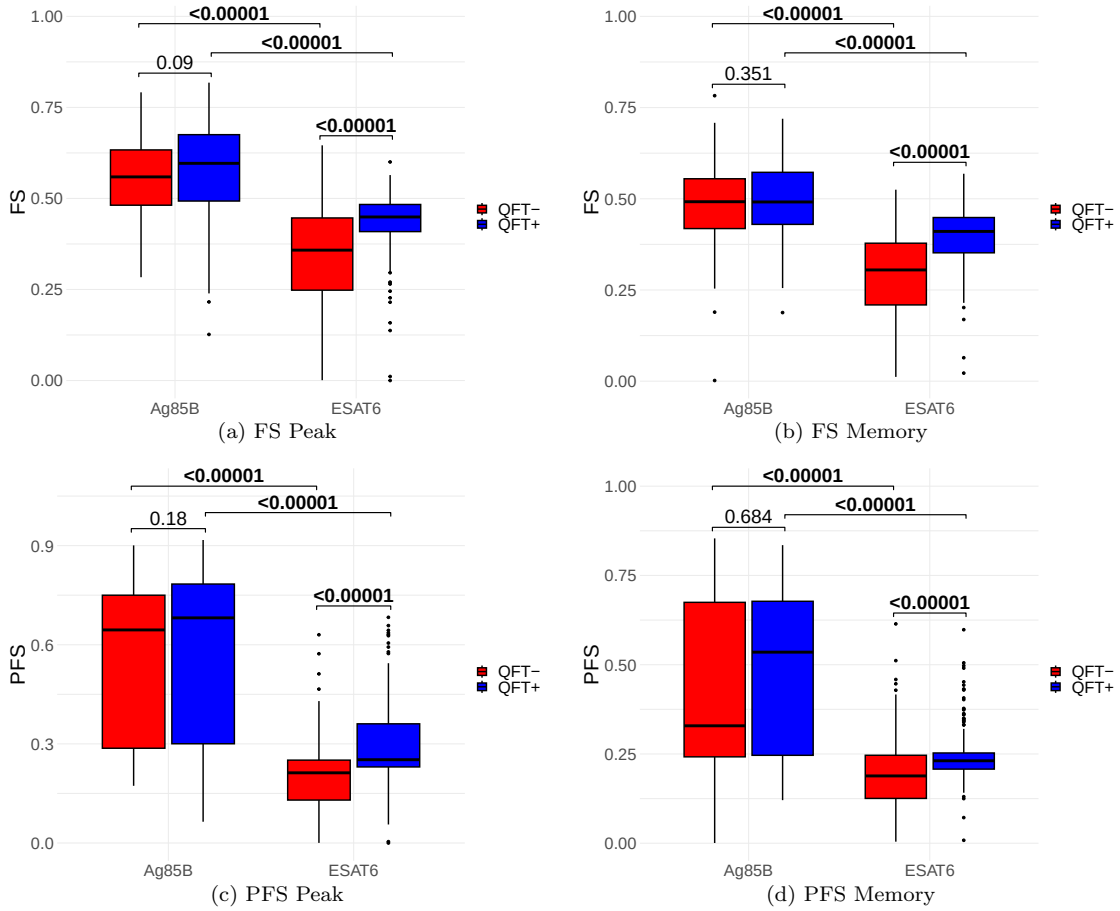
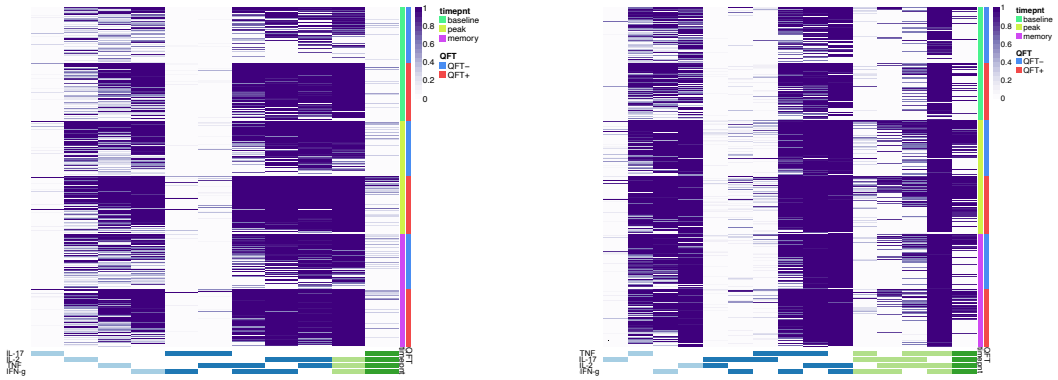


Figure 6.9: Box-plots of the FS and PFS for two administrations of 5 µg of the vaccine for different stimuli and QFT statuses. Benjamini Hochberg adjusted p-values are shown, where the p-values between different QFT statuses but within the same stimulus were determined by (unpaired) Mann Whitney tests while the p-values between different stimuli but within the same QFT status were determined by paired Wilcoxon tests.

Before moving onto the non-parametric approach where hypotheses around specific cytokine combinations shall be investigated, the posterior probabilities calculated using COMPASS can also be analysed to select which cytokine combinations may or may not be reliably detected. Recall that the posterior probabilities can be interpreted as the probability that a cytokine combination is antigen-specific for a given participant visit. Figure 6.10 shows the posterior probabilities produced by applying COMPASS to ESAT-6-stimulated observations and Ag85B-stimulated observations for all participant visits, even those that did not receive two administrations of 5 µg of the vaccine. Referring to the heatmap for the ESAT-6 stimulated observations, cytokine combinations IFN-γ+IL2+TNF−IL17+, IFN-γ+IL2-TNF+IL17+, IFN-γ−IL2+TNF+IL17+ and IFN-γ−IL2+TNF−IL17+ were excluded from the plot because their posterior probabilities were below the specified threshold and thus were identified as biologically irrelevant. Cells expressing IL-17 and IFN-γ only as well as cells expressing IL-17 and TNF only are associated with very low posterior probabilities, and thus had low antigen-specific responses and may also be biologically irrelevant. Cells expressing all cytokines as well as cells expressing only IL-17 are also associated with low posterior probabilities, indicating that they had a low ESAT-6-specific response.

After Ag85B stimulation, no cytokine combinations were excluded from the plot. Cells expressing IL-17 and IL-2 only as well as cells expressing IL-17 only are associated with very low posterior probabilities, and thus had low antigen-specific responses and may be biologically irrelevant. After Ag85B stimulation, more low posterior probabilities are seen at baseline compared to peak and memory time points. At baseline, there seems to be slightly more low posterior probabilities seen for QFT- participants compared to QFT+ participants, however at peak and memory time points there does not seem to be any obvious difference between QFT- and QFT+ participants.



(a) ESAT-6 (b) Ag85B

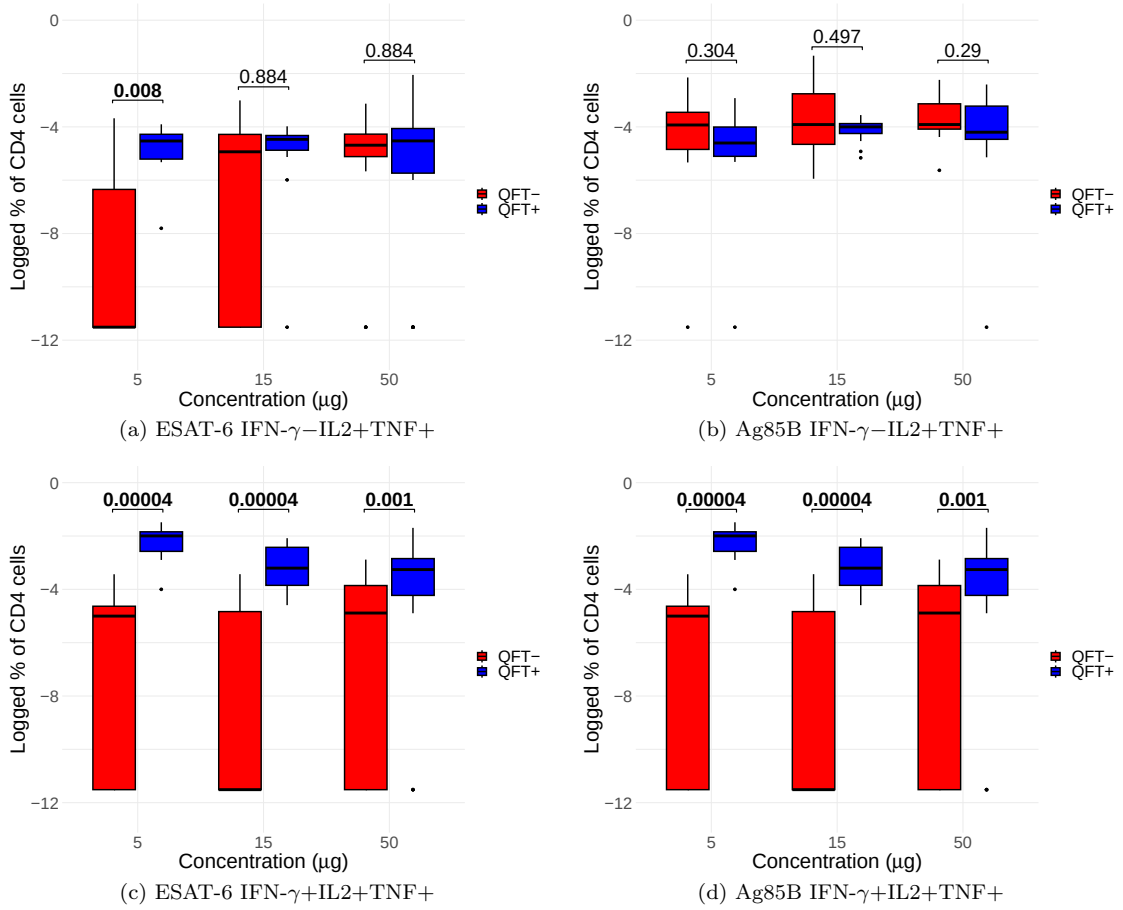
Figure 6.10: Heatmap of COMPASS posterior probabilities for the (a) ESAT-6 and (b) Ag85B data. The columns correspond to the different cell subsets modelled by COMPASS, colour-coded by the cytokines they express (white = "not expressed", shaded = "expressed" and colour = "degree of functionality"), and ordered by degree of functionality from one function on the left to four functions on the right. Rows correspond to participant visits ordered by QFT status and time point. Each cell of the heatmap shows the probability that a given cytokine combination (column) has an antigen-specific response in the corresponding participant visit (column), where the probability is colour-coded from white (zero) to purple (one). Four cytokine combinations, IFN-γ+IL2+TNF−IL17+, IFN-γ+IL2−TNF+IL17+, IFN-γ−IL2+TNF+IL17+ and IFN-γ−IL2+TNF−IL17+, were excluded from the ESAT-6 results as they had a mean posterior probability that was less than 0.01.

6.1.2 Non-parametric approach

This section will focus on the response of the individual cytokine combinations rather than the overall functional quality of the response. The optimal regimen will not be investigated again using individual cytokine combination frequencies, but rather the regimen of two administrations of 5 µg shall be used in order to avoid an excessive amount of plots and repetition.

In order to investigate the hypothesis that high frequencies of T cells expressing IL-2 and TNF together and T cells expressing IL-2, TNF and IFN-γ together are associated with lower concentrations of the vaccine in *M.tb*-sensitised individuals compared to unsensitised individuals, the frequency of cells expressing IFN-γ−IL2+TNF+ and IFN-γ+IL2+TNF+ after stimulation with either ESAT-6 or Ag85B was plot for all vaccine concentrations, as seen in Figure 6.11. Only participants that received two administrations of the vaccine, as chosen in Section 6.1.1, are shown. After ESAT-6 stimulation, QFT+ individuals had higher frequencies of cells expressing IFN-γ−IL2+TNF+ compared to QFT- individuals, at the lowest vaccine concentration (5 µg) only. After Ag85B stimulation, there was no difference in the frequency of cells expressing IFN-

γ−IL2+TNF+ between QFT+ and QFT- participants. For both stimuli, QFT+ individuals had much higher frequencies of cells expressing IFN-γ+IL2+TNF+ compared to QFT- individuals at all vaccine concentrations. However, the difference does seem to be strongest at lower concentrations. Thus, there is minimal evidence that QFT+ individuals had higher frequencies of cells expressing IFN-γ−IL2+TNF+ compared to QFT- individuals. However, there is strong evidence that QFT+ individuals had higher frequencies of cells expressing IL2+TNF+IFN-γ+ compared to QFT- individuals after both Ag85B and ESAT-6 stimulation, particularly at lower concentrations.



(a) ESAT-6 IFN-γ−IL2+TNF+

(b) Ag85B IFN-γ−IL2+TNF+

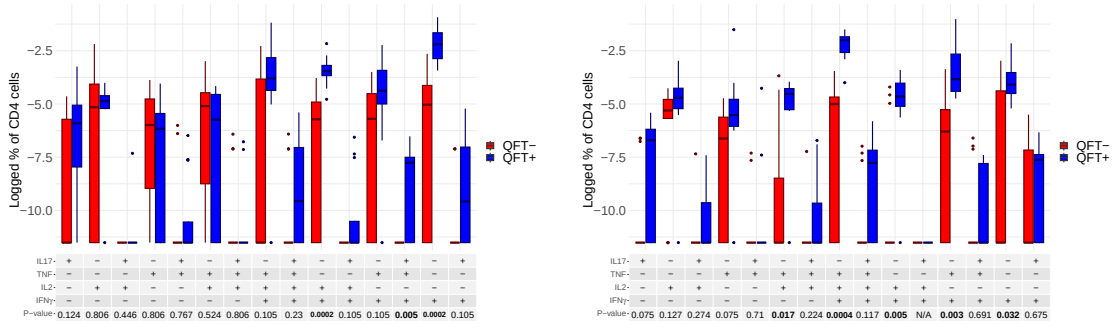
(c) ESAT-6 IFN-γ+IL2+TNF+

(d) Ag85B IFN-γ+IL2+TNF+

Figure 6.11: Box-plots of the frequency of CD4 T cells expressing IFN-γ−IL2+TNF+ and IFN-γ+IL2+TNF+ after stimulation with either ESAT-6 or Ag85B are shown for different concentrations. Only observations from participants who received two administrations of the vaccine were used. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing QFT+ and QFT- observations for the same vaccine concentrations are shown.

The plots in Figures 6.12 and 6.13 were created to further investigate the hypothesis that after vaccination, the functional profiles of T cells recognising Ag85B or ESAT6 differ between QFT+ and QFT- individuals. Figure 6.12 was used to compare QFT+ and QFT- individuals while Figure 6.13 was used to compare the two stimuli. Referring to Figure 6.12, after ESAT-6 stimulation at the peak time point there is evidence that QFT+ participants had higher frequencies of cells expressing only IL-2 and IFN-γ, cells expressing only IFN-γ and cells expressing all cytokines except IL-2 compared to QFT- participants. This is again seen at the memory

time point, where additionally the frequency of cells expressing only IL-2 and TNF as well as cells expressing all cytokines except IL-17 was higher in QFT+ participants compared to QFT- participants. After Ag85B stimulation, there is evidence that QFT- participants had higher frequencies of cells expressing all cytokines except IFN-γ. Otherwise, there is no evidence that a difference in frequency existed between QFT- and QFT+ participants after Ag85B stimulation. Overall, there is evidence that the functional profiles of T cells recognising ESAT-6 differed between QFT+ and QFT- individuals, however there is little evidence that this difference exists after Ag85B stimulation.



(a) ESAT6 Peak

(b) ESAT6 Memory

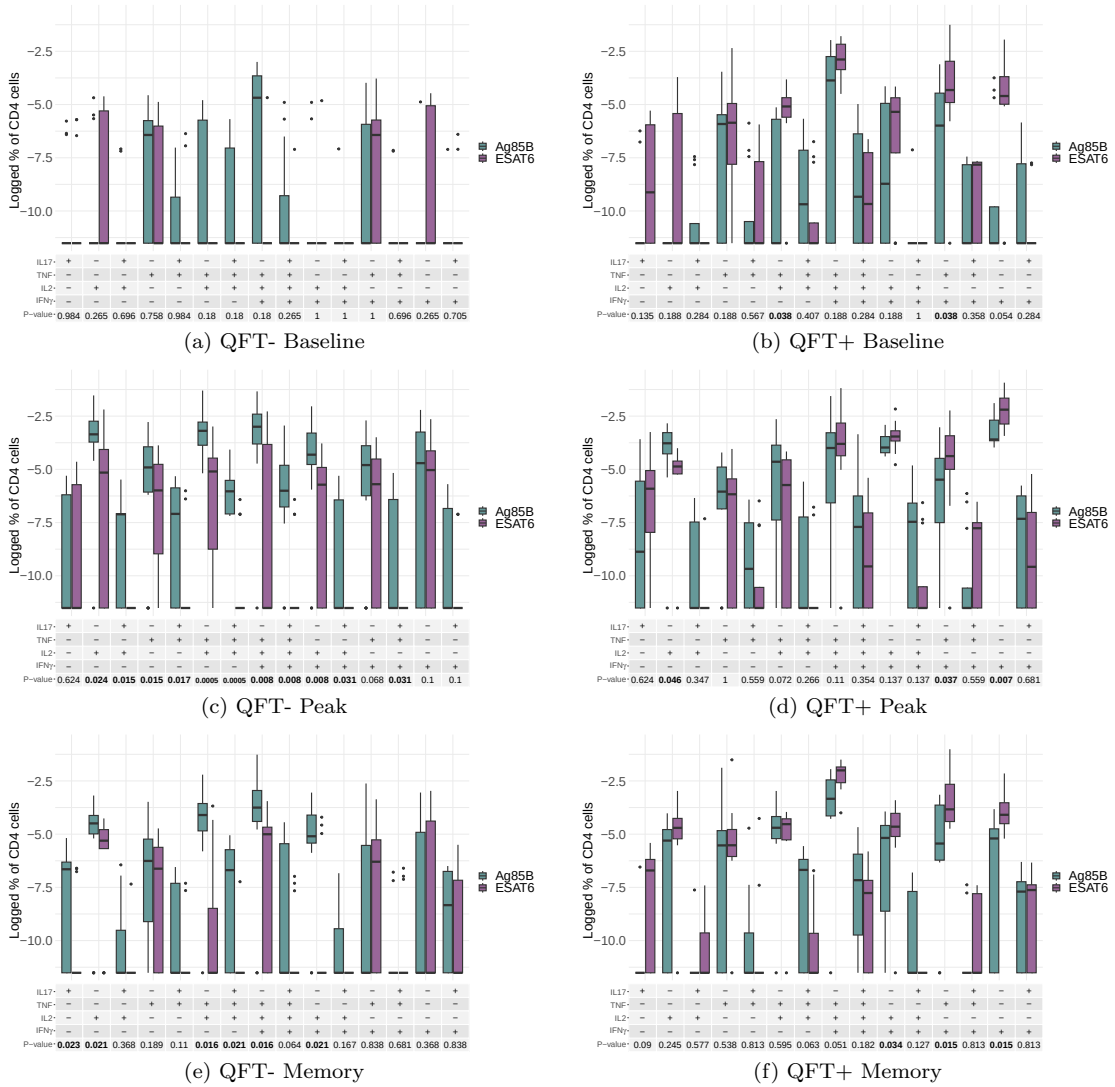
(c) Ag85B Peak

(d) Ag85B Memory

Figure 6.12: Box-plots of the frequency of CD4 T cells expressing various marker combinations after two administrations of 5 µg of the vaccine. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing the different QFT status for the same marker combination are shown in the table.

Referring to Figure 6.13, there is evidence that for QFT- participants at baseline, there was no difference in the frequency of any of the cytokine combinations between the two stimuli. In general, very low frequencies were observed after ESAT-6 stimulation. For QFT+ participants at baseline, there is evidence that the frequency of cells expressing only TNF and IL-2 as well the frequency of cells expressing only TNF and IFN-γ were higher after ESAT-6 stimulation compared to Ag85B stimulation. For QFT- participants at peak, almost all cytokine combinations were expressed at higher frequencies after Ag85B stimulation compared to after ESAT-6 stimulation. Generally, the frequencies after ESAT-6 stimulation were much higher than they were at baseline. For QFT+ participants at peak, again the frequency of cells expressing only TNF and IFN-γ was higher after ESAT-6 stimulation compared to Ag85B stimulation. The frequency of cells expressing IFN-γ only was also higher after ESAT-6 stimulation compared to Ag85B stimulation. However, the frequency of cells expressing only IL-2 was lower after ESAT-6 stimulation compared to Ag85B stimulation. For QFT- participants at memory, many

the cytokine combinations were expressed at much higher frequencies after Ag85B stimulation compared to after ESAT-6 stimulation, although not as many combinations as was seen at peak. Generally, the frequencies after ESAT-6 stimulation were much lower than they were at peak. For QFT+ participants at memory, again the frequency of cells expressing only TNF and IFN-γ as well as the frequency of cells expressing only IFN-γ were higher after ESAT-6 stimulation compared to Ag85B stimulation. The frequency of cells expressing IL-2 and IFN-γ only was higher after ESAT-6 stimulation compared to Ag85B stimulation. Thus, besides from QFT- participants at baseline, there is evidence that the functional profiles of cells recognising Ag85B were distinct from those recognising ESAT-6, especially in QFT- participants post-vaccination.



(a) QFT- Baseline

(b) QFT+ Baseline

(c) QFT- Peak

(d) QFT+ Peak

(e) QFT- Memory

(f) QFT+ Memory

Figure 6.13: Box-plots of the frequency of CD4 T cells expressing various marker combinations after two administrations of 5 µg of the vaccine. Benjamini Hochberg adjusted p-values, determined by paired Wilcoxon tests comparing the different stimuli for the same marker combination are shown in the table.

6.1.3 Comment on the standard approach

In this section, two different sets of response variables were explored: the FS and PFS produced using COMPASS and the CD4 T cell frequency of the 16 different cytokine combinations.

Both responses gave slightly different inference, with the FS and PFS providing a summary measure on functional quality of the immune response while the cytokine combinations frequency provided more insight into the exact functional profiles making up the response. Using the FS and PFS, it is far easier to identify which subset of observations produces higher quality immune responses, where this is harder to determine using the frequency of the cytokine combinations. While the FS and PFS had the advantage of reducing the dimensionality of the response and thus make inference easier, they did have the disadvantage of relying on COMPASS being implemented appropriately and it was not possible to depict the confidence intervals for each FS and PFS when they are analysed.

Using the FS and PFS, the optimal regimen of two administrations of 5 µg of the vaccine was selected. This was the same regimen that was selected when analysing the magnitude of the T cell response, suggesting that this regimen was not only optimal in terms of functional quality but also in terms of magnitude of the response. The FS and PFS provided evidence that QFT+ participants had higher quality immune responses after ESAT-6 stimulation at lower concentrations compared to QFT- participants but that there was no evidence of a difference under Ag85B stimulation. Using frequency of the cytokine combinations response, the same hypothesis was investigated but focusing on the two cytokine combinations IFN-γ+IL2+TNF+ and IFN-γ−IL2+TNF+. There was strong evidence that QFT+ individuals had higher frequencies of cells expressing IFN-γ+IL2+TNF+ after stimulation with either stimulus compared to QFT- participants, particularly at lower concentrations, but there was no evidence of a difference in expression of IFN-γ−IL2+TNF+. Both sets of response variables provided evidence that the functional profiles of T cells recognising ESAT-6 differed between QFT+ and QFT- individuals but that no difference existed for T cells recognising Ag85B. The FS and PFS was able to qualify this difference, providing evidence that QFT+ participants had higher quality immune responses. Both responses also found evidence that the functional profiles of cells recognising Ag85B were distinct from those recognising ESAT-6, with the FS and PFS further qualifying this difference by identifying that Ag85B induced higher quality immune responses.

6.2 GLMEM's

In order to make inference about specific cytokine combinations, the response variable of interest for this section is the frequency of CD4 T cells expressing cytokine combination j observed for participant i at time point k under stimulus s, $Y_{ijk}^{(s)}$. However, not all cytokine combinations shall be included in the response, since some are observed at very low frequencies and are biological irrelevant and would further zero-inflate the data substantially. Thus, the COMPASS output was used to inform the selection of the appropriate cytokine combinations. Specifically, the posterior probabilities produced by COMPASS were used to select the cytokine combinations. All cytokine combinations for which more than 30% of observations had posterior probabilities less than 0.1 were excluded from the response, a similar criteria to what has been used in previous literature [78, 68]. Cytokine combination IFN-γ-IL2-TNF-IL17- was also excluded since only cytokine positive combinations were of interest. Since COMPASS had to be applied separately to

each stimulus, different cytokine combinations were excluded depending on the stimulus. Under ESAT-6, only 7 cytokine combinations were retained and under Ag85B, 9 cytokine combinations were retained. The exact cytokine combinations that were retained for each stimulus can be seen in Table 6.1. Interestingly, no cytokine combinations that expressed IL-17 after ESAT-6 stimulation were retained as biologically relevant, while after Ag85B stimulation, only two cytokine combinations that expressed IL-17 were retained as biologically relevant. All cytokine combinations that did not express IL-17 were identified as biologically relevant, for both stimuli.

Separate two-level GLMEM's grouped by memory combination nested within participant shall be fit to each stimulus using $\boldsymbol{Y}_{ijk}$ as the response variable. Since different cytokine combinations were excluded depending on the stimulus, a three-level GLMEM grouped by cytokine combination nested in stimulus was not fit, as depending on the stimulus different cytokine combinations would need to be included into the model.

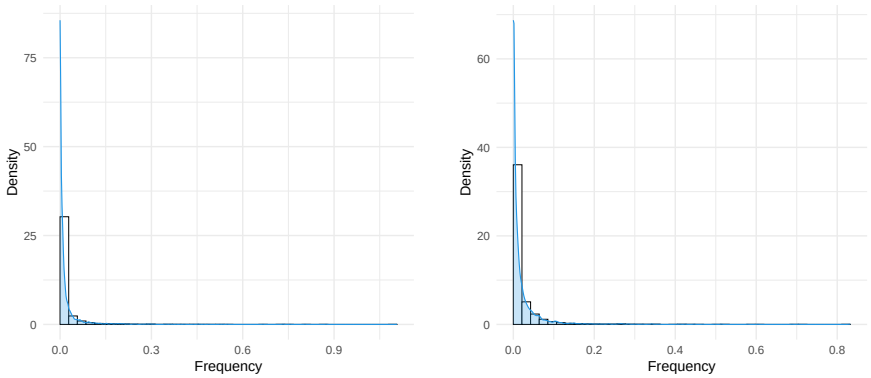
Table 6.1: Table showing the cytokine combinations that were identified using COMPASS as biologically relevant depending on the stimulation used, indicated with a tick.

Cytokine Combination	ESAT-6	Ag85B
IFN-γ+IL2+TNF+IL17+		✓
IFN-γ+IL2+TNF−IL17+		
IFN-γ+IL2+TNF+IL17−	✓	✓
IFN-γ+IL2+TNF−IL17−	✓	✓
IFN-γ+IL2−TNF+IL17+		
IFN-γ+IL2−TNF−IL17+		
IFN-γ+IL2−TNF+IL17−	✓	✓
IFN-γ+IL2−TNF−IL17−	✓	✓
IFN-γ−IL2+TNF+IL17+		✓
IFN-γ−IL2+TNF−IL17+		
IFN-γ−IL2+TNF+IL17−	✓	✓
IFN-γ−IL2+TNF−IL17−	✓	✓
IFN-γ−IL2−TNF+IL17+		
IFN-γ−IL2−TNF−IL17+		
IFN-γ−IL2−TNF+IL17−	✓	✓

6.2.1 Candidate distributions for the response

In order to choose an appropriate conditional distribution for the response variable, the marginal distribution of the response needs to be assessed. Figure 6.14 shows the density plots for each of the response variables for the two-level GLMEM's grouped by memory combination. Both response variables are continuous on the domain $[0, 100]$, highly zero-inflated and positively skewed. Possible candidate distributions are again the Gamma, log-normal and Weibull, since they all at most require two parameters and are suitable for continuous right-skewed data. Again, the Tweedie distribution shall also be considered because it is well-suited to skew data. For the same reasoning given in Section 5.2.1, neither the Sichel, ZIP or ZIPNB distributions

were considered.



(a) ESAT-6 (b) Ag85B

Figure 6.14: Density plots of the frequency of cells expressing any one of the cytokine combinations selected in Table 6.1 after stimulation with (a) ESAT-6 or (b) Ag85B. These frequencies will be used as the response variable for the two-level GLMEM's grouped by memory combination.

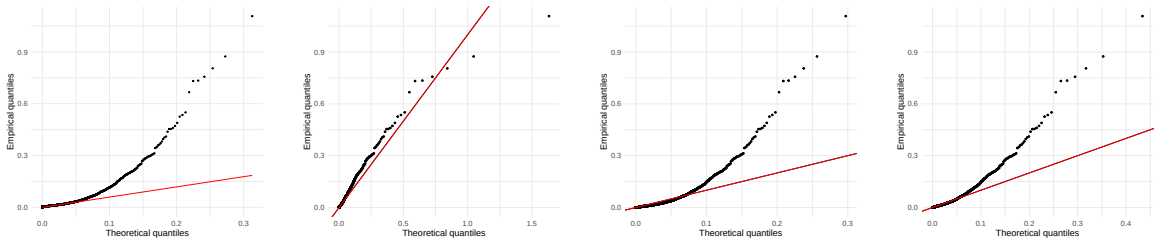
6.2.2 Two-level GLMEM after ESAT-6 stimulation

A log-normal, Gamma, Weibull and Tweedie distribution were all fit to response variable, the frequency of CD4 T cells which express certain cytokine combinations, selected using the COMPASS output, after stimulation with ESAT-6. The log-normal, Gamma and Weibull distributions were fit using the `fitdistrplus` [76] package in `R`, while the Tweedie distribution was fit using the `cplm` [77] package. Table 6.2 summaries the resultant AIC and BIC values from the fitted log-normal, Gamma and Weibull distributions. The log-normal had the lowest AIC and BIC compared to the other distributions, suggesting that out of the three distributions the log-normal fit the data best.

Table 6.2: The AIC and BIC after fitting log-normal, Gamma and Weibull distributions to the cytokine combination frequencies after ESAT-6 stimulation

Distribution	**AIC**	**BIC**
Log-normal	-35128.59	-35115.51
Gamma	-32846.64	-32833.55
Weibull	-33808.39	-33795.31

To further assess the fit, the empirical versus theoretical quantiles were plot, as seen in Figure 6.15. All four qq-plots suggest that neither of the distributions fit the data well, however the log-normal distribution does seem to fit the data far better than the other three distributions. Since the log-normal also had the lowest AIC and BIC compared to the Weibull and Gamma, it was decided to assume that the conditional response follows a log-normal distribution.



(a) Tweedie (b) Log-normal (c) Gamma (d) Weibull

Figure 6.15: QQ-plots of the cytokine combination frequencies after ESAT-6 stimulation fit to (a) Tweedie, (b) log-normal, (c) Gamma and (b) Weibull distributions.

Following the model building steps laid out in Section 4.3.1, the optimal model chosen is as follows

$$\begin{aligned}
&\log(\mu_{ijk}) = \boldsymbol{x}_{ijk}^T\boldsymbol{\beta} + b_{0i},\\
&\log(\sigma_{ijk}) = \boldsymbol{w}_{ijk}^T\boldsymbol{\alpha},\\
&\text{for} \quad i = 1, ..., I, \quad j = 1, 2, .., 7 \quad \text{and} \quad k = 1, 2, 3,
\end{aligned}$$

where

$$b_{0i} \sim N\left(0, \sigma_0^2\right) \qquad \text{and} \qquad \epsilon_{ijk} \sim N\left(0, \sigma_e^2\right).$$

Here i is the participant ID, j is the cytokine combination and k is the time point. $\hat{\sigma}_0 = 0.647$ with 95% confidence interval $(0.585; 0.714)$ and $\hat{\sigma}_e = 1.022$ with 95% confidence interval $(1.002; 1.043)$. Vector notation was used due to the large number of model parameters, and the exact explanatory variables included in the model can be seen in Table 6.3. No participant-specific random effect was included on the indicator variable reflecting the cytokine combinations due to convergence issues and it was not included on time point because it resulted in higher GAIC values, indicating an inferior fit. A three-way interaction between number of administrations, concentration and time point was included in the model as it improved the model fit.

Model diagnostics can be found in Appendix E.1, where there was evidence that the normality assumption of the random effects is violated, but that the normality and homogeneity assumptions of the residuals are valid.

Referring to the parameter estimates seen in Table 6.3, there is evidence of a strong interaction between QFT status and cytokine combination. QFT+ participants are associated with higher frequencies compared to QFT- participants, no matter the cytokine combination. However, the size of the difference is dependent on the cytokine combination. For example, the reference category IFN-γ+IL2+TNF+IL17$-$ is associated with the largest difference in frequency between QFT+ and QFT- participants, while IFN-$\gamma-$IL2+TNF$-$IL17$-$ is associated with the smallest difference. There is a strong interaction between time point, vaccine concentration and number of administrations. At baseline, number of administration and vaccine concentration does not

have an effect on cytokine frequency. This makes sense because baseline is a pre-vaccination time point, and thus future vaccine regimen should have no effect on the response. At peak, frequency decreases as the number of administrations increases, except for when a vaccine concentration of 50 µg is used, in which case the relationship is reversed. At peak, frequency decreases as the vaccine concentration decreases, except for when three administrations are received, in which case the relationship is reversed. Thus there is evidence that one administration of 5 µg produces the greatest immune response. At the memory time point, the frequency is highest when three administrations are given, regardless of concentration, and when 5 µg is administered, regardless of the number of administrations. Thus to induce a long-term immune response, there is evidence that one administration of 5 µg, the least resource intensive regimen, is optimal.

The variation in frequency is highest for cells expressing IFN-γ+IL2−TNF−IL17−, and secondly for cells expressing the reference category IFN-γ+IL2+TNF+IL17−. There is also evidence that variance increases with increasing number of administrations and decreases with increasing vaccine concentrations. QFT+ participants are associated with higher variation in cytokine combination frequencies compared to QFT- participants.

Table 6.3: Parameter estimates of the two-level GLMEM fit to the cytokine combination frequencies after ESAT-6 stimulation.

Variable	Parameter	Estimate	Std. Error	t-value	p-value
μ Coefficients:					
	β_0	-6.405	0.136	-47.039	$< 2 \times 10^{-16}$
Peak	β_1	1.945	0.180	10.826	$< 2 \times 10^{-16}$
Mem.	β_2	0.630	0.178	3.538	0.0004
QFT+	β_3	2.565	0.079	32.375	$< 2 \times 10^{-16}$
Admin.	β_4	-0.056	0.062	-0.903	0.366
Conc.	β_5	0.002	0.003	0.721	0.471
IFN-γ+IL2+TNF−IL17−	β_6	-0.285	0.068	-4.205	2.66×10^{-5}
IFN-γ+IL2−TNF+IL17−	β_7	-0.038	0.070	-0.543	0.587
IFN-γ+IL2−TNF−IL17−	β_8	0.167	0.085	1.959	0.050
IFN-γ−IL2+TNF+IL17−	β_9	0.255	0.070	3.619	0.0003
IFN-γ−IL2+TNF−IL17−	β_{10}	0.225	0.075	3.017	0.003
IFN-γ−IL2−TNF+IL17−	β_{11}	0.201	0.069	2.901	0.004
Admin.*Conc.	β_{12}	-0.001	0.002	-0.686	0.493
Admin.*Peak	β_{13}	-0.378	0.088	-4.311	1.66×10^{-5}
Admin.*Mem.	β_{14}	0.031	0.087	0.352	0.725
Conc.*Peak	β_{15}	-0.035	0.005	-7.159	9.33×10^{-13}
Conc.*Mem.	β_{16}	-0.020	0.005	-4.131	3.67×10^{-5}
IFN-γ+IL2+TNF−IL17−*QFT+	β_{17}	-1.006	0.101	-9.941	$< 2 \times 10^{-16}$
IFN-γ+IL2−TNF+IL17−*QFT+	β_{18}	-1.032	0.105	-9.867	$< 2 \times 10^{-16}$
IFN-γ+IL2−TNF−IL17−*QFT+	β_{19}	-0.930	0.127	-7.303	3.27×10^{-13}
IFN-γ−IL2+TNF+IL17−*QFT+	β_{20}	-1.694	0.105	-16.102	$< 2 \times 10^{-16}$
IFN-γ−IL2+TNF−IL17−*QFT+	β_{21}	-2.509	0.112	-22.456	$< 2 \times 10^{-16}$
IFN-γ−IL2−TNF+IL17−*QFT+	β_{22}	-1.616	0.1035	-15.604	$< 2 \times 10^{-16}$
Admin.*Conc.*Peak	β_{23}	0.014	0.002	5.566	2.74×10^{-8}
Admin.*Conc.*Mem.	β_{24}	0.007	0.002	3.012	0.003
σ Coefficients:					
	α_0	-0.035	0.042	-0.850	0.395
IFN-γ+IL2+TNF−IL17−	α_1	-0.228	0.037	-6.132	9.35×10^{-10}
IFN-γ+IL2−TNF+IL17−	α_2	-0.147	0.037	-3.952	7.85×10^{-5}
IFN-γ+IL2−TNF−IL17−	α_3	0.228	0.037	6.146	8.59×10^{-10}
IFN-γ−IL2+TNF+IL17−	α_4	-0.135	0.037	-3.629	0.0003
IFN-γ−IL2+TNF−IL17−	α_5	-0.007	0.037	-0.196	0.844
IFN-γ−IL2−TNF+IL17−	α_6	-0.173	0.037	-4.674	3.04×10^{-6}
QFT+	α_7	0.129	0.020	6.385	1.87×10^{-10}
Conc.	α_8	-0.002	0.0005	-4.398	1.12×10^{-5}
Admin.	α_9	0.058	0.014	4.147	3.43×10^{-5}

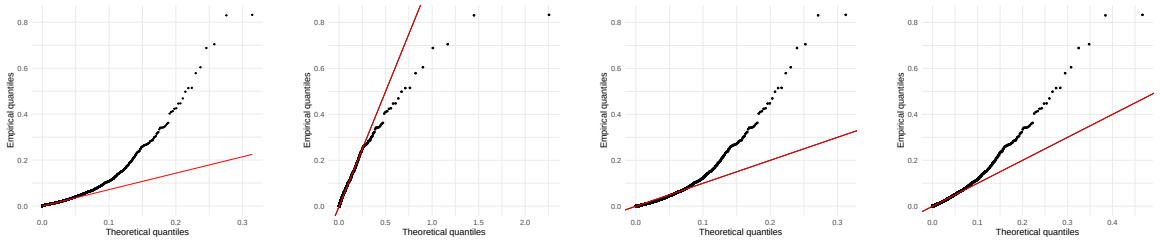
6.2.3 Two-level GLMEM after Ag85B stimulation

A log-normal, Gamma, Weibull and Tweedie distribution were all fit to response variable, the frequency of CD4 T cells which express certain cytokine combinations, selected using the COMPASS output, after stimulation with Ag85B. The log-normal, Gamma and Weibull distributions were fit using the `fitdistrplus` [76] package in `R`, while the Tweedie distribution was fit using the `cplm` [77] package. Table 6.4 summaries the resultant AIC and BIC values from the fitted log-normal, Gamma and Weibull distributions. The log-normal had the lowest AIC and BIC compared to the other distributions, suggesting that out of the three distributions the log-normal fit the data best.

Table 6.4: The AIC and BIC after fitting log-normal, Gamma and Weibull distributions to the cytokine combination frequencies after Ag85B stimulation

Distribution	AIC	BIC
Log-normal	-42721.26	-42707.68
Gamma	-40567.19	-40553.61
Weibull	-41430.89	-41417.3

To further assess the fit, the empirical versus theoretical quantiles were plot, as seen in Figure 6.16. All four qq-plots suggest that neither of the distributions fit the data well, however the log-normal distribution does seem to fit the data far better than the other three distributions. Since the log-normal also had the lowest AIC and BIC compared to the Weibull and Gamma, it was decided to assume that the conditional response follows a log-normal distribution.



(a) Tweedie (b) Log-normal (c) Gamma (d) Weibull

Figure 6.16: QQ-plots of the cytokine combination frequencies after Ag85B stimulation fit to (a) Tweedie, (b) log-normal, (c) Gamma and (b) Weibull distributions.

Following the model building steps laid out in Section 4.3.1, the optimal model chosen is as follows

$$\begin{aligned}
&\log(\mu_{ijk}) = \boldsymbol{x}_{ijk}^T\boldsymbol{\beta} + b_{0i},\\
&\log(\sigma_{ijk}) = \boldsymbol{w}_{ijk}^T\boldsymbol{\alpha},\\
&\text{for} \quad i = 1, ..., I, \quad j = 1, 2, .., 9 \quad \text{and} \quad k = 1, 2, 3,
\end{aligned}$$

where

$$b_{0i} \sim N\left(0, \sigma_0^2\right) \qquad \text{and} \qquad \epsilon_{ijk} \sim N\left(0, \sigma_e^2\right).$$

Here i is the participant ID, j is the cytokine combination and k is the time point. $\hat{\sigma}_0 = 0.488$ with 95% confidence interval $(0.442; 0.538)$ and $\hat{\sigma}_e = 1.017$ with 95% confidence interval $(1.000; 1.035)$. Vector notation was used due to the large number of model parameters, and the exact explanatory variables included in the model can be seen in Table 6.5. No participant-specific random effect was included on the indicator variable reflecting the cytokine combinations due to convergence issues and it was not included on time point because it resulted in higher GAIC values, indicating an inferior fit. No three-way interactions were included in the model as they did not improve the model fit.

Model diagnostics can be found in Appendix E.2, where there was evidence that the normality assumption of the random effects is violated, but that the normality and homogeneity assumptions of the residuals are valid.

Referring to the parameter estimates seen in Tables 6.5 and 6.6, it is clear that this model is very complicated and potentially over-fit due to the many parameters required. All of the variables of interest, time point, QFT status, number of administrations and vaccine concentration, interact strongly with cytokine combination. This suggests that the relationship between these variables of interest and the response variable is highly dependent on the cytokine combination under consideration. This motivates the decision to model this response using a LVM, since a LVM would allow cytokine combination-specific coefficients on these variables of interest. Thus, a LVM would be able to capture the cytokine-combination dependant relationship between the variables of interest and the response variable without the need for interaction variables which require more model parameters. LVM's coefficients would also potentially be easier to interpret compared to these interaction coefficients. This model is very difficult to interpret, so only inference for the reference category IFN-γ+IL2+TNF+IL17+ shall be made. For QFT- participants, the frequency of IFN-γ+IL2+TNF+IL17+ at both post-vaccination time points decreases with increasing vaccine concentration, regardless of number of vaccine administrations. For a fixed concentration, the frequency of IFN-γ+IL2+TNF+IL17+ at peak increases with increasing numbers of administrations for QFT- participants. However at the memory time point, the frequency of IFN-γ+IL2+TNF+IL17+ at peak increases with increasing administrations, regardless of concentration. Thus, three administrations of 5 µg induces the greatest frequency of IFN-γ+IL2+TNF+IL17+ in QFT- participants at both post-vaccination time points.

In terms of variation, all of the cytokine combinations are associated with higher variation in frequency compared to the reference combination IFN-γ+IL2+TNF+IL17+. Baseline observations are also associated with higher variation in frequency and QFT+ participants are associated with higher variation in cytokine combination frequencies compared to QFT- participants.

Table 6.5: Parameter estimates of the two-level GLMEM fit to the cytokine combination frequencies after Ag85B stimulation.

Variable	Parameter	Estimate	Std. Error	t-value	p-value
μ Coefficients:					
	β_0	-6.863	0.090	-76.160	$< 2 \times 10^{-16}$
Peak	β_1	0.508	0.096	5.309	1.14×10^{-7}
Mem.	β_2	0.018	0.096	0.190	0.849
QFT+	β_3	0.230	0.051	4.529	6.05×10^{-6}
Admin.	β_4	0.069	0.036	1.912	0.056
Conc.	β_5	-0.003	0.001	-2.580	0.010
IFN-γ+IL2+TNF+IL17−	β_6	1.255	0.187	6.693	2.38×10^{-11}
IFN-γ+IL2+TNF−IL17−	β_7	0.991	0.147	6.757	1.53×10^{-11}
IFN-γ+IL2−TNF+IL17−	β_8	0.993	0.166	5.972	2.47×10^{-9}
IFN-γ+IL2−TNF−IL17−	β_9	1.708	0.205	8.316	$< 2 \times 10^{-16}$
IFN-γ−IL2+TNF+IL17+	β_{10}	0.278	0.105	2.654	0.008
IFN-γ−IL2+TNF+IL17−	β_{11}	0.345	0.141	2.438	0.015
IFN-γ−IL2+TNF−IL17−	β_{12}	1.065	0.157	6.773	1.37×10^{-11}
IFN-γ−IL2−TNF+IL17−	β_{13}	1.016	0.158	6.437	1.31×10^{-10}
Admin.*IFN-γ+IL2+TNF+IL17−	β_{14}	0.145	0.071	2.047	0.041
Admin.*IFN-γ+IL2+TNF−IL17−	β_{15}	-0.205	0.055	-3.693	0.0002
Admin.*IFN-γ+IL2−TNF+IL17−	β_{16}	-0.029	0.063	-0.466	0.641
Admin.*IFN-γ+IL2−TNF−IL17−	β_{17}	-0.500	0.078	-6.444	1.25×10^{-10}
Admin.*IFN-γ−IL2+TNF+IL17+	β_{18}	-0.122	0.040	-3.098	0.002
Admin.*IFN-γ−IL2+TNF+IL17−	β_{19}	0.056	0.053	1.042	0.298
Admin.*IFN-γ−IL2+TNF−IL17−	β_{20}	-0.263	0.059	-4.430	9.58×10^{-6}
Admin.*IFN-γ−IL2−TNF+IL17−	β_{21}	-0.121	0.060	-2.023	0.043
IFN-γ+IL2+TNF+IL17−*Conc.	β_{22}	-0.003	0.003	-1.250	0.211
IFN-γ+IL2+TNF−IL17−*Conc.	β_{23}	0.003	0.002	1.399	0.162
IFN-γ+IL2−TNF+IL17−*Conc.	β_{24}	-0.001	0.002	-0.470	0.638
IFN-γ+IL2−TNF−IL17−*Conc.	β_{25}	0.002	0.003	0.860	0.390
IFN-γ−IL2+TNF+IL17+*Conc.	β_{26}	0.002	0.001	1.311	0.190
IFN-γ−IL2+TNF+IL17−*Conc.	β_{27}	0.006	0.002	2.974	0.003
IFN-γ−IL2+TNF−IL17−*Conc.	β_{28}	0.015	0.002	6.963	3.66×10^{-12}
IFN-γ−IL2−TNF+IL17−*Conc.	β_{29}	0.007	0.002	3.307	0.0009
Conc.*Peak	β_{30}	-0.005	0.001	-3.791	0.0002
Conc.*Mem.	β_{31}	0.0004	0.001	0.274	0.784
Admin.*Peak	β_{32}	0.198	0.036	5.536	3.22×10^{-8}
Admin.*Mem.	β_{33}	0.224	0.036	6.247	4.46×10^{-10}

Table 6.6: Parameter estimates of the two-level GLMEM fit to the cytokine combination frequencies after Ag85B stimulation continued.

Variable	Parameter	Estimate	Std. Error	t-value	p-value
μ Coefficients:					
IFN-γ+IL2+TNF+IL17−*Peak	β_{34}	0.776	0.125	6.206	5.77×10^{-10}
IFN-γ+IL2+TNF−IL17−*Peak	β_{35}	1.112	0.098	11.367	$< 2 \times 10^{-16}$
IFN-γ+IL2−TNF+IL17−*Peak	β_{36}	0.333	0.111	2.996	0.003
IFN-γ+IL2−TNF−IL17−*Peak	β_{37}	1.136	0.137	8.297	$< 2 \times 10^{-16}$
IFN-γ−IL2+TNF+IL17+*Peak	β_{38}	-0.216	0.070	-3.090	0.002
IFN-γ−IL2+TNF+IL17−*Peak	β_{39}	1.599	0.095	16.915	$< 2 \times 10^{-16}$
IFN-γ−IL2+TNF−IL17−*Peak	β_{40}	1.531	0.105	14.597	$< 2 \times 10^{-16}$
IFN-γ−IL2−TNF+IL17−*Peak	β_{41}	0.319	0.105	3.025	0.002
IFN-γ+IL2+TNF+IL17−*Mem.	β_{42}	0.603	0.126	4.802	1.61×10^{-6}
IFN-γ+IL2+TNF−IL17−*Mem.	β_{43}	0.870	0.098	8.855	$< 2 \times 10^{-16}$
IFN-γ+IL2−TNF+IL17−*Mem.	β_{44}	0.001	0.111	0.012	0.990
IFN-γ+IL2−TNF−IL17−*Mem.	β_{45}	0.386	0.137	2.805	0.005
IFN-γ−IL2+TNF+IL17+*Mem.	β_{46}	-0.090	0.070	-1.287	0.198
IFN-γ−IL2+TNF+IL17−*Mem.	β_{47}	1.420	0.095	14.942	$< 2 \times 10^{-16}$
IFN-γ−IL2+TNF−IL17−*Mem.	β_{48}	1.203	0.105	11.421	$< 2 \times 10^{-16}$
IFN-γ−IL2−TNF+IL17−*Mem.	β_{49}	0.007	0.106	0.071	0.944
IFN-γ+IL2+TNF+IL17−*QFT+	β_{50}	0.775	0.100	7.722	1.32×10^{-14}
IFN-γ+IL2+TNF−IL17−*QFT+	β_{51}	0.329	0.079	4.193	2.79×10^{-5}
IFN-γ+IL2−TNF+IL17−*QFT+	β_{52}	0.595	0.089	6.682	2.56×10^{-11}
IFN-γ+IL2−TNF−IL17−*QFT+	β_{53}	0.494	0.110	4.495	7.09×10^{-6}
IFN-γ−IL2+TNF+IL17+*QFT+	β_{54}	-0.099	0.056	-1.760	0.078
IFN-γ−IL2+TNF+IL17−*QFT+	β_{55}	0.133	0.076	1.751	0.080
IFN-γ−IL2+TNF−IL17−*QFT+	β_{56}	-0.134	0.084	-1.594	0.111
IFN-γ−IL2−TNF+IL17−*QFT+	β_{57}	0.243	0.085	2.871	0.004
Peak*QFT+	β_{58}	-0.159	0.051	-3.152	0.002
Mem.*QFT+	β_{59}	-0.224	0.051	-4.402	1.09×10^{-5}
σ Coefficients:					
	α_0	-0.592	0.032	-18.612	$< 2 \times 10^{-16}$
IFN-γ+IL2+TNF+IL17−	α_1	0.844	0.037	22.655	$< 2 \times 10^{-16}$
IFN-γ+IL2+TNF−IL17−	α_2	0.536	0.037	14.443	$< 2 \times 10^{-16}$
IFN-γ+IL2−TNF+IL17−	α_3	0.700	0.037	18.846	$< 2 \times 10^{-16}$
IFN-γ+IL2−TNF−IL17−	α_4	0.950	0.037	25.583	$< 2 \times 10^{-16}$
IFN-γ−IL2+TNF+IL17+	α_5	0.002	0.037	0.048	0.962
IFN-γ−IL2+TNF+IL17−	α_6	0.488	0.037	13.120	$< 2 \times 10^{-16}$
IFN-γ−IL2+TNF−IL17−	α_7	0.628	0.037	16.857	$< 2 \times 10^{-16}$
IFN-γ−IL2−TNF+IL17−	α_8	0.633	0.037	17.059	$< 2 \times 10^{-16}$
Peak	α_9	-0.105	0.022	-4.834	1.37×10^{-6}
Mem.	α_{10}	-0.098	0.022	-4.508	6.65×10^{-6}
QFT+	α_{11}	0.069	0.018	3.890	0.0001

6.2.4 Comment on the GLMEM's

There was evidence that the ESAT-6 GLMEM fit the data well, with evidence that only the normality assumption of the random effects was violated. As such, inference can be made using this model. There was evidence that one administration of 5 µg was the optimal regimen for inducing long-term immune responses. With regards to the hypotheses of interest, there was no evidence of an interaction between QFT status and concentration, and thus no evidence that higher quality immune responses are associated with lower concentrations of the vaccine in QFT+ individuals compared to QFT- individuals. There was also evidence that the frequency of all 7 cytokine combinations was higher in QFT+ participants compared to QFT- participants, and there was evidence of an interaction between QFT status and cytokine combination. The effect size of QFT on the frequency of cells expressing IFN-γ+IL2+TNF+IL17− was relatively large, supporting the hypothesis that the frequency of T cells expressing IFN-γ, IL-2 and TNF together is higher in QFT+ participants. Although the effect size of QFT on the frequency of cells expressing IFN-γ−IL2+TNF+IL17− was not as large, there was still evidence that the frequency of T cells expressing IL-2 and TNF together is higher in QFT+ participants.

There was evidence from diagnostic plots that the Ag85B GLMEM fit the data adequately, however the large number of parameters in the model suggested that it may be over-fit to the data. As such, the Ag85B GLMEM should not be relied upon for inference. However, the general interpretability of the model can be discussed. It was difficult to determine which regimen was optimal using the parameter estimates due to the large number of interaction terms. The output would probably be more beneficial to an immunologist with specialist knowledge as to which cytokine combinations are most important when trying to induce a long-term immune response. By focusing on these cytokine combinations, it may be possible to infer an optimal regimen. Modelling the FS and PFS would also be helpful for selecting a vaccine regimen, as these scores act as useful summary statistics of the cytokine frequencies. Perhaps it would be beneficial to model the FS and PFS specifically with the goal of selecting an optimal hypothesis, and then modelling the individual cytokine combinations as done here can be used to explore the effects of QFT status, stimulus and vaccine regimen on the individual cytokine combinations to better understand the quality of the immune response.

6.3 LVM's

Since the complexity of the Ag85B GLMEM seemed to suggest that a LVM would be more appropriate while the simpler ESAT-6 GLMEM seemed to fit the data well, only a Ag85B LVM shall be fit. A further reason for only considering Ag85B LVM's and not ESAT-6 LVM's as well is to avoid unnecessary repetition, as this is an illustrative example only. The structure of the Ag85B GLMEM shall be used to inform the structure of the LVM's fit. Based off of the Ag85B GLMEM, all LVM's fit shall include a random effect on the intercept only, main effects time point, QFT status, number of administrations and dose and interactions between time point and number of administrations, concentration and QFT status. The conditional distribution

of each cytokine combination response was assumed to be log-normal with standard deviation σ_j and an identity link function was used. First, a LVM without traits shall be fit followed by a LVM with traits. The two models shall then be compared using their diagnostic plots and parameter estimates.

6.3.1 LVM without traits

The model specification for the LVM fit without traits is as follows

$$\begin{aligned}\mu_{ijk} = {} & \beta_{0j} + \beta_{1j} I_k^{Peak} + \beta_{2j} I_k^{Mem} + \beta_{3j} I_i^{QFT+} + \beta_{4j} x_i^{Admin} + \beta_{5j} x_i^{Conc} + \\ & \beta_{6j} x_{ik}^{Admin*Peak} + \beta_{7j} x_{ik}^{Admin*Mem} + \beta_{8j} x_{ik}^{Conc*Peak} + \beta_{9j} x_{ik}^{Conc*Mem} + \\ & \beta_{10j} I_{ik}^{Peak*QFT+} + \beta_{11j} I_{ik}^{Mem*QFT+} + \\ & b_{0i} + \theta_{1j} u_{1ik} + \theta_{2j} u_{2ik}, \\ & i = 1, ..., I, \qquad j = 1, ..., 9, \qquad k = 1, 2, 3\end{aligned}$$

where

$$b_{0i} \sim N\left(0, \sigma_0^2\right) \qquad \text{and} \qquad \boldsymbol{u}_{ik} \sim N\left(\boldsymbol{0}, \boldsymbol{I}_{2\times 2}\right).$$

The participant-specific random effects b_{0i} are independent of the latent random variables $\boldsymbol{u}_{ik}$. The model was fit using the `boral` package in `R` [75].

The plots of the Dunn-Smyth residuals seen in Figure 6.17 were created in order to assess the model fit. Referring to Figure 6.17(a), their is a definite fan shape where the variation of the residuals decreases as the linear predictors increase. In Figure 6.17(b), a random scatter is observed, indicating that the model is appropriate across participant visits. In Figure 6.17(c), all cytokine combinations have very similar residual variation, except for the seventh combination, IFN-$\gamma-$IL2+TNF+IL17$-$, for which the residual variation is much less. Finally, Figure 6.17(d) suggests that the assumption of normally distributed residuals may be violated. Overall, there is evidence that some of the model assumptions may be violated and that the model does not fit the data that well.

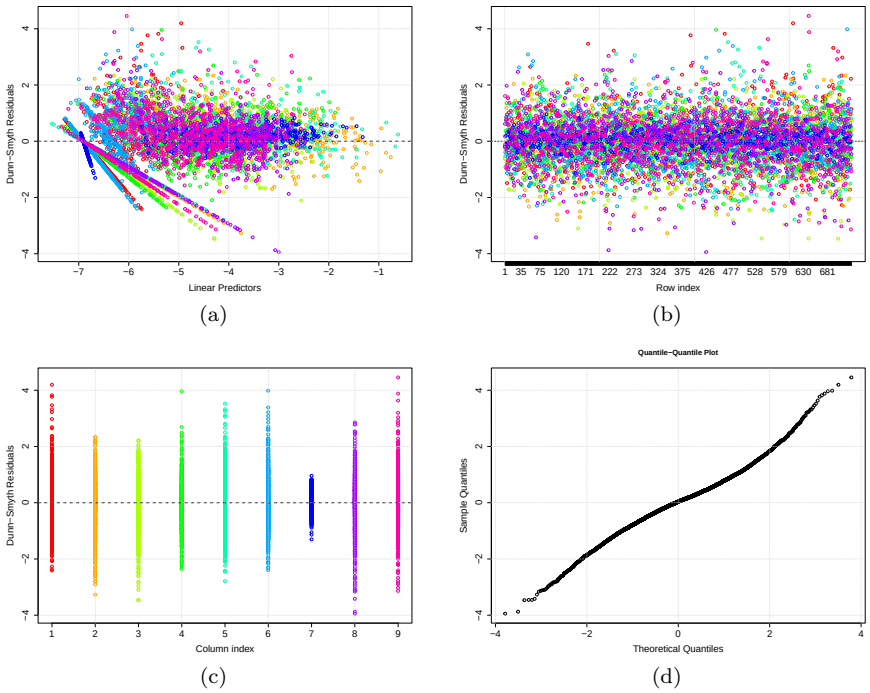
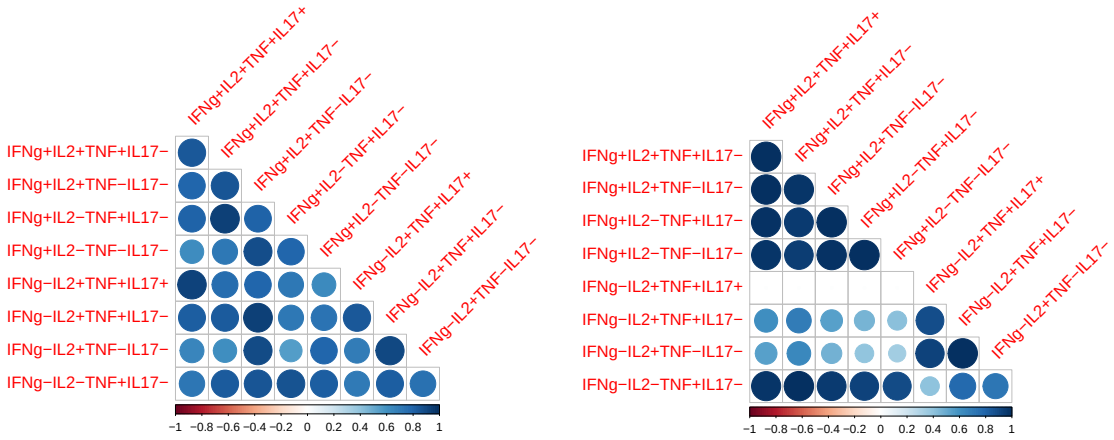


Figure 6.17: Plots for residual analysis of the Ag85B LVM without traits. Each colour represents a different cytokine combination. (a) Dunn– Smyth residuals versus linear predictors, (b) Dunn–Smyth residuals versus row index, (c) Dunn–Smyth residuals versus column index and (d) normal quantile plot of Dunn–Smyth residuals.

Referring to Figure 6.18(a), all correlations between the cytokine combinations due to the model covariates are strong and positive. In other words, these cytokine combinations are strongly and positively correlated due to being observed for the same QFT status, number of administrations, concentrations and at the same time points. Referring to Figure 6.18(b), most of the residual correlations are positive and strong, suggesting that there is a large number of strong correlations between the cytokine combinations even after controlling for covariates. This is likely due to some underlying immunological relationship between the cytokine combinations, since it is clear that many of the cytokine combinations are correlated due to reasons other than QFT status, time point and vaccine regimen. However, for IFN-γ−IL2+TNF+IL17+ there seems to be no significant correlation between it and other IFN-γ-expressing cytokine combinations after controlling for covariates QFT status, time point and vaccine regimen.



(a) Correlation due to covariates

(b) Residual correlations

Figure 6.18: Plots of the correlations between cytokine combinations due to the model covariates (a) and residual correlations (b), based on the correlated response model. Only the significant correlations, as based on the 95% credible intervals excluding zero, have been plotted.

The parameter estimates for the fitted LVM can be found in Tables 6.7, 6.8 and 6.9. Referring to Tables 6.7 and 6.8, the peak time point coefficients are all positive and with confidence intervals that do not contain zero, suggesting that for all cytokine combinations, the frequency that they are expressed is higher at peak compared to baseline. Many of the confidence intervals for the coefficients of the interactions between time point and concentration and time point and QFT status do not include zero, suggesting that perhaps the model could do without these interactions. At baseline and peak, for all cytokine combinations except three, the QFT+ participants have a higher frequency compared to QFT-. For the other cytokine combinations, there is no difference in frequency between QFT+ and QFT- participants. At memory, QFT+ participants have higher frequencies of all cytokine combinations except four, and have lower frequencies of IFN-$\gamma-$IL2+TNF+IL17− and IFN-$\gamma-$IL2+TNF+IL17+. Many of the administration and concentration confidence intervals include zero, which is to be expected since there should not be much difference in immune response between the different treatment arms at baseline, prior to vaccination. After vaccination, the frequency of more than half of the cytokine combinations (all of which express IL-2) increases with increasing numbers of vaccine administrations, except for IFN-$\gamma+$IL2$-$TNF$-$IL17− which decreases. The frequency of IFN-$\gamma+$IL2+TNF+IL17+ , IFN-$\gamma+$IL2+TNF+IL17− and IFN-$\gamma-$IL2+TNF+IL17+ decreases with increasing vaccine concentration, while for all other cytokine combinations there is no change, except for IFN-$\gamma-$IL2+TNF$-$IL17− which increases slightly with increasing concentration.

Referring to Table 6.9, the confidence intervals of the factor loadings for all of the first latent variable and most of the second latent variable do not contain zero, suggesting that there is a lot of residual covariation not explained by the model covariates that the latent variables are accounting for.

Table 6.7: Parameter estimates of the LVM without traits fit to the cytokine combination frequencies after Ag85B stimulation. Values in the brackets correspond to the 95% highest posterior density (credible) intervals.

	Parameter:				
	Peak	Memory	QFT+	Admin.	Conc.
Response j**:**	$\hat{\beta}_{1j}$	$\hat{\beta}_{2j}$	$\hat{\beta}_{3j}$	$\hat{\beta}_{4j}$	$\hat{\beta}_{5j}$
IFN-γ+IL2+TNF+IL17+	**0.398 (0.126, 0.686)**	0.005 (-0.271, 0.270)	0.107 (-0.033, 0.270)	0.057 (-0.035, 0.169)	-0.003 (-0.007, 0.001)
IFN-γ+IL2+TNF+IL17−	**0.988 (0.136, 1.834)**	0.264 (-0.539, 1.161)	**1.137 (0.785, 1.483)**	0.071 (-0.165, 0.311)	**-0.011 (-0.021, -0.001)**
IFN-γ+IL2+TNF−IL17−	**1.156 (0.515, 1.762)**	0.074 (-0.572, 0.726)	**0.532 (0.213, 0.806)**	**-0.384 (-0.593, -0.204)**	-0.003 (-0.010, 0.005)
IFN-γ+IL2−TNF+IL17−	**0.892 (0.143, 1.624)**	-0.133 (-0.843, 0.568)	**0.745 (0.415, 1.073)**	0.027 (-0.211, 0.234)	-0.006 (-0.015, 0.002)
IFN-γ+IL2−TNF−IL17−	**2.062 (1.176, 2.947)**	0.055 (-0.848, 0.925)	**0.488 (0.076, 0.853)**	**-0.441 (-0.704, -0.163)**	0.001 (-0.010, 0.011)
IFN-γ−IL2+TNF+IL17+	**0.401 (0.113, 0.681)**	0.200 (-0.068, 0.496)	0.103 (-0.057, 0.253)	0.010 (-0.096, 0.121)	-0.001 (-0.005, 0.002)
IFN-γ−IL2+TNF+IL17−	**1.593 (1.033, 2.172)**	**1.225 (0.661, 1.863)**	**0.691 (0.423, 0.981)**	-0.064 (-0.250, 0.114)	-0.002 (-0.009, 0.005)
IFN-γ−IL2+TNF−IL17−	**1.782 (1.200, 2.405)**	**0.607 (0.007, 1.277)**	-0.101 (-0.381 ,0.169)	**-0.342 (-0.514, -0.134)**	**0.012 (0.005 ,0.020)**
IFN-γ−IL2−TNF+IL17−	**0.916 (0.282, 1.631)**	0.209 (-0.448, 0.869)	**0.554 (0.251, 0.864)**	0.017 (-0.189, 0.204)	0.000 (-0.008, 0.008)

Table 6.8: Parameter estimates of the LVM without traits fit to the cytokine combination frequencies after Ag85B stimulation continued. Values in the brackets correspond to the 95% highest posterior density (credible) intervals.

	Parameter:					
	Admin*Peak	Admin*Mem	Conc*Peak	Conc*Mem	Peak*QFT+	Mem*QFT+
Response j**:**	$\hat{\beta}_{6j}$	$\hat{\beta}_{7j}$	$\hat{\beta}_{8j}$	$\hat{\beta}_{9j}$	$\hat{\beta}_{10j}$	$\hat{\beta}_{11j}$
IFN-γ+IL2+TNF+IL17+	**0.219 (0.107, 0.339)**	**0.201 (0.090, 0.321)**	**-0.005 (-0.009, -0.001)**	0.000 (-0.005, 0.004)	0.001 (-0.164, 0.189)	-0.069 (-0.252, 0.093)
IFN-γ+IL2+TNF+IL17−	0.295 (-0.091, 0.620)	**0.446 (0.113, 0.799)**	0.003 (-0.011, 0.016)	0.002 (-0.010, 0.017)	-0.346 (-0.844, 0.133)	-0.457 (-0.910, 0.066)
IFN-γ+IL2+TNF−IL17−	**0.459 (0.219, 0.722)**	**0.575 (0.309, 0.839)**	-0.006 (-0.015, 0.004)	0.006 (-0.005, 0.015)	-0.176 (-0.561, 0.198)	-0.194 (-0.563, 0.209)
IFN-γ+IL2−TNF+IL17−	0.124 (-0.190, 0.421)	0.246 (-0.054, 0.519)	-0.004 (-0.016, 0.007)	0.002 (-0.011, 0.012)	-0.095 (-0.512, 0.353)	-0.091 (-0.515, 0.359)
IFN-γ+IL2−TNF−IL17−	0.077 (-0.318, 0.440)	0.244 (-0.146, 0.583)	-0.014 (-0.029, 0.000)	0.002 (-0.012, 0.015)	-0.034 (-0.551, 0.500)	0.283 (-0.236, 0.816)
IFN-γ−IL2+TNF+IL17+	**0.126 (0.005, 0.253)**	0.095 (-0.015, 0.219)	**-0.005 (-0.010, -0.001)**	-0.001 (-0.005, 0.004)	-0.101 (-0.270, 0.058)	**-0.221 (-0.377, -0.053)**
IFN-γ−IL2+TNF+IL17−	**0.440 (0.154, 0.652)**	**0.436 (0.179, 0.672)**	0.003 (-0.006, 0.012)	0.004 (-0.005, 0.014)	**-0.508 (-0.829, -0.143)**	**-0.828 (-1.168, -0.509)**
IFN-γ−IL2+TNF−IL17−	**0.367 (0.113, 0.645)**	**0.430 (0.161, 0.673)**	-0.008 (-0.018, 0.001)	0.002 (-0.009, 0.011)	-0.070 (-0.430, 0.323)	0.187 (-0.155, 0.581)
IFN-γ−IL2−TNF+IL17−	0.034 (-0.228, 0.319)	0.151 (-0.153, 0.393)	0.003 (-0.007, 0.015)	0.001 (-0.010, 0.012)	-0.252 (-0.661, 0.155)	**-0.398 (-0.801, -0.007)**

Table 6.9: Parameter estimates of the LVM without traits fit to the cytokine combination frequencies after Ag85B stimulation continued. Values in the brackets correspond to the 95% highest posterior density (credible) intervals.

	Parameter:			
	Intercept	Factor Loadings		Standard Deviation
Response j**:**	$\hat{\beta}_{0j}$	$\hat{\theta}_{1j}$	$\hat{\theta}_{2j}$	$\hat{\sigma}_j$
IFN-γ+IL2+TNF+IL17+	**-6.781 (-7.022, -6.517)**	**0.085 (0.042, 0.129)**	0.000 (0.000, 0.000)	**0.481 (0.454, 0.511)**
IFN-γ+IL2+TNF+IL17−	**-5.272 (-5.860, -4.680)**	**0.687 (0.418, 1.104)**	**0.814 (0.313, 1.030)**	**0.972 (0.908 ,1.027)**
IFN-γ+IL2+TNF−IL17−	**-5.326 (-5.831, -4.855)**	**0.369 (0.151, 0.747)**	**0.652 (0.382, 0.788)**	**0.756 (0.704, 0.810)**
IFN-γ+IL2−TNF+IL17−	**-5.748 (-6.212, -5.184)**	**0.413 (0.123, 0.977)**	**0.949 (0.592, 1.103)**	**0.701 (0.629, 0.774)**
IFN-γ+IL2−TNF−IL17−	**-5.060 (-5.700, -4.365)**	**0.400 (0.032, 1.056)**	**1.096 (0.724, 1.266)**	**0.997 (0.916, 1.072)**
IFN-γ−IL2+TNF+IL17+	**-6.683 (-6.937, -6.425)**	**0.136 (0.072, 0.196)**	**-0.085 (-0.156, -0.025)**	**0.446 (0.421, 0.474)**
IFN-γ−IL2+TNF+IL17−	**-6.195 (-6.650, -5.773)**	**0.963 (0.802, 1.054)**	0.072 (-0.535, 0.357)	**0.139 (0.004, 0.393)**
IFN-γ−IL2+TNF−IL17−	**-5.443 (-5.902, -4.972)**	**0.361 (0.268, 0.452)**	-0.021 (-0.254, 0.127)	**0.994 (0.944, 1.048)**
IFN-γ−IL2−TNF+IL17−	**-5.876 (-6.366, -5.391)**	**0.583 (0.372, 0.841)**	**0.540 (0.122, 0.742)**	**0.844 (0.798, 0.897)**

Referring to the ordination plots seen in Figure 6.19, all the participant visit are grouped together in a single cluster in the middle of all the cytokine combinations. This indicates that the different variable covariates induce responses across similar cytokine combinations. One cytokine combination, IFN-$\gamma-$IL2$-$TNF$+$IL17$-$, is very close to many of the observations, suggesting that many observations are more associated with the IFN-$\gamma-$IL2$-$TNF$+$IL17$-$ response compared to other cytokine combinations. Comparatively, IFN-$\gamma-$IL2$+$TNF$+$IL17$-$ is located furthest away from the observations, suggesting that the observations are least associated with IFN-$\gamma-$IL2$+$TNF$+$IL17$-$. Cytokine combinations far from the origin and in the same direction are correlated [73]. For example, IFN-$\gamma+$IL2$-$TNF$+$IL17$-$ and IFN-$\gamma+$IL2$-$TNF$-$IL17$-$ are highly correlated with each other, and IFN-$\gamma-$IL2$+$TNF$+$IL17$+$ and IFN-$\gamma+$IL2$+$TNF$+$IL17$+$ are highly correlated with each other and fairly highly correlated with IFN-$\gamma-$IL2$+$TNF$-$IL17$-$ as well.

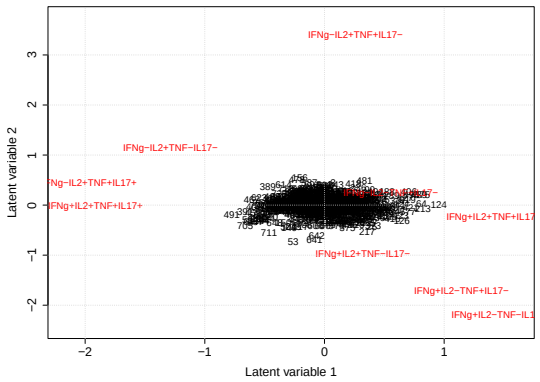


Figure 6.19: Model based unconstrained ordination biplots for the LVM without traits. Each of the participant visits is labelled by its row number, while the 9 cytokine combinations are shown in red.

6.3.2 LVM with traits

In this section, a LVM with traits for the responses IFN-$\gamma+$IL2$+$TNF$+$IL17$+$, ..., IFN-$\gamma-$IL2$-$TNF$+$IL17$-$ was fit. Traits were included in order to help explain the differences in the cytokine combinations' immune responses to the model covariates [74]. Since no cytokine combination-specific variables were available, it was decided to use the cytokines as traits. Since there are four cytokines, each response's trait variables form a 4×1 vector of indicator variables which describe which cytokines are expressed by that response. For example, a the trait vector associated with response IFN-$\gamma+$IL2$-$TNF$+$IL17$-$ is as follows

$$\boldsymbol{T}_{\text{IFN}\gamma+\text{IL2}-\text{TNF}+\text{IL17}-} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

where the first element is 1 indicating that IFN-γ was expressed, the second element is 0 indicating that IL-2 was not expressed, the next element 1 indicates that TNF was expressed

and the final element 0 indicates that IL-17 was not expressed.

Thus, the model specification for the LVM fit with traits is as follows

$$\begin{aligned}\mu_{ijk} = {} & \beta_{0j} + \beta_{1j} I_k^{Peak} + \beta_{2j} I_k^{Mem} + \beta_{3j} I_i^{QFT+} + \beta_{4j} x_i^{Admin} + \beta_{5j} x_i^{Conc} + \\ & \beta_{6j} x_{ik}^{Admin*Peak} + \beta_{7j} x_{ik}^{Admin*Mem} + \beta_{8j} x_{ik}^{Conc*Peak} + \beta_{9j} x_{ik}^{Conc*Mem} + \\ & \beta_{10j} I_{ik}^{Peak*QFT+} + \beta_{11j} I_{ik}^{Mem*QFT+} + \\ & b_{0i} + \theta_{1j} u_{1ik} + \theta_{2j} u_{2ik}, \\ & i = 1, ..., I, \qquad j = 1, ..., 9, \qquad k = 1, 2, 3.\end{aligned}$$

Due to the inclusion of traits in the model, the cytokine combination-specific variable coefficients, β_{wj} where $w = 0, ..., 11$, are now modelled as random effects drawn from the normal distributions

$$\begin{aligned}\beta_{0j} &\sim N\left(\kappa_{0,0} + \boldsymbol{T}_j' \boldsymbol{\kappa}_0, \sigma_0^2\right) \\ \beta_{1j} &\sim N\left(\kappa_{0,1} + \boldsymbol{T}_j' \boldsymbol{\kappa}_1, \sigma_1^2\right) \\ &... \\ \beta_{11j} &\sim N\left(\kappa_{0,11} + \boldsymbol{T}_j' \boldsymbol{\kappa}_1 1, \sigma_1 1^2\right)\end{aligned}$$

where $\kappa_{0,w}$ and $\boldsymbol{\kappa}_w$ are the regression parameters relating the traits to the cytokine combination-specific variable coefficients. $\boldsymbol{\kappa}_w$ is a 4×1 vector formulated as follows

$$\boldsymbol{\kappa}_w = \begin{pmatrix} \kappa_{\text{IFN-}\gamma,w} \\ \kappa_{\text{IL2},w} \\ \kappa_{\text{TNF},w} \\ \kappa_{\text{IL17},w} \end{pmatrix}$$

where $\kappa_{\text{IFN-}\gamma,w}$ is the regression parameter relating trait IFN-γ to variable coefficient β_{wj}, and similarly so for $\kappa_{\text{IL2},w}$, $\kappa_{\text{TNF},w}$ and $\kappa_{\text{IL17},w}$. Note that $\boldsymbol{\kappa}_w$ does not depend on j, the cytokine combination. In other words, $\kappa_{0,w}$ and $\boldsymbol{\kappa}_w$ are shared across the different cytokine combinations.

The reason for including the β coefficients as random effects rather than fixed effects is that it allows the β's to absorb any additional between cytokine combination differences in immune responses not explained by traits [75].

The inclusion of traits in a LVM can be considered a simplification of a LVM without traits [74]. For the LVM without traits, 12×9 response-specific β coefficients had to be estimated, one for each covariate and cytokine combination pair, where there were a total of 12 covariates (including the intercept) and 9 cytokine combinations. For the LVM with traits, this is reduced

to 12×6 parameter estimates, specifically one regression coefficient κ for each covariate and trait combination pair and one σ for each covariate, where there are a total of 12 covariates and 5 traits (including the intercepts).

Referring to Figure 6.20, the diagnostic plots for the LVM with traits are very similar to those seen in Figure 6.17 for the model without traits, and thus the same comments apply. Thus, there is no evidence that the more complicated model (the LVM without traits) fits the data better than the simpler model (the LVM with traits).

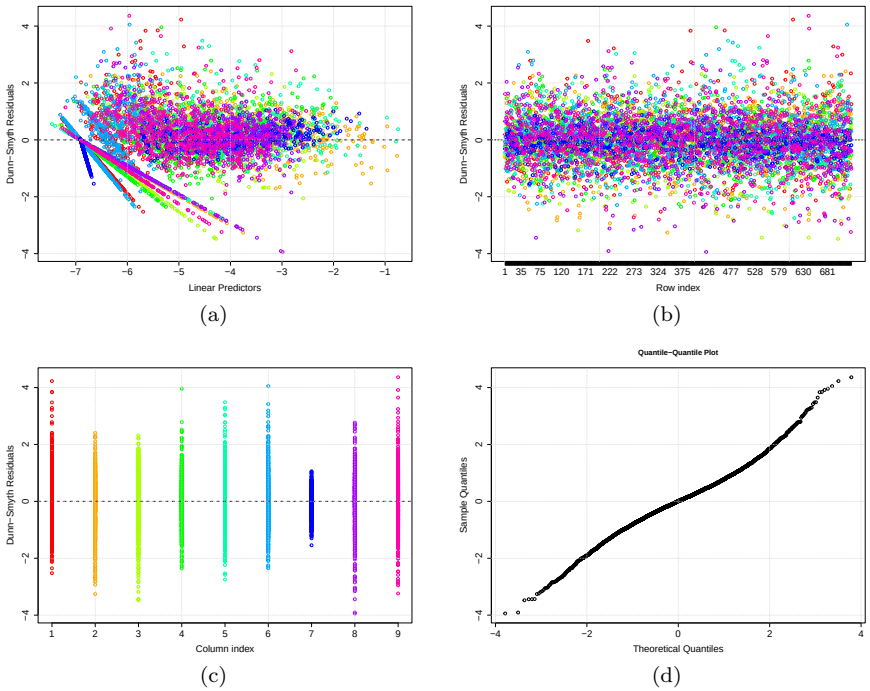


Figure 6.20: Plots for residual analysis of the Ag85B LVM with traits. Each colour represents a different cytokine combination. (a) Dunn– Smyth residuals versus linear predictors, (b) Dunn–Smyth residuals versus row index, (c) Dunn–Smyth residuals versus column index and (d) normal quantile plot of Dunn–Smyth residuals.

Referring to Figure 6.21, the correlation plots are almost identical to those seen in Figure 6.18 for the LVM without traits, and thus the same comments and analysis apply here.

(a) Correlation due to covariates

(b) Residual correlations

Figure 6.21: Plots of the correlations between cytokine combinations due to the model covariates (a) and residual correlations (b), based on the correlated response model. Only the significant correlations, as based on the 95% credible intervals excluding zero, have been plotted.

The parameter estimates for the fitted LVM can be found in Tables 6.10 and 6.11. Referring to Table 6.10, a couple of the confidence intervals for the traits' coefficients do not include zero, suggesting that the inclusion of traits in the model did help explain the differences in the cytokine combinations' immune responses to the model covariates. Notably, the confidence intervals for the coefficients for trait TNF did not contain zero for covariates QFT status, administrations and vaccine concentration as well as two of the interaction variables, suggesting that including TNF as a trait in the model was necessary as it helps to explain the cytokine combinations' different responses to these covariates. However, there is evidence that IFN-γ could be omitted from the model since all of it's coefficients' confidence intervals contained zero. All cytokine combinations are expected to be observed at higher frequencies at peak time point compared to baseline, based on the $\hat{\kappa}_{0,Peak}$. At all time points, all cytokine combinations expressing IL-17 are expected to be observed at lower frequencies in QFT+ participants compared to QFT- participants. However, at baseline and peak, all cytokine combinations expressing TNF but not IL-17 are expected to be observed at higher frequencies in QFT+ participants compared to QFT- participants and at memory all cytokine combinations expressing TNF are expected to be observed at lower frequencies in QFT+ participants compared to QFT- participants. All cytokine combinations that do not express TNF are expected to decrease in frequency with increasing administrations, while all cytokines combinations that do express TNF are expected to increase in frequency, even more so at peak and memory time points if IL-2 is expressed as well. All cytokine combinations, are expected to increase in frequency with increasing concentration, although less so at baseline and memory for combinations expressing TNF.

Referring to Table 6.11, the confidence intervals of the factor loadings for all of the first latent variable and most of the second latent variable do not contain zero, suggesting that much of the residual covariation is not explained by the model covariates that the latent variables are accounting for.

Table 6.10: Parameter estimates of the LVM with traits fit to the cytokine combination frequencies after Ag85B stimulation. Values in the brackets correspond to the 95% highest posterior density (credible) intervals.

Variable	$\hat{\beta}_{wj}$	$\hat{\kappa}_{0,w}$	$\hat{\kappa}_{\text{IFN-}\gamma,w}$	$\hat{\kappa}_{\text{IL2},w}$	$\hat{\kappa}_{\text{TNF},w}$	$\hat{\kappa}_{\text{IL17},w}$	$\hat{\sigma}_w$
Intercept	β_{0j}	**-5.512 (-6.534, -4.262)**	0.181 (-0.605, 1.084)	-0.064 (-0.937, 0.805)	-0.536 (-1.505, 0.295)	-0.750 (-1.752, 0.488)	**0.401 (0.142, 1.044)**
Peak	β_{1j}	**1.879 (0.699, 3.116)**	-0.210 (-1.066, 0.713)	-0.038 (-0.981, 0.914)	-0.551 (-1.384, 0.571)	-0.739 (-1.838, 0.470)	**0.459 (0.142, 1.153)**
Memory	β_{2j}	0.346 (-0.766, 1.423)	-0.424 (-1.126, 0.299)	0.433 (-0.407, 1.254)	0.161 (-0.725, 0.985)	-0.612 (-1.717, 0.441)	**0.360 (0.117, 0.913)**
QFT+	β_{3j}	-0.022 (-0.494, 0.598)	0.313 (-0.116, 0.749)	0.173 (-0.230, 0.606)	**0.532 (0.119, 1.016)**	**-0.728 (-1.298, -0.270)**	**0.193 (0.038, 0.525)**
Admin.	β_{4j}	**-0.369 (-0.617, -0.121)**	0.047 (-0.105, 0.177)	0.005 (-0.165, 0.179)	**0.410 (0.236, 0.595)**	-0.024 (-0.241, 0.198)	0.042 (0.000, 0.158)
Conc.	β_{5j}	**0.012 (0.001, 0.023)**	-0.007 (-0.014, 0.000)	-0.003 (-0.011, 0.006)	**-0.009 (-0.018, -0.001)**	0.001 (-0.008, 0.012)	0.003 (0.000, 0.009)
Admin*Peak	β_{6j}	0.028 (-0.344, 0.346)	0.043 (-0.162, 0.253)	**0.326 (0.081, 0.586)**	-0.012 (-0.304, 0.203)	-0.210 (-0.505, 0.098)	0.059 (0.000, 0.230)
Admin*Mem	β_{7j}	0.161 (-0.137, 0.441)	0.081 (-0.086, 0.245)	**0.273 (0.043, 0.472)**	-0.047 (-0.270, 0.144)	-0.285 (-0.530, -0.023)	0.033 (0.000, 0.125)
Conc*Peak	β_{8j}	-0.012 (-0.022, 0.001)	-0.002 (-0.011, 0.005)	0.004 (-0.005, 0.012)	**0.009 (0.001, 0.018)**	-0.006 (-0.016, 0.006)	0.002 (0.000, 0.009)
Conc*Mem	β_{9j}	0.000 (-0.011, 0.013)	0.000 (-0.008, 0.007)	0.003 (-0.006, 0.011)	-0.001 (-0.010, 0.009)	-0.002 (-0.013, 0.009)	0.002 (0.000, 0.008)
Peak*QFT+	β_{10j}	0.000 (-0.431, 0.434)	0.099 (-0.137, 0.390)	-0.231 (-0.533, 0.107)	-0.219 (-0.519, 0.082)	0.349 (-0.012, 0.742)	0.060 (0.000, 0.240)
Mem*QFT+	β_{11j}	0.268 (-0.270, 0.781)	0.147 (-0.212, 0.479)	-0.407 (-0.813, 0.054)	**-0.557 (-0.959, -0.161)**	0.471 (-0.056, 0.971)	0.124 (0.000, 0.443)

Table 6.11: Parameter estimates of the LVM with traits fit to the cytokine combination frequencies after Ag85B stimulation continued. Values in the brackets correspond to the 95% highest posterior density (credible) intervals.

	Parameter:		
	Factor Loadings		Standard Deviation
Response j:	$\hat{\theta}_{1j}$	$\hat{\theta}_{2j}$	$\hat{\sigma}_j$
IFN-γ+IL2+TNF+IL17+	**0.087 (0.050, 0.133)**	0.000 (0.000, 0.000)	**0.481 (0.452, 0.509)**
IFN-γ+IL2+TNF+IL17−	**0.922 (0.612, 1.156)**	**0.544 (0.065, 0.906)**	**0.968 (0.911, 1.035)**
IFN-γ+IL2+TNF−IL17−	**0.567 (0.312, 0.803)**	**0.510 (0.197, 0.735)**	**0.755 (0.703, 0.805)**
IFN-γ+IL2−TNF+IL17−	**0.717 (0.349, 1.063)**	**0.777 (0.330, 1.014)**	**0.703 (0.634, 0.775)**
IFN-γ+IL2−TNF−IL17−	**0.753 (0.326, 1.178)**	**0.911 (0.447, 1.195)**	**0.996 (0.920, 1.076)**
IFN-γ−IL2+TNF+IL17+	**0.109 (0.038, 0.173)**	**-0.112 (-0.170, -0.057)**	**0.446 (0.421, 0.472)**
IFN-γ−IL2+TNF+IL17−	**0.940 (0.726, 1.048)**	-0.266 (-0.688, 0.134)	**0.166 (0.014, 0.343)**
IFN-γ−IL2+TNF−IL17−	**0.342 (0.229, 0.437)**	-0.130 (-0.314, 0.029)	**0.994 (0.947, 1.051)**
IFN-γ−IL2−TNF+IL17−	**0.729 (0.533, 0.874)**	0.323 (-0.077, 0.600)	**0.845 (0.797, 0.893)**

Referring to Figure 6.22, the correlation plots are almost identical to those seen in Figure 6.19 for the LVM without traits, and thus the same comments and analysis apply here.

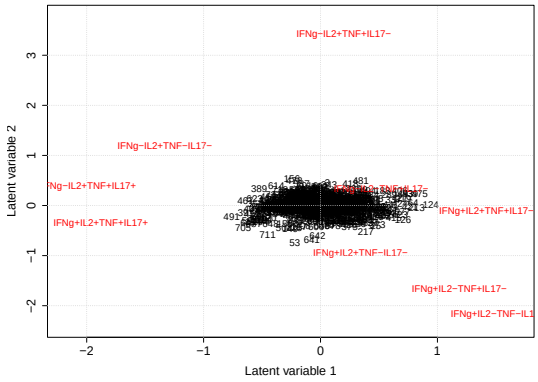


Figure 6.22: Model based unconstrained ordination biplots for the LVM with traits. Each of the participant visits is labelled by its row number, while the 9 cytokine combinations are shown in red.

6.3.3 Comment on the LVM's

Based on the the Dunn-Smyth residuals, there is evidence that both LVM's violated some of the model assumptions, specifically homogeneity and normality of the residuals, and thus should not be relied upon for inference. However, for the sake of illustration the results in terms of the objectives of this chapter shall be discussed with the caveat that they should be interpreted with caution.

For the LVM without traits, there was evidence that QFT+ participants had higher frequencies of many of the cytokine combinations compared to QFT- participants, thus supporting the hypothesis that the functional profiles of T cells recognising Ag85B differed between QFT+ and QFT- individuals. However, there was evidence against the hypothesis that QFT+ participants had higher frequencies of T cells expressing IL-2 and TNF together compared to QFT- participants. For T cells expressing IL-2, TNF and IFN-γ together, there was evidence that frequencies were higher in QFT+ participants only when IL-17 was not expressed as well. Most cytokine combinations increased in frequency with increasing number of administrations while only three cytokine combinations, all with high functionality, decreased in frequency with increasing vaccine concentration, while all other frequencies remained unaffected. This could suggest that three administrations of 5 µg is the optimal regimen.

For the LVM with traits, the effect of administration on the frequency of the cytokine combinations depended on their expression of TNF, while increasing concentration seemed to always result in an increase in frequency. Thus, 50 µg would be the optimal vaccine concentration but the number of administrations is inconclusive, although specialist knowledge about the individual cytokine combinations could change this. There was evidence supporting the hypothesis that QFT+ participants had higher frequencies of T cells expressing IL-2 and TNF together and T cells expressing IL-2, TNF and IFN-γ together compared to QFT- participants only at baseline and peak time points and only if IL-17 was not expressed. This relationship was mainly

dependant on the expression of TNF.

The LVM with traits allowed one to make inference more specific to the individual cytokines rather than the cytokine combinations. This made it more difficult to make inference about the overall effect of covariates on the functional quality of the response and thus made it more difficult to identify an optimal vaccine regimen. However, the inclusion of traits did have the advantage of simplifying the model without worsening the model fit. Thus, the choice of whether to use a LVM with traits or without traits mainly depends on the type of inference the investigator would like to make. If the focus is on the effect of the expression of individual cytokines on the relationship between the cytokine combinations and the the covariates, then a model with traits would be preferable. However if they are more interested in investigating the relationship between specific cytokine combinations and the covariates, then a model without traits is preferable. Note that a LVM with traits can be used to make inference about cytokine combinations, it just may be more difficult to interpret.

6.4 Discussion

Since the Ag85B GLMEM and both LVM's did not fit the data adequately, only the inference from the ESAT-6 GLMEM shall be compared to that of the standard approach. Both the ESAT-6 GLMEM and the standard approach provided evidence suggesting that 5 µg was the optimal vaccine concentration, however they disagreed on the optimal number of administrations. The standard approach provided evidence that higher quality immune responses are associated with lower concentrations of the vaccine in QFT+ individuals compared to QFT- individuals but the GLMEM provided evidence against this. Both the GLMEM and the standard approach showed evidence that QFT+ participants had higher quality immune responses compared to QFT- individuals. Both the standard approach and the GLMEM provided evidence that QFT+ individuals had higher frequencies of cells expressing IFN-γ+IL2+TNF+ after stimulation with ESAT-6 compared to QFT- participants, but the GLMEM and standard approach disagreed on whether IFN-$\gamma-$IL2+TNF+ was differentially expressed. Because the ESAT-6 GLMEM only used ESAT-6 stimulated observations, no comparisons could be made between stimuli.

As discussed in Section 5.3, a major disadvantage of the hypothesis testing used in the standard approach as well as the fitting of multiple models in the mixed effect modelling approach is the multiple testing which prevents one from making any strong inference from the results. The effects of multiple testing can be minimised though not eliminated in the case of the mixed effect models by fitting one model for each stimulus, either a GLMEM or a LVM, and focusing more on the model building procedure. Note that the use of COMPASS adds another source of multiple testing since it is used to select cytokine combinations that are differentially expressed, which can be considered a form of significance testing. In the case of the standard approach, again the effects can be minimised by formulating a clear statistical analysis plan and focusing on one or two hypotheses of interest. Thus all inference made in this chapter should be interpreted with caution and be used for exploratory analysis rather than confirmatory due to the issue of

multiple testing.

Again, the main advantage that both GLMEM's and LVM's have is that they provide effect sizes in the form of variable coefficients, where as significance tests focus more on hypothesis testing and p-values. Also, LVM's are able to account for the multivariate nature of the data and take into consideration correlations between different response variables.

In terms of which response variable to use, it depends on the aim of the investigator. If they are more interested in identifying which subsets have higher quality immune responses compared to others, then the FS and PFS are more appropriate. If they wish to make inference about specific cytokine combinations and how their expression is affected by different covariates, then the cytokine combination frequency should be used. Note that while both response variables can be used for all approaches, it would probably not be necessary to use LVM's for the FS and PFS response, as it is very low dimensional and the relationships between covariates and the FS and the PFS probably do not vary substantially.

As to whether a GLMEM or LVM is more appropriate when modelling the cytokine combinations frequencies, it depends very much on the relationship between the different covariates and the response variables. In the unlikely case that relationship between covariates and cytokine combinations is the same across all cytokine combinations, then it is suggested to rather model the cytokine combinations frequencies using GLMEM's as it requires less parameters, as was seen for ESAT-6 GLMEM. Otherwise, it is recommended that LVM's should be used as it makes the interpretation easier due to the large number of interaction terms in the GLMEM.

Key to both approaches in this chapter was the use of COMPASS for dimension reduction, whether that be by providing summary statistics FS and PFS, or by identifying biologically meaningful cytokine combinations. A disadvantage of using COMPASS is that no diagnostic checks are available, meaning that it is not possible to check model assumptions. While COMPASS does provide confidence intervals, p-values and q-values, it is not possible to represent these when visualising the FS and PFS as was done in this chapter, meaning that the reader is not provided with information as to the confidence of these estimates.

Chapter 7

Investigating the effects of vaccine regimen, *M.tb* sensitisation and antigen specificity on CD4 T cell *differentiation*

In this chapter, the third aim of this project, that is the effects of vaccine regimen, *M.tb* sensitisation and antigen specificity on CD4 T cell *differentiation*, shall be investigated. Along with exploring this aim, this chapter shall also investigate the hypothesis that a high vaccine concentration in *M.tb*-sensitised individuals drives T cells towards a more differentiated phenotype compared to unsensitised individuals. Another hypothesis of interest is that the differentiation profiles of T cells recognising Ag85B are distinct from those of T cells recognising ESAT-6 and that after vaccination, the differentiation profiles of T cells recognising Ag85B or ESAT-6 differ between *M.tb*-sensitised and *M.tb*-unsensitised individuals.

As discussed in Section 1.2, T cell differentiation depends on the specific combination of memory markers CCR7 and CD45RA expressed by the cell. Naive T cells are defined as CCR7+CD45RA+ and are the least differentiated T cells, followed by T_{CM} cells which are defined as CCR7+CD45RA−, then T_{EM} cells which are defined as CCR7−CD45RA− and finally the most differentiated T cells, T_{TE} cells which are defined as CCR7−CD45RA+.

In this chapter, two different response variables will be considered: the magnitude of the antigen-specific immune response and the qualitative antigen-specific immune response. Both responses shall be assessed using the standard approach of box-plots and Mann Whitney tests with Benjamini Hochberg adjusted p-values for multiple testing, but only the magnitude of the antigen-specific immune response shall be assessed using GLMEM's.

Numerically the magnitude of the antigen-specific immune response can be defined as follows

$$Y_{ijk}^{(s)} = \frac{n_{ijk^+}^{(s)}}{N_{ij}^{(s)}} - \frac{n_{ijk^+}^{(u)}}{N_{ij}^{(u)}}, \tag{7.1}$$

where $n_{ijk^+}^{(s)}$ and $n_{ijk^+}^{(u)}$ are the number of cytokine positive cells observed for subject i which express marker combination k^+ at time point j when stimulated with s and when unstimulated respectively. $N_{ij}^{(s)}$ and $N_{ij}^{(u)}$ are the total number of *CD4 T cells* observed for subject i at visit j when stimulated with s and when unstimulated respectively. Note that the markers forming the marker combinations k^+ considered here are IL-2, TNF, IFN-γ, CCR7 and CD45RA. Thus, cytokine IL-17 is ignored as its expression does not impact the differentiation profile of the cell.

Numerically the qualitative antigen-specific immune response can be defined as follows

$$Y_{ijm}^{(s)} = \frac{n_{ijm^+}^{(s)}}{N_{ij+}^{(s)}}, \tag{7.2}$$

where $n_{ijm^+}^{(s)}$ is the number of cytokine positive cells observed for subject i which express memory combination m at time point j when stimulated with s and $N_{ij+}^{(s)}$ is the number of *cytokine positive* cells observed for subject i at visit j.

Generally, upon antigen-stimulation a cell is only considered antigen-specific (i.e. it recognises the antigen) if it expresses at least one cytokine above the background levels measured in the unstimulated sample. To assess the quality of the antigen-specific immune response, one looks at the frequencies of different memory combinations. Thus, even though this is a numerical value, it is considered a qualitative measure because it describes the differentiation attributes of antigen-specific cells.

However, the qualitative antigen-specific immune response is only antigen-specific and hence meaningful when it is calculated for observations classified as responders. This response cannot simply be background subtracted because when a sample is unstimulated, there should be no antigen-specific response because no antigen stimulation took place, and thus $n_{ij}^{(u)} \approx 0$ meaning that the response cannot be calculated for the unstimulated case. This is why it is necessary to filter out non-responding observations so that the qualitative immune response is antigen-specific. Responders shall again be classified using MIMOSA as outlined in Section 4.2.2.

Section 7.1 shall employ the standard approach of box-plots and Mann Whitney tests to assess the magnitude of the antigen-specific immune response in Section 7.1.1, and the qualitative antigen-specific immune response in Section 7.1.2. Section 7.2 shall then fit single-level, two-level and three-level GLMEM's to the magnitude of the antigen-specific immune response and compare the results. Section 7.3 shall then discuss the similarities and differences between the standard approach and the GLMEM approach in terms of inference made as well as general interpretability and ease of application.

7.1 Standard approach

In this section, both the magnitude of the antigen-specific immune response and the qualitative antigen-specific immune response shall be assessed in order to investigate the aim and hypotheses of this chapter. First, the hypothesis that higher vaccine concentrations in *M.tb*-sensitised individuals drives T cells towards a more differentiated phenotype compared to unsensitised individuals shall be investigated. Secondly, the hypothesis that after vaccination, differentiation profiles differ between *M.tb*-sensitised and *M.tb*-unsensitised individuals shall be investigated. Finally, the hypothesis that differentiation profiles differ between ESAT-6 stimulation and Ag85B stimulation shall be investigated. In order to minimise the number of plots that

need to be created and to avoid unnecessary repetition, the last two hypotheses shall be investigated using only observations from participants who received two administrations of 5 µg of the vaccine, which was the optimal regimen determined by the standard approach in Sections 5.1 and 6.1 for both T cell magnitude and T cell functional quality. For the same reason, the first aim will only use observations from participants who received two administrations of the vaccine.

7.1.1 Magnitude of the antigen-specific immune response

The plots in Figure 7.1 were created to investigate the hypothesis that higher vaccine concentration in QFT+ individuals drives T cells towards more differentiated phenotypes compared to QFT- individuals. After Ag85B stimulation, there was evidence that QFT- and QFT+ participants only differed in terms of frequency of T_{CM} cells expressing IL-2 when administered 5 µg, where higher frequencies were observed in QFT- participants. After ESAT-6 stimulation for samples from participants who received 5 µg of the vaccine, the frequencies of almost all marker combinations were higher in QFT+ participants. After ESAT-6 stimulation for samples from participants who received 15 µg f the vaccine as well as those who received 50 µg of the vaccine, the frequencies of more highly differentiated cells was higher in QFT+ participants compared to QFT- participants, while there is mostly no difference in the frequencies of early differentiated cells. Overall, there does seem to be evidence supporting the hypothesis that higher concentrations in QFT+ individuals drives T cells towards more differentiated phenotypes compared to QFT- individuals. However, this is only seen after ESAT-6 stimulation, not Ag85B stimulation.

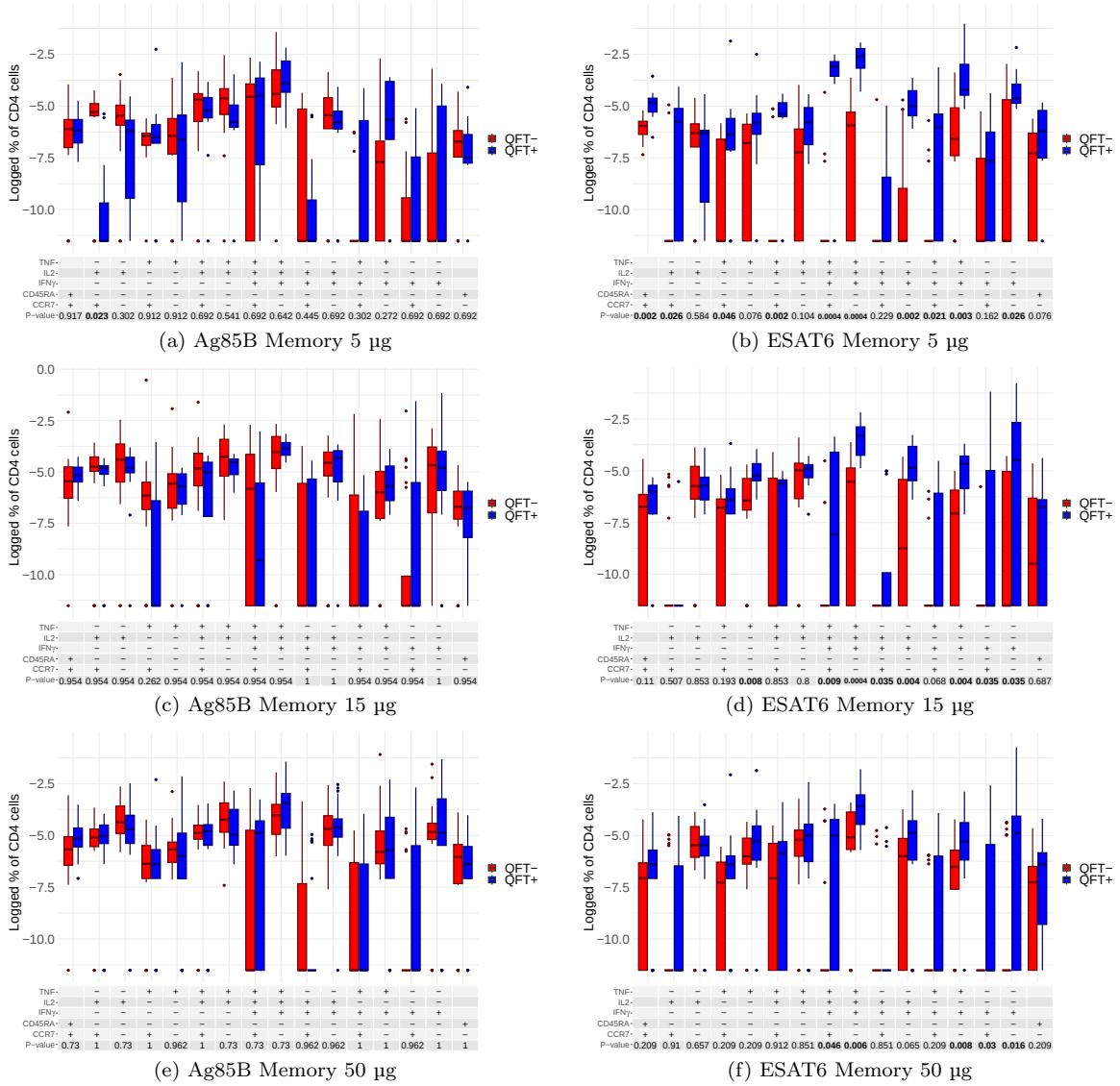
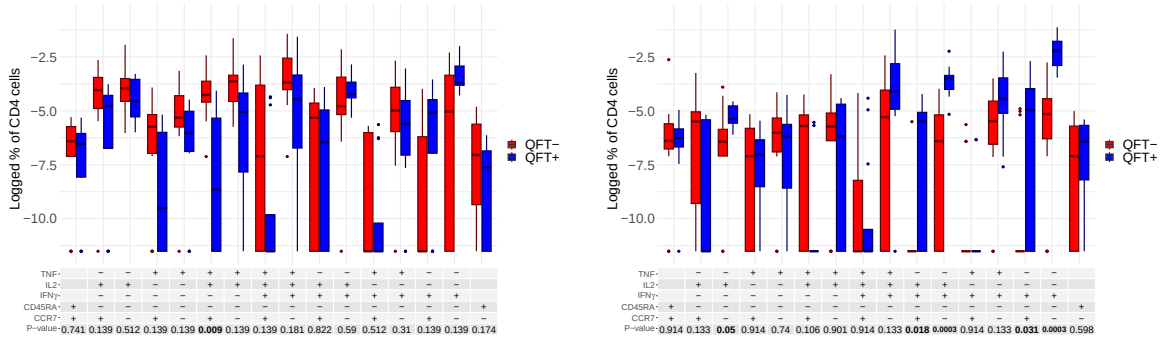


Figure 7.1: Box-plots of the logged background-subtracted frequencies of ESAT-6-stimulated CD4 T cells which express a certain combination of cytokines and memory markers at the memory time point. These frequencies are from all participants who received two administrations of the vaccine. Box-plots are plotted for different QFT statuses as well as for different stimuli. Benjamini Hochberg adjusted p-values, determined by Mann Whitney tests, are shown in the tables and compare the QFT- samples to the QFT+ samples. CCR7+CD45RA+ denotes cells that express at least one cytokine (IFN-γ, IL-2 and/or TNF) and express both memory markers CCR7 and CD45RA. Similarly, CCR7−CD45RA+ are cells that express at least one cytokine and express memory marker CD45RA but not CCR7.

Using only observations from participants who received two administrations of 5 µg of the vaccine, Figure 7.2 in conjunction with Figure 7.1 (a) and (b) can be used to investigate the hypothesis that after vaccination, differentiation profiles of T cells recognising Ag85B or ESAT-6 differ between QFT+ and QFT- individuals. As already seen in Figure 7.1, there is evidence that differentiation profiles of T cells recognising ESAT-6 differed between QFT+ and QFT- individuals. After Ag85B stimulation, the only difference seen was for the frequency of T_{CM} cells expressing IL-2 only, where frequencies were higher in QFT- participants compared to QFT+ participants. From Figure 7.2, there is evidence that at the peak time point, differentiation profiles of T cells recognising ESAT-6 differed between QFT+ and QFT- individuals, with

QFT+ participants having higher frequencies of T_{EM} cells expressing IL-2 only, T_{CM} and T_{EM} cells expressing IFN-γ but not TNF. After Ag85B stimulation, the only difference seen at the peak time point was for the frequency of T_{CM} cells expressing TNF and IL-2 but not IFN-γ, where frequencies were significantly higher in QFT- participants compared to QFT+ participants. Thus, there is evidence that after vaccination differentiation profiles of T cells recognising ESAT-6 differ between QFT+ and QFT- individuals, however there is limited evidence that this difference exists for Ag85B-recognising T cells.

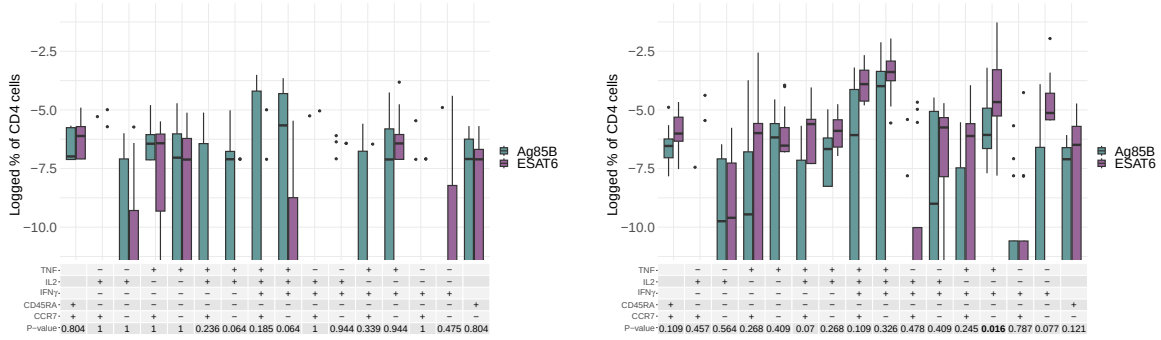


(a) Ag85B Peak

(b) ESAT6 Peak

Figure 7.2: Box-plots of the logged background-subtracted frequencies of ESAT-6-stimulated CD4 T cells which express a certain combination of cytokines and memory markers at the peak time point. These frequencies are from all participants who received two administrations of 5 µg. Box-plots are plotted for different QFT statuses as well as for both stimuli. Benjamini Hochberg adjusted p-values, determined by Mann Whitney tests, are shown in the tables and compare the QFT- samples to the QFT+ samples. CCR7+CD45RA+ denotes cells that express at least one cytokine (IFN-γ, IL-2 and/or TNF) and express both memory markers CCR7 and CD45RA. Similarly, CCR7−CD45RA+ are cells that express at least one cytokine and express memory marker CD45RA but not CCR7.

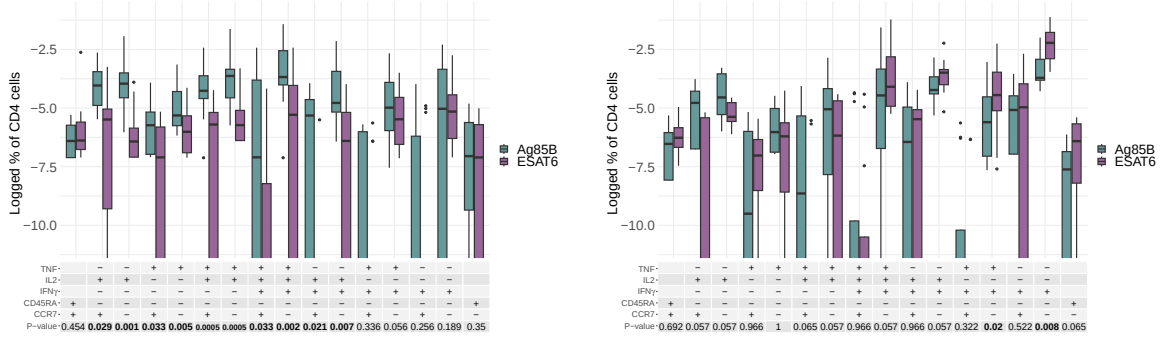
Using only observations from participants who received two administrations of 5 µg of the vaccine, Figures 7.3, 7.4 and 7.5 were created to investigate the hypothesis that the differentiation profiles of cells recognising Ag85B are distinct from those recognising ESAT6. Referring to Figure 7.3, there was no evidence that a difference existed between Ag85B and ESAT-6 at baseline for any of the marker combinations, except for T_{EM} cells expressing only IFN-γ and TNF in QFT+ participants. Specifically, for QFT+ participants, the frequency of T_{EM} cells expressing only IFN-γ and TNF is lower after Ag85B stimulation compared to after ESAT-6 stimulation. Generally, the frequencies after ESAT-6 stimulation were quite low for QFT- participants but not for QFT+ participants. Thus at baseline, there is little evidence that the differentiation profiles of cells recognising Ag85B are distinct from those recognising ESAT-6.



(a) QFT- Baseline Response (b) QFT+ Baseline Response

Figure 7.3: Box-plots of the logged background-subtracted frequencies of ESAT-6-stimulated CD4 T cells which express a certain combination of cytokines and memory markers at baseline. These frequencies are from all participants who received two administrations of 5 µg. Box-plots are plotted for different QFT statuses as well as for different stimuli. Benjamini Hochberg adjusted p-values, determined by Mann Whitney tests, are shown in the tables and compare the Ag85B samples to the ESAT-6 samples. CCR7+CD45RA+ denotes cells that express at least one cytokine (IFN-γ, IL-2 and/or TNF) and express both memory markers CCR7 and CD45RA. Similarly, CCR7−CD45RA+ are cells that express at least one cytokine and express memory marker CD45RA but not CCR7.

Referring to Figure 7.4, there is evidence that frequencies were lower after ESAT-6 stimulation compared to Ag85B stimulation for many early differentiated marker combinations for QFT- participants. Specifically, differences were seen between all T_{CM} and T_{EM} cells expressing IL-2 as well as T_{CM} and T_{EM} cells expressing only TNF. For QFT+ participants, higher frequencies were observed after ESAT-6 stimulation compared to Ag85B stimulation for only two marker combinations, both of which are highly differentiated. Generally, ESAT-6 frequencies are always lower compared to Ag85B frequencies in QFT- participants, while for QFT+ participants this relationship is reversed. Thus at peak, there is evidence that the differentiation profiles of cells recognising Ag85B are distinct from those recognising ESAT-6, more so in QFT- participants compared to QFT+ participants.

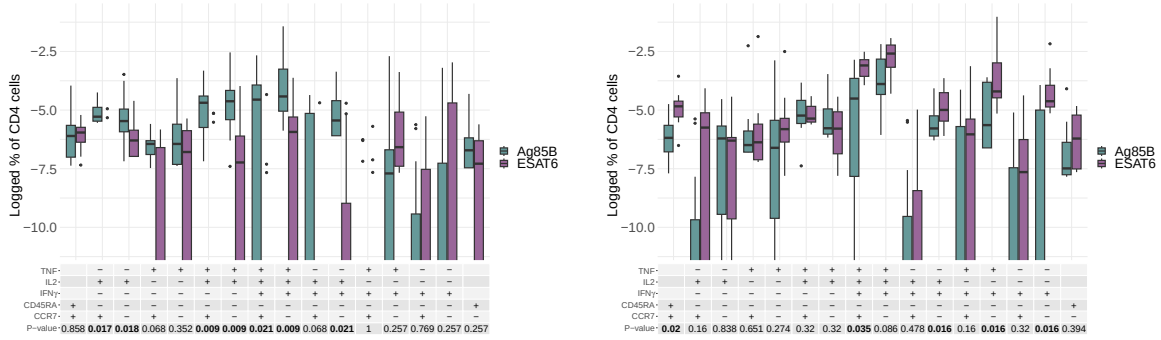


(a) QFT- Peak Response (b) QFT+ Peak Response

Figure 7.4: Box-plots of the logged background-subtracted frequencies of ESAT-6-stimulated CD4 T cells which express a certain combination of cytokines and memory markers at peak time point. These frequencies are from all participants who received two administrations of 5 µg. Box-plots are plotted for different QFT statuses as well as for different stimuli. Benjamini Hochberg adjusted p-values, determined by Mann Whitney tests, are shown in the tables and compare the Ag85B samples to the ESAT-6 samples. CCR7+CD45RA+ denotes cells that express at least one cytokine (IFN-γ, IL-2 and/or TNF) and express both memory markers CCR7 and CD45RA. Similarly, CCR7−CD45RA+ are cells that express at least one cytokine and express memory marker CD45RA but not CCR7.

Referring to Figure 7.5, there is evidence that for QFT- participants, the frequencies of almost all T_{CM} and T_{EM} cells expressing IL-2 is still higher after Ag85B stimulation compared to ESAT-6 stimulation. However, the frequency of T_{EM} cells expressing only TNF is no longer different between stimuli. For QFT+ participants, the frequencies of T_{EM} cells expressing IFN-γ and either IL-2 or TNF (but not both), T_{CM} cells expressing all three cytokines as well as naive T cells are higher after ESAT-6 stimulation compared to Ag85B stimulation. Generally, ESAT-6 frequencies are always lower compared to Ag85B frequencies in QFT- participants, while for QFT+ participants the opposite is true. Thus at memory, there is evidence that the differentiation profiles of cells recognising Ag85B are distinct from those recognising ESAT-6, more so in QFT- participants compared to QFT+ participants.

In conclusion, post-vaccination there is evidence that the differentiation profiles of cells recognising Ag85B are distinct from those recognising ESAT-6, more so in QFT- participants compared to QFT+ participants. However, pre-vaccination there is little evidence that a difference exists.



(a) QFT- Memory Response

(b) QFT+ Memory Response

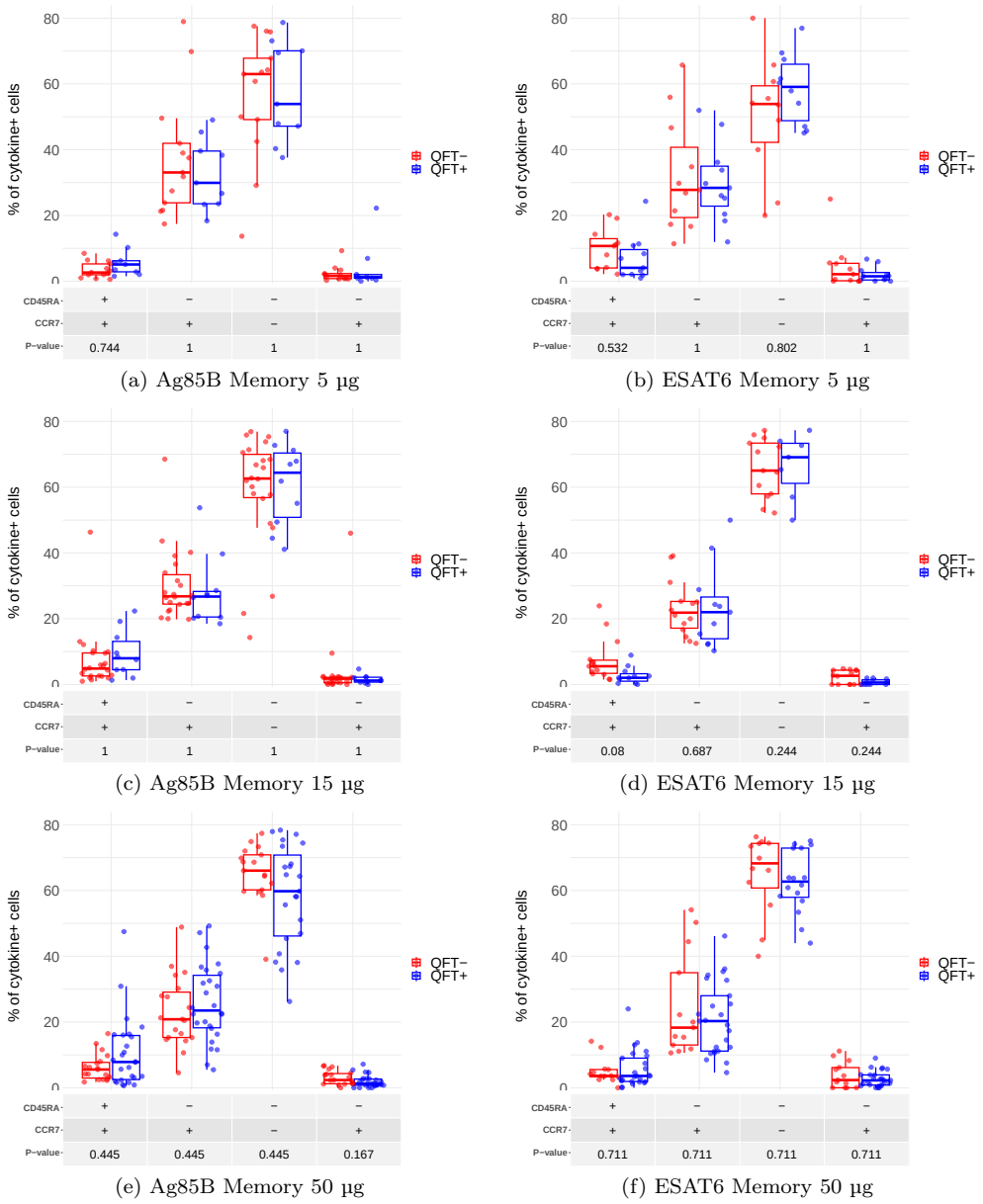
Figure 7.5: Box-plots of the logged background-subtracted frequencies of ESAT-6-stimulated CD4 T cells which express a certain combination of cytokines and memory markers at memory. These frequencies are from all participants who received two administrations of 5 µg. Box-plots are plotted for different QFT statuses as well as for different stimuli. Benjamini Hochberg adjusted p-values, determined by Mann Whitney tests, are shown in the tables and compare the Ag85B samples to the ESAT-6 samples. CCR7+CD45RA+ denotes cells that express at least one cytokine (IFN-γ, IL-2 and/or TNF) and express both memory markers CCR7 and CD45RA. Similarly, CCR7−CD45RA+ are cells that express at least one cytokine and express memory marker CD45RA but not CCR7.

7.1.2 Qualitative antigen-specific immune response

This section shall investigated the same hypotheses as the preceding section, however now the qualitative antigen-specific immune response shall be investigated rather than the magnitude of the antigen-specific immune response. Since the qualitative antigen-specific immune response is being used, the plots shall be subdivided by memory combination rather than marker combination.

The plots in Figure 7.6 were created to investigate the hypothesis that higher concentrations in QFT+ individuals drives T cells towards more differentiated phenotypes compared to QFT- individuals. There is no evidence of that a difference in frequency exists between QFT- and QFT+ participants for any of the memory combinations. For all plots, the frequency of T_{EM} cells is highest, followed by T_{CM} cells, then naive cells and finally terminally differentiated

cells. Thus, there is no evidence that higher concentrations in QFT+ individuals drives T cells towards more differentiated phenotypes compared to QFT- individuals.



(a) Ag85B Memory 5 µg

(b) ESAT6 Memory 5 µg

(c) Ag85B Memory 15 µg

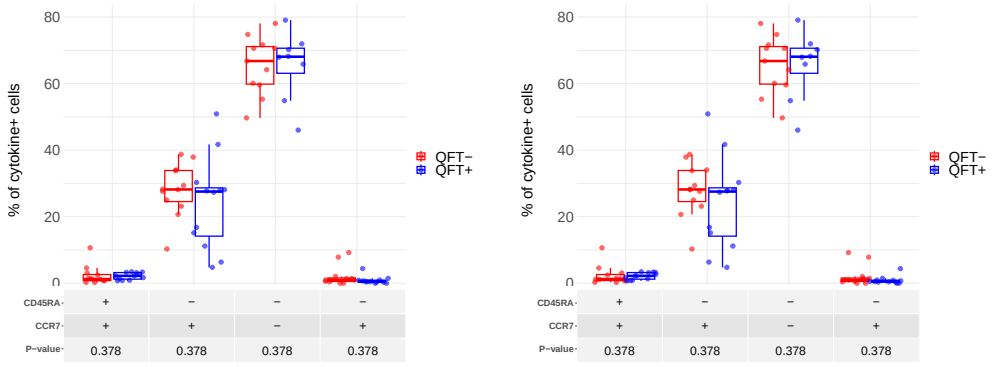
(d) ESAT6 Memory 15 µg

(e) Ag85B Memory 50 µg

(f) ESAT6 Memory 50 µg

Figure 7.6: Box-plots of the frequencies of cytokine positive CD4 T cells which are cytokine positive and express a certain memory combination at the memory time point. These frequencies are from all participants who received 2 administrations of the vaccine. Box-plots are plotted for different QFT statuses as well as for different stimuli. Benjamini Hochberg adjusted p-values, determined by Mann Whitney tests, are shown in the tables and compare the QFT- samples to the QFT+ samples.

The below plots seen in Figure 7.7 in conjunction with Figure 7.6 (a) and (b) can be used to investigate the hypothesis that after vaccination, differentiation profiles of T cells recognising Ag85B or ESAT-6 differ between QFT+ and QFT- individuals. For all four plots, there is no evidence that a difference in the frequencies of any of the memory combinations exists between QFT- and QFT+ participants. Thus, there is no evidence that after vaccination, differentiation profiles of T cells recognising Ag85B or ESAT-6 differ between QFT+ and QFT- individuals.

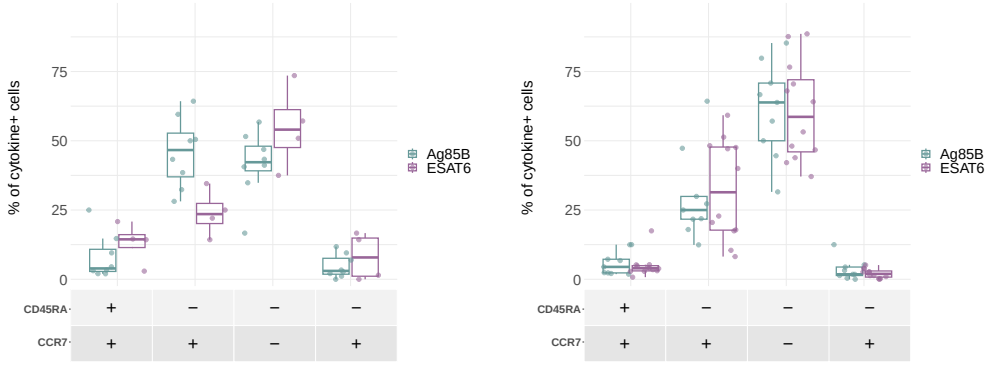


(a) Ag85B Peak (b) ESAT6 Peak

Figure 7.7: Box-plots of the frequencies of cytokine positive CD4 T cells which are cytokine positive and express a certain memory combination at the peak time point. These frequencies are from all participants who received 2 administrations of 5 µg. Box-plots are plotted for different QFT statuses as well as for both stimuli. Benjamini Hochberg adjusted p-values, determined by Mann Whitney tests, are shown in the tables and compare the QFT- samples to the QFT+ samples.

The below plots in Figures 7.8, 7.9 and 7.10 were created to investigate the hypothesis that the differentiation profiles of cells recognising Ag85B are distinct from those recognising ESAT-6. Note that because of the filtering out of non-responding observations, the number of observations available for ESAT-6 was often very different to the number available for Ag85B, and hence a paired Wilcoxon test could not be conducted and no p-values could be calculated. Instead, the box-plots shall be compared in terms of how much they overlap.

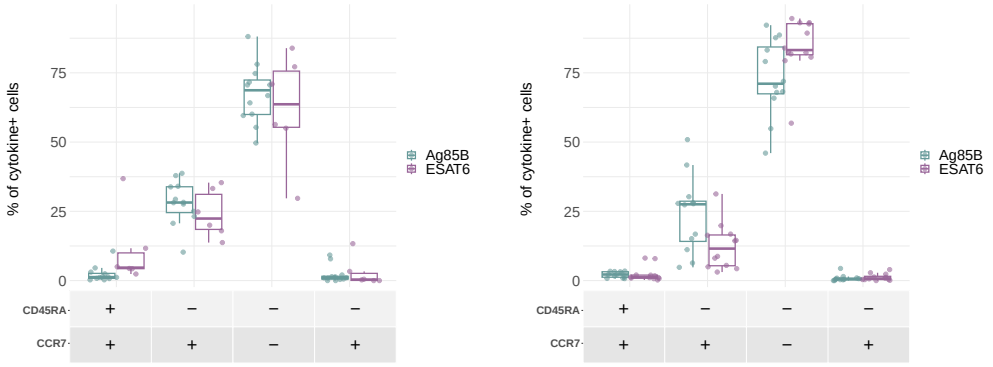
Referring to Figure 7.8, while there is not much overlap between the ESAT-6 and Ag85B box-plots for naive T cells and T_{CM} cells in QFT-, both of the ESAT-6 box-plots are only made up of four observations, and hence any comparison between the two stimuli for these cell subsets has very low power. For QFT+ participants, all plots overlap substantially, thus providing no evidence that the frequencies of any of the memory combinations differs between stimuli. Thus, there is no evidence that at baseline, the differentiation profiles of cells recognising Ag85B are distinct from those recognising ESAT-6.



(a) QFT- Baseline Response (b) QFT+ Baseline Response

Figure 7.8: Box-plots of the frequencies of cytokine positive CD4 T cells which are cytokine positive and express a certain memory combination at baseline. These frequencies are from all participants who received 2 administrations of 5 µg. Box-plots are plotted for different QFT statuses as well as for different stimuli.

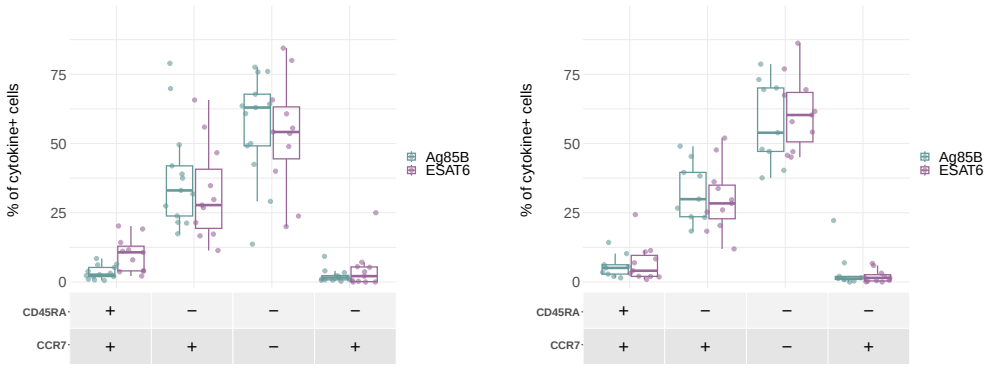
Referring to Figure 7.9, for both QFT- and QFT+ participants, there is substantial overlap between most of the box-plots, providing no evidence that the frequencies of any of the memory combinations differs between stimuli. However, for QFT- participants, the box-plots of the naive T cell frequencies do not overlap much, suggesting that the frequency of naive T cells may be higher under ESAT-6 stimulation compared to Ag85B stimulation in QFT- participants. Thus, there is no evidence that at the peak time point, the differentiation profiles of cells recognising Ag85B are distinct from those recognising ESAT-6.



(a) QFT- Peak Response (b) QFT+ Peak Response

Figure 7.9: Box-plots of the frequencies of cytokine positive CD4 T cells which are cytokine positive and express a certain memory combination at peak time point. These frequencies are from all participants who received 2 administrations of 5 µg. Box-plots are plotted for different QFT statuses as well as for different stimuli.

Referring to Figure 7.10, for both QFT- and QFT+ participants, there is substantial overlap between the box-plots, providing no evidence that the frequencies of any of the memory combinations differs between stimuli. Thus, there is no evidence at any of the time points that the differentiation profiles of cells recognising Ag85B are distinct from those recognising ESAT-6.



(a) QFT- Memory Response (b) QFT+ Memory Response

Figure 7.10: Box-plots of the frequencies of cytokine positive CD4 T cells which are cytokine positive and express a certain memory combination at memory. These frequencies are from all participants who received 2 administrations of 5 µg. Box-plots are plotted for different QFT statuses as well as for different stimuli.

7.1.3 Comment on the standard approach

This section analysed two different response variables, magnitude of the antigen-specific immune response and the qualitative antigen-specific immune response, to investigate the differentiation of the immune response. Both response variables provided no evidence that higher concentrations in QFT+ individuals drives T cells towards more differentiated phenotypes compared to QFT- individuals after Ag85B stimulation, but disagreed as to whether there was evidence of this after ESAT-6 stimulation. Similarly, both responses provided no evidence that after vaccination, differentiation profiles of T cells recognising Ag85B differ between QFT+ and QFT- individuals, but disagreed as to whether there was evidence of this for T cells recognising ESAT-6. While the magnitude of the antigen-specific immune response provided evidence that the differentiation profiles of cells recognising Ag85B were distinct from those recognising ESAT-6, the qualitative antigen-specific immune response provided no evidence that a difference existed. Thus, there was very little agreement in terms of the inference made between the two responses.

7.2 Generalised Linear Mixed Effect Models

The response variable of interest for this section is the magnitude of the antigen-specific immune response, $Y_{ijk}^{(s)}$. However, for the GLMEM's, k will be the memory combinations only, not the marker combination (i.e. the combinations of both cytokines and memory markers) as seen in Section 7.1.1.

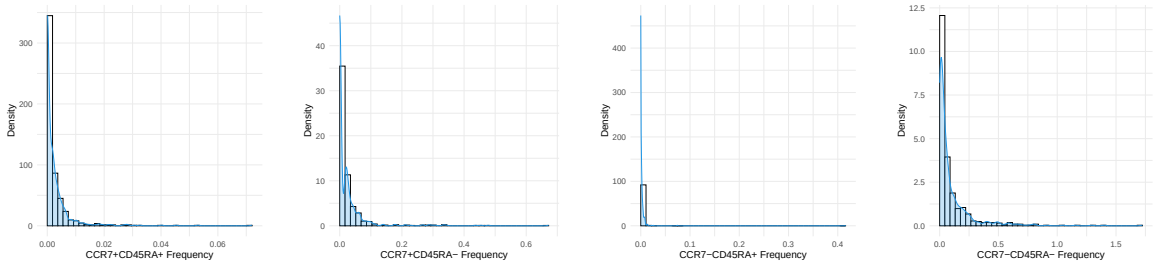
For the mixed effect modelling, the magnitude of the antigen-specific immune response is used rather than the qualitative antigen-specific immune response response because the filtering out of non-responding participant visits needed for the qualitative response results in incomplete observations for the majority of participants. Mixed effect models were not fit to incomplete data because any form of imputation, including model-based imputation, would be inappropriate since the observations are removed rather than missing, the qualitative response cannot be used for the mixed effect models.

Thus, 8 single-level GLMEM's shall be fit using $\boldsymbol{Y}_k^{(s)}$ as the response variable. For example, a single-level model will be fit to $\boldsymbol{Y}_{CCR7+CD45RA-}^{(ESAT-6)}$, the magnitude of the ESAT-6-specific immune response for cells expressing CCR7+CD45RA−. 2 two-level models grouped by memory combination nested within participant shall be fit using $\boldsymbol{Y}^{(s)}$ as the response variable and another 4 two-level models grouped by stimulus nested within participant shall be fit using $\boldsymbol{Y}_k$ as the response variable. Here, $\boldsymbol{Y}^{(s)}$ is the vector of all observed CD4 T cell frequencies observed under stimulus s, and $\boldsymbol{Y}_k$ is the vector of all observed CD4 T cell frequencies of cells expressing memory combination k. Finally, a three-level model grouped by memory combination nested in stimulus nested in participant shall be fit to using $\boldsymbol{Y}$ as the response variable.

7.2.1 Candidate distributions for the response

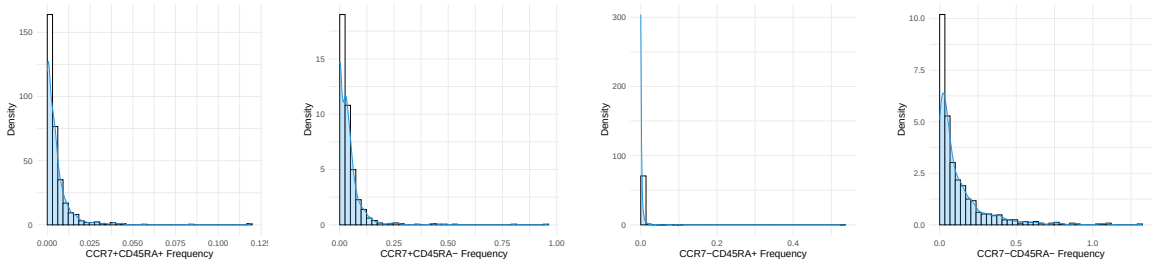
In order to choose an appropriate conditional distribution for the response variable, the marginal distribution of the response needs to be assessed. Figures 7.11 and 7.12 show the density plots for each of the 8 response variables for the single-level GLMEM's and Figure 7.13 shows the density plots for each of the 4 response variables for the two-level GLMEM's grouped by stimulus. Figures 7.14 and 7.15 show the density plots for each of the 2 response variables for the two-level GLMEM's grouped by memory combination using all memory combinations as well as the density plots of both response variables when only memory combinations CCR7+CD45RA− and CCR7−CD45RA− are considered. Similarly, Figure 7.16 shows the density plot for the response variable for the three-level GLMEM using all memory combinations as well as the response when only considering memory combinations CCR7+CD45RA− and CCR7−CD45RA−. All variables are continuous on the domain $[0, 100]$, highly zero-inflated and positively skewed. Possible candidate distributions are the Gamma, log-normal and Weibull, since they all at most require two parameters and are suitable for continuous right-skewed data. Again, the Tweedie distribution shall also be considered because it is well-suited to skew data. For the same reasoning given in Section 5.2.1, neither the Sichel, ZIP or ZIPNB distributions were considered.

As seen in Figures 7.11, 7.12 and 7.13, the density plots of the frequencies of cells expressing CCR7+CD45RA+ as well as that for cells expressing CCR7−CD45RA+ are very zero-inflated compared to that for the other two marker combinations. This made it very difficult to find a suitable distribution for the conditional response and thus made these two responses very difficult to model. Similarly, the density plots for observations from all memory combinations are much more zero-inflated compared to those using only observations for memory combinations CCR7+CD45RA− and CCR7−CD45RA−, as seen in Figures 7.14, 7.15 and 7.16. Thus, it was decided to only look at the frequencies for cells expressing CCR7+CD45RA− or CCR7−CD45RA−, since it was not possible to find a suitable conditional distribution when using all memory combinations.



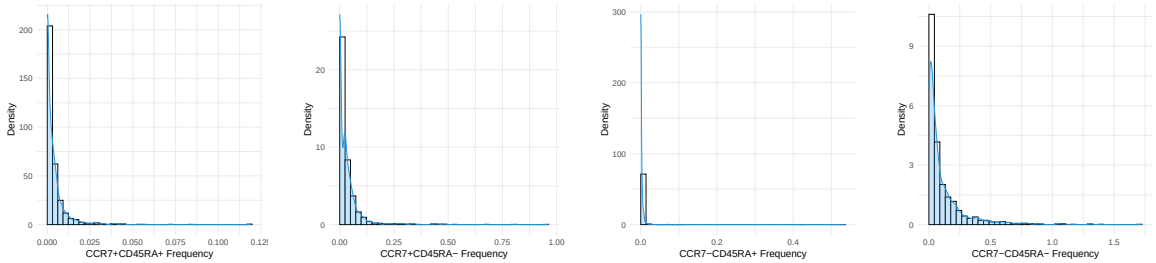
(a) CCR7+CD45RA+ (b) CCR7+CD45RA- (c) CCR7−CD45RA+ (d) CCR7−CD45RA-

Figure 7.11: Density plots of the frequency of cells expressing memory combinations (a) CCR7+CD45RA+, (b) CCR7+CD45RA−, (c) CCR7−CD45RA+ and (d) CCR7−CD45RA− after stimulation with ESAT-6, which will be used as the response variable for the ESAT-6 single-level GLMEM's.



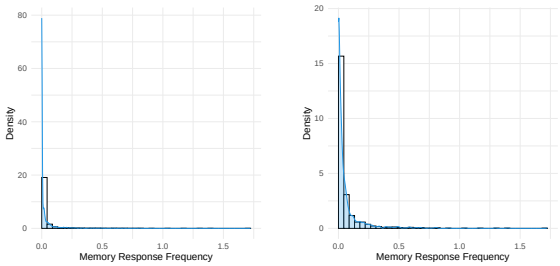
(a) CCR7+CD45RA+ (b) CCR7+CD45RA− (c) CCR7−CD45RA+ (d) CCR7−CD45RA-

Figure 7.12: Density plots of the frequency of cells expressing memory combinations (a) CCR7+CD45RA+, (b) CCR7+CD45RA−, (c) CCR7−CD45RA+ and (d) CCR7−CD45RA− after stimulation with Ag85B, which will be used as the response variable for the Ag85B single-level GLMEM's.



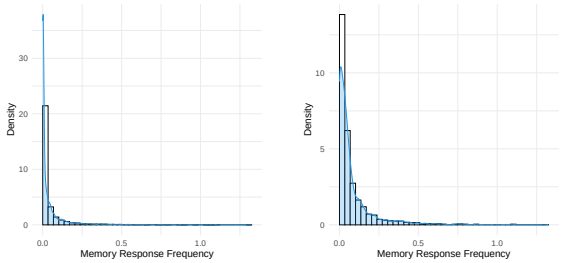
(a) CCR7+CD45RA+ (b) CCR7+CD45RA− (c) CCR7−CD45RA+ (d) CCR7−CD45RA−

Figure 7.13: Density plots of the frequency of cells expressing memory combinations (a) CCR7+CD45RA+, (b) CCR7+CD45RA−, (c) CCR7−CD45RA+ and (d) CCR7−CD45RA− after stimulation with either ESAT-6 or Ag85B, which will be used as the response variable for two-level GLMEM's grouped by stimulus.



(a) All memory combinations (b) CCR7+CD45RA− and CCR7−CD45RA−

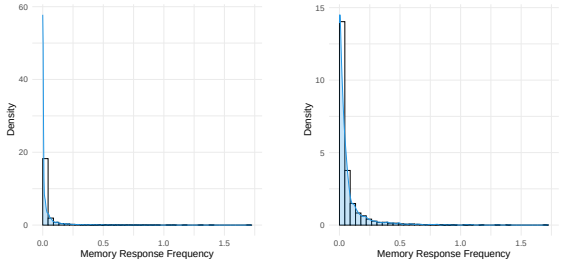
Figure 7.14: Density plots of the frequency of cells expressing a certain memory combination after stimulation with ESAT-6, which will be used as the response variable for two-level GLMEM's grouped by memory combination. Plot (a) uses observations from all four memory combinations while (b) uses observations from memory combinations CCR7+CD45RA− and CCR7-CD45RA− only.



(a) All memory combinations

(b) CCR7+CD45RA− and CCR7−CD45RA−

Figure 7.15: Density plots of the frequency of cells expressing a certain memory combination after stimulation with Ag85B, which will be used as the response variable for two-level GLMEM's grouped by memory combination. Plot (a) uses observations from all four memory combinations while (b) uses observations from memory combinations CCR7+CD45RA− and CCR7−CD45RA− only.



(a) All memory combinations

(b) CCR7+CD45RA− and CCR7−CD45RA−

Figure 7.16: Density plots of the frequency of cells expressing a certain memory combination after stimulation with either ESAT-6 or Ag85B, which will be used as the response variable for three-level GLMEM grouped by memory combination nested within stimulus. Plot (a) uses observations from all four memory combinations while (b) uses observations from memory combinations CCR7+CD45RA− and CCR7−CD45RA− only.

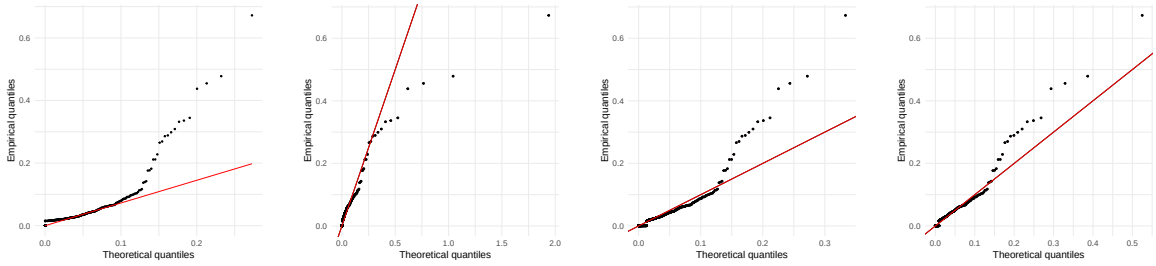
7.2.2 Single-level GLMEM: frequency of CCR7+CD45RA− after ESAT-6 stimulation

A log-normal, Gamma, Weibull and Tweedie distribution were all fit to response variable, the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− after stimulation with ESAT-6. The log-normal, Gamma and Weibull distributions were fit using the `fitdistrplus` [76] package in `R`, while the Tweedie distribution was fit using the `cplm` [77] package. Table 7.1 summaries the resultant AIC and BIC values from the fitted log-normal, Gamma and Weibull distributions. The log-normal had the lowest AIC and BIC compared to the other distributions, suggesting that out of the three distributions the log-normal fit the data best.

Table 7.1: The AIC and BIC after fitting log-normal, Gamma and Weibull distributions to the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− after stimulation with ESAT-6.

Distribution	AIC	BIC
Log-normal	-4815.481	-4806.286
Gamma	-4563.668	-4554.474
Weibull	-4668.539	-4659.344

To further assess the fit, the empirical versus theoretical quantiles were plot, as seen in Figure 7.17. All four qq-plots suggest that neither of the distributions fit the data well, however the log-normal distribution does seem to fit the data far better than the other three distributions. Since the log-normal also had the lowest AIC and BIC compared to the Weibull and Gamma, it was decided to assume that the conditional response follows a log-normal distribution.



(a) Tweedie (b) Log-normal (c) Gamma (d) Weibull

Figure 7.17: QQ-plots of the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− after stimulation with ESAT-6 fit to (a) Tweedie, (b) log-normal, (c) Gamma and (b) Weibull distributions.

Following the model building steps laid out in Section 4.3.1, the optimal model chosen is as follows

$$\mu_{ij} = \beta_0 + \beta_1 I_j^{Peak} + \beta_2 I_j^{Mem} + \beta_3 I_i^{QFT+} + \beta_4 x_i^{Admin} + \beta_5 x_i^{Conc} + b_{0i},$$

$$\log(\sigma_{ij}) = \alpha_0 + \alpha_1 I_i^{QFT+} + \alpha_2 x_i^{Admin},$$

$$\text{for} \quad i = 1, ..., I \quad \text{and} \quad j = 1, 2, 3,$$

where

$$b_{0i} \sim N\left(0, \sigma_0^2\right) \qquad \text{and} \qquad \epsilon_{ij} \sim N\left(0, \sigma_e^2\right).$$

Here i is the participant and j is the time point. The standard deviation estimates were $\hat{\sigma}_0 = 0.923$ with 95% confidence interval $(0.793; 1.074)$ and $\hat{\sigma}_e = 1.124$ with 95% confidence interval $(1.055; 1.198)$. No participant-specific random effect was included on the time point variable because doing so resulted in a model that would not converge.

Model diagnostics can be found in Appendix F.1, where there was some evidence that the homogeneity assumption of the residuals and normality assumption of the random effects are violated.

Referring to the parameter estimates in Table 7.2, there is evidence that frequency increases from baseline to peak, and slightly less from baseline to memory. QFT+ participants seem to have higher frequencies of cells expressing CCR7+CD45RA− compared to QFT- participants. There is also evidence that frequency decreases with increasing vaccine concentration but that

number of administrations does not have a strong effect on frequency. QFT+ participants are also associated with a higher variation in frequency compared to QFT- participants, and variances seems to increase with increasing numbers of administrations.

Table 7.2: Parameter estimates of the single-level GLMEM fit to the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− after stimulation with ESAT-6.

Variable	Parameter	Estimate	Std. Error	t-value	p-value
μ Coefficients:					
	β_0	-6.635	0.142	-46.880	$< 2 \times 10^{-16}$
Peak	β_1	0.722	0.0962	7.502	2.42×10^{-13}
Mem.	β_2	0.349	0.096	3.625	0.0003
QFT+	β_3	1.876	0.084	22.421	$< 2 \times 10^{-16}$
Admin.	β_4	0.098	0.057	1.729	0.084
Conc.	β_5	-0.007	0.002	-3.268	0.001
σ Coefficients:					
	α_0	-0.264	0.080	-3.282	0.001
QFT+	α_1	0.414	0.054	7.718	5.28×10^{-14}
Admin	α_2	0.079	0.037	2.119	0.035

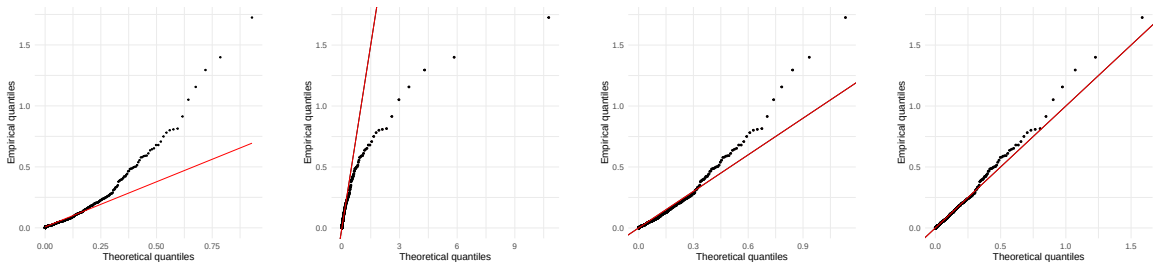
7.2.3 Single-level GLMEM: frequency of CCR7−CD45RA− after ESAT-6 stimulation

A log-normal, Gamma, Weibull and Tweedie distribution were all fit to response variable, the frequency of CD4 T cells which are cytokine positive and express CCR7−CD45RA− after stimulation with ESAT-6. Again, the log-normal, Gamma and Weibull distributions were fit using the `fitdistrplus` [76] package in `R`, while the Tweedie distribution was fit using the `cplm` [77] package. Table 7.3 summaries the resultant AIC and BIC values from the fitted log-normal, Gamma and Weibull distributions. The Weibull distribution had the lowest AIC and BIC compared to the other distributions, suggesting that out of the three distributions the Weibull fit the data best.

Table 7.3: The AIC and BIC after fitting log-normal, Gamma and Weibull distributions to the frequency of CD4 T cells which are cytokine positive and express CCR7−CD45RA− after stimulation with ESAT-6.

Distribution	**AIC**	**BIC**
Log-normal	-2193.888	-2184.694
Gamma	-2191.159	-2181.965
Weibull	-2227.587	-2218.393

Referring to the empirical versus theoretical quantiles seen in Figure 7.18, the Weibull clearly fits the data far better than the other three distributions. Thus, based off the qq-plots and the AIC and BIC, it was decided to assume that the conditional response follows a Weibull distribution.



(a) Tweedie (b) Log-normal (c) Gamma (d) Weibull

Figure 7.18: QQ-plots of the frequency of CD4 T cells which are cytokine positive and express CCR7−CD45RA− after stimulation with ESAT-6 fit to (a) Tweedie, (b) log-normal, (c) Gamma and (b) Weibull distributions.

Following the model building steps laid out in Section 4.3.1, the optimal model chosen is as follows

$$\log(\mu_{ij}) = \beta_0 + \beta_1 I_j^{Peak} + \beta_2 I_j^{Mem} + \beta_3 I_i^{QFT+} + \beta_4 x_i^{Admin} + \beta_5 x_i^{Conc} + \beta_6 x_i^{Admin*QFT+} + b_{0i},$$

$$\log(\sigma_{ij}) = \alpha_0,$$

$$\text{for} \quad i = 1, ..., I \quad \text{and} \quad j = 1, 2, 3,$$

where

$$b_{0i} \sim N\left(0, \sigma_0^2\right) \qquad \text{and} \qquad \epsilon_{ij} \sim N\left(0, \sigma_e^2\right).$$

Here i is the participant and j is the time point. The standard deviation estimates were $\hat{\sigma}_0 = 0.986$ with 95% confidence interval $(0.891; 1.090)$ and $\hat{\sigma}_e = 1.013$ with 95% confidence interval $(0.952; 1.079)$. The scale parameter σ_{ij} was modelled as a constant, and again no participant-specific random effect was added on time point due to convergences issues.

Model diagnostics can be found in Appendix F.2, where there was slight evidence that the homogeneity assumption of the residuals and substantial evidence that the normality assumption of the residuals and random effects are violated.

Referring to the parameter estimates in Table 7.4, there is again evidence that frequency increases from baseline to peak, and slightly less from baseline to memory. There is also evidence that frequency decreases with increasing vaccine concentration and that a strong interaction between number of administrations and QFT status exists, which is depicted in Figure 7.19. While frequency was always greater for QFT+ participants compared to QFT- participants, for QFT+ participants frequency decreases as number of administrations increases while for QFT- participants frequency increases as number of administrations increases.

Table 7.4: Parameter estimates of the single-level GLMEM fit to the frequency of CD4 T cells which are cytokine positive and express CCR7−CD45RA− after stimulation with ESAT-6.

Variable	Parameter	Estimate	Std. Error	t-value	p-value
μ Coefficients:					
	β_0	-5.012	0.113	-44.365	$< 2 \times 10^{-16}$
Peak	β_1	1.307	0.058	22.366	$< 2 \times 10^{-16}$
Mem.	β_2	0.736	0.059	12.566	$< 2 \times 10^{-16}$
QFT+	β_3	2.984	0.132	22.583	$< 2 \times 10^{-16}$
Admin.	β_4	0.303	0.0507	5.985	4.08×10^{-9}
Conc.	β_5	-0.011	0.001	-8.512	$< 2 \times 10^{-16}$
Admin*QFT+	β_6	-0.611	0.068	-9.043	$< 2 \times 10^{-16}$

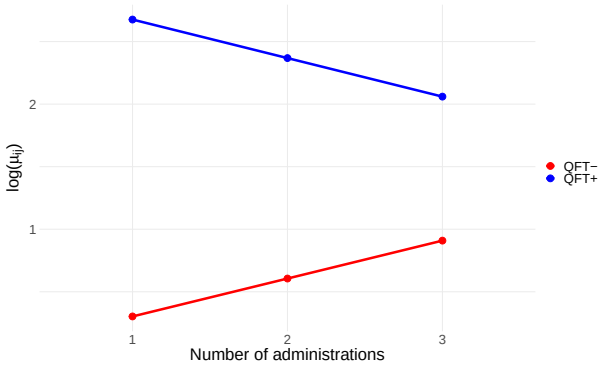


Figure 7.19: The expected value of logged mean $\log(\mu_{ij})$ for changing QFT status and number of administrations while keeping all other variables constant.

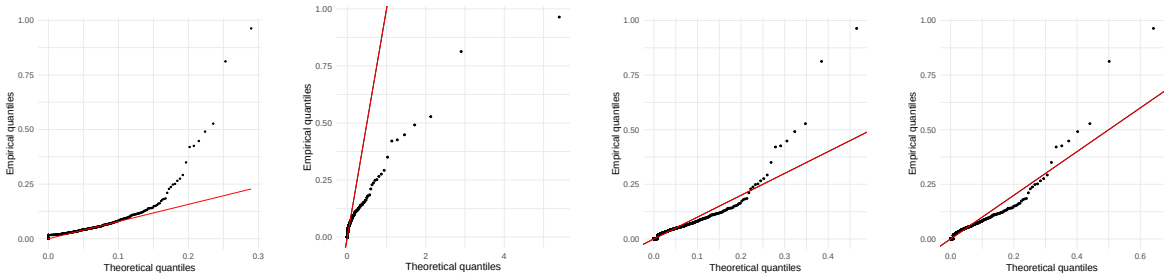
7.2.4 Single-level GLMEM: frequency of CCR7+CD45RA− after Ag85B stimulation

A log-normal, Gamma, Weibull and Tweedie distribution were all fit to response variable, the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− after stimulation with Ag85B. Again, the log-normal, Gamma and Weibull distributions were fit using the `fitdistrplus` [76] package in `R`, while the Tweedie distribution was fit using the `cplm` [77] package. Table 7.5 summaries the resultant AIC and BIC values from the fitted log-normal, Gamma and Weibull distributions. The Weibull distribution had the lowest AIC and BIC compared to the other distributions, suggesting that out of the three distributions the Weibull fit the data best.

Table 7.5: The AIC and BIC after fitting log-normal, Gamma and Weibull distributions to the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− after stimulation with Ag85B.

Distribution	**AIC**	**BIC**
Log-normal	-3315.677	-3306.488
Gamma	-3389.29	-3380.101
Weibull	-3391.507	-3382.319

Referring to the empirical versus theoretical quantiles seen in Figure 7.20, the Weibull clearly fits the data far better than the other three distributions. Thus, based off the qq-plots and the AIC and BIC,it was decided to assume that the conditional response follows a Weibull distribution.



(a) Tweedie (b) Log-normal (c) Gamma (d) Weibull

Figure 7.20: QQ-plots of the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− after stimulation with Ag85B fit to (a) Tweedie, (b) log-normal, (c) Gamma and (b) Weibull distributions.

Following the model building steps laid out in Section 4.3.1, the optimal model chosen is as follows

$$\begin{aligned}\log(\mu_{ij}) = \beta_0 &+ \beta_1 I_j^{Peak} + \beta_2 I_j^{Mem} + \beta_3 I_i^{QFT+} + \beta_4 x_i^{Admin} + \beta_5 x_i^{Conc} \\ &+ \beta_6 I_{ij}^{Peak*QFT+} + \beta_7 I_{ij}^{Mem*QFT+} + b_{0i},\end{aligned}$$

$$\log(\sigma_{ij}) = \alpha_0 + \alpha_1 x_i^{Admin},$$

$$\text{for} \quad i = 1, ..., I \quad \text{and} \quad j = 1, 2, 3,$$

where

$$b_{0i} \sim N\left(0, \sigma_0^2\right) \qquad \text{and} \qquad \epsilon_{ij} \sim N\left(0, \sigma_e^2\right).$$

Here i is the participant and j is the time point. The standard deviation estimates were $\hat{\sigma}_0 = 1.038$ with 95% confidence interval $(0.931; 1.157)$ and $\hat{\sigma}_e = 1.005$ with 95% confidence interval $(0.944; 1.071)$. Again, no participant-specific random effect was included on time point due to convergences issues.

Model diagnostics can be found in Appendix F.3, where there was some evidence that the homogeneity and normality assumptions of the residuals are violated.

Referring to the parameter estimates seen in Table 7.6, there is evidence that frequency increases with increasing number of vaccine administrations and that vaccine concentration does not have a strong effect on frequency. Variation in frequency also increases with increasing number of administrations. There is evidence of a strong interaction between time point and QFT status, as depicted in Figure 7.21. While frequency is always higher for QFT+ participants compared to

QFT- participants, the profiles over time are flatter for QFT+ participants compared to QFT- participants. In other words, the increase in frequency from baseline to both post-vaccination time points is more pronounced in QFT- participants compared to QFT+ participants.

Table 7.6: Parameter estimates of the single-level GLMEM fit to the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− after stimulation with Ag85B.

Variable	Parameter	Estimate	Std. Error	t-value	p-value
μ Coefficients:					
	β_0	-5.883	0.122	-48.414	$< 2 \times 10^{-16}$
Peak	β_1	2.015	0.102	19.843	$< 2 \times 10^{-16}$
Mem.	β_2	1.625	0.101	16.086	$< 2 \times 10^{-16}$
QFT+	β_3	1.260	0.101	12.486	$< 2 \times 10^{-16}$
Admin.	β_4	0.296	0.042	7.055	5.57e-12
Conc.	β_5	-0.002	0.001	-1.317	0.188
Peak*QFT+	β_6	-0.788	0.143	-5.529	5.13×10^{-8}
Mem*QFT+	β_7	-0.875	0.142	-6.186	1.26×10^{-9}
σ Coefficients:					
	α_0	-0.066	0.085	-0.779	0.437
Admin	α_1	0.165	0.044	3.757	0.0002

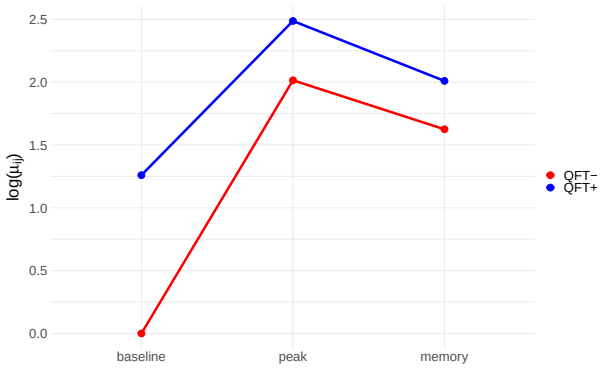


Figure 7.21: The expected value of logged mean $\log(\mu_{ij})$ for changing QFT status and time point while keeping all other variables constant.

7.2.5 Single-level GLMEM: frequency of CCR7−CD45RA− after Ag85B stimulation

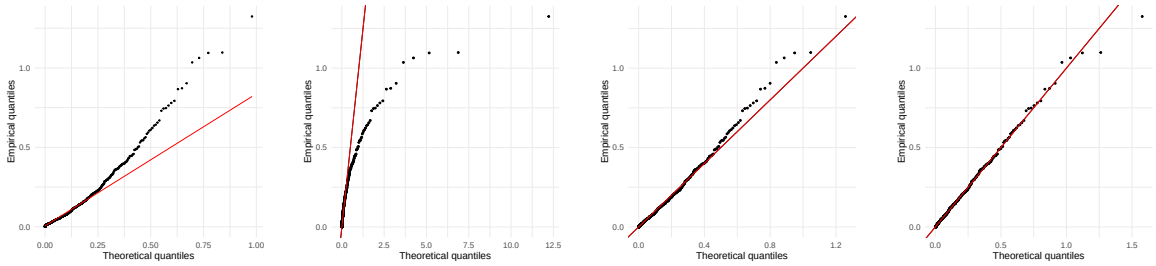
A log-normal, Gamma, Weibull and Tweedie distribution were all fit to response variable, the frequency of CD4 T cells which are cytokine positive and express CCR7-CD45RA- after stimulation with Ag85B. Again, the log-normal, Gamma and Weibull distributions were fit using the `fitdistrplus` [76] package in `R`, while the Tweedie distribution was fit using the `cplm` [77] package. Table 7.7 summaries the resultant AIC and BIC values from the fitted log-normal, Gamma and Weibull distributions. The Weibull distribution had the lowest AIC and BIC compared to the other distributions, suggesting that out of the three distributions the

Weibull fit the data best.

Table 7.7: The AIC and BIC after fitting log-normal, Gamma and Weibull distributions to the frequency of CD4 T cells which are cytokine positive and express CCR7−CD45RA− after stimulation with Ag85B.

Distribution	**AIC**	**BIC**
Log-normal	-1580.604	-1571.416
Gamma	-1688.453	-1679.264
Weibull	-1693.197	-1684.008

Referring to the empirical versus theoretical quantiles seen in Figure 7.22, the Weibull clearly fits the data far better than the other three distributions. Thus, based off the qq-plots and the AIC and BIC, it was decided to assume that the conditional response follows a Weibull distribution.



(a) Tweedie (b) Log-normal (c) Gamma (d) Weibull

Figure 7.22: QQ-plots of the frequency of CD4 T cells which are cytokine positive and express CCR7−CD45RA− after stimulation with Ag85B fit to (a) Tweedie, (b) log-normal, (c) Gamma and (b) Weibull distributions.

Following the model building steps laid out in Section 4.3.1, the optimal model chosen is as follows

$$\log(\mu_{ij}) = \beta_0 + \beta_1 I_j^{Peak} + \beta_2 I_j^{Mem} + \beta_3 I_i^{QFT+} + \beta_4 x_i^{Admin} + \beta_5 x_i^{Conc} + \beta_6 I_{ij}^{Peak*QFT+} + \beta_7 I_{ij}^{Mem*QFT+} + b_{0i},$$

$$\log(\sigma_{ij}) = \alpha_0,$$

$$\text{for} \quad i = 1, ..., I \quad \text{and} \quad j = 1, 2, 3,$$

where

$$b_{0i} \sim N\left(0, \sigma_0^2\right) \qquad \text{and} \qquad \epsilon_{ij} \sim N\left(0, \sigma_e^2\right).$$

Here i is the participant and j is the time point. The standard deviation estimates were $\hat{\sigma}_0 = 1.016$ with 95% confidence interval $(0.922; 1.119)$ and $\hat{\sigma}_e = 0.960$ with 95% confidence interval $(0.902; 1.022)$. The scale parameter σ_{ij} was modelled as a constant, and again no

participant-specific random effect was added on time point due to convergences issues.

Model diagnostics can be found in Appendix F.4, where there was some evidence that the homogeneity and normality assumptions of the residuals are violated.

Referring to the parameter estimates seen in Table 7.8, there is evidence that frequency increases with increasing number of administrations but decreases with higher vaccine concentrations. There is again evidence of a strong interaction between time point and QFT status, as depicted in Figure 7.23. The interpretation here is the same as for Figure 7.21: frequency is always higher for QFT+ participants compared to QFT- participants, but the profiles over time are flatter for QFT+ participants compared to QFT- participants.

Table 7.8: Parameter estimates of the single-level GLMEM fit to the frequency of CD4 T cells which are cytokine positive and express CCR7−CD45RA− after stimulation with Ag85B.

Variable	Parameter	Estimate	Std. Error	t-value	p-value
μ Coefficients:					
	β_0	-4.192	0.082	-51.130	$< 2 \times 10^{-16}$
Peak	β_1	2.098	0.074	28.516	$< 2 \times 10^{-16}$
Mem.	β_2	1.293	0.073	17.613	$< 2 \times 10^{-16}$
QFT+	β_3	1.185	0.073	16.288	$< 2 \times 10^{-16}$
Admin.	β_4	0.078	0.030	2.605	0.009
Conc.	β_5	-0.004	0.001	-3.614	0.0003
Peak*QFT+	β_6	-0.801	0.103	-7.807	3.46×10^{-14}
Mem*QFT+	β_7	-0.691	0.103	-6.732	4.61×10^{-11}

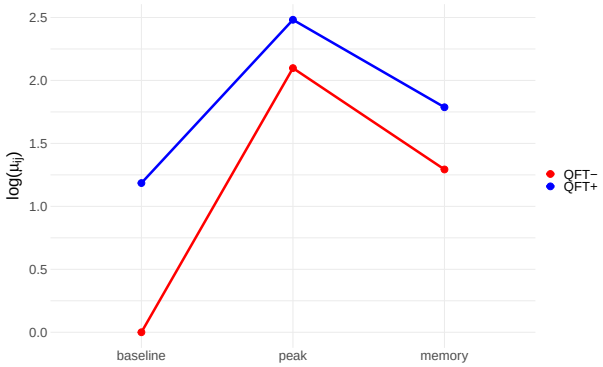


Figure 7.23: The expected value of logged mean $\log(\mu_{ij})$ for changing QFT status and time point while keeping all other variables constant.

7.2.6 Two-level GLMEM: Frequency of CCR7+CD45RA−

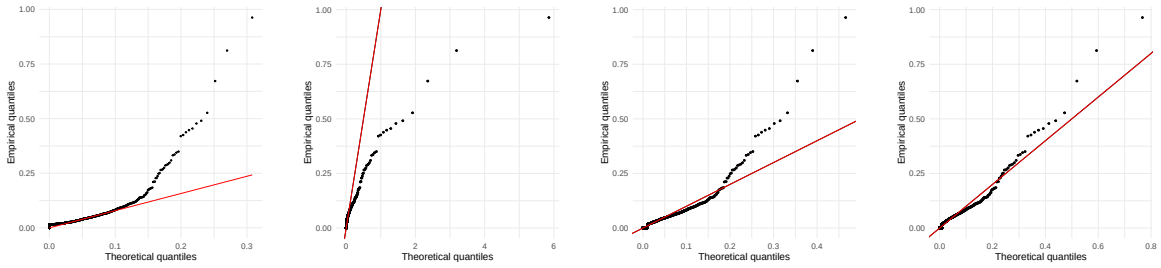
A log-normal, Gamma, Weibull and Tweedie distribution were all fit to response variable, the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− after stimulation with either Ag85B or ESAT-6. Again, the log-normal, Gamma and Weibull distributions were fit using the `fitdistrplus` [76] package in `R`, while the Tweedie distribution was fit using

the `cplm` [77] package. Table 7.9 summaries the resultant AIC and BIC values from the fitted log-normal, Gamma and Weibull distributions. The log-normal distribution had the lowest AIC and BIC compared to the other distributions, suggesting that out of the three distributions the log-normal fit the data best.

Table 7.9: The AIC and BIC after fitting log-normal, Gamma and Weibull distributions to the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− after stimulation with either Ag85B or ESAT-6.

Distribution	**AIC**	**BIC**
Log-normal	-8031.532	-8020.954
Gamma	-7885.24	-7874.663
Weibull	-7963.073	-7952.496

Referring to the empirical versus theoretical quantiles seen in Figure 7.24, the Weibull clearly fits the data far better than the other three distributions, including the log-normal. Thus, since the Weibull and log-normal have very similar AIC and BIC values, it was decided to assume that the conditional response follows a Weibull distribution.



(a) Tweedie (b) Log-normal (c) Gamma (d) Weibull

Figure 7.24: QQ-plots of the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− after stimulation with either Ag85B or ESAT-6 fit to (a) Tweedie, (b) log-normal, (c) Gamma and (b) Weibull distributions.

Following the model building steps laid out in Section 4.3.1, the optimal model chosen is as follows

$$\begin{aligned}\log(\mu_{ijk}) = \beta_0 &+ \beta_1 I_k^{Peak} + \beta_2 I_k^{Mem} + \beta_3 I_i^{QFT+} + \beta_4 x_i^{Admin} + \beta_5 x_i^{Conc} + \beta_6 I_i^{ESAT6} \\ &+ \beta_7 I_{ik}^{Peak*QFT+} + \beta_8 I_{ik}^{Mem*QFT+} + \beta_9 I_{jk}^{Peak*ESAT6+} + \beta_{10} I_{jk}^{Mem*ESAT6+} \\ &+ \beta_{11} I_{ij}^{ESAT6*QFT+} + b_{0i},\end{aligned}$$

$$\log(\sigma_{ijk}) = \alpha_0,$$

$$\text{for} \quad i = 1, ..., I, \quad j = 1, 2 \quad \text{and} \quad k = 1, 2, 3,$$

where

$$b_{0i} \sim N\left(0, \sigma_0^2\right) \qquad \text{and} \qquad \epsilon_{ijk} \sim N\left(0, \sigma_e^2\right).$$

Here i is the participant, j is the stimulus and k is the time point. The standard deviation estimates were $\hat{\sigma}_0 = 0.908$ with 95% confidence interval $(0.816; 1.011)$ and $\hat{\sigma}_e = 1.023$ with 95% confidence interval $(0.983; 1.065)$. The scale parameter σ_{ij} was modelled as a constant, and again no participant-specific random effect was added on time point due to convergences issues. No participant-specific random effect was included on stimulus since it resulted in higher GAIC values, indicating an inferior fit. No three-way interactions were included in the model since they did not improve the model fit.

Model diagnostics can be found in Appendix F.5, where there was some evidence that the homogeneity assumption of the residuals and normality assumption of the residuals and random effects are violated.

Referring to the parameter estimates seen in Table 7.10, there is evidence that frequency increases with increasing number of vaccine administrations but decreases with higher vaccine concentrations. There is also evidence of strong interactions between time point and QFT status, time point and stimulus and QFT status and stimulus, although a three-way interaction between these variables did not improve the model fit. Referring to the interactions represented in Figure 7.25, QFT+ participants always had higher frequencies than QFT- participants when comparing responses to the same stimulus. At post-vaccination time points, frequencies are always higher after Ag85B stimulation compared to ESAT-6 stimulation. The difference in frequency between ESAT-6 and Ag85B samples is greater for QFT- participants compared to QFT+ participants at all time points, especially at baseline. Again, profiles are much flatter for QFT+ participants compared to QFT- participants.

Table 7.10: Parameter estimates of the two-level GLMEM fit to the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− after stimulation with either Ag85B or ESAT-6.

Variable	Parameter	Estimate	Std. Error	t-value	p-value
μ Coefficients:					
	β_0	-5.508	0.118	-46.858	$< 2 \times 10^{-16}$
Peak	β_1	2.029	0.113	17.910	$< 2 \times 10^{-16}$
Mem.	β_2	1.556	0.112	14.057	$< 2 \times 10^{-16}$
QFT+	β_3	1.365	0.105	12.955	$< 2 \times 10^{-16}$
Admin.	β_4	0.202	0.037	5.385	8.62×10^{-8}
Conc.	β_5	-0.006	0.001	-4.346	1.5×10^{-5}
ESAT6	β_6	-1.142	0.106	-10.794	$< 2 \times 10^{-16}$
Peak*QFT+	β_7	-0.993	0.129	-7.697	2.81×10^{-14}
Mem*QFT+	β_8	-0.930	0.129	-7.218	9.12×10^{-13}
Peak*ESAT6	β_9	-0.877	0.129	-6.788	1.75×10^{-11}
Mem*ESAT6	β_{10}	-0.730	0.129	-5.645	2.04×10^{-8}
ESAT6*QFT+	β_{11}	1.370	0.105	13.025	$< 2 \times 10^{-16}$

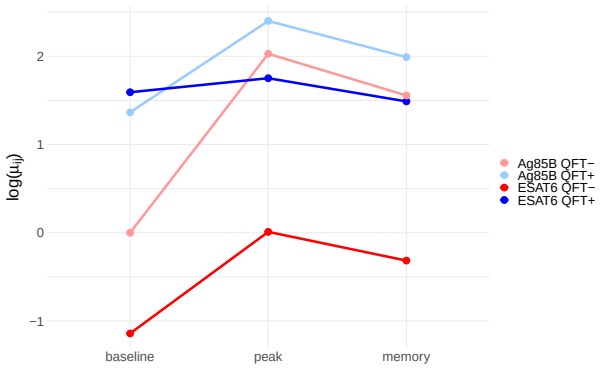


Figure 7.25: The expected value of logged mean $\log(\mu_{ij})$ for changing QFT status, stimulus and number of administrations while keeping all other variables constant.

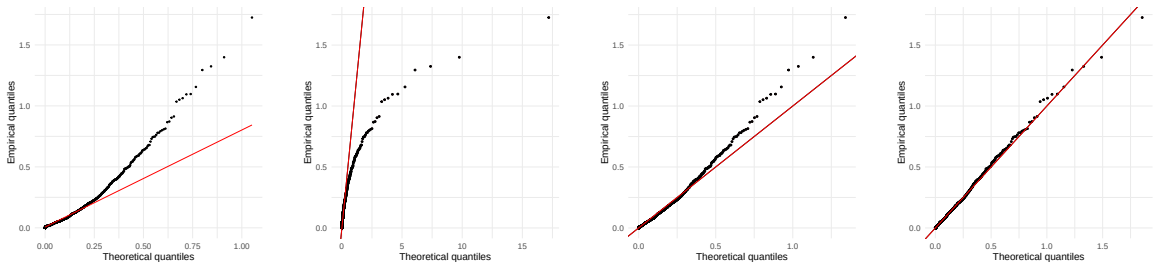
7.2.7 Two-level GLMEM: Frequency of CCR7−CD45RA−

A log-normal, Gamma, Weibull and Tweedie distribution were all fit to response variable, the frequency of CD4 T cells which are cytokine positive and express CCR7−CD45RA− after stimulation with either Ag85B or ESAT-6. Again, the log-normal, Gamma and Weibull distributions were fit using the `fitdistrplus` [76] package in `R`, while the Tweedie distribution was fit using the `cplm` [77] package. Table 7.11 summaries the resultant AIC and BIC values from the fitted log-normal, Gamma and Weibull distributions. The Weibull distribution had the lowest AIC and BIC compared to the other distributions, suggesting that out of the three distributions the Weibull fit the data best.

Table 7.11: The AIC and BIC after fitting log-normal, Gamma and Weibull distributions to the frequency of CD4 T cells which are cytokine positive and express CCR7−CD45RA− after stimulation with either Ag85B or ESAT-6.

Distribution	**AIC**	**BIC**
Log-normal	-3750.196	-3739.618
Gamma	-3862.865	-3852.287
Weibull	-3897.445	-3886.867

Referring to the empirical versus theoretical quantiles seen in Figure 7.26, the Weibull clearly fits the data far better than the other three distributions. Thus, based off the qq-plots and the AIC and BIC, it was decided to assume that the conditional response follows a Weibull distribution.



(a) Tweedie (b) Log-normal (c) Gamma (d) Weibull

Figure 7.26: QQ-plots of the frequency of CD4 T cells which are cytokine positive and express CCR7−CD45RA− after stimulation with either Ag85B or ESAT-6 fit to (a) Tweedie, (b) log-normal, (c) Gamma and (b) Weibull distributions.

Following the model building steps laid out in Section 4.3.1, the optimal model chosen is as follows

$$\begin{aligned}\log(\mu_{ijk}) = \beta_0 + \beta_1 I_k^{Peak} + \beta_2 I_k^{Mem} + \beta_3 I_i^{QFT+} + \beta_4 x_i^{Admin} + \beta_5 x_i^{Conc} + \beta_6 I_i^{ESAT6} \\ + \beta_7 I_{ik}^{Peak*QFT+} + \beta_8 I_{ik}^{Mem*QFT+} + \beta_9 I_{ij}^{ESAT6*QFT+} + b_{0i},\end{aligned}$$

$$\log(\sigma_{ijk}) = \alpha_0,$$

$$\text{for} \quad i = 1, ..., I, \quad j = 1, 2, \quad \text{and} \quad k = 1, 2, 3,$$

where

$$b_{0i} \sim N\left(0, \sigma_0^2\right) \qquad \text{and} \qquad \epsilon_{ijk} \sim N\left(0, \sigma_e^2\right).$$

Here i is the participant, j is the stimulus and k is the time point. The standard deviation estimates were $\hat{\sigma}_0 = 0.857$ with 95% confidence interval $(0.773; 0.949)$ and $\hat{\sigma}_e = 1.008$ with 95% confidence interval $(0.969; 1.048)$. The scale parameter σ_{ij} was modelled as a constant, and again no participant-specific random effect was added on time point due to convergences issues. No participant-specific random effect was included on stimulus since it resulted in higher GAIC values, indicating an inferior fit. No three-way interactions were included in the model since they did not improve the model fit.

Model diagnostics can be found in Appendix F.6, where there was some evidence that the homogeneity assumption of the residuals and normality assumption of the random effects are violated.

Referring to the parameter estimates seen in Table 7.12, there is evidence that frequency decreases with higher vaccine concentrations but that number of administrations does not have a strong effect. There is also evidence of strong interactions between time point and QFT status and QFT status and stimulus, as represented in Figure 7.27. QFT+ participants consistently have higher frequencies compared to QFT- participants and again profiles are flatter for QFT+

participants compared to QFT- participants. Frequency is higher under ESAT-6 stimulation compared to Ag85B stimulation for QFT+ individuals, however the opposite is true for QFT- individuals. The difference in frequency between ESAT-6 and Ag85B samples is greater for QFT- participants compared to QFT+ participants at all time points, with virtually no difference between the two stimuli for QFT+ participants.

Table 7.12: Parameter estimates of the two-level GLMEM fit to the frequency of CD4 T cells which are cytokine positive and express CCR7−CD45RA− after stimulation with either Ag85B or ESAT-6.

Variable	Parameter	Estimate	Std. Error	t-value	p-value
μ Coefficients:					
	β_0	-3.723	0.089	-41.612	$< 2 \times 10^{-16}$
Peak	β_1	1.752	0.073	23.971	$< 2 \times 10^{-16}$
Mem.	β_2	1.103	0.073	15.120	$< 2 \times 10^{-16}$
QFT+	β_3	1.114	0.085	13.093	$< 2 \times 10^{-16}$
Admin.	β_4	0.030	0.030	1.000	0.318
Conc.	β_5	-0.008	0.001	-7.406	2.40×10^{-13}
ESAT6	β_6	-1.048	0.060	-17.566	$< 2 \times 10^{-16}$
Peak*QFT+	β_7	-0.620	0.102	-6.067	1.73×10^{-9}
Mem*QFT+	β_8	-0.587	0.102	-5.755	1.09×10^{-08}
ESAT6*QFT+	β_9	1.174	0.083	14.085	$< 2 \times 10^{-16}$

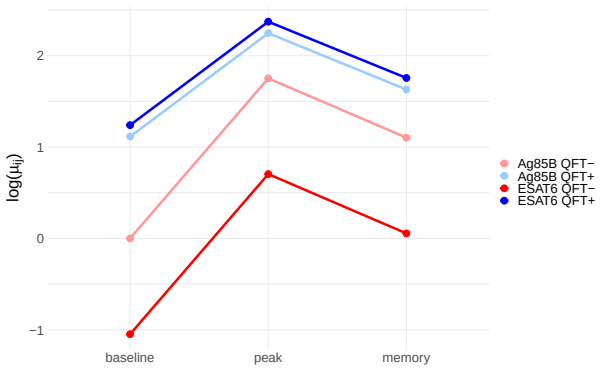


Figure 7.27: The expected value of logged mean $\log(\mu_{ij})$ for changing QFT status, time point and stimulus while keeping all other variables constant.

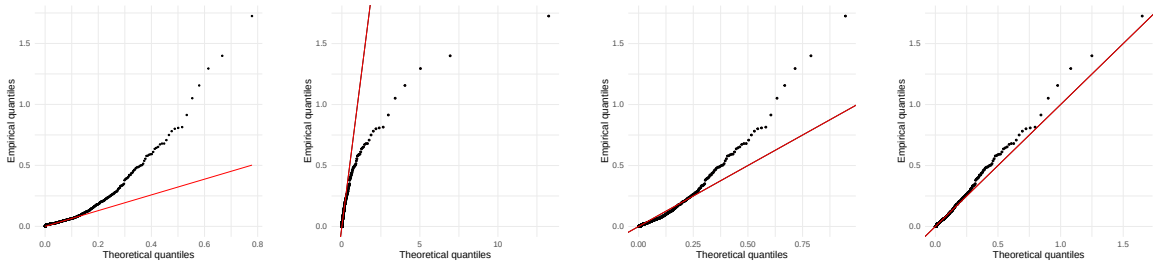
7.2.8 Two-level GLMEM: frequency after ESAT-6 stimulation

A log-normal, Gamma, Weibull and Tweedie distribution were all fit to response variable, the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− or CCR7−CD45RA− after stimulation with ESAT-6. Again, the log-normal, Gamma and Weibull distributions were fit using the `fitdistrplus` [76] package in `R`, while the Tweedie distribution was fit using the `cplm` [77] package. Table 7.13 summaries the resultant AIC and BIC values from the fitted log-normal, Gamma and Weibull distributions. The log-normal distribution had the lowest AIC and BIC compared to the other distributions, suggesting that out of the three distributions the log-normal fit the data best.

Table 7.13: The AIC and BIC after fitting log-normal, Gamma and Weibull distributions to the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− or CCR7−CD45RA− after stimulation with ESAT-6.

Distribution	**AIC**	**BIC**
Log-normal	-6703.374	-6692.794
Gamma	-6462.285	-6451.705
Weibull	-6607.771	-6597.19

Referring to the empirical versus theoretical quantiles seen in Figure 7.28, the Weibull clearly fits the data far than the other three distributions, including the log-normal distribution. Thus, despite the lower AIC and BIC of the log-normal, it was decided based off the qq-plots to assume that the conditional response follows a Weibull distribution.



(a) Tweedie (b) Log-normal (c) Gamma (d) Weibull

Figure 7.28: QQ-plots of the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− or CCR7−CD45RA− after stimulation with ESAT-6 fit to (a) Tweedie, (b) log-normal, (c) Gamma and (b) Weibull distributions.

Following the model building steps laid out in Section 4.3.1, the optimal model chosen is as follows

$$\begin{aligned}\log(\mu_{ijk}) = \beta_0 &+ \beta_1 I_k^{Peak} + \beta_2 I_k^{Mem} + \beta_3 I_i^{QFT+} + \beta_4 x_i^{Admin} + \beta_5 x_i^{Conc} \\ &+ \beta_6 I_j^{CCR7-CD45RA-} + \beta_7 I_{jk}^{Peak*CCR7-CD45RA-} + \beta_8 I_{jk}^{Mem*CCR7-CD45RA-} \\ &+ \beta_9 I_{ik}^{Peak*QFT+} + \beta_{10} I_{ik}^{Mem*QFT+} + b_{0i},\end{aligned}$$

$$\log(\sigma_{ijk}) = \alpha_0 + \alpha_1 I_j^{CCR7-CD45RA-},$$

$$\text{for} \quad i = 1, ..., I, \quad j = 1, 2 \quad \text{and} \quad k = 1, 2, 3,$$

where

$$b_{0i} \sim N\left(0, \sigma_0^2\right) \qquad \text{and} \qquad \epsilon_{ijk} \sim N\left(0, \sigma_e^2\right).$$

Here i is the participant, j is the memory combination and k is the time point. The standard deviation estimates were $\hat{\sigma}_0 = 0.943$ with 95% confidence interval $(0.854; 1.041)$ and $\hat{\sigma}_e = 1.048$

with 95% confidence interval $(1.008; 1.089)$. No participant-specific random effect was included on time point due to convergences issues and it was not included on stimulus because it resulted in higher GAIC values, indicating an inferior fit. No three-way interactions were included in the model since they did not improve the model fit.

Model diagnostics can be found in Appendix F.7, where there was evidence that the normality assumption of the random effects is violated.

Referring to the parameter estimates seen in Table 7.14, there is evidence that frequency decreases for with higher vaccine concentrations but that number of administrations does not have a strong effect. Variation in the frequency of cells expressing CCR7−CD45RA− seems to be higher than that for cells expressing CCR7+CD45RA−. There is also evidence of strong interactions between time point and QFT status and QFT status and memory combination, as represented in Figure 7.29. Frequency is highest at peak and lowest at baseline for all except for cells expressing CCR7+CD45RA− from QFT+ participants which is lowest at memory. QFT+ participants always have higher frequencies compared to QFT- participants when comparing the same memory combination, and again the profiles of QFT+ participants are flatter. For the same QFT status, the frequency of cells expressing CCR7+CD45RA− is always less than those expressing CCR7−CD45RA−.

Table 7.14: Parameter estimates of the two-level GLMEM fit to the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− or CCR7−CD45RA− after stimulation with ESAT-6.

Variable	Parameter	Estimate	Std. Error	t-value	p-value
μ Coefficients:					
	β_0	-5.905	0.108	-53.725	$< 2 \times 10^{-16}$
Peak	β_1	0.959	0.115	8.340	$< 2 \times 10^{-16}$
Mem.	β_2	0.577	0.117	4.950	8.45×10^{-7}
QFT+	β_3	2.367	0.079	29.795	$< 2 \times 10^{-16}$
Admin.	β_4	-0.004	0.032	-0.126	0.9
Conc.	β_5	-0.010	0.001	-8.932	$< 2 \times 10^{-16}$
CCR7−CD45RA−	β_6	1.182	0.086	13.788	$< 2 \times 10^{-16}$
Peak*CCR7−CD45RA−	β_7	0.774	0.119	6.523	1×10^{-10}
Mem*CCR7−CD45RA−	β_8	0.569	0.119	4.788	1.89×10^{-6}
Peak*QFT+	β_9	-0.699	0.112	-6.249	5.66×10^{-10}
Mem*QFT+	β_{10}	-0.671	0.111	-6.069	1.71×10^{-9}
σ Coefficients:					
	α_0	-0.077	0.029	-2.658	0.008
CCR7−CD45RA−	α_1	0.391	0.042	9.308	$< 2 \times 10^{-16}$

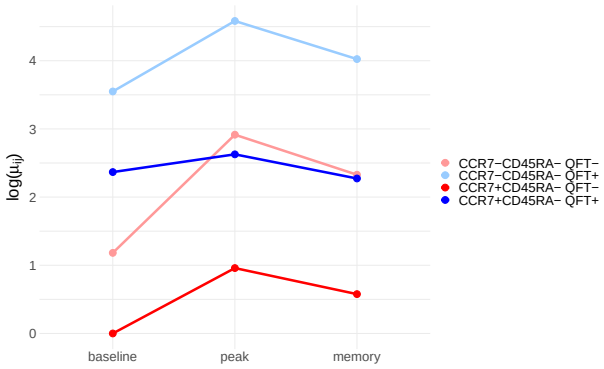


Figure 7.29: The expected value of logged mean $\log(\mu_{ij})$ for changing QFT status, time point and memory combination while keeping all other variables constant.

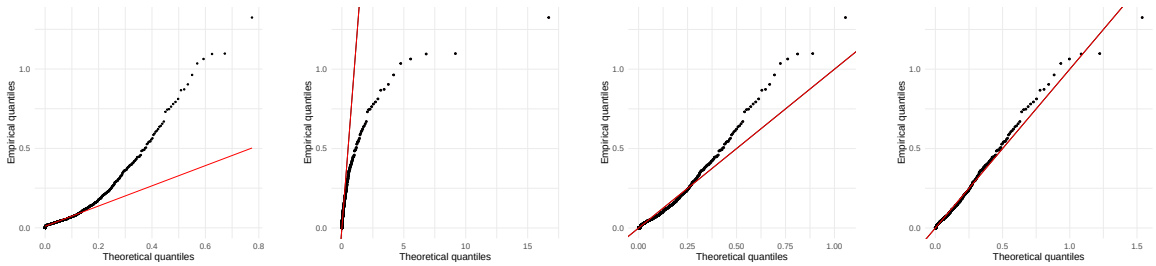
7.2.9 Two-level GLMEM: frequency after Ag85B stimulation

A log-normal, Gamma, Weibull and Tweedie distribution were all fit to response variable, the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− or CCR7−CD45RA− after stimulation with Ag85B. Again, the log-normal, Gamma and Weibull distributions were fit using the `fitdistrplus` [76] package in `R`, while the Tweedie distribution was fit using the `cplm` [77] package. Table 7.15 summaries the resultant AIC and BIC values from the fitted log-normal, Gamma and Weibull distributions. The Weibull distribution had the lowest AIC and BIC compared to the other distributions, suggesting that out of the three distributions the Weibull fit the data best.

Table 7.15: The AIC and BIC after fitting log-normal, Gamma and Weibull distributions to the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− or CCR7−CD45RA− after stimulation with Ag85B.

Distribution	**AIC**	**BIC**
Log-normal	-4727.231	-4716.656
Gamma	-4849.644	-4839.069
Weibull	-4882.598	-4872.023

Referring to the empirical versus theoretical quantiles seen in Figure 7.30, the Weibull clearly fits the data far better than the other three distributions. Thus, based off the qq-plots and the AIC and BIC, it was decided to assume that the conditional response follows a Weibull distribution.



(a) Tweedie (b) Log-normal (c) Gamma (d) Weibull

Figure 7.30: QQ-plots of the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− or CCR7−CD45RA− after stimulation with Ag85B fit to (a) Tweedie, (b) log-normal, (c) Gamma and (b) Weibull distributions.

Following the model building steps laid out in Section 4.3.1, the optimal model chosen is as follows

$$\begin{aligned}\log(\mu_{ijk}) = \beta_0 &+ \beta_1 I_k^{Peak} + \beta_2 I_k^{Mem} + \beta_3 I_i^{QFT+} + \beta_4 x_i^{Admin} + \beta_5 x_i^{Conc} \\ &+ \beta_6 I_j^{CCR7-CD45RA-} + \beta_7 I_{jk}^{Peak*CCR7-CD45RA-} \\ &+ \beta_8 I_{jk}^{Mem*CCR7-CD45RA-} + b_{0i},\end{aligned}$$

$$\log(\sigma_{ij}) = \alpha_0 + \alpha_1 I_j^{CCR7-CD45RA-} + \alpha_2 x_i^{Admin},$$

$$\text{for} \quad i = 1, ..., I, \quad j = 1, 2 \quad \text{and} \quad k = 1, 2, 3,$$

where

$$b_{0i} \sim N\left(0, \sigma_0^2\right) \qquad \text{and} \qquad \epsilon_{ijk} \sim N\left(0, \sigma_e^2\right).$$

Here i is the participant, j is the memory combination and k is the time point. The standard deviation estimates were $\hat{\sigma}_0 = 0.965$ with 95% confidence interval $(0.880; 1.058)$ and $\hat{\sigma}_e = 0.979$ with 95% confidence interval $(0.942; 1.019)$. No participant-specific random effect was included on time point due to convergences issues and it was not included on stimulus because it resulted in higher GAIC values, indicating an inferior fit.

Model diagnostics can be found in Appendix F.8, where there was some evidence that the homogeneity assumption of the residuals and normality assumption of the residuals and random effects are violated.

Referring to the parameter estimates seen in Table 7.16, frequency increases with increasing number of administrations but decreases with higher vaccine concentrations. Cells seem to express CCR7−CD45RA− at higher frequencies compared to CCR7+CD45RA− and there is a higher variation in frequency of cells expressing CCR7−CD45RA−. The variation in cell frequency also increases as number of administrations increases. There is evidence that a strong

interaction exists between time point and QFT status, as represented in Figure 7.31. Frequency is still always higher in QFT+ participants compared to QFT- participants and QFT+ participants still have flatter profiles over time.

Table 7.16: Parameter estimates of the two-level GLMEM fit to the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− or CCR7−CD45RA− after stimulation with Ag85B.

Variable	Parameter	Estimate	Std. Error	t-value	p-value
μ Coefficients:					
	β_0	-5.414	0.087	-62.038	$< 2 \times 10^{-16}$
Peak	β_1	2.064	0.069	29.938	$< 2 \times 10^{-16}$
Mem.	β_2	1.388	0.069	20.024	$< 2 \times 10^{-16}$
QFT+	β_3	1.207	0.068	17.624	$< 2 \times 10^{-16}$
Admin.	β_4	0.149	0.029	5.169	2.75×10^{-7}
Conc.	β_5	-0.003	0.001	-3.181	0.002
CCR7−CD45RA−	β_6	1.061	0.046	23.302	$< 2 \times 10^{-16}$
Peak*QFT+	β_7	-0.783	0.097	-8.099	1.32×10^{-15}
Mem*QFT+	β_8	-0.702	0.097	-7.258	6.95×10^{-13}
σ Coefficients:					
	α_0	-0.160	0.061	-2.642	0.008
CCR7−CD45RA−	α_1	0.440	0.043	10.249	$< 2 \times 10^{-16}$
Admin.	α_2	0.097	0.030	3.250	0.001

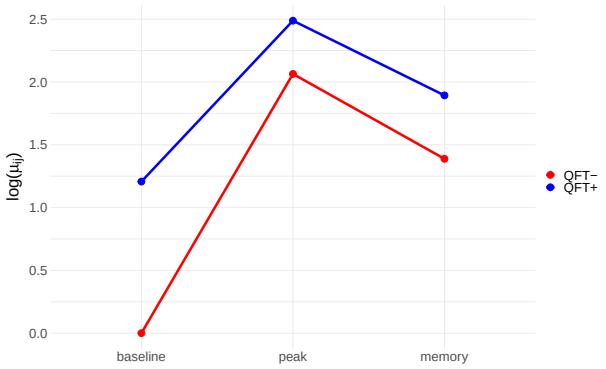


Figure 7.31: The expected value of logged mean $\log(\mu_{ij})$ for changing QFT status and time point while keeping all other variables constant.

7.2.10 Three-level GLMEM: frequency of all memory combinations

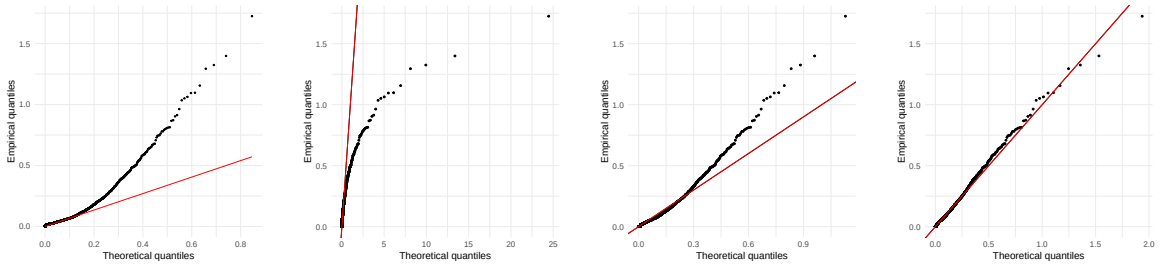
A log-normal, Gamma, Weibull and Tweedie distribution were all fit to response variable, the frequency of CD4 T cells which are cytokine positive and express either CCR7+CD45RA− or CCR7−CD45RA− after stimulation with either Ag85B or ESAT-6. Again, the log-normal, Gamma and Weibull distributions were fit using the `fitdistrplus` [76] package in `R`, while the Tweedie distribution was fit using the `cplm` [77] package. Table 7.17 summaries the resultant AIC and BIC values from the fitted log-normal, Gamma and Weibull distributions. The log-

normal distribution had the lowest AIC and BIC compared to the other distributions, suggesting that out of the three distributions the log-normal fit the data best.

Table 7.17: The AIC and BIC after fitting log-normal, Gamma and Weibull distributions to the frequency of CD4 T cells which are cytokine positive and express CCR7+CD45RA− or CCR7−CD45RA− after stimulation with either Ag85B or ESAT-6.

Distribution	**AIC**	**BIC**
Log-normal	-11325.36	-11313.4
Gamma	-11246.85	-11234.89
Weibull	-11395.16	-11383.2

Referring to the empirical versus theoretical quantiles seen in Figure 7.32, the Weibull clearly fits the data far better than the other three distributions. Thus, based off the qq-plots and the AIC and BIC, it was decided to assume that the conditional response follows a Weibull distribution.



(a) Tweedie (b) Log-normal (c) Gamma (d) Weibull

Figure 7.32: QQ-plots of the frequency of CD4 T cells which are cytokine positive and express either CCR7+CD45RA− or CCR7−CD45RA− after stimulation with either Ag85B or ESAT-6 fit to (a) Tweedie, (b) log-normal, (c) Gamma and (b) Weibull distributions.

Following the model building steps laid out in Section 4.3.1, the optimal model chosen is as follows

$$
\begin{aligned}
\log(\mu_{ijkl}) = {} & \beta_0 + \beta_1 I_l^{Peak} + \beta_2 I_l^{Mem} + \beta_3 I_i^{QFT+} + \beta_4 x_i^{Admin} \\
& + \beta_5 x_i^{Conc} + \beta_6 I_j^{ESAT6} + \beta_7 I_k^{CCR7-CD45RA-} \\
& + \beta_8 I_{il}^{Peak*QFT+} + \beta_9 I_{il}^{Mem*QFT+} + \beta_{10} I_{jl}^{Peak*ESAT6} + \beta_{11} I_{jl}^{Mem*ESAT6} \\
& + \beta_{12} I_{kl}^{Peak*CCR7-CD45RA-} + \beta_{13} I_{kl}^{Mem*CCR7-CD45RA-} \\
& + \beta_{14} I_{il}^{Peak*Admin} + \beta_{15} I_{il}^{Mem*Admin} + \beta_{16} I_{ij}^{ESAT6*QFT+} + \beta_{17} I_{ij}^{ESAT6*Conc} \\
& + \beta_{18} I_{jk}^{ESAT6*CCR7-CD45RA-} + \beta_{19} I_{ik}^{Admin*CCR7-CD45RA-} \\
& + b_{0i},
\end{aligned}
$$

$$
\log(\sigma_{ijkl}) = \alpha_0 + \alpha_1 I_k^{CCR7-CD45RA-} + \alpha_2 x_j^{ESAT6},
$$

$$
\text{for} \quad i = 1, ..., I, \quad j = 1, 2, \quad k = 1, 2 \quad \text{and} \quad l = 1, 2, 3,
$$

where

$$
b_{0i} \sim N\left(0, \sigma_0^2\right) \qquad \text{and} \qquad \epsilon_{ijkl} \sim N\left(0, \sigma_e^2\right).
$$

Here i is the participant, j is the stimulus, k is the memory combination and l is the time point. The standard deviation estimates were $\hat{\sigma}_0 = 0.827$ with 95% confidence interval $(0.749; 0.913)$ and $\hat{\sigma}_e = 1.024$ with 95% confidence interval $(0.997; 1.052)$. No participant-specific random effect was included on time point due to convergences issues and it was not included on stimulus and/or memory combination because it resulted in models with higher GAIC values, indicating inferior fits. No three-way interactions were included in the model since they did not improve the model fit.

Model diagnostics can be found in Appendix F.9, where there was some evidence that the homogeneity assumption of the residuals and normality assumption of the random effects are violated.

Referring to the parameter estimates seen in Table 7.18, the variation in frequency is greater for CCR7−CD45RA− compared to CCR7+CD45RA− and also greater after Ag85B stimulation compared to ESAT-6 stimulation. There are a large number of interaction variables in the model, which does make it difficult to interpret. Notably, time point interacts with all covariates except for concentration and stimulus interacts with all covariates except for number of administrations. As concentration increases, frequency decreases, with the difference even more pronounced after ESAT-6 stimulation compared to Ag85B stimulation. Number of vaccine administrations only has a strong effect on frequency at the memory time point, where more administrations is associated with higher frequencies, although more so for cells expressing CCR7+CD45RA− compared to CCR7−CD45RA−. QFT status has a very strong effect on frequency, with QFT+ participants associated with higher frequencies compared to QFT- participants at all time points, although this difference is more pronounced at baseline.

The frequency of cells expressing CCR7−CD45RA− is always higher than the frequency of cells expressing CCR7+CD45RA−, although this difference decreases as number of administrations increases. Stimulus has a very strong effect on frequency, but it depends a lot on the other covariates. For example, for low vaccine concentrations at baseline, the frequency of cells expressing CCR7+CD45RA− is lower after ESAT-6 stimulation compared to after Ag85B stimulation when looking at QFT- participants, but when looking at QFT+ participants this relationship is reversed.

Table 7.18: Parameter estimates of the three-level GLMEM fit to the frequency of CD4 T cells which are cytokine positive and express either CCR7+CD45RA− or CCR7−CD45RA− after stimulation with either Ag85B or ESAT-6.

Variable	Parameter	Estimate	Std. Error	t-value	p-value
μ Coefficients:					
	β_0	-5.120	0.129	-39.774	$< 2 \times 10^{-16}$
Peak	β_1	1.561	0.145	10.768	$< 2 \times 10^{-16}$
Mem.	β_2	0.759	0.144	5.280	1.4×10^{-7}
QFT+	β_3	1.248	0.068	18.275	$< 2 \times 10^{-16}$
Admin.	β_4	0.083	0.054	1.532	0.126
Conc.	β_5	-0.003	0.001	-2.711	0.007
ESAT6	β_6	-1.080	0.103	-10.518	$< 2 \times 10^{-16}$
CCR7−CD45RA−	β_7	1.277	0.120	10.670	$< 2 \times 10^{-16}$
ESAT6*QFT+	β_8	1.237	0.072	17.237	$< 2 \times 10^{-16}$
Peak*QFT+	β_9	-0.775	0.087	-8.915	$< 2 \times 10^{-16}$
Mem*QFT+	β_{10}	-0.759	0.087	-8.751	$< 2 \times 10^{-16}$
Peak*Admin	β_{11}	0.077	0.060	1.284	0.199
Mem*Admin	β_{12}	0.266	0.061	4.379	1.24×10^{-5}
ESAT6*Conc	β_{13}	-0.008	0.002	-4.157	3.32×10^{-5}
Peak*CCR7−CD45RA−	β_{14}	0.358	0.091	3.956	7.81×10^{-5}
Mem*CCR7−CD45RA−	β_{15}	0.139	0.090	1.536	0.125
ESAT6*CCR7−CD45RA−	β_{16}	0.505	0.075	6.783	1.45×10^{-11}
Peak*ESAT6	β_{17}	-0.469	0.089	-5.298	1.26×10^{-7}
Mem*ESAT6	β_{18}	-0.367	0.088	-4.165	3.21×10^{-5}
Admin*CCR7−CD45RA−	β_{19}	-0.170	0.051	-3.320	0.001
σ Coefficients:					
	α_0	-0.034	0.026	-1.328	0.184
CCR7−CD45RA−	α_1	0.270	0.029	9.188	$< 2 \times 10^{-16}$
ESAT6	α_2	-0.153	0.029	-5.221	1.92×10^{-7}

7.2.11 Comment on the GLMEM's

Deviation from normality and constant variance assumptions for some models may impact on accuracy of the statistical inference. Hence, it is preferred to look at what relationships are con-

sistently seen across all nine models. As such, after assessing the results, the following inference was common amongst all models. For all models, there was no evidence that the demographic variables age, sex and ethnicity had an effect on frequency. For all single- and two-level models except for one, frequency decreased with increasing vaccine concentration. Thus, there is strong evidence that the frequency of CCR7+CD45RA− cells and highly differentiated cells CCR7−CD45RA− decreases with increasing concentration. There was evidence that frequency is higher for QFT+ participants compared to QFT- participants for all single- and two-level models under all conditions, except for the two-level GLMEM of CCR7+CD45RA− where this relationship depended on time point. Thus, there is strong evidence that frequency of highly differentiated cells is greater in QFT+ samples compared to QFT- samples. There was conflicting evidence on the effect of number of administrations on frequency for both memory combinations. All models except for the one showed evidence of an interaction between time point and QFT status. Across all of these models (except the last model), QFT+ samples had flatter profiles over time, i.e. the difference in frequency post-vaccination compared to pre-vaccination was smaller for QFT+ participants compared to QFT- participants, which is expected since QFT+ participants had pre-existing responses to both antigens included in the vaccine. For all two-level GLMEM's grouped by stimulus and the three-level GLMEM, there was evidence of an interaction between QFT status and stimulus. However, the way that these variables interact seem to be differ between the models. For the two-level GLMEM's grouped by cytokine combination and the three-level GLMEM, there was strong evidence that a difference in frequency between the two stimuli existed. However, the nature of the difference differs between models. None of the models showed evidence of an interaction between QFT and concentration.

Thus, in terms of the hypotheses specific to this chapter, there is evidence that the lowest vaccine concentration, 5 µg, should be considered as the preferred vaccine concentration. Since the models had conflicting results on the effect of number of vaccine administrations, it is inconclusive as to which number of vaccine administrations is preferred. There was strong evidence supporting the hypothesis that after vaccination, differentiation profiles differed between QFT+ and QFT- participants, and specifically, there was evidence that frequency of highly differentiated cells was greater in QFT+ samples compared to QFT- samples. There was also evidence supporting the hypothesis that differentiation profiles differed between ESAT-6 and Ag85B stimulation, although there was conflicting evidence as to how they differed. There was no evidence that higher concentrations in QFT+ individuals drives T cells towards more differentiated phenotypes compared to QFT- individuals

In terms of the interpretability of the models, as more levels of grouping were included in the model, the more interaction terms were required to account for the sharing of covariate coefficients between Ag85B and ESAT-6 and/or between memory combinations. The three-level model in particular had a large number of interaction terms and was a very complex model. The advantage of these higher-level models is that they allow for inference related to specific hypotheses, such as comparison between ESAT-6 and Ag85B as well as between memory combinations. A notable disadvantage is the increase in the number of interaction terms and model complexity (although there is also an increase in the number of observations), which can

result in a model that is difficult to interpret.

7.3 Discussion

The standard approach produced conflicting results and very much depended on the response of interest. Given the discordance in results between the two approaches, and the fact that many observations were lost with the filtering step when doing the qualitative analysis, and that the literature supports the hypothesis that T cells recognizing Ag85B and ESAT-6 have different differentiation profiles [79], it seems as if the first approach, which used the magnitude of the antigen-specific immune response, is more informative for the field. In particular, the first approach has the advantage of being easier to implement since there are less filtering steps, retaining all the data and thus having more power and it seems to reflect known biology in a more comprehensive way.

As discussed in Sections 5.3 and 6.4, multiplicity is very much an issue in this chapter as well, and all inference should be treated as exploratory. With this caveat, the inference made using both methods shall be compared. The GLMEM's showed strong evidence supporting the hypothesis that after vaccination, differentiation profiles differed between QFT+ and QFT- participants regardless of stimulus. However, the magnitude of the antigen-specific immune response only provided evidence of this for ESAT-6 stimulated observations, and both of the standard approach response variables provided evidence against this hypothesis for Ag85B stimulated observations. Both the GLMEM's and standard approach using the magnitude of the antigen-specific immune response provided evidence that the differentiation profiles of cells recognising Ag85B were distinct from those recognising ESAT-6, while the qualitative antigen-specific immune response provided evidence against this. The GLMEM's and both standard approaches provided no evidence that higher concentrations in QFT+ individuals drives T cells towards more differentiated phenotypes compared to QFT- individuals after Ag85B stimulation, but disagreed as to whether this was the case after ESAT-6 stimulation as well. Thus there was very little agreement between the different methods, possibly at least in part due to the fact that they all used slightly different responses.

Again, the main advantage that the GLMEM's have is that they provide effect sizes in the form of variable coefficients, where as significance tests focus more on hypothesis testing and p-values. A key disadvantage of the qualitative antigen-specific immune response is that it is very dependant on the fit of MIMOSA for which their are no diagnostic plots available. While MIMOSA, like COMPASS, does provide confidence intervals, p-values and q-values, there is no way to include this in the standard approach plots. Thus, the reader is not provided with a measure of the uncertainty associated with the selected responders. However, GLMEM's are mainly interpreted using the estimated variable coefficients, all of which have associated standard errors, p-values and confidence intervals.

Chapter 8

Concluding Remarks and Suggestions For Future Work

This dissertation used many plots, hypothesis tests and statistical models to investigated several different hypotheses under three different aims. The three aims were to investigate the effects of vaccine regimen, *M.tb* sensitization and antigen specificity on CD4 T cell magnitude, functional quality and differentiation. For each of these aims, an analysis using standard procedures that would typically be employed in the immunology research community was used as well as a mixed effect modelling approach. As well as investigating the three immunological aims, the two approaches were implemented in order to also investigate the statistical aim to compare the standard approach and the mixed effect modelling approach in terms of inference and ease of interpretation. The standard approach consisted of non-parametric pair-wise tests and the data analysis pipelines MIMOSA and COMPASS, while the mixed effect modelling approach made use of GLMEM's with various hierarchical structures as well as LVM's.

When comparing the standard approach to the GLMEM approach in Chapter 5, there was strong agreement between the two methods in terms of the inference made about the CD4 T cell magnitude, with the only disagreement being on the choice of the number of vaccine administrations. In Chapter 6, both GLMEM's and LVM's formed the mixed effect modelling approach, however only one of the four models, the two-level GLMEM modelling ESAT-6 stimulated observations, fit the data adequately. Thus only inference from this model could be compared to the standard approach. Between the two approaches, there was a roughly equal mixture of hypotheses that the two methods agreed on and disagreed on. In Chapter 7, there was a lot of disagreement between the two methods. Based on the results of all three aims, there seems to be consensus that 5 µg of the vaccine should be considered as the preferred vaccine concentration. There was also evidence that QFT+ participants had higher quality immune responses and different differentiation profiles compared to QFT- participants, but that no difference in magnitude of the immune response existed between the two QFT statuses.

Both the hypothesis testing of the standard approach and the mixed effect models struggle with the issue of multiple testing, which prevents one from making any strong inference from the results and consequently the results reported in this project should be treated as exploratory only. The effects of multiple testing can be minimised, although not eliminated, in the case of the mixed effect models by fitting fewer models and focusing more on the model building procedure. In the case of the standard approach, the effects can be minimised by formulating a clear statistical analysis plan and focusing on one or two hypotheses of interest, rather than trying to use one dataset to answer many different hypotheses as done in this project.

Both data analysis pipelines COMPASS and MIMOSA were very useful in terms of dimension reduction and COMPASS in particular could be used in both the standard approach as well as the mixed effect modelling approach. However, while both pipelines provide confidence

intervals, p-values and q-values, these measures of uncertainty could not be included in the plots used in the standard approach nor could they be represented in any practical way for the selection of the biologically meaningful cytokine combinations for the GLMEM's and LVM's. The GLMEM's that did not use these data analysis pipelines had the advantage of providing measures of uncertainty when reporting results through the inclusion of standard errors and p-values with the option to also include confidence intervals. COMPASS and MIMOSA also were not able to provide diagnostic plots to assess the model fit and adherence to model assumptions, while the mixed effect models could provide this.

The main advantage that the GLMEM's and LVM's provided over the standard approach was that they provided effect sizes rather than just focusing on p-values as the hypothesis testing does. Mixed effect models also have the convenience of being able to compare the effects of many different covariates on (sometimes) many different response variables using a single model, while for the standard approach, often multiple plots and hypothesis tests are required in order to answer a single question, making them both tedious to create and tedious to interpret. In general, statistical models have the advantage of describing the data generating process of the entire complex system, while hypothesis testing can only probe one isolated dimension of the system [80]. Thus, GLMEM's and LVM's should be considered for use in future studies investigating immune responses to *M.tb*.

Suggestions for future work include investigating the use of quantile regression for modelling immune responses. The skewness and zero-inflation of the data may be a reason for the inadequate fit of most of the models in this project. Quantile regression is suited to such data because it relates the chosen quantile rather than the mean to a linear combination of covariates [81]. Another area of future work is the use of GLMEM's and/or LVM's for confirmatory analysis by prioritising model fit and carefully planning how the model should be formulated with an immunologist. It may also be worth further investigating the applicability of LVM's to immune responses. LVM's showed promise as an alternative to GLMEM's, especially when the dimensionality of the response variable is large, and it would be interesting to apply them to a similar data set but perform a more rigorous model building procedure then was done in this project. It would also be interesting to include other traits besides the cytokines.

Bibliography

1. Tuberculosis. 2023 Apr. Available from: https://www.who.int/news-room/fact-sheets/detail/tuberculosis
2. The Global Plan to End TB 2023-2030. Geneva, 2022
3. Harris RC, Sumner T, Knight GM, Zhang H, and White RG. Potential impact of tuberculosis vaccines in China, South Africa, and India. Science Translational Medicine 2020; 12:eaax4607
4. Stenger S, Hanson DA, Teitelbaum R, Dewan P, Niazi KR, Froelich CJ, Ganz T, Thoma-Uszynski S, Melián A, Bogdan C, et al. An antimicrobial activity of cytolytic T cells mediated by granulysin. Science 1998; 282:121–5
5. Perez-Velez CM and Marais BJ. Tuberculosis in children. New England Journal of Medicine 2012; 367:348–61
6. Nyabadza F and Winkler D. A simulation age-specific tuberculosis model for the Cape Town metropole. South African Journal of Science 2013; 109:1–7
7. Hesseling A, Caldwell J, Cotton M, Eley B, Jaspan H, Jennings K, Marais B, Nuttall J, Rabie H, Roux P, and Schaaf H. BCG vaccination in South African HIV-exposed infants: risks and benefits. South African medical journal 2009 Feb; 99:88–91
8. BCG - Immunization coverage estimates by country. 2021. Available from: https://apps.who.int/gho/data/view.main.80500?lang=en
9. Dye C, Glaziou P, Floyd K, and Raviglione M. Prospects for tuberculosis elimination. Annual review of public health 2013; 34
10. Dierig A, Tebruegge M, Krivec U, Heininger U, Ritz N, group PTNET, et al. Current status of Bacille Calmette Guérin (BCG) immunisation in Europe–a ptbnet survey and review of current guidelines. Vaccine 2015; 33:4994–9
11. Ellis R, Hatherill M, Tait D, Snowden M, Churchyard G, Hanekom W, Evans T, and Ginsberg A. Innovative clinical trial designs to rationalize TB vaccine development. Tuberculosis 2015; 95:352–7
12. Knight GM, Griffiths UK, Sumner T, Laurence YV, Gheorghe A, Vassall A, Glaziou P, and White RG. Impact and cost-effectiveness of new tuberculosis vaccines in low-and middle-income countries. Proceedings of the National Academy of Sciences 2014; 111:15520–5
13. Demangel C, Garnier T, Rosenkrands I, and Cole ST. Differential effects of prior exposure to environmental mycobacteria on vaccination with Mycobacterium bovis BCG or a recombinant BCG strain expressing RD1 antigens. Infection and immunity 2005; 73:2190–6
14. Mahomed H, Hughes E, Hawkridge T, Minnies D, Simon E, Little F, Hanekom W, Geiter L, and Hussey G. Comparison of Mantoux skin test with three generations of a whole blood IFN-γ assay for tuberculosis infection. The International Journal of Tuberculosis and Lung Disease 2006; 10:310–6

15. Malik A, Gupta M, Mani R, and Bhatnagar R. Single-dose Ag85B-ESAT6–loaded poly (lactic-co-glycolic acid) nanoparticles confer protective immunity against tuberculosis. International Journal of Nanomedicine 2019 :3129–43

16. Seder RA, Darrah PA, and Roederer M. T-cell quality in memory and protection: implications for vaccine design. Nature Reviews Immunology 2008; 8:247–58

17. NIH. macrophage. Accessed: 19-06-2023. Available from: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/macrophage

18. Srivastava S, Ernst JD, and Desvignes L. Beyond macrophages: the diversity of mononuclear cells in tuberculosis. Immunological reviews 2014; 262:179–92

19. Raviglione M and Sulis G. Tuberculosis 2015: burden, challenges and strategy for control and elimination. Infectious disease reports 2016; 8:6570

20. Wolf AJ, Linas B, Trevejo-Nuñez GJ, Kincaid E, Tamura T, Takatsu K, and Ernst JD. Mycobacterium tuberculosis infects dendritic cells with high frequency and impairs their function in vivo. The Journal of Immunology 2007; 179:2509–19

21. Ernst JD. Mechanisms of M. tuberculosis immune evasion as challenges to TB vaccine design. Cell host & microbe 2018; 24:34–42

22. Urdahl KB. Understanding and overcoming the barriers to T cell-mediated immunity against tuberculosis. 2014; 26:578–87

23. Flynn JL. Mutual attraction: does it benefit the host or the bug? Nature immunology 2004; 5:778–9

24. Cadena AM, Fortune SM, and Flynn JL. Heterogeneity in tuberculosis. Nature Reviews Immunology 2017; 17:691–702

25. Bean AG, Roach DR, Briscoe H, France MP, Korner H, Sedgwick JD, and Britton WJ. Structural deficiencies in granuloma formation in TNF gene-targeted mice underlie the heightened susceptibility to aerosol Mycobacterium tuberculosis infection, which is not compensated for by lymphotoxin. The Journal of Immunology 1999; 162:3504–11

26. Chakravarty SD, Zhu G, Tsai MC, Mohan VP, Marino S, Kirschner DE, Huang L, Flynn J, and Chan J. Tumor necrosis factor blockade in chronic murine tuberculosis enhances granulomatous inflammation and disorganizes granulomas in the lungs. Infection and immunity 2008; 76:916–26

27. Caruso AM, Serbina N, Klein E, Triebold K, Bloom BR, and Flynn JL. Mice deficient in CD4 T cells have only transiently diminished levels of IFN-γ, yet succumb to tuberculosis. The Journal of Immunology 1999; 162:5407–16

28. Flynn JL. Immunology of tuberculosis and implications in vaccine development. Tuberculosis 2004; 84:93–101

29. CDC. Adjuvants and Vaccines. Accessed: 14-06-2023. 2022. Available from: https://www.cdc.gov/vaccinesafety/concerns/adjuvants.html#:~:text=An%5C%20adjuvant%5C%20is%5C%20an%5C%20ingredient,adjuvants%5C%20help%5C%20vaccines%5C%20work%5C%20better.

30. Agger EM, Rosenkrands I, Olsen AW, Hatch G, Williams A, Kritsch C, Lingnau K, Von Gabain A, Andersen CS, Korsholm KS, et al. Protective immunity to tuberculosis with Ag85B-ESAT-6 in a synthetic cationic adjuvant system IC31. Vaccine 2006; 24:5452–60

31. CDC. Antigenic Characterization. Accessed: 14-06-2023. 2022. Available from: `https://www.cdc.gov/flu/about/professionals/antigenic.htm`

32. Aagaard C, Hoang T, Dietrich J, Cardona PJ, Izzo A, Dolganov G, Schoolnik GK, Cassidy JP, Billeskov R, and Andersen P. A multistage tuberculosis vaccine that confers efficient protection before and after exposure. Nature medicine 2011; 17:189–94

33. Mustafa AS, Shaban FA, Abal AT, Al-Attiyah R, Wiker HG, Lundin KE, Oftung F, and Huygen K. Identification and HLA restriction of naturally derived Th1-cell epitopes from the secreted Mycobacterium tuberculosis antigen 85B recognized by antigen-specific human CD4+ T-cell lines. Infection and immunity 2000; 68:3933–40

34. Brandt L, Elhay M, Rosenkrands I, Lindblad EB, and Andersen P. ESAT-6 subunit vaccination against Mycobacterium tuberculosis. Infection and immunity 2000; 68:791–5

35. Mustafa A, Oftung F, Amoudy H, Madi N, Abal A, Shaban F, Rosen Krands I, and Andersen P. Multiple epitopes from the Mycobacterium tuberculosis ESAT-6 antigen are recognized by antigen-specific human T cell lines. Clinical infectious diseases 2000; 30:S201–S205

36. Suliman S, Luabeya AKK, Geldenhuys H, Tameris M, Hoff ST, Shi Z, Tait D, Kromann I, Ruhwald M, Rutkowski KT, et al. Dose optimization of H56: IC31 vaccine for tuberculosis-endemic populations. A double-blind, placebo-controlled, dose-selection trial. American journal of respiratory and critical care medicine 2019; 199:220–31

37. Lewinsohn DA, Lewinsohn DM, and Scriba TJ. Polyfunctional CD4+ T cells as targets for tuberculosis vaccination. Frontiers in immunology 2017; 8:1262

38. Ahmed R and Gray D. Immunological memory and protective immunity: understanding their relation. Science 1996; 272:54–60

39. Lanzavecchia A and Sallusto F. Understanding the generation and function of memory T cell subsets. Current opinion in immunology 2005; 17:326–32

40. Sallusto F, Lanzavecchia A, Araki K, and Ahmed R. From vaccines to memory and back. Immunity 2010; 33:451–63

41. Sakai S, Kauffman KD, Schenkel JM, McBerry CC, Mayer-Barber KD, Masopust D, and Barber DL. Cutting edge: control of Mycobacterium tuberculosis infection by a subset of lung parenchyma–homing CD4 T cells. The Journal of Immunology 2014; 192:2965–9

42. Reiley WW, Shafiani S, Wittmer ST, Tucker-Heard G, Moon JJ, Jenkins MK, Urdahl KB, Winslow GM, and Woodland DL. Distinct functions of antigen-specific CD4 T cells during murine Mycobacterium tuberculosis infection. Proceedings of the National Academy of Sciences 2010; 107:19408–13

43. Sologuren I, Boisson-Dupuis S, Pestano J, Vincent QB, Fernández-Pérez L, Chapgier A, Cárdenes M, Feinberg J, García-Laorden MI, Picard C, et al. Partial recessive IFN-γR1 deficiency: genetic, immunological and clinical features of 14 patients from 11 kindreds. Human molecular genetics 2011; 20:1509–23

44. Williams MA, Tyznik AJ, and Bevan MJ. Interleukin-2 signals during priming are required for secondary expansion of CD8+ memory T cells. Nature 2006; 441:890–3

45. Gopal R, Monin L, Slight S, Uche U, Blanchard E, A. Fallert Junecko B, Ramos-Payan R, Stallings CL, Reinhart TA, Kolls JK, et al. Unexpected role for IL-17 in protective immunity against hypervirulent Mycobacterium tuberculosis HN878 infection. PLoS pathogens 2014; 10:e1004099

46. Darrah PA, Zeppa JJ, Wang C, Irvine EB, Bucsan AN, Rodgers MA, Pokkali S, Hackney JA, Kamath M, White AG, et al. Airway T cells are a correlate of iv Bacille Calmette-Guerin-mediated protection against tuberculosis in rhesus macaques. Cell Host & Microbe 2023

47. QIAGEN. QuantiFERON®-TB Gold by QIAGEN The most accurate test for TB infection. Accessed: 14-06-2023. 2016. Available from: `https://www.quantiferon.com/wp-content/uploads/2017/04/PROM-7540-003_1099497_BRO_QFT_0914_WW.pdf`

48. Pai M, Denkinger CM, Kik SV, Rangaka MX, Zwerling A, Oxlade O, Metcalfe JZ, Cattamanchi A, Dowdy DW, Dheda K, et al. Gamma interferon release assays for detection of Mycobacterium tuberculosis infection. Clinical microbiology reviews 2014; 27:3–20

49. Wikell A, Jonsson J, Dyrdak R, Henningsson A, Eringfält A, Kjerstadius T, Hålldin E, Baqir H, Kholod V, Sturegård E, et al. The impact of borderline Quantiferon-TB Gold Plus results for latent tuberculosis screening under routine conditions in a low-endemicity setting. Journal of clinical microbiology 2021; 59:10–1128

50. Lin L, Finak G, Ushey K, Seshadri C, Hawn TR, Frahm N, Scriba TJ, Mahomed H, Hanekom W, Bart PA, et al. COMPASS identifies T-cell subsets correlated with clinical outcomes. Nature biotechnology 2015; 33:610–6

51. Finak G, McDavid A, Chattopadhyay P, Dominguez M, De Rosa S, Roederer M, and Gottardo R. Mixture models for single-cell assays with applications to vaccine studies. Biostatistics 2014; 15:87–101

52. Mearns H, Geldenhuys HD, Kagina BM, Musvosvi M, Little F, Ratangee F, Mahomed H, Hanekom WA, Hoff ST, Ruhwald M, et al. H1: IC31 vaccination is safe and induces long-lived TNF-α+ IL-2+ CD4 T cell responses in M. tuberculosis infected and uninfected adolescents: A randomized trial. Vaccine 2017; 35:132–41

53. Luabeya AKK, Kagina BM, Tameris MD, Geldenhuys H, Hoff ST, Shi Z, Kromann I, Hatherill M, Mahomed H, Hanekom WA, et al. First-in-human trial of the post-exposure tuberculosis vaccine H56: IC31 in Mycobacterium tuberculosis infected and non-infected healthy adults. Vaccine 2015; 33:4130–40

54. Staats J, Divekar A, McCoy JP, and Maecker HT. Guidelines for gating flow cytometry data for immunological assays. Immunophenotyping: Methods and Protocols 2019 :81–104

55. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria, 2021. Available from: https://www.R-project.org/

56. Pourahmad J and Salimi A. Isolated human peripheral blood mononuclear cell (PBMC), a cost effective tool for predicting immunosuppressive effects of drugs and xenobiotics. Iranian journal of pharmaceutical research: IJPR 2015; 14:979

57. Ellis B, Haaland P, Hahne F, Le Meur N, Gopalakrishnan N, Spidlen J, Jiang M, and Finak G. flowCore: Basic structures for flow cytometry data. R package version 2.7.1. 2022

58. Finak G. flowWorkspaceData: A data package containing two flowJo, one diva xml workspace and the associated fcs files as well as three GatingSets for testing the flowWorkspace, openCyto and CytoML packages. R package version 3.6.0. 2021

59. Finak G, Jiang W, and Gottardo R. CytoML for cross-platform cytometry data sharing. Cytometry Part A 2018; 93. R package version 2.6.0. Available from: https://doi.org/10.1002/cyto.a.23663

60. Myors B, Murphy KR, and Wolach A. Statistical power analysis: A simple and general model for traditional and modern hypothesis tests. Routledge, 2010

61. Gelman A, Hill J, and Yajima M. Why we (usually) don't have to worry about multiple comparisons. Journal of research on educational effectiveness 2012; 5:189–211

62. Bender R and Lange S. Adjusting for multiple testing—when and how? Journal of clinical epidemiology 2001; 54:343–9

63. Multiple Testing. Accessed: 26-11-2023. 2012. Available from: https://www.researchgate.net/post/Applying_Bonferronis_correction_correctly_in_several_non-parametric_tests_with_different_DVs/5d329261979fdc0c3a1766cc/citation/download

64. Benjamini Y and Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) 1995; 57:289–300

65. Finak G. The MIMOSA Package. 2020 Nov

66. Storey JD and Tibshirani R. Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences 2003; 100:9440–5

67. Lloyd T, Steigler P, Mpande CA, Rozot V, Mosito B, Schreuder C, Reid TD, Hatherill M, Scriba TJ, Little F, et al. Multidimensional analysis of immune responses identified biomarkers of recent Mycobacterium tuberculosis infection. PLoS Computational Biology 2021; 17:e1009197

68. Mpande CA, Rozot V, Mosito B, Musvosvi M, Dintwe OB, Bilek N, Hatherill M, Scriba TJ, and Nemes E. Immune profiling of Mycobacterium tuberculosis-specific T cells in recent and remote infection. EBioMedicine 2021; 64

69. Stasinopoulos MD, Rigby RA, Heller GZ, Voudouris V, and De Bastiani F. Flexible regression and smoothing: using GAMLSS in R. CRC Press, 2017

70. Rigby RA and Stasinopoulos DM. Generalized additive models for location, scale and shape,(with discussion). Applied Statistics 2005; 54:507–54

71. Pinheiro J and Bates D. Mixed-effects models in S and S-PLUS. Springer science & business media, 2006

72. Loy A, Hofmann H, and Cook D. Model choice and diagnostics for linear mixed-effects models using statistics on street corners. Journal of Computational and Graphical Statistics 2017; 26:478–92

73. Warton DI, Blanchet FG, O'Hara RB, Ovaskainen O, Taskinen S, Walker SC, and Hui FK. So many variables: joint modeling in community ecology. Trends in ecology & evolution 2015; 30:766–79

74. Hui FK. boral–Bayesian ordination and regression analysis of multivariate abundance data in R. Methods in Ecology and Evolution 2016; 7:744–50

75. Hui FKC. boral: Bayesian Ordination and Regression AnaLysis. 2024. R package version 2.0.2

76. Delignette-Muller ML and Dutang C. fitdistrplus: An R Package for Fitting Distributions. Journal of Statistical Software 2015; 64:1–34. DOI: 10.18637/jss.v064.i04

77. Zhang Y. Likelihood-based and Bayesian Methods for Tweedie Compound Poisson Linear Mixed Models. *Statistics and Computing*, 23, 743-757. 2013

78. Lloyd T. Understanding the immune response upon acquisition of Mycobacterium Tuberculosis from Longitudinal Data. University of Cape Town, 2019

79. Moguche AO, Musvosvi M, Penn-Nicholson A, Plumlee CR, Mearns H, Geldenhuys H, Smit E, Abrahams D, Rozot V, Dintwe O, et al. Antigen availability shapes T cell differentiation and function during tuberculosis. Cell host & microbe 2017; 21:695–706

80. Lang KM, Sweet SJ, and Grandfield EM. Getting beyond the null: Statistical modeling as an alternative framework for inference in developmental science. Research in Human Development 2017; 14:287–304

81. Yirga AA, Melesse SF, Mwambi HG, and Ayele DG. Additive quantile mixed effects modelling with application to longitudinal CD4 count data. Scientific reports 2021; 11:17945

82. Dharmadhikari AS, Basaraba RJ, Van Der Walt ML, Weyer K, Mphahlele M, Venter K, Jensen PA, First MW, Parsons S, McMurray DN, et al. Natural infection of guinea pigs exposed to patients with highly drug-resistant tuberculosis. Tuberculosis 2011; 91:329–38

83. Seiler C, Ferreira AM, Kronstad LM, Simpson LJ, Le Gars M, Vendrame E, Blish CA, and Holmes S. CytoGLMM: conditional differential analysis for flow and mass cytometry experiments. BMC bioinformatics 2021; 22:1–14

84. Sallin MA, Kauffman KD, Riou C, Du Bruyn E, Foreman TW, Sakai S, Hoft SG, Myers TG, Gardina PJ, Sher A, et al. Host resistance to pulmonary Mycobacterium tuberculosis infection requires CD153 expression. Nature microbiology 2018; 3:1198–205

85. Du Bruyn E, Ruzive S, Lindestam Arlehamn CS, Sette A, Sher A, Barber DL, Wilkinson RJ, and Riou C. Mycobacterium tuberculosis-specific CD4 T cells expressing CD153 inversely associate with bacterial load and disease severity in human tuberculosis. Mucosal immunology 2021; 14:491–9

86. Chattopadhyay PK, Yu J, and Roederer M. Live-cell assay to detect antigen-specific CD4+ T-cell responses by CD154 expression. Nature protocols 2006; 1:1–6

Appendices

Appendix A

Code

All code created and used in this thesis can be found in this GitHub repository: `https://github.com/Kelly-A-W/Masters-Thesis-`

Appendix B

Data processing diagrams

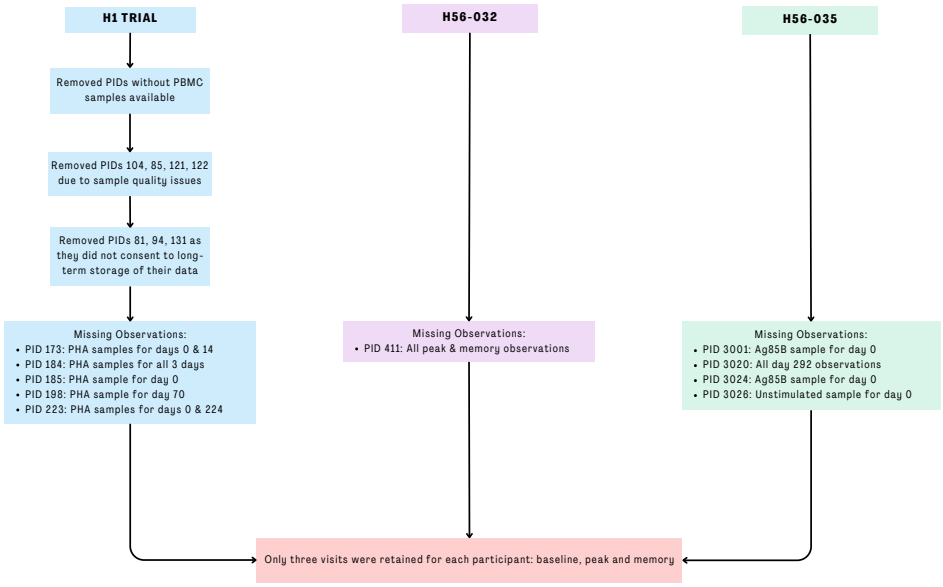


Figure B.1: Diagram showing the data integration process and which observations were removed or missing.

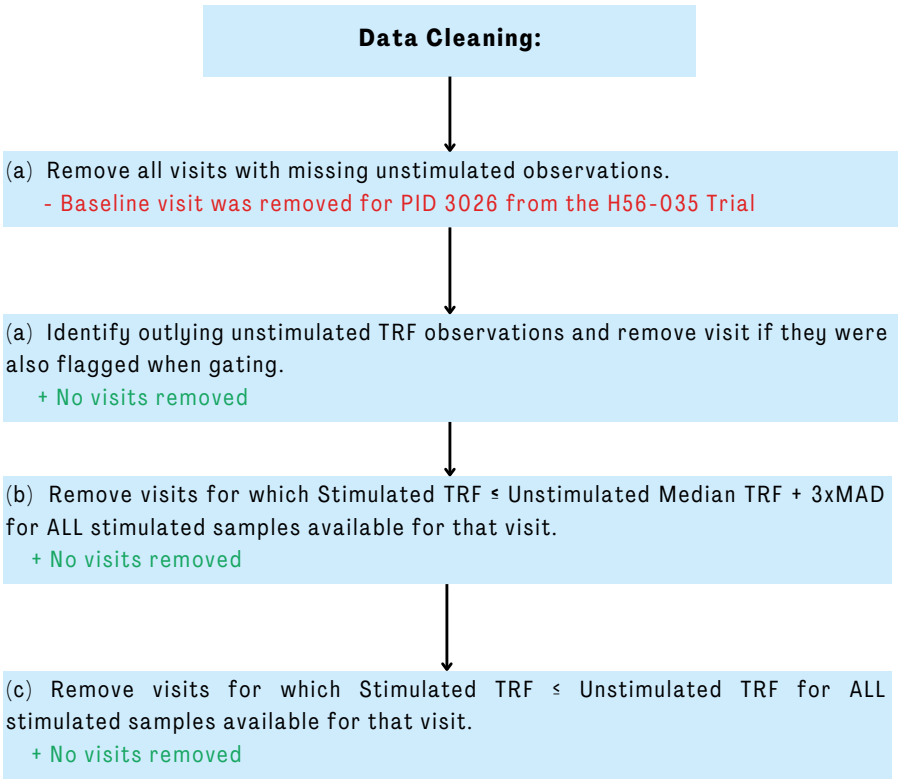
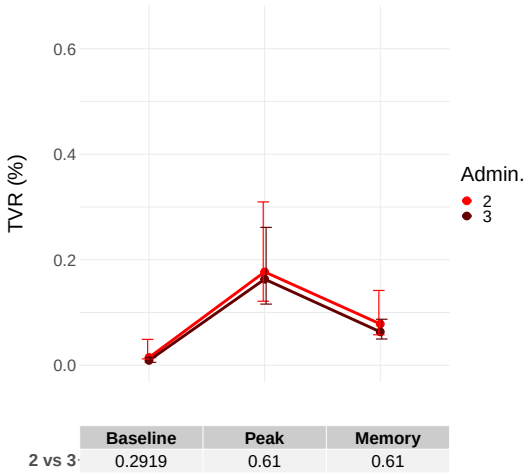


Figure B.2: Diagram showing the data cleaning process for the dataset, indicating which observations were removed.

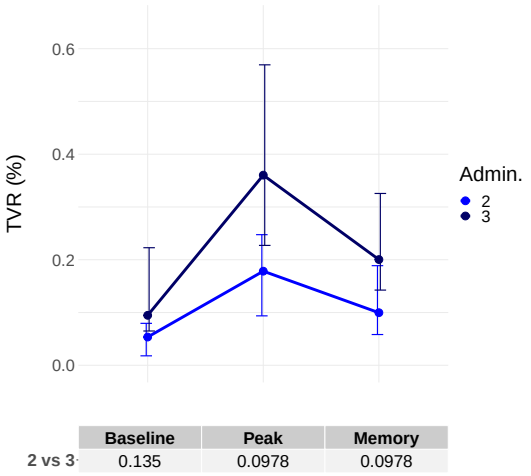
Appendix C

TRF non-parametric approach plots



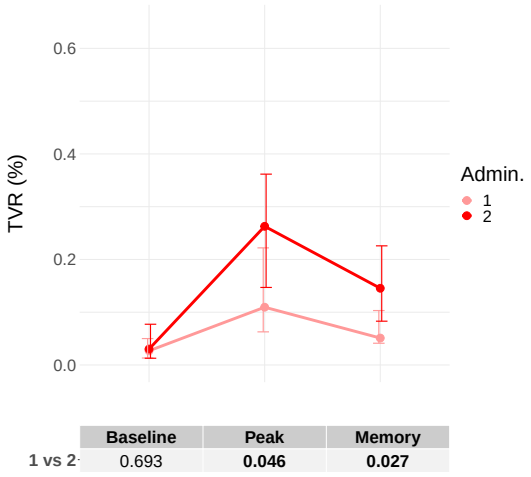
	Baseline	Peak	Memory
2 vs 3	0.2919	0.61	0.61

(a) QFT- with 5 µg



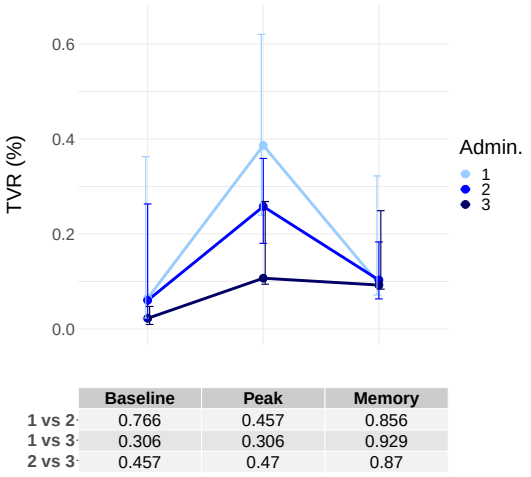
	Baseline	Peak	Memory
2 vs 3	0.135	0.0978	0.0978

(b) QFT- with 15 µg



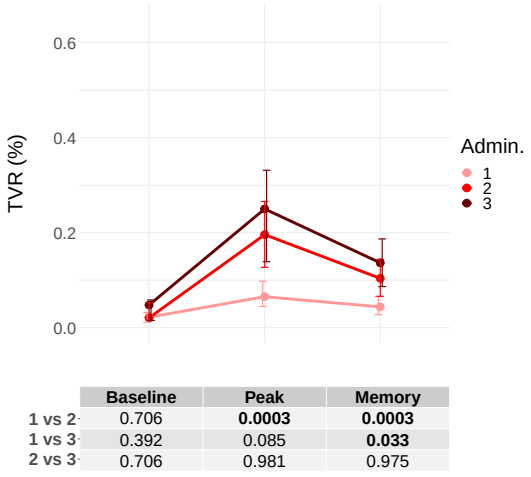
	Baseline	Peak	Memory
1 vs 2	0.693	**0.046**	**0.027**

(c) QFT- with 50 µg



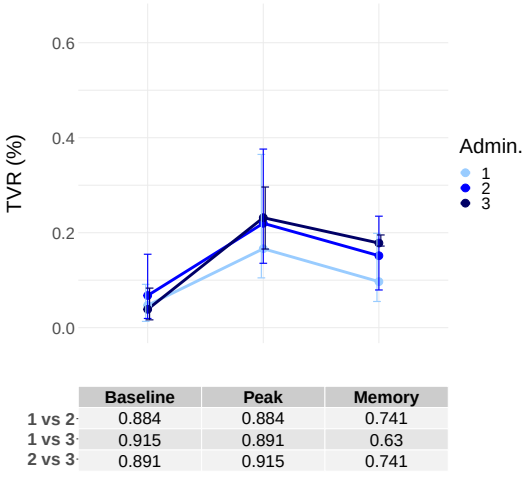
	Baseline	Peak	Memory
1 vs 2	0.766	0.457	0.856
1 vs 3	0.306	0.306	0.929
2 vs 3	0.457	0.47	0.87

(d) QFT+ with 5 µg



	Baseline	Peak	Memory
1 vs 2	0.706	**0.0003**	**0.0003**
1 vs 3	0.392	0.085	**0.033**
2 vs 3	0.706	0.981	0.975

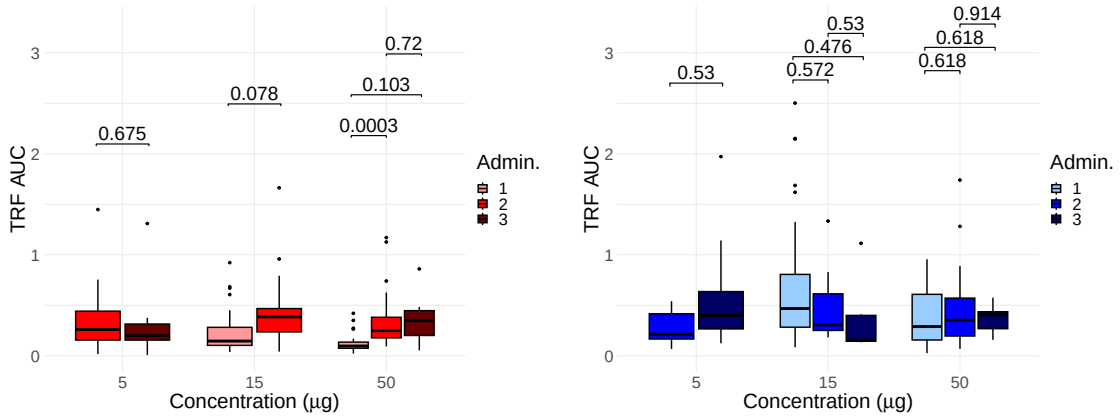
(e) QFT+ with 15 µg



	Baseline	Peak	Memory
1 vs 2	0.884	0.884	0.741
1 vs 3	0.915	0.891	0.63
2 vs 3	0.891	0.915	0.741

(f) QFT+ with 50 µg

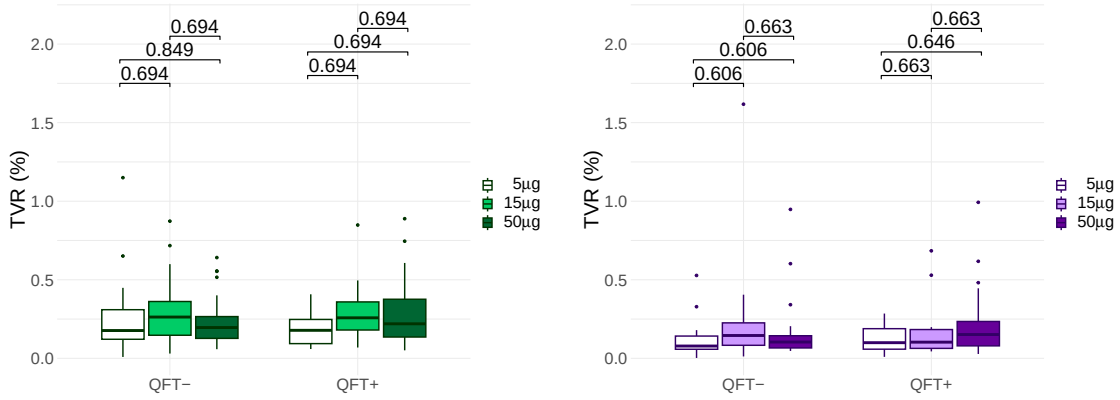
Figure C.1: Shown are the longitudinal profiles of the median Ag85B TRF of CD4 T cells for 1, 2 or 3 administrations of the vaccine (No. of Admin.). The error bars denote the interquartile range of the TVR frequencies. The medians have been plotted for three different time points: baseline, peak and memory. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing the different numbers of administrations for the same time point are shown in the table.



(a) QFT-

(b) QFT+

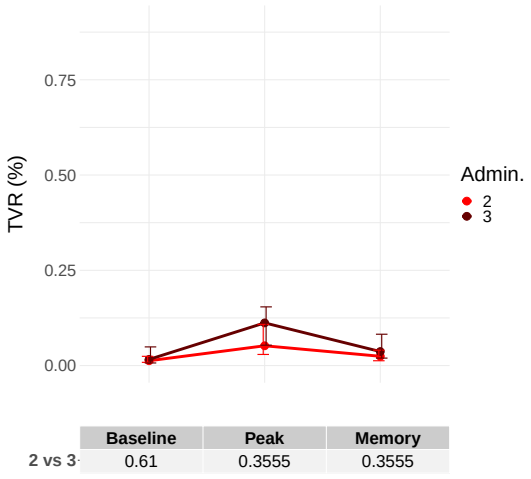
Figure C.2: Shown are box-plots of AUC for each individual's CD4 T cell Ag85B TRF over time. Box-plots of AUCs are plotted for different QFT statuses, vaccine concentrations and number of vaccine administrations received. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing the different numbers of administrations for the same vaccine concentration are shown.



(a) Peak Response

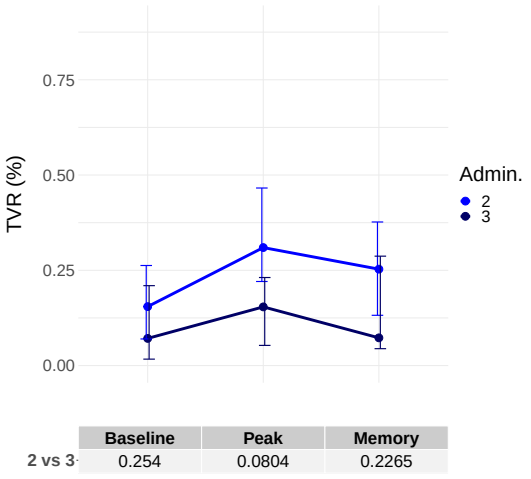
(b) Memory Response

Figure C.3: Shown are the box-plots for the peak and memory CD4 T cell Ag85B TRF for the participants who received two administrations of the vaccine. Box-plots are plotted for different QFT statuses as well as for different vaccine concentrations. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing the different vaccine concentrations for the same QFT status are shown.



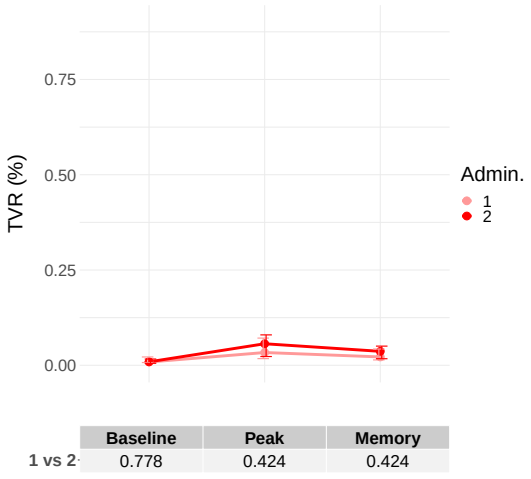
	Baseline	Peak	Memory
2 vs 3	0.61	0.3555	0.3555

(a) QFT- with 5 µg



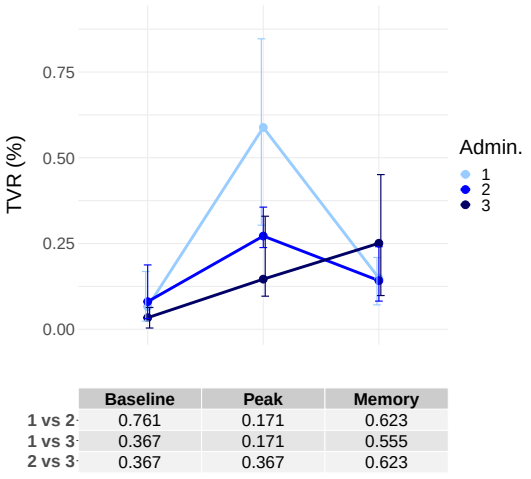
	Baseline	Peak	Memory
2 vs 3	0.254	0.0804	0.2265

(b) QFT- with 15 µg



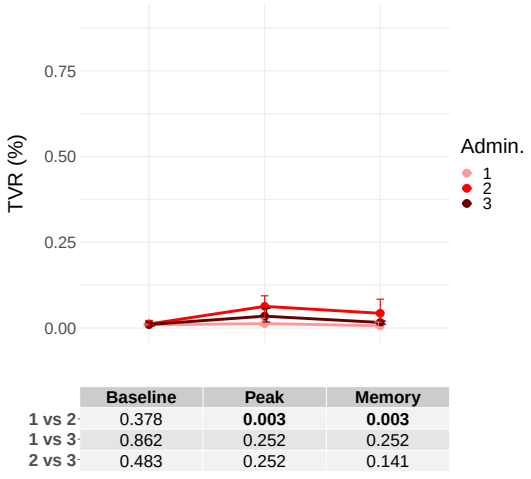
	Baseline	Peak	Memory
1 vs 2	0.778	0.424	0.424

(c) QFT- with 50 µg



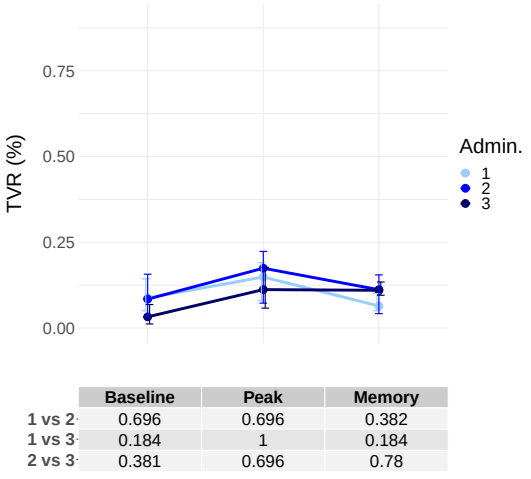
	Baseline	Peak	Memory
1 vs 2	0.761	0.171	0.623
1 vs 3	0.367	0.171	0.555
2 vs 3	0.367	0.367	0.623

(d) QFT+ with 5 µg



	Baseline	Peak	Memory
1 vs 2	0.378	**0.003**	**0.003**
1 vs 3	0.862	0.252	0.252
2 vs 3	0.483	0.252	0.141

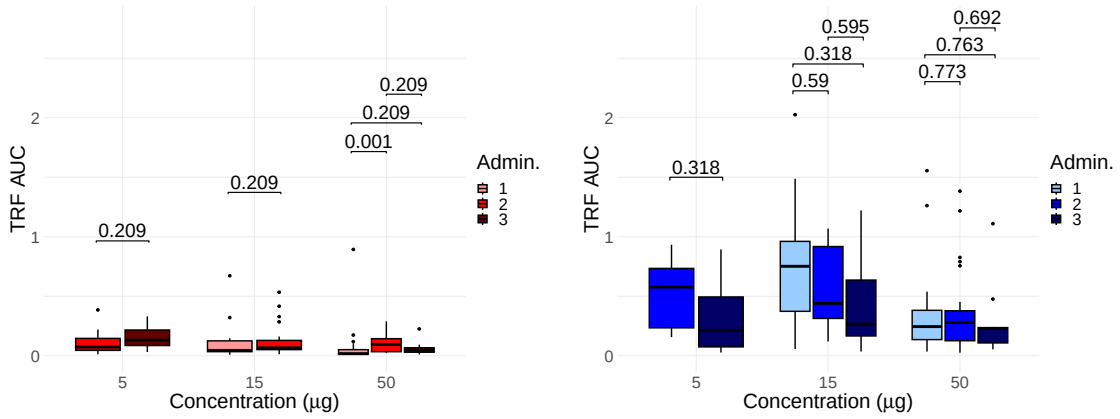
(e) QFT+ with 15 µg



	Baseline	Peak	Memory
1 vs 2	0.696	0.696	0.382
1 vs 3	0.184	1	0.184
2 vs 3	0.381	0.696	0.78

(f) QFT+ with 50 µg

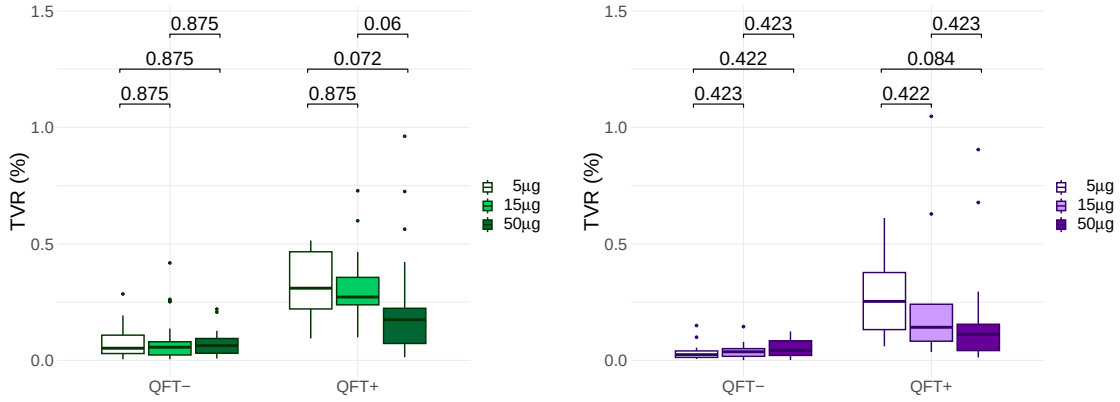
Figure C.4: Shown are the longitudinal profiles of the median ESAT-6 TRF of CD4 T cells for 1, 2 or 3 administrations of the vaccine (No. of Admin.). The error bars denote the interquartile range of the TVR frequencies. The medians have been plotted for three different time points: baseline, peak and memory. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing the different numbers of administrations for the same time point are shown in the table.



(a) QFT-

(b) QFT+

Figure C.5: Shown are box-plots of AUC for each individual's CD4 T cell ESAT-6 TRF over time. Box-plots of AUCs are plotted for different QFT statuses, vaccine concentrations and number of vaccine administrations received. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing the different numbers of administrations for the same vaccine concentration are shown.



(a) Peak Response

(b) Memory Response

Figure C.6: Shown are the box-plots for the peak and memory CD4 T cell ESAT-6 TRF for the participants who received two administrations of the vaccine. Box-plots are plotted for different QFT statuses as well as for different vaccine concentrations. Benjamini Hochberg adjusted p-values, determined by (unpaired) Mann Whitney tests comparing the different vaccine concentrations for the same QFT status are shown.

Appendix D

TRF GLMEM's diagnostic plots

D.1 Single-level GLMEM: TRF after ESAT-6 stimulation

The homogeneity of the within-participant residual variance across the model covariates was assessed using box-plots of the normalised quantile residuals, as seen in Figure D.1. The variation of the residuals seems to decrease as time point increases, suggesting that there is a lot of variation at baseline that is unexplained by the model and consequently that the assumption of homoscedastic within-participant errors may be invalid. There also seems to be a slight difference in residual variation for different vaccine concentrations and doses, with the lowest concentration and the lowest dose being associated with smaller variance. There does not seem to be a difference in variation between QFT+ and QFT- participants until subdivided by number of administrations, where QFT- participants seem to be associated with a higher residual variance when administered two or three doses compared to QFT+ participants. For all plots, the residuals are centered around zero, suggesting that the assumption that the within-participant errors have a mean of zero is valid. Plot (a) also provides insight into whether the parameterization of time point in the model was appropriate. Since the box-plots are linearly arranged, this suggests that the parameterization of time point in the model is appropriate.

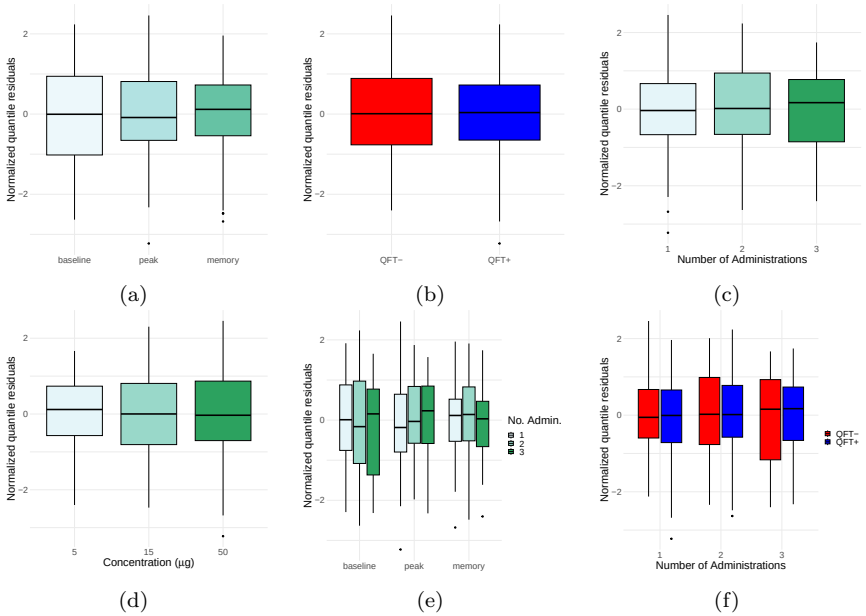


Figure D.1: Box-plots of the normalised quantile residuals plot against the model covariates for the single-level GLMEM modelling the TRF after ESAT-6 stimulation.

To further assess the within-group errors assumptions, the residuals versus fitted values were plot, as seen in Figure D.2. There is no obvious pattern in the plot and no clear change in variation of the residuals with increasing fitted values, providing no evidence against the assumption of constant variance of the within-group errors.

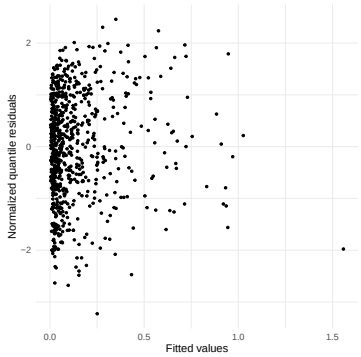
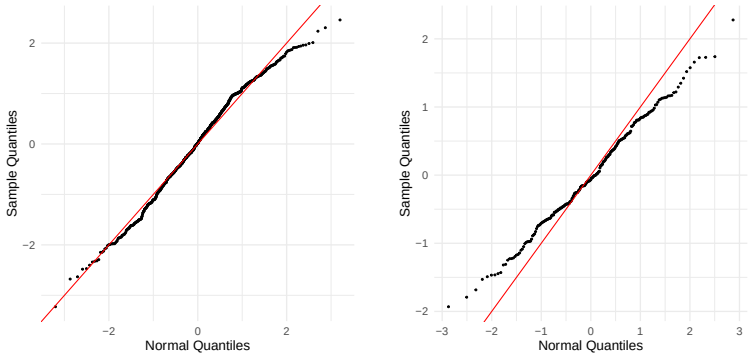


Figure D.2: Scatter plot of the normalised quantile residuals plot against the fitted values of the single-level GLMEM modelling the TRF after ESAT-6 stimulation.

The normality assumptions of the within-participant errors as well as the random effects were checked using qq-plots, as seen in Figure D.3. The residuals keep fairly close to the reference line, suggesting that the assumption of normality may be valid for the chosen model. However, the estimated random effects deviate a significantly from the reference line, suggesting that the random effects may not be normally distributed. However, this also does not compromise model fit, although it may effect model inference.



(a) Residuals (b) Random Effects

Figure D.3: QQ-plots of (a) the residuals and (b) the random effects of the single-level GLMEM modelling the TRF after ESAT-6 stimulation.

D.2 Single-level GLMEM: TRF after Ag85B stimulation

The homogeneity of the within-participant residual variance across the model covariates was assessed using the box-plots of the normalised quantile residuals seen in Figure D.4. The variation of the residuals at baseline is much larger than that at peak and memory, suggesting that there is a lot of unexplained variation associated with baseline observations and that the assumption of homoscedastic within-participant errors may be invalid. However, since the box-plots of the time points are linearly arranged, there is evidence that parameterization of time point in the model was appropriate. With regards to the other variables, there does not seem to be much difference in variation between the different variable levels. For all plots, the residuals are centered around zero, suggesting that the assumption that the within-participant errors

have a mean of zero is valid.

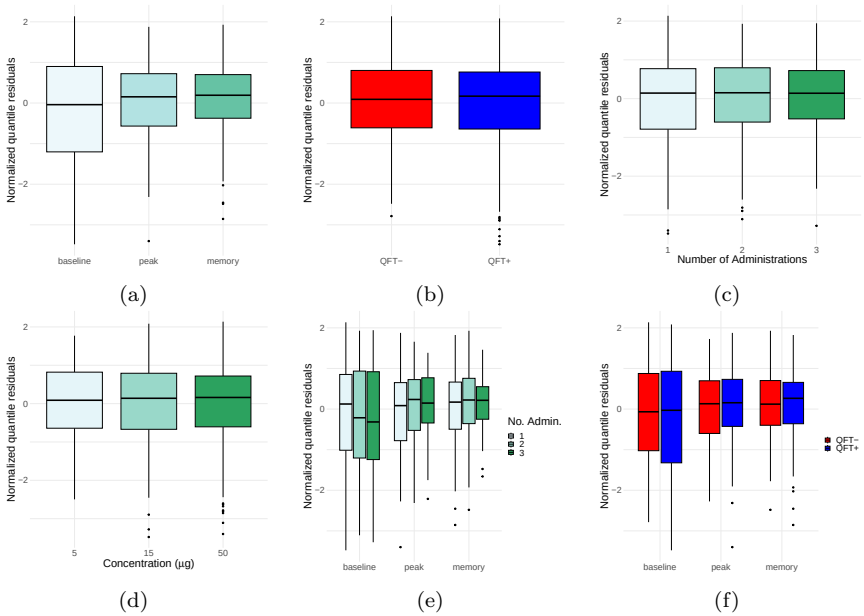


Figure D.4: Box-plots of the normalised quantile residuals plot against the model covariates for the single-level GLMEM modelling the TRF after Ag85B stimulation.

To further assess the within-group errors assumptions, the residuals versus fitted values were plot, as seen in Figure D.5. There does seem to be a slight funnel effect with variation decreasing as fitted values increase. This suggests that the model better captures larger, more extreme frequencies.

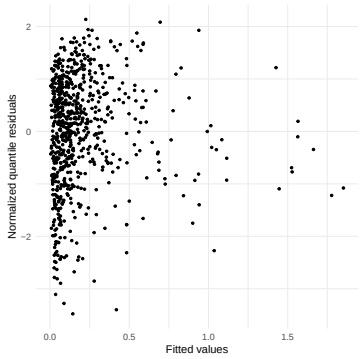
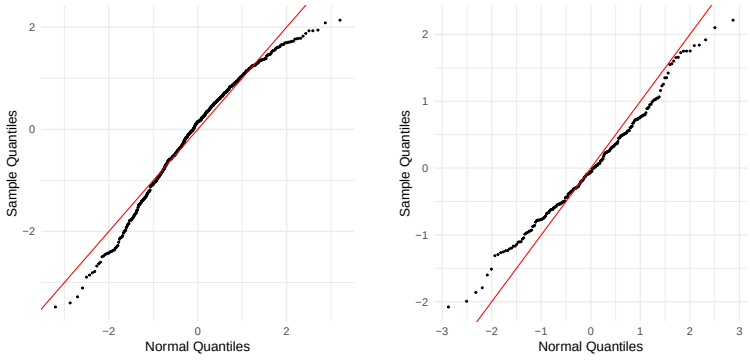


Figure D.5: Scatter plot of the normalised quantile residuals plot against the fitted values of the single-level GLMEM modelling the TRF after Ag85B stimulation.

The normality assumptions of the within-participant errors as well as the random effects were checked using qq-plots, as seen in Figure D.6. The residuals and the random effects deviate substantially from the reference line, suggesting that the assumption of normality is violated for both the residuals and the random effects.

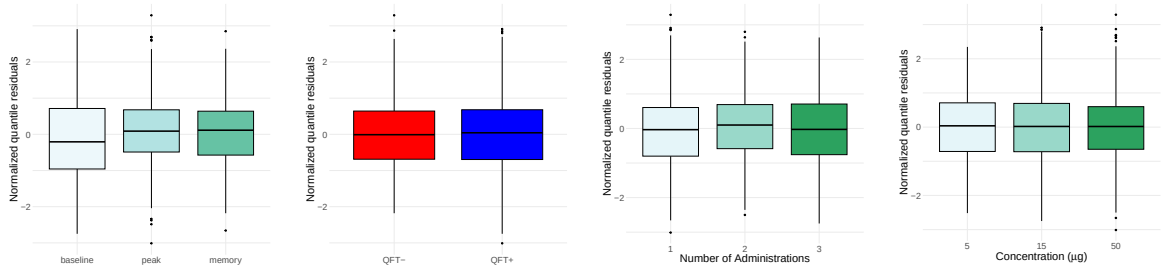


(a) Residuals (b) Random Effects

Figure D.6: QQ-plots of (a) the residuals and (b) the random effects of the single-level GLMEM modelling the TRF after Ag85B stimulation.

D.3 Two-level GLMEM: TRF after stimulation with either ESAT-6 or Ag85B

The homogeneity of the within-participant residual variance across the model covariates was assessed using box-plots of the normalised quantile residuals, as seen in Figure D.7. The variation of the residuals at baseline is again slightly larger than that at peak and memory, suggesting that there is a lot of unexplained variation associated with baseline observations and that the assumption of homoscedastic within-participant errors may be invalid. Otherwise, there is no obvious difference in variation between the levels of the remaining variables.



(a) (b) (c) (d)

Figure D.7: Box-plots of the normalised quantile residuals plot against the model covariates for the two-level GLMEM modelling the TRF.

To further assess the within-group errors assumptions, the residuals versus fitted values were plot, as seen in Figure D.8. There is a prominent funnel effect with variation decreasing as fitted values increase. This suggests that the model better captures larger, more extreme frequencies.

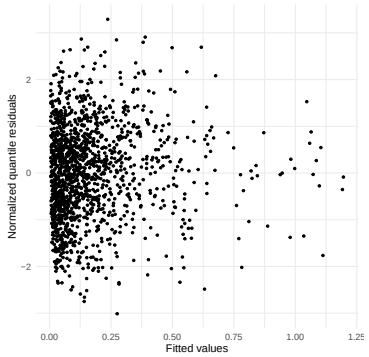
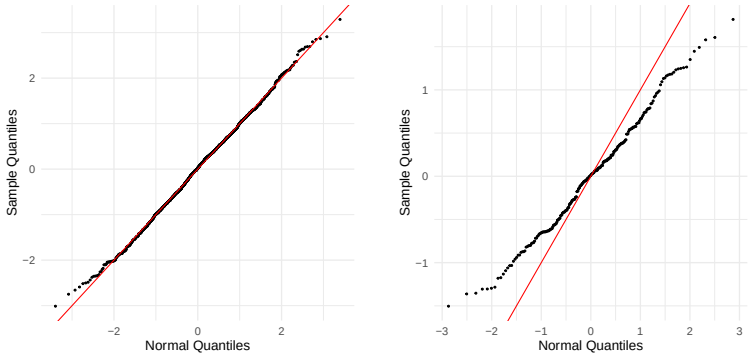


Figure D.8: Scatter plot of the normalised quantile residuals plot against the fitted values of the two-level GLMEM modelling the TRF.

The normality assumptions of the within-participant errors as well as the random effects were checked using qq-plots, as seen in Figure D.9. The residuals keep very close to the reference line, suggesting that the assumption of normality is valid for the chosen model. However, the estimated random effects deviate a significantly from the reference line, suggesting that the random effects may not be normally distributed.



(a) Residuals (b) Random Effects

Figure D.9: QQ-plots of (a) the residuals and (b) the random effects of the two-level GLMEM modelling the TRF.

Appendix E

Cytokine combination GLMEM's diagnostic plots

E.1 Two-level GLMEM after ESAT-6 stimulation

The homogeneity of the within-participant residual variance across the model covariates was assessed using box-plots of the normalised quantile residuals, as seen in Figure E.1. There is no obvious difference in variation between the levels of any of the variables, suggesting that the assumption of homoscedastic within-participant errors may be valid.

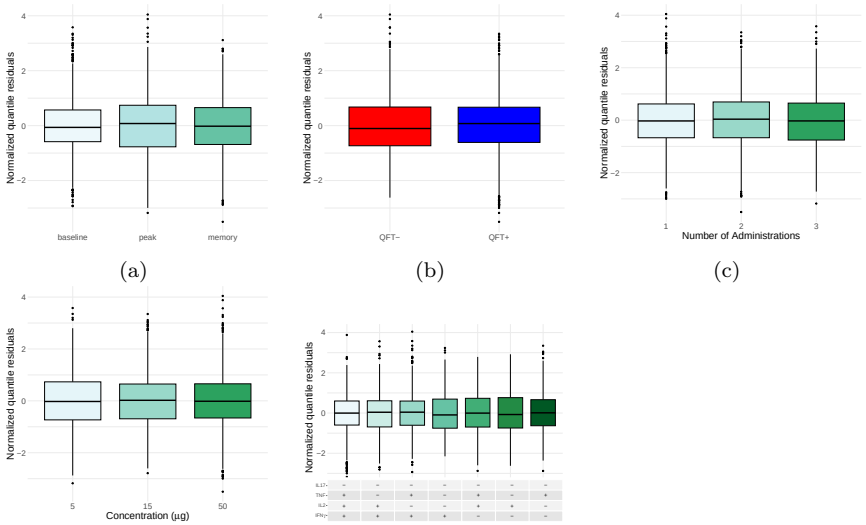


Figure E.1: Box-plots of the normalised quantile residuals plot against the model covariates for the two-level GLMEM modelling the cytokine combination frequency after ESAT-6 stimulation.

To further assess the within-group errors assumptions, the residuals versus fitted values were plot, as seen in Figure E.2. There is no obvious pattern in the residuals and no clear change in variation with increasing fitted values, providing no evidence against the assumption of constant variance of the within-group errors.

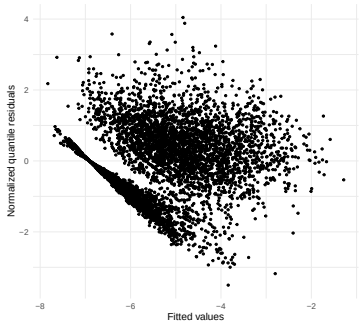
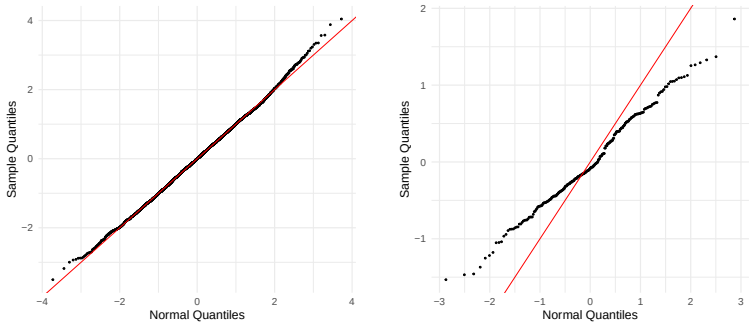


Figure E.2: Scatter plot of the normalised quantile residuals plot against the fitted values of the two-level GLMEM modelling the cytokine combination frequency after ESAT-6 stimulation.

The normality assumptions of the within-participant errors as well as the random effects were checked using qq-plots, as seen in Figure E.3. The random effects deviate substantially from the reference line, suggesting that the assumption of normal random effects may not be valid. The residuals however lie very close to the reference line, suggesting that the assumption of normal residuals is valid.



(a) Residuals (b) Random Effects

Figure E.3: QQ-plots of (a) the residuals and (b) the random effects of the two-level GLMEM modelling the cytokine combination frequency after ESAT-6 stimulation.

E.2 Two-level GLMEM after Ag85B stimulation

The homogeneity of the within-participant residual variance across the model covariates was assessed using box-plots of the normalised quantile residuals, as seen in Figure E.4. There is no obvious difference in variation between the levels of any of the variables, suggesting that the assumption of homoscedastic within-participant errors may be valid.

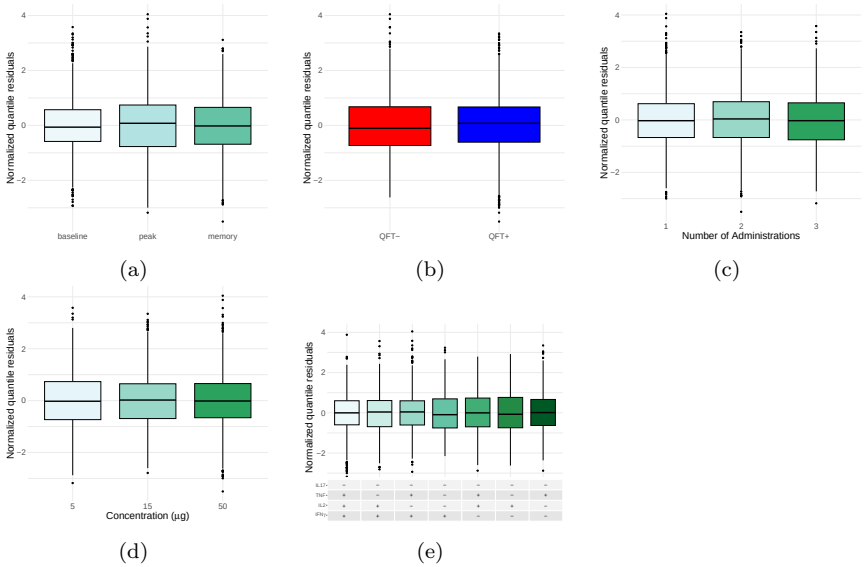


Figure E.4: Box-plots of the normalised quantile residuals plot against the model covariates for the two-level GLMEM modelling the cytokine combination frequency after Ag85B stimulation.

To further assess the within-group errors assumptions, the residuals versus fitted values were plot, as seen in Figure E.5. There is no obvious pattern in the residuals and no clear change in variation with increasing fitted values, providing no evidence against the assumption of constant variance of the within-group errors.

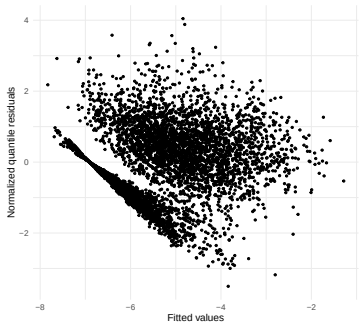
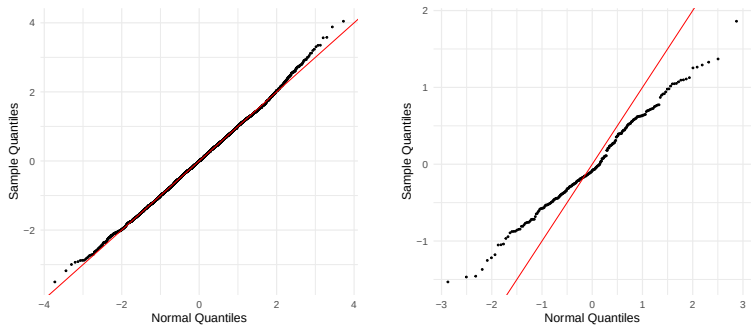


Figure E.5: Scatter plot of the normalised quantile residuals plot against the fitted values of the two-level GLMEM modelling the cytokine combination frequency after Ag85B stimulation.

The normality assumptions of the within-participant errors as well as the random effects were checked using qq-plots, as seen in Figure E.6. The random effects deviate substantially from the reference line, suggesting that the assumption of normal random effects may not be valid. The residuals however lie very close to the reference line, suggesting that the assumption of normal residuals is valid.



(a) Residuals (b) Random Effects

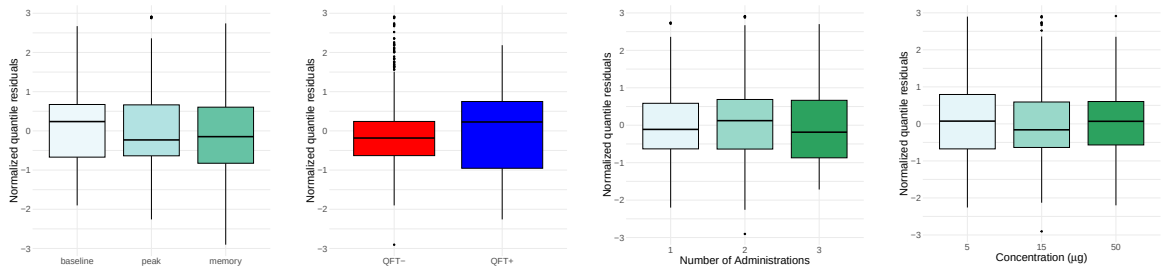
Figure E.6: QQ-plots of (a) the residuals and (b) the random effects of the two-level GLMEM modelling the cytokine combination frequency after Ag85B stimulation.

Appendix F

Memory combination GLMEM's diagnostic plots

F.1 Single-level GLMEM: frequency of CCR7+CD45RA− after ESAT-6 stimulation

The homogeneity of the within-participant residual variance across the model covariates was assessed using box-plots of the normalised quantile residuals, as seen in Figure F.1. The variation of the residuals for QFT+ participants is much larger than that of QFT- participants, suggesting that there is a lot of variation associated with QFT+ participants that is unexplained by the model and consequently that the assumption of homoscedastic within-participant errors may be invalid. However, this does not compromise model fit, especially since there does not seem to be much difference in variation between the different levels of the other variables. For all plots, the residuals are centered around zero, suggesting that the assumption that the within-participant errors have a mean of zero is valid. Plot (a) also provides insight into whether the parameterization of time point in the model was appropriate. Since the box-plots are linearly arranged, this suggests that the parameterization of time point in the model is appropriate.



(a) (b) (c) (d)

Figure F.1: Box-plots of the normalised quantile residuals plot against the model covariates for the single-level GLMEM modelling the frequency of CCR7+CD45RA− expression after ESAT-6 stimulation.

To further assess the within-group errors assumptions, the residuals versus fitted values were plot, as seen in Figure F.2. The diagonal lines seen in the plot are due to the zero-inflation in the observed response as well as because the response variable can only take on positive values, and thus zeros acts as a cut-off. Otherwise, there is no obvious pattern in the plot and no clear change in variation of the residuals with increasing fitted values, providing no evidence against the assumption of constant variance of the within-group errors.

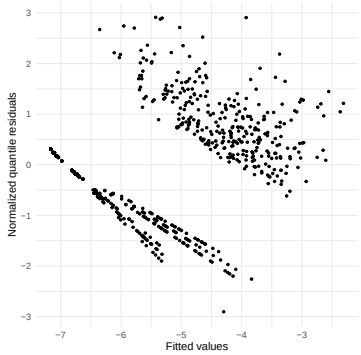
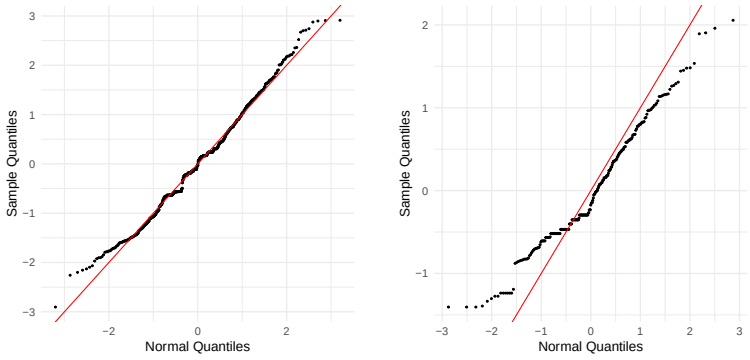


Figure F.2: Scatter plot of the normalised quantile residuals plot against the fitted values of the single-level GLMEM modelling the frequency of CCR7+CD45RA− expression after ESAT-6 stimulation.

The normality assumptions of the within-participant errors as well as the random effects were checked using qq-plots, as seen in Figure F.3. The residuals keep fairly close to the reference line, suggesting that the assumption of normality may be valid for the chosen model. However, the estimated random effects deviate a significantly from the reference line, suggesting that the random effects may not be normally distributed. However, this also does not compromise model fit, although it may effect model inference.



(a) Residuals (b) Random Effects

Figure F.3: QQ-plots of (a) the residuals and (b) the random effects of the single-level GLMEM modelling the frequency of CCR7+CD45RA− expression after ESAT-6 stimulation.

F.2 Single-level GLMEM: frequency of CCR7−CD45RA− after ESAT-6 stimulation

The homogeneity of the within-participant residual variance across the model covariates was assessed using the box-plots of the normalised quantile residuals seen in Figure F.4. The variation of the residuals at baseline is much larger than that at peak and memory, suggesting that there is a lot of unexplained variation associated with baseline observations and that the assumption of homoscedastic within-participant errors may be invalid. However, since the box-plots of the time points are linearly arranged, there is evidence that parameterization of time point in the model was appropriate. With regards to the other variables, there does not seem to be much

difference in variation between the different variable levels. For all plots, the residuals are centered around zero, suggesting that the assumption that the within-participant errors have a mean of zero is valid.

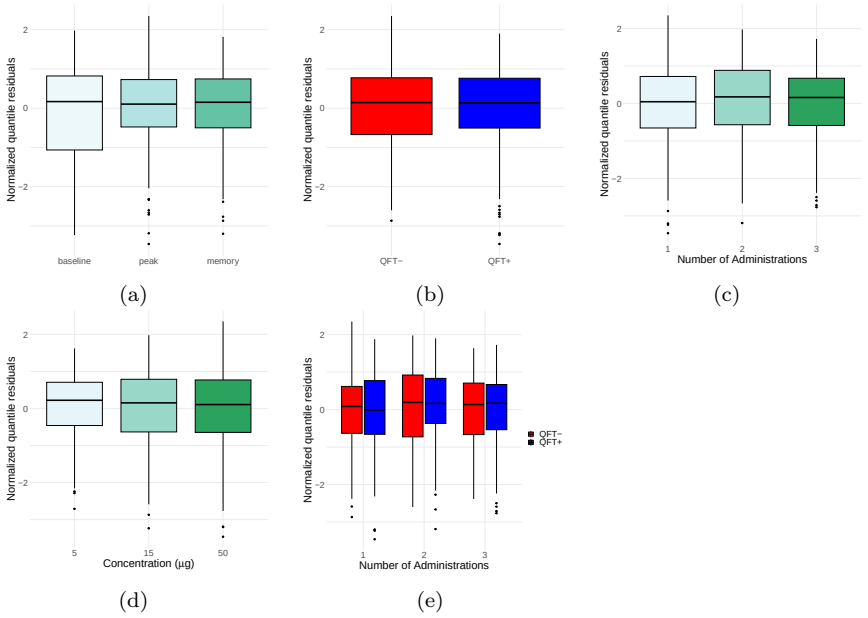


Figure F.4: Box-plots of the normalised quantile residuals plot against the model covariates for the single-level GLMEM modelling the frequency of CCR7−CD45RA− expression after ESAT-6 stimulation.

To further assess the within-group errors assumptions, the residuals versus fitted values were plot, as seen in Figure F.5. There is no obvious pattern in the residuals and no clear change in variation with increasing fitted values, providing no evidence against the assumption of constant variance of the within-group errors.

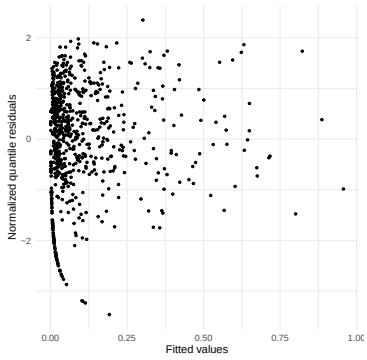
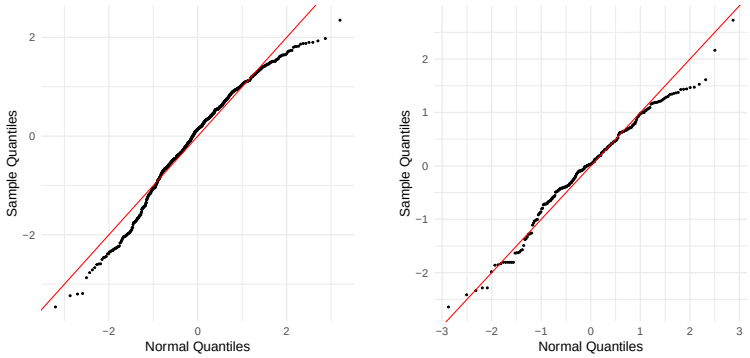


Figure F.5: Scatter plot of the normalised quantile residuals plot against the fitted values of the single-level GLMEM modelling the frequency of CCR7−CD45RA− expression after ESAT-6 stimulation.

The normality assumptions of the within-participant errors as well as the random effects were checked using qq-plots, as seen in Figure F.15. The points in both plots deviate from the reference line substantially at the tails, suggesting that the assumption of normality may not be valid for both the residuals and the random effects.



(a) Residuals (b) Random Effects

Figure F.6: QQ-plots of (a) the residuals and (b) the random effects of the single-level GLMEM modelling the frequency of CCR7−CD45RA− expression after ESAT-6 stimulation.

F.3 Single-level GLMEM: frequency of CCR7+CD45RA− after Ag85B stimulation

The homogeneity of the within-participant residual variance across the model covariates was assessed using box-plots of the normalised quantile residuals, as seen in Figure F.7. The variation of the residuals at baseline is again much larger than that at peak and memory, suggesting suggesting that there is a lot of unexplained variation associated with baseline observations and that the assumption of homoscedastic within-participant errors may be invalid. The residual variation also seems to be slightly larger for QFT- participants compared to QFT+ participants and this difference becomes more pronounced when subdivided by time point. It seems as if at baseline the unexplained variation is larger for QFT+ participants compared to QFT- participants, but at peak and memory the opposite is true. With regards to the other variables, there does not seem to be much difference in variation between the different variable levels. For all plots, the residuals are centered around zero, suggesting that the assumption that the within-subject errors have a mean of zero is valid.

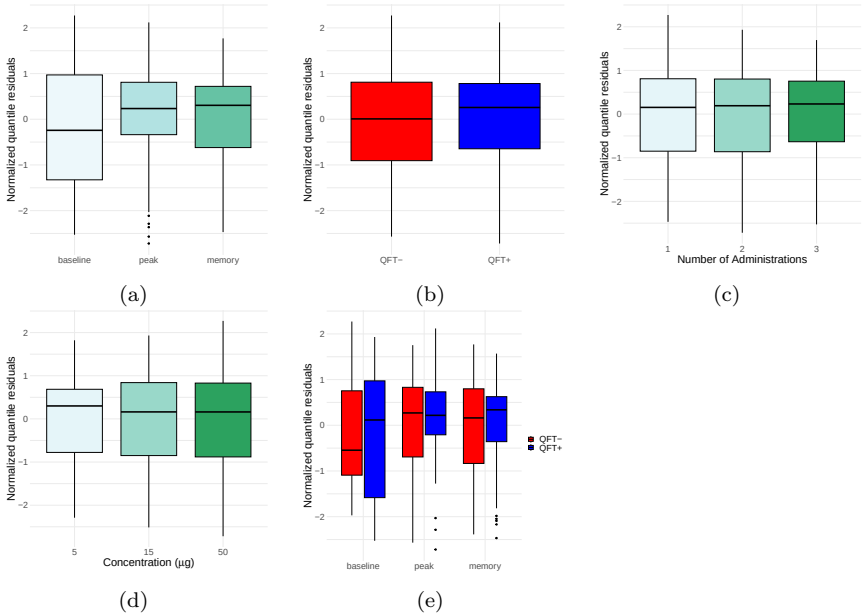


Figure F.7: Box-plots of the normalised quantile residuals plot against the model covariates for the single-level GLMEM modelling the frequency of CCR7+CD45RA− expression after Ag85B stimulation.

To further assess the within-group errors assumptions, the residuals versus fitted values were plot, as seen in Figure F.8. There is no obvious pattern in the residuals and no clear change in variation with increasing fitted values, providing no evidence against the assumption of constant variance of the within-group errors.

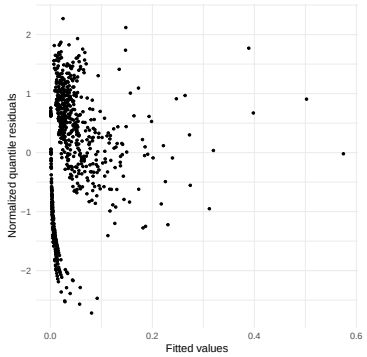
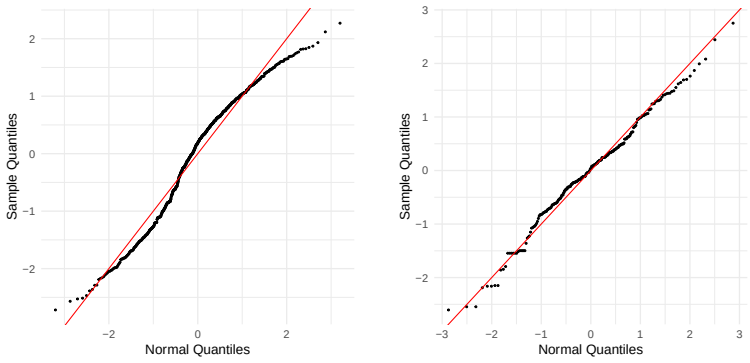


Figure F.8: Scatter plot of the normalised quantile residuals plot against the fitted values of the single-level GLMEM modelling the frequency of CCR7+CD45RA− expression after Ag85B stimulation.

The normality assumptions of the within-participant errors as well as the random effects were checked using qq-plots, as seen in Figure F.9. The residuals deviate substantially from the reference line, suggesting that the assumption of normal residuals may not be valid. The random effects however lie fairly close to the reference line, suggesting that the assumption of normal random effects is valid.

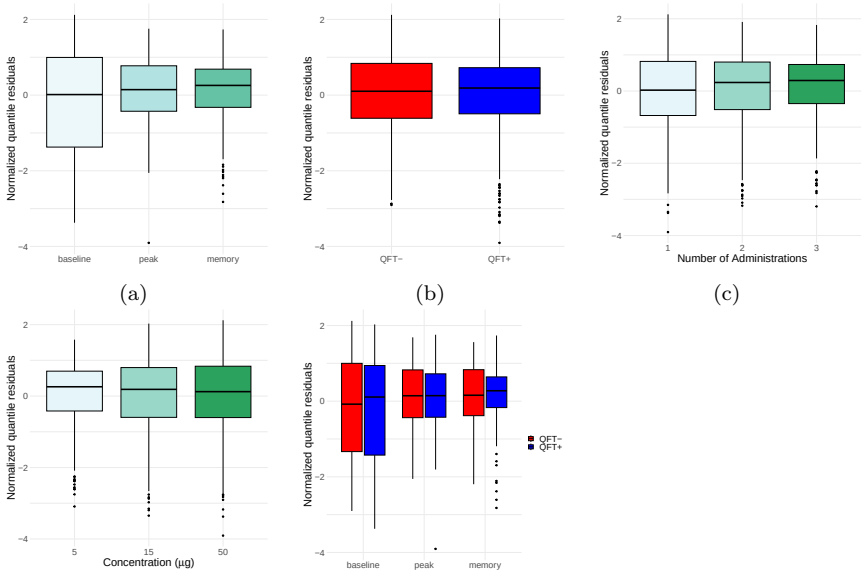


(a) Residuals (b) Random Effects

Figure F.9: QQ-plots of (a) the residuals and (b) the random effects of the single-level GLMEM modelling the frequency of CCR7+CD45RA− expression after Ag85B stimulation.

F.4 Single-level GLMEM: frequency of CCR7−CD45RA− after Ag85B stimulation

The homogeneity of the within-participant residual variance across the model covariates was assessed using box-plots of the normalised quantile residuals, as seen in Figure F.10. The variation of the residuals at baseline is again much larger than that at peak and memory, suggesting suggesting that there is a lot of unexplained variation associated with baseline observations and that the assumption of homoscedastic within-participant errors may be invalid. There is not much difference between the levels of the other variables, although when subdivided by time, there does seem to be more unexplained variation associated with QFT- participants at memory time point compared to QFT+ participants. For all plots, the residuals are centered around zero, suggesting that the assumption that the within-subject errors have a mean of zero is valid.



(a) (b) (c)

(d) (e)

Figure F.10: Box-plots of the normalised quantile residuals plot against the model covariates for the single-level GLMEM modelling the frequency of CCR7−CD45RA− expression after Ag85B stimulation.

To further assess the within-group errors assumptions, the residuals versus fitted values were plot, as seen in Figure F.11. There does seem to be a slight funnel effect with variation decreasing as fitted values increase. This suggests that the model better captures larger, more extreme frequencies.

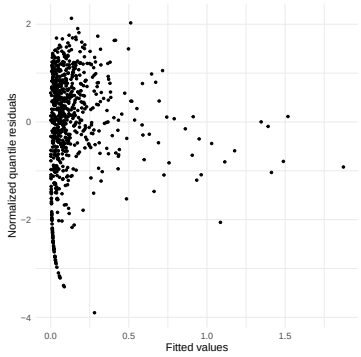
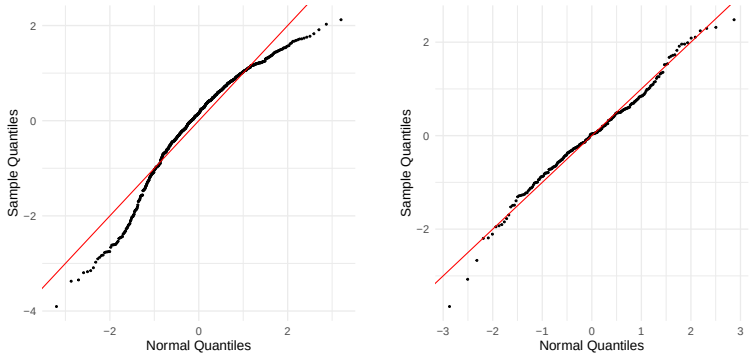


Figure F.11: Scatter plot of the normalised quantile residuals plot against the fitted values of the single-level GLMEM modelling the frequency of CCR7−CD45RA− expression after Ag85B stimulation.

The normality assumptions of the within-participant errors as well as the random effects were checked using qq-plots, as seen in Figure F.12. The residuals deviate substantially from the reference line, suggesting that the assumption of normal residuals may not be valid. The random effects however lie fairly close to the reference line, suggesting that the assumption of normal random effects is valid.



(a) Residuals (b) Random Effects

Figure F.12: QQ-plots of (a) the residuals and (b) the random effects of the single-level GLMEM modelling the frequency of CCR7−CD45RA− expression after Ag85B stimulation.

F.5 Two-level GLMEM: Frequency of CCR7+CD45RA−

The homogeneity of the within-participant residual variance across the model covariates was assessed using box-plots of the normalised quantile residuals, as seen in Figure F.10. The variation of the residuals at baseline is again much larger than that at peak and memory, suggesting suggesting that there is a lot of unexplained variation associated with baseline observations and that the assumption of homoscedastic within-participant errors may be invalid. There also a

difference in unexplained variation between QFT+ and QFT- participants and this difference depends on time point as well as stimulus. At baseline, the unexplained variation is larger for QFT+ participants compared to QFT- participants and the opposite is true at peak, while at memory there is no difference. Under ESAT-6 stimulation, the unexplained variation is also larger for QFT+ participants compared to QFT- participants. The unexplained variation between Ag85B and ESAT-6 seems to depend on time point, with larger unexplained variation associated with Ag85B at baseline compared to ESAT-6, but the opposite is true at peak.

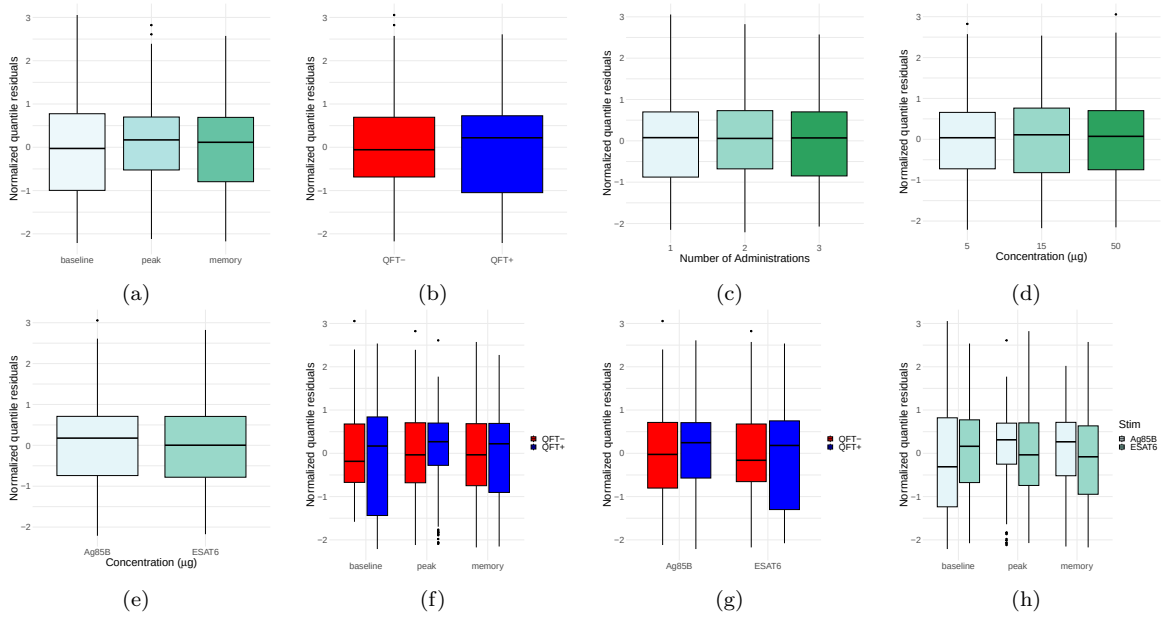


Figure F.13: Box-plots of the normalised quantile residuals plot against the model covariates for the two-level GLMEM grouped by stimulus modelling the frequency of CCR7+CD45RA− expression.

To further assess the within-group errors assumptions, the residuals versus fitted values were plot, as seen in Figure F.14. There is no obvious pattern in the residuals and no clear change in variation with increasing fitted values, providing no evidence against the assumption of constant variance of the within-group errors.

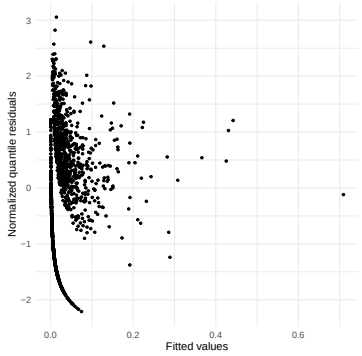
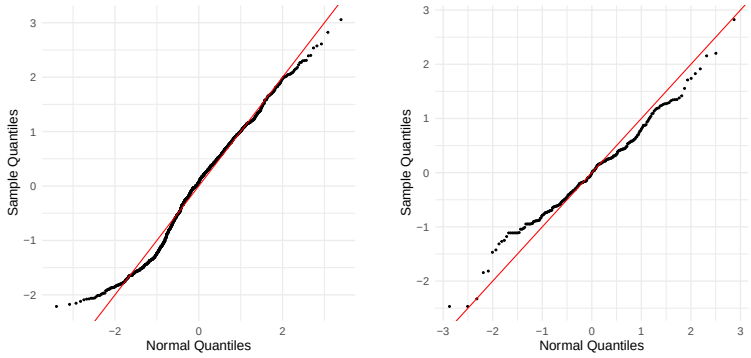


Figure F.14: Scatter plot of the normalised quantile residuals plot against the fitted values of the two-level GLMEM grouped by stimulus modelling the frequency of CCR7+CD45RA− expression.

The normality assumptions of the within-participant errors as well as the random effects were

checked using qq-plots, as seen in Figure F.15. The points in both plots deviate from the reference line substantially at the tails, suggesting that the assumption of normality may not be valid for both the residuals and the random effects.



(a) Residuals (b) Random Effects

Figure F.15: QQ-plots of (a) the residuals and (b) the random effects of the two-level GLMEM grouped by stimulus modelling the frequency of CCR7+CD45RA− expression.

F.6 Two-level GLMEM: Frequency of CCR7−CD45RA−

The homogeneity of the within-participant residual variance across the model covariates was assessed using box-plots of the normalised quantile residuals, as seen in Figure F.16. The variation of the residuals at baseline is again much larger than that at peak and memory, suggesting suggesting that there is a lot of unexplained variation associated with baseline observations and that the assumption of homoscedastic within-participant errors may be invalid. There is not much difference between the levels of the other variables. There is notably no obvious difference in variation between QFT- and QFT+ participants, even when subdivided by time. For all plots, the residuals are centered around zero, suggesting that the assumption that the within-subject errors have a mean of zero is valid.

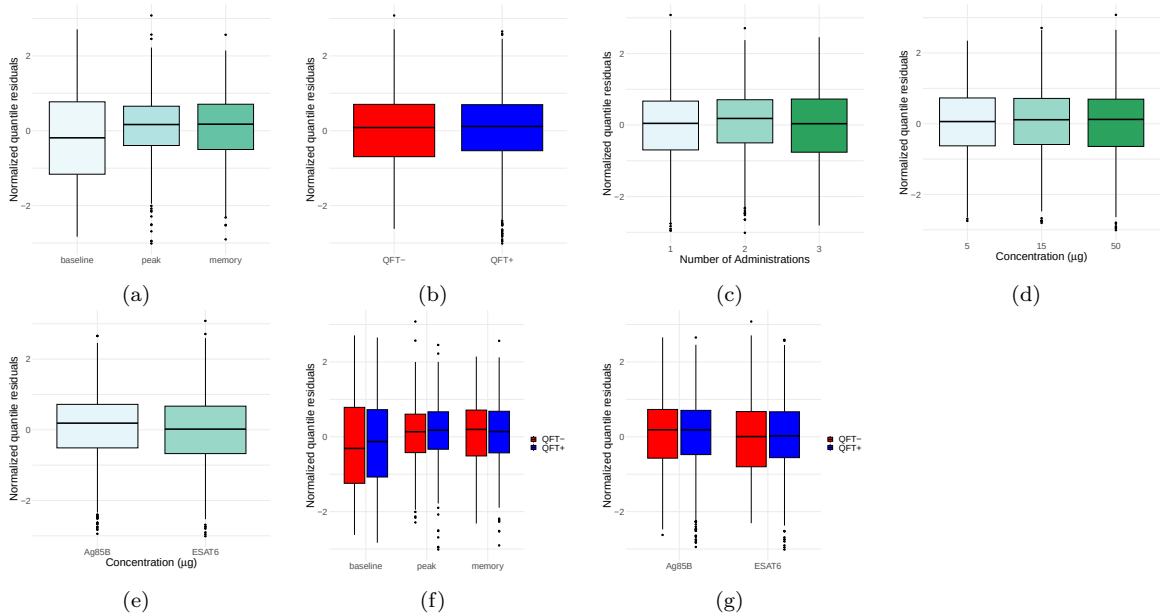


Figure F.16: Box-plots of the normalised quantile residuals plot against the model covariates for the two-level GLMEM grouped by stimulus modelling the frequency of CCR7−CD45RA− expression.

To further assess the within-group errors assumptions, the residuals versus fitted values were plot, as seen in Figure F.17. There does seem to be a slight funnel effect with variation decreasing as fitted values increase. This suggests that the model better captures larger, more extreme frequencies.

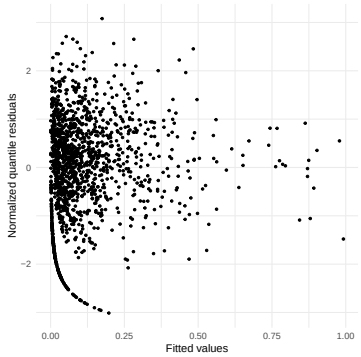
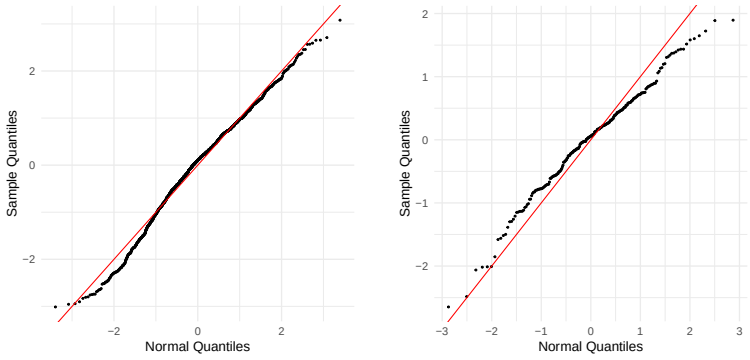


Figure F.17: Scatter plot of the normalised quantile residuals plot against the fitted values of the two-level GLMEM grouped by stimulus modelling the frequency of CCR7−CD45RA− expression.

The normality assumptions of the within-participant errors as well as the random effects were checked using qq-plots, as seen in Figure F.18. The residuals deviate from the reference lines quite substantially at the tails while the the random effects deviate severely from the reference line. This suggests that the assumption of normality may not be valid for both the residuals and the random effects.

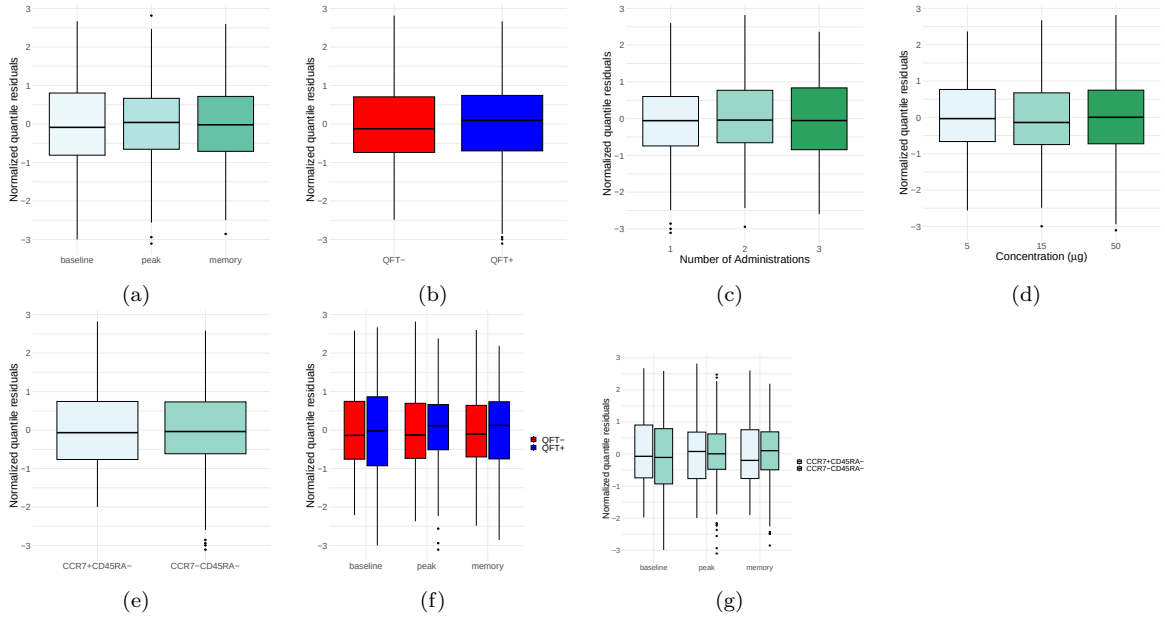


(a) Residuals (b) Random Effects

Figure F.18: QQ-plots of (a) the residuals and (b) the random effects of the two-level GLMEM grouped by stimulus modelling the frequency of CCR7−CD45RA− expression.

F.7 Two-level GLMEM: frequency after ESAT-6 stimulation

The homogeneity of the within-participant residual variance across the model covariates was assessed using box-plots of the normalised quantile residuals, as seen in Figure F.19. There is no obvious difference in variation between the levels of any of the variables, suggesting that the assumption of homoscedastic within-participant errors may be valid.



(a) (b) (c) (d)

(e) (f) (g)

Figure F.19: Box-plots of the normalised quantile residuals plot against the model covariates for the two-level GLMEM grouped by memory combination modelling the frequency after ESAT-6 stimulation.

To further assess the within-group errors assumptions, the residuals versus fitted values were plot, as seen in Figure F.20. There is no obvious pattern in the residuals and no clear change in variation with increasing fitted values, providing no evidence against the assumption of constant variance of the within-group errors.

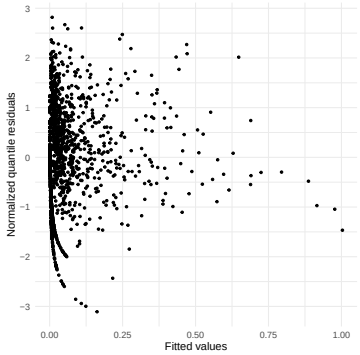
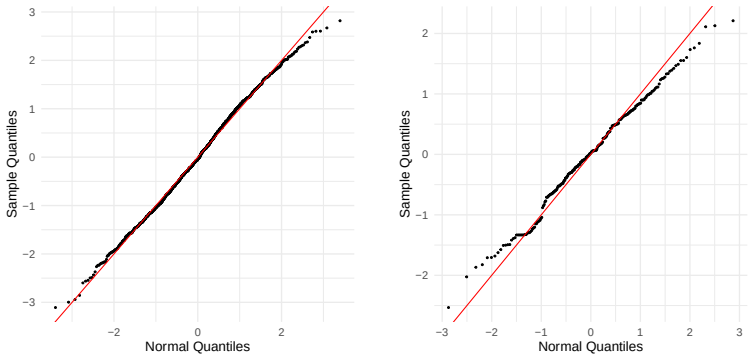


Figure F.20: Scatter plot of the normalised quantile residuals plot against the fitted values of the two-level GLMEM grouped by memory combination modelling the frequency after ESAT-6 stimulation.

The normality assumptions of the within-participant errors as well as the random effects were checked using qq-plots, as seen in Figure F.21. The random effects deviate substantially from the reference line, suggesting that the assumption of normal random effects may not be valid. The residuals however lie fairly close to the reference line, suggesting that the assumption of normal residuals is valid.



(a) Residuals (b) Random Effects

Figure F.21: QQ-plots of (a) the residuals and (b) the random effects of the two-level GLMEM grouped by memory combination modelling the frequency after ESAT-6 stimulation.

F.8 Two-level GLMEM: frequency after Ag85B stimulation

The homogeneity of the within-participant residual variance across the model covariates was assessed using box-plots of the normalised quantile residuals, as seen in Figure F.22. The variation of the residuals at baseline is again much larger than that at peak and memory, suggesting that there is a lot of unexplained variation associated with baseline observations and that the assumption of homoscedastic within-participant errors may be invalid. There also a slight difference in unexplained variation at baseline between QFT+ and QFT- participants, where QFT- participants are associated with larger unexplained variance compared to QFT+ participants. Otherwise, there is no obvious difference in variation between the levels of the remaining variables.

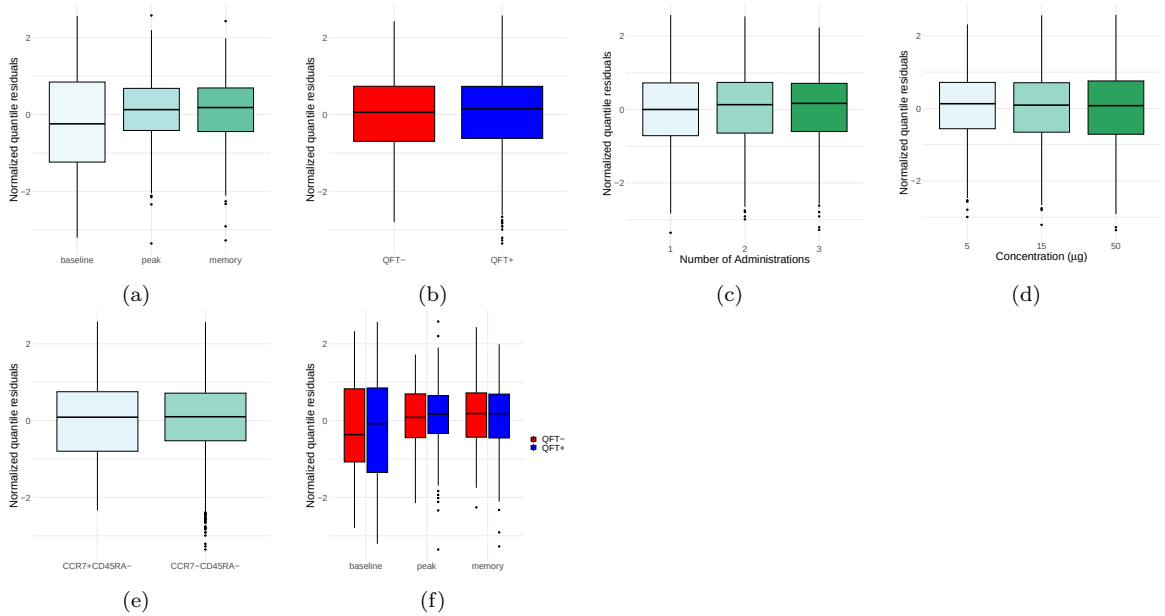


Figure F.22: Box-plots of the normalised quantile residuals plot against the model covariates for the two-level GLMEM grouped by memory combination modelling the frequency after Ag85B stimulation.

To further assess the within-group errors assumptions, the residuals versus fitted values were plot, as seen in Figure F.23. There does seem to be a slight funnel effect with variation decreasing as fitted values increase. This suggests that the model better captures larger, more extreme frequencies.

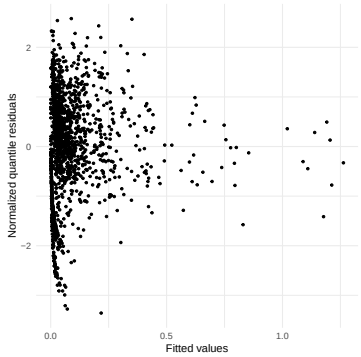
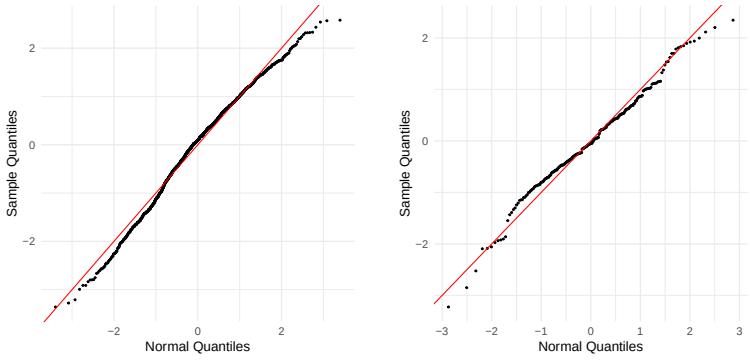


Figure F.23: Scatter plot of the normalised quantile residuals plot against the fitted values of the two-level GLMEM grouped by memory combination modelling the frequency after ESAT-6 stimulation.

The normality assumptions of the within-participant errors as well as the random effects were checked using qq-plots, as seen in Figure F.24. The residuals deviate from the reference lines quite substantially at the tails while the the random effects deviate severely from the reference line. This suggests that the assumption of normality may not be valid for both the residuals and the random effects.



(a) Residuals (b) Random Effects

Figure F.24: QQ-plots of (a) the residuals and (b) the random effects of the two-level GLMEM grouped by memory combination modelling the frequency after ESAT-6 stimulation.

F.9 Three-level GLMEM: frequency of all memory combinations

The homogeneity of the within-participant residual variance across the model covariates was assessed using box-plots of the normalised quantile residuals, as seen in Figure F.25. The variation of the residuals at baseline is again slightly larger than that at peak and memory, suggesting that there is a lot of unexplained variation associated with baseline observations and that the assumption of homoscedastic within-participant errors may be invalid. Otherwise, there is no obvious difference in variation between the levels of the remaining variables.

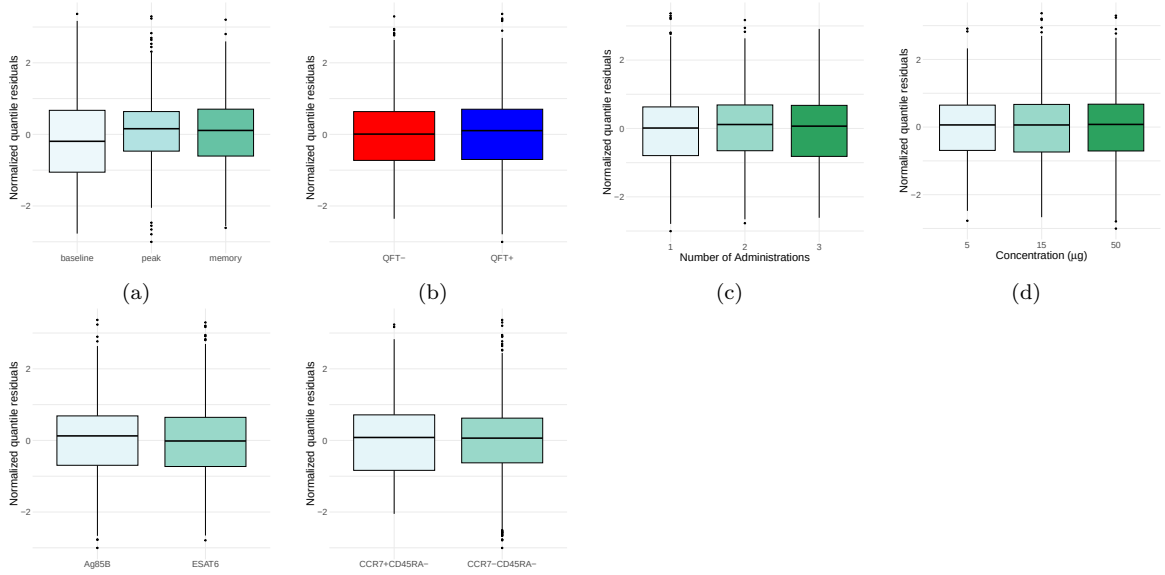


Figure F.25: Box-plots of the normalised quantile residuals plot against the model covariates for the three-level GLMEM modelling memory combination frequency.

To further assess the within-group errors assumptions, the residuals versus fitted values were plot, as seen in Figure F.26. There does seem to be a funnel effect with variation decreasing

as fitted values increase. This suggests that the model better captures larger, more extreme frequencies.

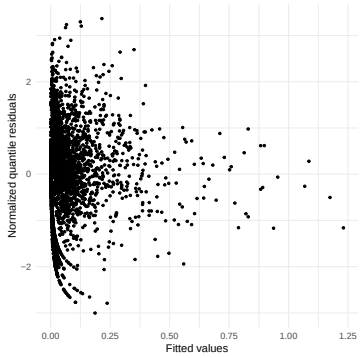
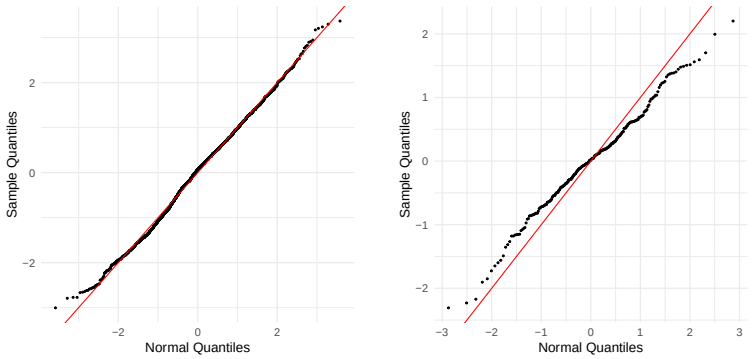


Figure F.26: Scatter plot of the normalised quantile residuals plot against the fitted values of the three-level GLMEM modelling memory combination frequency.

The normality assumptions of the within-participant errors as well as the random effects were checked using qq-plots, as seen in Figure F.27. The random effects deviate substantially from the reference line, suggesting that the assumption of normal random effects may not be valid. The residuals however lie fairly close to the reference line, suggesting that the assumption of normal residuals is valid.



(a) Residuals (b) Random Effects

Figure F.27: QQ-plots of (a) the residuals and (b) the random effects of the three-level GLMEM modelling memory combination frequency.