UNIVERSITY OF CAPE TOWN

DEPARTMENT OF STATISTICAL SCIENCES

Minor Dissertation (MSc Biostatistics)

In collaboration with the South African Tuberculosis Vaccine Initiative

---

OUTCOME SELECTION IN LONGITUDINAL ANALYSIS OF IMMUNOLOGICAL DATA

Author:
Shannon HOLCROFT

Supervisor:
Professor Francesca LITTLE

Co-Supervisor:
Professor Elisa NEMES

17th June 2025

Abstract

Immunological research often compares subgroups defined by exposure variables known (or hypothesised) to influence continuous immune responses. Many immune outcomes are measured over time, often in a small number of patients. Effective outcome selection ensures that research focuses on immune outcomes with the strongest signals for subgroup differences. This dissertation explores an outcome selection technique for longitudinal immunological data, addressing current methodological limitations and proposing improvements. The approach integrates statistical modelling with dimension reduction to identify immune outcomes with the most evidence for subgroup differences. By focusing on these subsets, fewer statistical hypotheses are tested simultaneously, preserving power when stricter significance thresholds are applied to reduce Type I error inflation.

The dissertation examines the suitability of different longitudinal modelling frameworks. Generalised linear mixed-effects models are better suited to the characteristics of immunological data and research than linear mixed-effects models. Two dimension reduction techniques are compared: principal component analysis (PCA) and hierarchical cluster analysis (HCA) followed by PCA. PCA identifies the largest sources of variance across all outcomes, while HCA followed by PCA identifies variance within groups of similar outcomes. These techniques influence the definition of families of tests for false discovery rate (FDR) corrections.

When outcomes are selected via PCA-only dimension reduction, more tests are performed simultaneously and require correction. It was hypothesised that HCA followed by PCA would yield more significant discoveries after FDR control. However, fewer simultaneous comparisons did not reliably correspond with more statistically significant discoveries.

The methodology was applied to a data set from the South African Tuberculosis Vaccine Initiative (SATVI), focusing on 33 immune outcomes and three exposures: MVA85A priming, maternal *Mycobacterium tuberculosis* sensitisation (measured by a positive QuantiFERON TB Gold test), and combinations of feeding practices and cotrimoxazole treatment. The analysis shows that different dimension reduction techniques lead to different outcome selections and families of tests, emphasising the need to align analysis objectives with outcome selection techniques. This dissertation contributes to outcome selection methodology in high-dimensional, longitudinal settings, with broader applications in biomedical research.

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. Francesca Little, for her guidance, support, and mentorship throughout my academic journey. You have helped me find a balance between questioning everything and appreciating the value of straightforward solutions. I aim to carry this lesson forward in my future career as a statistician.

I would also like to thank my co-supervisor, Prof. Elisa Nemes, and the South African Tuberculosis Vaccine Initiative (SATVI) for presenting me with a challenging research problem that has greatly contributed to my growth as a statistician. Gaining insight into your world of immunology and vaccine development has been invaluable. Most importantly, this experience has improved my ability to communicate statistical concepts to non-statisticians, an ability I plan to build on in the future.

Second, I extend my appreciation to the Department of Statistical Sciences for providing a supportive and engaging academic environment.

Third, I would like to acknowledge the support of UCT's Disability Service during my Honours in Statistical Sciences. Being allowed to write my examinations in a quiet, secluded environment made it possible for me to continue to Master's research. This accommodation was difficult to obtain, but it allowed me to access the curriculum and supported my growth as a researcher.

The support of the DSI-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS), South Africa towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the CoE.

The code supporting this dissertation is available on my GitHub repository: https://github.com/ShannonHolcroft/MScDissertation.

Dedication

To my parents (Louise and Richard Holcroft), my sister (Kathleen), and my grandparents (Jeannette and Robbie Crawford-Brunt; Richard and Marjorie Holcroft).

To my second family, although our paths have since diverged: Jono, Louise, John, and Catherine Makepeace.

To Jared Tavares, my classmate, collaborator, co-conspirator, and friend.

To Dilys Wyngaard.

To Frank Lin, who was late to join this journey but changed everything for the better.

To my Social Happiness colleagues, especially Tamsin Haley and Carmen Rieth.

This would not have been possible without your support, encouragement, and belief in me and my goals.

Plagiarism 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 an appropriate convention for citation and referencing (APA). Each significant contribution to, and quotation in, this dissertation from the work of other people has been attributed, and has been cited and referenced.
3. This dissertation is my own work.

Shannon Holcroft
17th June 2025

Contents

List of Figures

List of Tables

List of Key Terms

***Mtb* sensitisation** The immune system's development of a response to *Mycobacterium tuberculosis* exposure. It indicates prior exposure or latent infection but not necessarily active TB disease. 3

BCG vaccination Immunisation using the Bacillus Calmette-Guérin (BCG) vaccine, primarily administered to prevent severe forms of tuberculosis disease, particularly in high-risk populations. It is derived from a live attenuated strain of *Mycobacterium bovis*. 2

BCG-specific immune responses Activation of the immune system, including T-cells and cytokine production, in response to antigens from the Bacillus Calmette-Guérin (BCG) vaccine, reflecting immunity or sensitisation to the vaccine. 3

exposure variables Measurable factors or conditions that study participants experience or are subjected to. These variables are thought to influence the outcomes of interest. Their assignment may or may not be controlled by the research design. 1

immune outcomes Measurable responses or effects within the immune system, typically assessed as part of a study, to evaluate immune function, status, or reaction to an exposure, intervention, or condition. Examples include cytokine expression levels, antibody titers, or the frequency of specific immune cell phenotypes. 1

independent hypotheses In multiple hypothesis testing, these hypotheses are not influenced by or correlated with each other, meaning the outcome of testing one hypothesis does not affect the testing of others. 18

MVA85A priming A vaccine strategy where MVA85A is administered to enhance the immune response initiated by the Bacille Calmette-Guérin (BCG) vaccine, aiming to improve protection against tuberculosis. MVA85A uses a modified vaccinia Ankara virus to deliver the antigen 85A. 2

negatively dependent hypotheses In multiple hypothesis testing, the rejection of one of these hypotheses decreases the likelihood of rejecting others, often due to shared opposing influences between the tests. 18

positively dependent hypotheses In multiple hypothesis testing, the rejection of one of these hypotheses increases the likelihood of rejecting others, often due to shared underlying factors or correlations among the tests. 18

QuantiFERON-TB Gold test A blood-based assay that measures interferon-gamma release from T-cells in response to *Mycobacterium tuberculosis*-specific antigens, used to detect latent or active tuberculosis infection. 3

tuberculosis (TB) disease Infection caused by *Mycobacterium tuberculosis*. It primarily affects the lungs but can spread to other organs. 2

Chapter 1

Introduction

1.1 Overview

Immunological research studies the molecular, cellular, and genetic characteristics of the immune system. High-throughput technologies allow clinical researchers to generate and analyse large amounts of immunological data (Genser et al., 2007). One key application is vaccine development, where data is generated and analysed to assess the ability of a candidate vaccine to confer long-term protection against a target pathogen (Mahanty et al., 2015). This immunogenicity is evaluated by tracking changes in immune responses over time after immunisation. These analyses often compare subgroups that are defined by exposures to factors known or hypothesised to influence immune responses (Pineda et al., 2016).

Statistical models are useful tools for representing and interpreting the longitudinal relationships between immune responses and these exposure variables for different subgroups. These exposures of interest may include experimental factors (e.g., randomisation to a treatment group) or baseline characteristics (e.g., sex or age). A model uses statistical assumptions to describe how relationships between these exposures of interest and additional covariates may have generated the observed immunological data. Model coefficients estimate the magnitude and direction of these relationships from the observed data (Harrell et al., 2001). Contrasts may be computed from fitted models to estimate subgroup differences at particular time points. Confidence intervals and other probability statements assess the evidence for these differences in broader populations (Harrell et al., 2001).

However, analysing longitudinal profiles for many immune responses increases the risk of false detection of at least one statistically significant result due to multiple hypothesis testing (Goeman & Solari, 2014). Correcting for Type I error inflation imposes stricter significance thresholds for individual hypotheses, reducing false positives but also reducing the power to detect true effects, especially for small or moderate effect sizes (Chow & Liu, 2013). Many immune outcomes may have weak associations with the exposures of interest. Correcting probability statements for these superfluous statistical hypotheses may lead to overly stringent adjustments to the significance threshold.

Immunological data is often high-dimensional, with the number of immune responses exceeding the number of samples collected (Forlin et al., 2023). Larger samples produce more precise estimates and greater power to detect effects, especially for small or moderate sizes (Chow & Liu, 2013). This increases the power to detect true effects, especially for small or moderate effect sizes. In small-sample studies where multiple statistical hypotheses are investigated simultaneously, it is crucial to limit the number of immune outcomes that are considered in order to preserve statistical power.

If knowledge of the underlying biological processes is limited (as is often the case in exploratory research), a reduced set of outcomes can be identified for further analysis using outcome selection techniques. These techniques restrict a higher-dimensional problem with

a large number (y) of outcomes to lower dimensions by selecting a smaller set of outcomes that are most strongly associated with one or more of the k exposures of interest.

By restricting the scope of the analysis to the subset(s) of immune outcomes identified by outcome selection, a smaller number of statistical hypotheses are investigated. In doing so, the analysis preserves statistical power when stricter significance thresholds are imposed to account for Type I error inflation.

1.2 Context

This dissertation applies an outcome selection technique to longitudinal immunological data generated by a Phase II Randomised Controlled Trial (NCT01650389) designed to evaluate safety and immunogenicity of MVA85A priming and selective, delayed Bacille Calmette-Guerin (BCG) vaccination in infants of mothers living with human immunodeficiency virus (HIV) who remain HIV uninfected. The trial was conducted by the South African Tuberculosis Vaccine Initiative (SATVI).

Childhood immunisation with BCG confers protection against the *Mycobacterium tuberculosis* (*Mtb*) pathogen and development of tuberculosis (TB) disease. BCG is generally administered at birth, but this live attenuated vaccine can lead to severe adverse events in HIV-infected infants. In HIV-exposed infants born to mothers living with HIV, there is a risk of in utero and peripartum HIV infection. MVA85A prime vaccination at birth was proposed to induce *Mtb*-specific immune responses in this vulnerable period prior to BCG vaccination (Nemes et al., 2018). The trial found that MVA85A prime vaccination was safe and induced an early modest antigen-specific immune response that did not interfere with, or enhance, immunogenicity of subsequent BCG vaccination (Nemes et al., 2018).

It is important to note that there are no sufficiently validated biomarkers to support TB vaccine development (Walzl et al., 2011). There is considerable heterogeneity in immune responses to infection with *Mtb* (Cadena et al., 2017; Walzl et al., 2011). Analysing the dissertation data set therefore requires careful consideration of measured and unmeasured sources of variability that may contribute to subgroup differences in immune outcomes over time.

The trial measured many immune outcomes over time. Appendix A summarises their roles within the immune system. To reduce the amount of blood collected from infants, infants were sequentially enrolled in two groups, A ($n = 65$) and B ($n = 58$).

Different immunological assays were performed on samples from each group at different time points. BCG-specific T cells, proliferation, cytotoxic potential, and T cell differentiation were quantified in Group A. Cellular blood composition was quantified in Group B. As different immune outcomes are measured in different patients at different time points, Group A and Group B may be analysed separately as if generated by separate studies.

For the purposes of this dissertation, the outcome selection technique was applied to the Group A data only. This allowed the technique to be demonstrated clearly and more succinctly, with the understanding that the outcome selection technique could be applied to the Group B data in a more comprehensive analysis for SATVI.

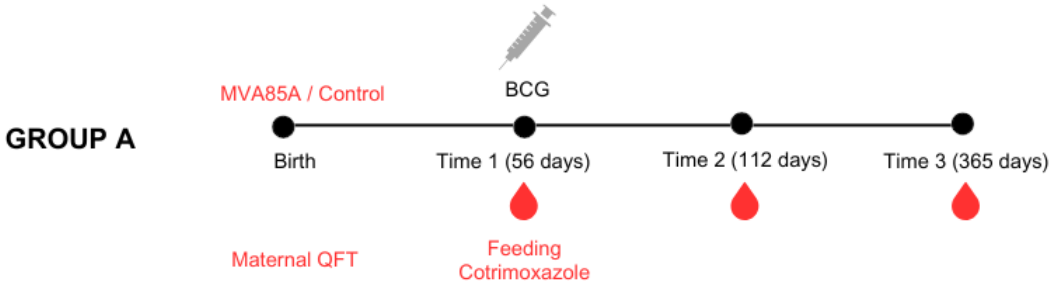


Figure 1.1: Schema of study design and time points at which exposures of interest were measured, and immunological outcomes were generated in Group A.

There are 33 immune outcomes ($m_1 = 33$) measured in Group A (y_i with $i \in 1, ..., m_1$). Measurements taken at the time of BCG vaccination ($Time = 1$) estimate baseline. The peak immune response to BCG vaccination is anticipated at the second measurement ($Time = 2$). For Group A immune outcomes, this peak is anticipated 8 weeks after BCG vaccination.

The peak immune response declines over time, but may not return to baseline. Group A estimates a long-term immune response ≈ 44 weeks after BCG vaccination ($Time = 3$). Long-term changes from baseline provide sustained protection by allowing the immune system to produce a memory response when subsequently exposed to the pathogen.

The longitudinal immunological data generated by the clinical trial describes a number of subgroups that can differ with respect to their immune responses over time, including:

1. **Infants receiving MVA85A priming at birth**

   In the main study, MVA85A prime vaccination induced an early modest antigen-specific immune response that did not interfere with, or enhance, immunogenicity of subsequent BCG vaccination (Nemes et al., 2018).

   The main study did not consider the immune outcomes listed in Appendix A. Hence, it is possible that infants receiving MVA85A priming may differ from the control group with respect to BCG-specific immune outcomes and cellular blood composition.

2. **Infants born to mothers with *Mtb* sensitisation**

   A positive QuantiFERON-TB Gold test indicates sensitisation to the *Mtb* pathogen. *Mtb*-sensitised mothers can transfer mycobacterial antigens, *Mtb*-specific antibodies, and immune cells to infants (Lubyayi et al., 2020; Rahman et al., 2010). This could inhibit BCG replication and compromise the development of BCG-specific immune responses.

3. **Infants fed breast milk compared to formula**

   Breast milk contains bioactive ingredients that support the development of the infant immune system. Immune cells, antibodies and other components of breast milk provide both direct and indirect protection against pathogens (Victora et al., 2016). Improved vaccine-induced immune responses have been observed in infants who are

breastfed compared to infants who are formula fed (Zimmermann & Curtis, 2019). Similar improvements might also be observed for BCG-specific immune responses.

4. **Infants treated with the antibiotic cotrimoxazole**

   Antibiotic drugs prevent and treat bacterial infection by inhibiting bacterial growth. This leads to changes in the gut microbiome. At the time of the trial, treatment with the antibiotic cotrimoxazole was recommended to prevent opportunistic infections in HIV-exposed infants. These guidelines specified that cotrimoxazole should be taken until the mother is no longer breastfeeding, and HIV infection is excluded by negative virological testing (World Health Organisation, 2004).

   Alongside its antimicrobial effects, cotrimoxazole has been observed to inhibit pro-inflammatory responses and alter immune cell phenotypes (Bourke et al., 2019). This may limit BCG replication and persistence after administration. Therefore, differences in BCG-specific immune responses or pro-inflammatory response may be observed in infants treated with cotrimoxazole.

These patient subgroups are described by three exposures of interest (EOI):

1. *MVA85A*: A binary indicator for an infant's randomisation to receive MVA85A priming or placebo at birth (EOI_1).
   - $MVA85A = 1$ describes the subgroup of infants who received MVA85A priming before BCG vaccination.
2. *QFT*: A binary indicator for maternal *Mtb* sensitisation, measured after delivery (EOI_2).
   - $QFT = 1$ describes the subgroup of infants born to a mother sensitised to the *Mtb* pathogen, measured by a positive QFT test.
3. *FA*: A four-level categorical variable describing different combinations of feeding practices and cotrimoxazole treatment. This information was recorded at the time of BCG vaccination (56 days after delivery) (EOI_3).
   - $FA = 0$ describes the subgroup of infants fed formula and treated with the antibiotic cotrimoxazole.
   - $FA = 1$ describes the subgroup of infants fed formula and not treated with the antibiotic cotrimoxazole.
   - $FA = 2$ describes the subgroup of infants fed breast milk and not treated with the antibiotic cotrimoxazole.
   - $FA = 3$ describes the subgroup of infants fed breast milk and treated with the antibiotic cotrimoxazole.

For each of these three EOI and the subgroups they define, SATVI would like to compare how its different levels modify the expectations of the immune outcomes over time. *MVA85A*, *QFT*, and *FA* are initially assumed to have independent effects on the immune outcomes. However, SATVI anticipates that multiple EOI may contribute to observed subgroup differences in some immune outcomes. There is limited prior knowledge of the relationships between the EOI and immune outcomes. However, it may be necessary to

disentangle the effects of two or more *EOI* on a given immune outcome before describing or making inferences about subgroup differences.

SATVI specifies that subgroup differences should be adjusted for the average effect of sex, across male and female patients, on the immune outcomes. Although sex differences are not observed consistently, differences in both vaccine-specific and non-specific immune responses have been described for a number of childhood vaccines (St. Clair et al., 2023). For example, in children younger than 5 years, BCG-specific immune responses have been observed to be greater in males. A number of non-specific immune responses are also greater in male children of the same age, including Natural Killer cells and inflammatory responses (St. Clair et al., 2023).

In the dissertation data set, patient sex is described by a binary indicator (*Sex*) for whether an infant is male or female. Cross tables describing the distribution of *MVA85A*, *QFT*, and *FA* by *Sex* are included in Appendix B.

As many immune outcomes are measured in Group A, an outcome selection technique will be applied to identify the immune outcome subset(s) with the most evidence for subgroup differences with respect to MVA85A priming, maternal QFT, and the combination of feeding practices and cotrimoxazole treatment. A smaller number of immune outcomes will then be selected for further analysis. In doing so, the analysis will preserve statistical power when applying corrections for multiple hypothesis testing.

1.3 Implementation

This dissertation presents an outcome selection technique to select the subset(s) of immune outcomes with the most evidence for subgroup differences over time. There is no literature making use of the term 'outcome selection', but the concept overlaps with (i) data filtering for biomarker discovery and (ii) variable selection. The similarities and differences between these techniques are reviewed in Chapter 2.

The outcome selection technique investigated in this dissertation is loosely based on the RM-ASCA+ framework - the extended ANOVA simultaneous component analysis (ASCA+) framework for repeated measures. RM-ASCA+ combines linear mixed-effect models with principal component analysis (PCA) to decompose and visualise the separate effects of experimental factors over time (Madssen et al., 2021). Although not explicitly an outcome selection technique, this dissertation argues that elements of the RM-ASCA+ procedure are well suited to outcome selection in longitudinal immunological data. Chapter 2 describes the RM-ASCA+ framework in more detail.

Similar to the RM-ASCA+ technique, this dissertation applies longitudinal statistical modelling to estimate subgroup differences described by *MVA85A*, *QFT* and *FA* at different time points for the set of immune outcomes.

- A model is specified to describe the relationship between every immune outcome (y_i with $i \in 1, ..., m$) and every exposure (EOI_i with $i \in 1, ..., p$) over time.
- In every model, a single immune outcome (y_i) is the dependent variable. Study time ($Time$) and a single exposure (EOI_i with $i \in 1, ..., p$) are the independent variables.

- The model compares how different levels of the exposure modify the expectation of the immune outcome over time ($EOI_i * Time$). These are the comparisons of interest.

For every exposure (EOI_i with $i \in 1, ..., p$), model coefficient estimates ($\hat{\beta}_i$ with $i \in 1, ..., q$) of the comparisons of interest ($EOI_i * Time$) are extracted from the fitted models for every immune outcome (y_i with $i \in 1, ..., m$).

- This produces p data sets with m rows and q columns.
- Each row corresponds to an immune outcome (y_i with $i \in 1, ..., m$).
- Each column corresponds to a model coefficient estimate (β_i with $i \in 1, ..., q$).

A dimension reduction technique is applied to each of the p data sets containing model coefficient estimates. This technique identifies the subset(s) of immune outcomes with the most evidence for subgroup differences. These immune outcomes are then selected for further analysis. The workflow is summarised in the diagram below.

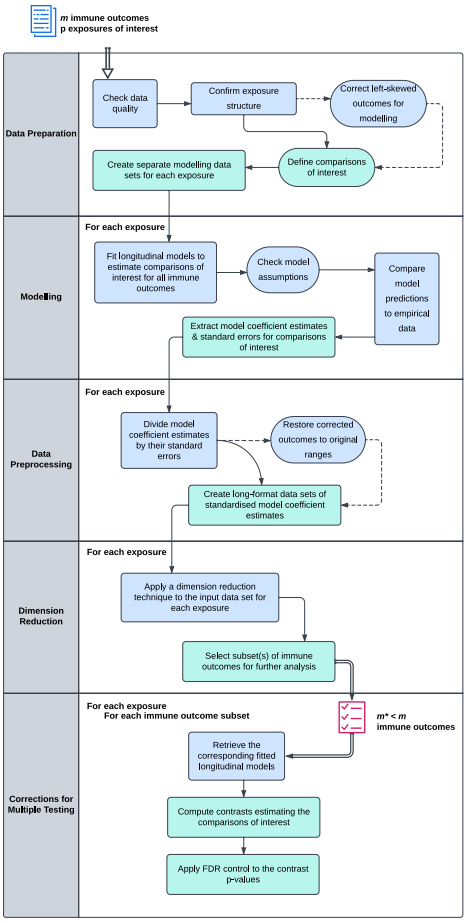


Figure 1.2: An overview of the outcome selection workflow from data preparation to corrections for multiple hypothesis testing.

Immune outcomes tend to have skew, long-tailed distributions and many null responses (Genser et al., 2007). The dissertation data set contains both left- and right-skewed immune outcomes. Several longitudinal modelling techniques, including the mixed-effect modelling framework investigated in the dissertation, require a specified parametric form for the

conditional distribution of the outcome. Few parametric distributions accommodate left-skewed outcome distributions. Hence, data preparation for modelling (Step 1 in Figure 1.2) identifies left-skewed immune outcomes and transforms them to right-skewed distributions. The data preparation process is summarised in Chapter 3.

Although the RM-ASCA+ framework implements a linear mixed-effect modelling framework, more flexible frameworks may be required to accommodate the characteristics of longitudinal immunological data. Chapter 4 therefore describes and compares linear mixed-effect models (LMM) and generalised linear mixed-effect models (GLMM) as longitudinal modelling frameworks for outcome selection. After identifying the most appropriate longitudinal modelling framework, model coefficient estimates and their standard errors are extracted and prepared for dimension reduction. For different levels of the exposure variables, the model coefficients estimate subgroup differences over time for the set of immune outcomes.

Two dimension reduction techniques are considered in this dissertation: first, PCA (as implemented in RM-ASCA+), and second, agglomerative hierarchical cluster analysis (HCA) followed by PCA. The techniques offer distinct strategies to select the subset(s) of immune outcomes with the most evidence of subgroup differences over time. Chapter 5 describes both data preprocessing (Step 3 in Figure 1.2) and dimension reduction (Step 4 in Figure 1.2) with PCA only and HCA followed by PCA.

The workflow for dimension reduction by PCA only is summarised in the diagram below.

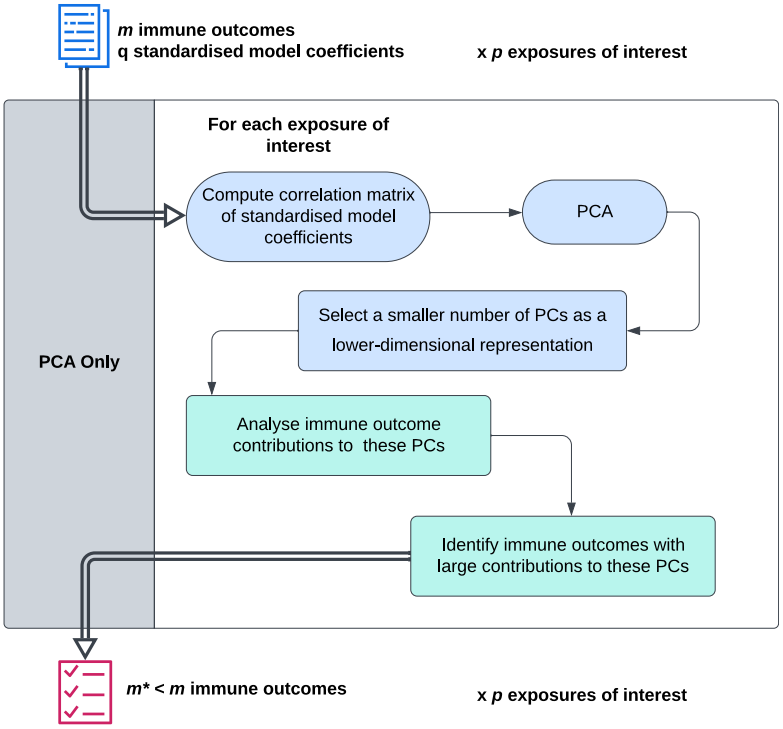


Figure 1.3: An overview of the outcome selection workflow when applying PCA only as a dimension reduction technique.

For an exposure of interest, PCA is applied to the correlation matrix of standardised model coefficient estimates. It decomposes the total variance of these model coefficient estimates into linear combinations called principal components (PCs). These PCs are

ranked by their contributions to this variance. A smaller number of PCs is selected to represent the variance, reducing dimensionality while preserving as much information in the input data as possible.

- By analysing immune outcome contributions to these PCs, one can identify the immune outcomes that contribute most to the total variance of model coefficient estimates across all immune outcomes for an exposure of interest.
- Based on these contributions, a single immune outcome subset is selected for further analysis. This outcome subset contains the immune outcomes with the most evidence of subgroup differences described by the levels of an exposure of interest.

For an exposure of interest, contrasts estimating subgroup differences are then computed for selected outcomes only. Subgroup differences are estimated within a smaller group of immune outcomes, preserving statistical power when applying corrections for multiple hypothesis testing. The workflow for dimension reduction by HCA followed by PCA is summarised in the diagram below.

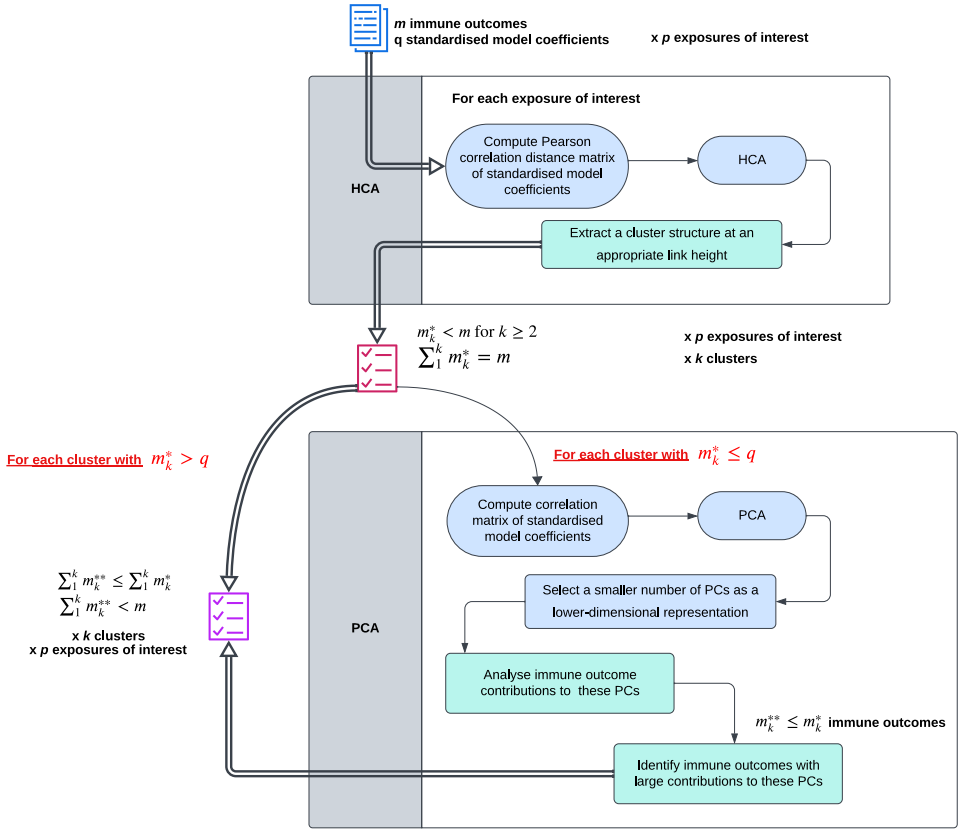


Figure 1.4: An overview of the outcome selection workflow when applying HCA followed by PCA as a dimension reduction technique.

For an exposure of interest, HCA organises immune outcomes into nested groups that are similar with respect to the standardised model coefficient estimates. These coefficients estimate subgroup differences over time for different levels of the exposure in all immune outcomes. If a minimum of two clusters ($k \geq 2$) are extracted from the hierarchy of groupings, every cluster will contain fewer immune outcomes than the complete set ($m_k^* <$

m). The number of immune outcomes in all clusters sums to the complete number of immune outcomes in the data set ($\sum_{i=1}^{k} m_k^* = m$).

For an exposure of interest, these k clusters then become the units of analysis. Dimension reduction by PCA can be performed on clusters containing a number of immune outcomes greater than or equal to the number of standardised model coefficients ($m_k^* \geq q$). In these cases, a PCA is performed on the correlation matrix of model coefficient estimates for the immune outcomes contained within the cluster. A smaller number of PCs is then selected to provide a lower-dimensional representation of the cluster variance in standardised model coefficient estimates.

By analysing immune outcome contributions to these PCs, the immune outcomes that contribute most to the cluster variance of model coefficient estimates can be selected as a subset for further analysis. If a cluster contains more immune outcomes than the number of standardised model coefficients ($m_k^* > q$), PCA cannot be applied. All immune outcomes contained in the cluster are then selected as an immune outcome subset for further analysis.

Unlike the PCA-only approach, multiple immune outcome subsets (one for each cluster) can be identified per exposure of interest. Contrasts estimating subgroup differences are computed within these immune outcome subsets. Similarly, corrections for multiple hypothesis testing are applied within these immune outcome subsets.

Depending on the dimensionality of the clusters, fewer statistical hypotheses may be investigated simultaneously when following this approach compared to dimension reduction by PCA only. More contrasts may be computed overall than within the PCA-only approach, but statistical power could be preserved by making the clusters (groups of similar immune outcomes) the units of analysis.

Chapter 6 presents the outcome selections for each exposure of interest by PCA only and by HCA followed by PCA. As these techniques represent different approaches to select the subset(s) of immune outcomes, they may select slightly different immune outcome subsets for further analysis. The size and contents of the immune outcome subset(s) selected by the dimension reduction techniques are compared. By comparing how the selected immune outcome subsets overlap for different variables of interest, Chapter 6 also identifies which longitudinal models may need to be adjusted for the effects of multiple exposures before describing and making inference on subgroup differences.

Chapter 7 describes and demonstrates how corrections for multiple hypothesis testing (Step 5 in Figure 1.2) may be applied to contrasts estimating subgroup differences. These contrasts estimate the expected difference in a given immune outcome, at different time points, across different levels of an exposure variable. They are computed from the longitudinal models corresponding to the outcome subset(s) selected by PCA only and by HCA followed by PCA. The contrast p-values quantify the evidence for these subgroup differences in wider populations beyond the observed data.

As multiple contrasts are estimated simultaneously for multiple immune outcomes, the contrast p-values must be adjusted for multiple hypothesis testing before reporting the subgroup differences. Two False Discovery Rate (FDR) control methods are compared: the Benjamini-Hochberg (BH) procedure (Benjamini & Hochberg, 1995) and the Benjamini-Yekutieli (BY) procedure (Benjamini & Yekutieli, 2001).

The BH procedure assumes that the hypotheses investigated by the longitudinal model contrasts are independent or positively correlated, while the BY procedure accommodates arbitrary dependence, including negative correlations (Stevens et al., 2017). Negatively correlated hypotheses are plausible in immunological research as the expression of one immune response may downregulate the expression of another immune response (see Vigano et al. (2012)). Although the BH procedure is generally robust to negative correlation, strong negative dependencies can interfere with exact FDR control. Chapter 7 therefore compares the discoveries and q-values obtained by applying the BH and BY procedures to the contrast p-values.

Chapter 7 also compares the discoveries and q-values obtained after outcome selection by PCA only and by HCA followed by PCA. First, these techniques may select different immune outcomes for further analysis. Second, the size of the immune outcome subset(s) may differ between techniques. Third, the techniques may estimate different numbers of contrasts simultaneously. PCA selects a single immune outcome subset per exposure; HCA followed by PCA selects multiple immune outcome subsets per exposure. These differences have implications for the statistical power of the analysis.

This chapter also investigates how adjusting longitudinal models for the effects of multiple exposures changes the contrasts, discoveries and q-values. In the absence of prior knowledge, these investigations are guided by immune outcomes selected to subsets for two or more exposures of interest (see Chapter 6). Chapter 8 provides an overview of the implementation of the outcome selection technique.

1.4 Aims & Objectives

In summary, this dissertation aims to identify an outcome selection technique that is appropriate for longitudinal immunological data. The chosen approach combines statistical modelling with a dimension reduction technique. A number of longitudinal modelling frameworks and dimension reduction techniques may be suitable for this purpose.

- **AIM 1:** To identify an appropriate longitudinal modelling framework for outcome selection which is well suited to the distributional characteristics of immune outcome distributions.
  - Linear mixed-effect and generalised linear mixed-effect modelling frameworks are applied to three immune outcomes that exemplify the characteristics of immunological data described in Chapter 2.
    - The RM-ASCA+ framework, on which the dissertation's outcome selection technique is loosely based, implements linear mixed-effect models to estimate the comparisons of interest for each outcome.
    - This modelling framework assumes that the conditional distribution of the dependent variable, given the fixed and random effects, is normally distributed with a constant variance.
    - Immunological outcomes tend to have skew, long-tailed distributions. These characteristics make it challenging to uphold an assumption of constant variance, as required by the linear mixed-effect modelling framework.

    - The generalised mixed-effect modelling framework, on the other hand, has more flexible assumptions to accommodate skew, long-tailed immune outcomes.
  - *Hypothesis:* Model assumptions will be better met within the generalised linear mixed-effect model framework than in the linear mixed-effect model framework.

- **AIM 2:** To explore differences in immune outcome subsets identified by different dimension reduction techniques.
  - Principal component analysis (PCA) and agglomerative Hierarchical Cluster Analysis (HCA) followed by PCA will be considered.
  - *Hypothesis:* Compared to PCA alone, combining HCA and PCA is expected to identify *more* immune outcomes in total for further analysis per exposure of interest. However, these immune outcomes will be contained in multiple smaller subsets and not within a single larger subset per exposure of interest.

- **AIM 3:** To explore differences in the discoveries and q-values when FDR-control procedures are applied to p-values for contrasts computed to estimate subgroup differences in selected immune outcomes.
  - *Hypothesis:* The number of discoveries is expected to be higher (for 15% and 5% false discovery rates) when contrasts are estimated from the immune outcome subsets obtained by HCA followed by PCA. Groups of similar immune outcomes are the units of analysis, and this is expected to preserve statistical power compared to estimating contrasts from a single immune outcome subset obtained by PCA only.
  - *Hypothesis:* For similar reasons, the q-values obtained by applying FDR-control procedures to p-values of contrasts are expected to be smaller for the immune outcome subset(s) obtained by HCA followed by PCA.
    - As negative dependencies can interfere with exact FDR control, results obtained by applying the Benjamini-Hochberg and Benjamini-Yekutieli procedures will be compared.

These aims will be achieved by applying outcome selection techniques to a longitudinal immunological data set produced by SATVI. Their own objectives for outcome selection are as follows:

- **OBJECTIVE 1:** To identify immune outcome subsets for which there is evidence of subgroup differences with respect to MVA85A priming, maternal *Mtb* sensitisation, and/or combinations of feeding practices and cotrimoxazole treatment. These subgroups are described by the levels of three exposures of interest (*MVA85A*, *QFT*, and *FA*).
- **OBJECTIVE 2:** To disentangle the effects of these exposures of interest so that differences in immune outcomes over time can be uniquely attributed to the levels of a single exposure of interest.
- **OBJECTIVE 3:** To apply appropriate corrections for multiple hypothesis testing to contrasts computed from the longitudinal models corresponding with these immune outcome subsets.

Chapter 2

Literature Review

Immunological research often generates and analyses large amounts of longitudinal immunological data (Genser et al., 2007). Comprehensive immunological assays are available that allow researchers to easily quantify the frequency and proportion of many different immune cells and responses over time (Walzl et al., 2011). Regression models are useful tools for representing and interpreting longitudinal relationships between immune responses and exposures of interest. However, the theory that underlies regression models is often unreliable in high dimensions, especially when the number of inputs far exceeds the number of empirical data points (Remeseiro & Bolon-Canedo, 2019). Parameter estimates may not converge to true values, as optimization is challenging under these conditions (Remeseiro & Bolon-Canedo, 2019).

Immunological data is typically produced by longitudinal study designs that take repeated measurements of an immune outcome from the same patient at different points in time (Genser et al., 2007). Longitudinal data can be measured at regular or irregular time intervals (Fitzmaurice et al., 2012). Repeated measurements are clustered within patients, and assumptions of residual independence are inappropriate (Fitzmaurice et al., 2012). This affects the regression modelling frameworks that can be applied to immunological data.

The characteristics of immunological data also make it challenging to specify an appropriate regression modelling framework. Immune responses tend to have skewed distributions and are often long or heavy-tailed (Genser et al., 2007). Log transformations are often applied to improve the symmetry of immune outcome data (Genser et al., 2007). However, if a transformation is applied to the immune outcome, the regression model will describe the relationship between the exposures of interest and this transformed immune outcome, rather than with the immune outcome on its original measured scale (Choi et al., 2022). Additionally, since the skewness of the immunological data is often profound, applying transformations to the immune outcomes rarely offers a practical improvement of the symmetry of the observed distribution (Genser et al., 2007).

Immunological data often contains relatively large proportions of null responses. Null responses may represent true absences of an immune outcome, but often represent left-censored measurements that lie below some detection threshold (Arboretti et al., 2020). Zero values are often imputed for these non-detects (Arboretti et al., 2020). This creates more probability mass at zero than would be expected in most continuous probability distributions (Arboretti et al., 2020). Long-tailed outcome distributions and/or large proportions of null responses make it challenging to specify an appropriate mean-variance relationship in a regression model of immunological data. Many parametric distributions do not allow enough spread to capture extreme observations in the tails (Rojo, 2013). Similarly, many parametric distributions do not allow many zeroes or zero inflation (Liu et al., 2019).

Within a single immunological analysis, multiple regression models may be formulated to investigate a family of statistical hypotheses about immune outcomes. However, it is plausible that several of these hypotheses will be non-informative to the research questions (Goeman & Solari, 2014). Each hypothesis test has the potential to falsely detect a significant difference between compared groups and/or time

points. When analysing many immune outcomes simultaneously, it is important to adjust the confidence levels of the statistical tests for these false discoveries (Goeman & Solari, 2014). However, as the number of statistical tests increases, the threshold for declaring a result statistically significant is often adjusted to be more stringent. Without reducing the number of immune outcomes considered in the analysis, it becomes challenging to identify potentially clinically meaningful relationships from the observed data (Goeman & Solari, 2014). If subject matter knowledge is limited (as is often the case in an exploratory setting), a smaller outcome set can be identified using outcome selection techniques.

Outcome selection techniques restrict a higher-dimensional analysis with many outcomes to lower dimensions by identifying a smaller set of immune outcomes for further analysis. Working with this immune outcome subset ensures that corrections for multiple hypothesis testing are applied only to statistical hypotheses that are highly informative to the research questions. Literature explicitly referring to 'outcome selection' is limited. Much of the existing literature concerns outcome or endpoint selection within the context of clinical trials (Williamson et al., 2012). By pre-selecting appropriate clinical outcomes within the study design phase, treatment effects of different clinical trials can be compared in a way that minimises bias. Outcome selection techniques for binary outcomes, for example, outcomes reflecting infection with a disease or the occurrence of a clinical symptom or clinical sign, are inherently different from when the research question considers continuous clinical outcomes.

When considering continuous immunological outcomes, the concept of 'outcome selection' overlaps with biomarker discovery. Biomarkers are measurements of biological or pathogenic processes that have been identified as appropriate proxies for some overarching clinical objective (Baumgartner et al., 2011). In vaccine development, a biomarker is a measurable indicator of response to a vaccine (Weiner 3rd & Kaufmann, 2014). Biomarkers are used to assess whether a candidate vaccine elicits the desired immune response without significant safety concerns (Weiner 3rd & Kaufmann, 2014).

Biomarker discovery is often formulated as a variable selection problem. This is a well-defined research area in multivariate analysis (Heinze et al., 2018; Remeseiro & Bolon-Canedo, 2019) These techniques identify a minimum number of input variables data features that are predictive of a clinical outcome or endpoint. There are three general categories of variable selection techniques: filter methods, wrapper methods, and embedded methods (Heinze et al., 2018; Remeseiro & Bolon-Canedo, 2019). Filter, wrapper, and embedded methods can also be combined to form hybrid selection methods (Heinze et al., 2018). These categories will be briefly defined using examples from the literature and contrasted with each other in terms of their strengths and weaknesses as they relate to the focus of this dissertation.

Filter methods rank variables based on their score on some evaluation metric that measures their association with the target response. Variables with scores below a predetermined threshold are then excluded from further analysis (Heinze et al., 2018; Remeseiro & Bolon-Canedo, 2019). No learning method is used to evaluate the performance of any candidate variable sets; instead, these methods are based on the general characteristics of the data (Remeseiro & Bolon-Canedo, 2019). The result is a ranked list of variables rather than a best subset (Heinze et al., 2018; Remeseiro & Bolon-Canedo, 2019). As an example, consider Grissa et al. (2016)'s comparison of different feature selection methods in metabolomic biomarker discov-

ery. The comparison utilises a data set produced by a case-control study within the GAZEL French population-based cohort with measurements of many metabolites. Metabolite concentrations are known to be highly correlated. A filter selection method was applied to identify highly correlated feature pairs with the intention of forming a reduced feature set for further investigation. The methodology evaluates various feature selection approaches, particularly focusing on the performance of machine learning techniques, such as random forest (RF) and support vector machines (SVM).

Wrapper methods select an optimal set of variables by applying a search strategy to explore all possible sets and then evaluating the performance of each set as a predictor of the target response (Heinze et al., 2018; Remeseiro & Bolon-Canedo, 2019). Wrapper methods tend to be very computationally intensive compared to filter and embedded methods since they involve a search step for candidate sets and then an evaluation step to assess candidate set performance. Heuristics can be applied to reduce large search spaces (Heinze et al., 2018). As an example, consider Maghsoudloo et al. (2020)'s genetic biomarker identification method for chronic inflammatory lung diseases. Several wrapper methods, including the meta-heuristic Particle Swarm Optimisation (PSO) algorithm, were then applied to identify potential biomarkers from a candidate set of genes. In embedded methods, the selection and evaluation steps are integrated within the chosen learning method. As an example, consider Li and Liu (2021)'s prognostic genetic biomarker discovery for breast cancer using regularised Cox proportional hazards (RCPH) models. Several different penalty functions were considered, including the L1 penalty.

As filter methods return a ranked list of variables rather than a best subset, these methods are able to convey information about the relative importance of variables within the overall set (Heinze et al., 2018). It is also possible to investigate the implications of different thresholds for exclusion. This flexibility is often advantageous in an exploratory setting (Heinze et al., 2018). Unlike filter methods, wrapper and embedded methods identify a single best-performing variable subset with respect to the target response. These techniques do not typically supply a measure of the relative importance of variables within the best-performing subset (Heinze et al., 2018). However, relative importance can be calculated or extracted for some embedded methods (Heinze et al., 2018). This characteristic means that wrapper and embedded methods may be less appropriate for exploratory research than filter methods, but it also relieves researchers and statisticians of the burden of predetermining thresholds for variable exclusion, as is required when implementing filter methods (Heinze et al., 2018). It is important to note that knowledge of biological and pathogenic pathways still play an important role in biomarker selection. Any identified variable rankings or variable sets should always be interpreted in these contexts (Heinze et al., 2018).

Beyond the three broad methodological categories, variable selection can follow a model-based approach or can be performed without imposing a model structure (Heinze et al., 2018). When variable selection follows a model-based approach, the reduced set of variables is the by-product of an optimisation or statistical model of the relationships between inputs and output(s) (Heinze et al., 2018). Both Li and Liu (2021)'s embedded method for breast cancer biomarker discovery and Maghsoudloo et al. (2020)'s wrapper methods for chronic inflammatory lung disease biomarker identification are examples of model-based approaches. When performed without specifying an optimisation or statistical model, variable selection methods can incor-

porate unsupervised learning techniques to identify underlying structures within a set of clinical measurements (Solorio-Fernandez et al., 2020). If these structures reflect differences in the levels of exposure variables, they can be used to identify candidate biomarkers for further study. These methods are typically located within the filter approach to variable selection (Solorio-Fernandez et al., 2020). As an example, Luckett et al. (2023) used hierarchical cluster analysis to identify biomarker groups for autosomal dominant Alzheimer's disease from a variety of inputs, including clinical assessments, neuroimaging, and measurements of various biomolecules.

Metwally et al. (2022) identifies a need for statistical frameworks that can identify both group-based differences in omics features *and* temporal differences. These frameworks should be able to account for the characteristics of longitudinal data, including repeated measurements with differing numbers of samples per patient. Model-based variable selection methods are well-suited to this purpose, especially methods incorporating longitudinal regression modelling frameworks. Mixed-effect models can account for dependent observations created by repeated measures (Fitzmaurice et al., 2012). They can also explain multiple sources of patient-level heterogeneity in immune responses that are not relevant to the research question (Fitzmaurice et al., 2012). Additionally, mixed-effect models can accommodate differing numbers of samples per patient (Fitzmaurice et al., 2012). These longitudinal modelling frameworks are discussed in more detail in Chapter 4.

Longitudinal model-based variable selection typically attaches an L1 penalty term to the objective function of the model Reisetter and Breheny (2021). This penalty constrains the size of the model coefficients. After the estimation routine is completed, non-zero fixed effects correspond with the most informative outcomes that describe the most variability in a clinical outcome or endpoint. As an example, consider Reisetter and Breheny (2021)'s simulation study of penalised linear mixed-effect models for structured genetic data. However, outcome selection by penalised longitudinal modelling requires the outcome selection problem to be restructured as a variable selection problem. In other words, immune outcomes must be defined as variables that are predictive of a clinical endpoint.

Ultimately, outcome selection and variable selection have very different objectives. The purpose of variable selection is to support accurate prediction of a clinical endpoint or outcome. On the other hand, outcome selection aims to narrow the scope of an analysis to preserve statistical power. Nevertheless, given the established collection of variable selection methodologies, it is tempting to restructure an outcome selection problem as a variable selection problem. However, this changes the direction of causality should inference be the objective of the analysis. For this reason, it is important to develop specific outcome selection techniques that do not require this restructuring. In the literature, no dedicated outcome selection techniques for continuous outcomes are described, but this dissertation applies the core principles of the RM-ASCA+ framework, the extended ANOVA simultaneous component analysis (ASCA+) framework for repeated measures, to outcome selection in longitudinal immunological data.

The RM-ASCA+ framework combines linear mixed-effect models (LMMs) with principal component analysis (PCA) to decompose and visualise the separate effects of experimental factors over time (Madssen et al., 2021). It extends the ANOVA-simultaneous component analysis (ASCA) framework to handle repeated measures, unbalanced designs, and patient-level heterogeneity that is not of interest to the

research questions (Erdős et al., 2023). The method decomposes the multivariate response matrix into effect matrices that estimate comparisons of interest over time (Madssen et al., 2021). Performing a PCA identifies underlying patterns and trends in these longitudinal multivariate effects, which can be visualised to contribute to understandings of dynamic biological processes. Permutation testing allows for robust, non-parametric estimation of p-values for comparisons of interest over time (Madssen et al., 2021). RM-ASCA+ may apply a multiple hypothesis testing correction, such as the Benjamini-Hochberg procedure, to control the false discovery rate (Madssen et al., 2021).

The RM-ASCA+ steps as implemented in the *ALASCA* package for *R* (Jarmund et al., 2022) are summarised below:

1. Fit linear mixed-effect models to estimate the comparisons of interest for each outcome. The comparisons of interest should be informed by the research question and study design:
   - For example, different time points (or segments of a functional form of time) paired with a random intercept can be modelled as predictors of the average time development of an outcome (Jarmund et al., 2022).
   - Alternatively, different time points (or segments of a functional form of time) can be modelled as having different effects on the average time development of an outcome - dependent on group level. Time, group and their interaction, paired with a random intercept, are modelled as predictors of the average time development of an outcome (Jarmund et al., 2022).
2. Extract fixed-effect coefficient estimates (including the intercept) and place them in a multivariate response matrix.
3. Decompose this response matrix into a baseline mean matrix, an experimental design matrix, a time-effect matrix, a group-effect matrix, and a group-time-effect matrix.
4. Select the appropriate effect matrices for further analysis.
   - For example, performing a PCA on the combination of the time-effect, group-effect, and time-group effect matrices describes *the common time development of the groups* (Jarmund et al., 2022).
   - On the other hand, performing a PCA on the combination of the group-effect and time-group-effect matrices describes *how the time development of other group levels differs from the reference* (Jarmund et al., 2022).
5. Perform a PCA on individual or combined effect matrices, depending on the comparison of time development required by the research question.
   - For example, performing a PCA on the combination of the time-effect, group-effect, and time-group effect matrices describes the common time development of the groups (Jarmund et al., 2022).
   - On the other hand, performing a PCA on the combination of the group-effect and time-group-effect matrices describes how the time development of other group levels differs from the reference (Jarmund et al., 2022).
6. Extract scores for the comparisons of interest from the PCs, plot them, and interpret the results to understand how experimental factors contribute to variability of the outcomes.

- For example, when decomposing the common time development of the groups, the first few PCs may capture trends over time (Jarmund et al., 2022).
- On the other hand, when decomposing the time development of other group levels relative to the reference, the first few PCs may reflect differences between group levels (Jarmund et al., 2022).

7. Extract loadings for the outcomes from the PCs, plot them, and interpret the results to understand which outcomes are most variable with respect to the comparisons of interest.
8. Use permutation testing to evaluate the significance of the variance explained by the principal components. The observed variance of each effect's components is compared to those generated from permuted data. This generates p-values that indicate whether the trends and differences are statistically meaningful.
9. If the trends and differences are statistically meaningful, visualise the marginal means from the underlying regression models (Jarmund et al., 2022).

RM-ASCA+ aims to make statistical inferences about dynamic biological processes (Madssen et al., 2021). However, this dissertation applies its approach to outcome selection in longitudinal immunological data. As in RM-ASCA+, the model coefficients for the comparisons of interest are extracted from longitudinal models of each immune outcome and supplied as inputs to a PCA. By extracting and interpreting scores for comparisons of interest and loadings for immune outcomes, it may be possible to narrow the scope of further analysis by identifying smaller, targeted sets of immune outcomes for which there is evidence of subgroup differences.

The RM-ASCA+ framework combines longitudinal statistical modelling with dimension reduction techniques. Although the *ALASCA* package implements linear mixed-effect models, other longitudinal modelling frameworks may also be suitable (Jarmund et al., 2022). Some possibilities include generalised estimating equations (GEEs), generalised linear mixed-effect models (GLMM), and linear quantile mixed-effect models (LQMM) (Fitzmaurice et al., 2012). Chapter 4 compares the LMM and GLMM frameworks with application to the dissertation data set.

Although the RM-ASCA+ framework makes use of PCA, other dimension reduction techniques may also be appropriate for outcome selection. Chapter 5 and 6 compare outcome selection by PCA only to outcome selection by agglomerative hierarchical cluster analysis (HCA) followed by PCA. HCA organises input data into a hierarchy based on a prespecified (dis)similarity metric and linkage method for cluster formation (Murtagh & Contreras, 2017). In agglomerative HCA, each input is initially contained in its own cluster. The most similar clusters are iteratively merged until all inputs are contained in a single cluster. This clustering process generates a dendrogram that visually represents the relationships between the inputs (Murtagh & Contreras, 2017). The dendogram can be trimmed at different (dis)similarity values to retrieve different cluster structures. PCA can then be applied to model coefficient estimates for each cluster to identify smaller, targeted sets of immune outcomes for which there is evidence of subgroup differences. There were no examples of a similar methodology in the literature. However, Hierarchical Clustering on Principal Components (HCPC) combines PCA and agglomerative hierarchical cluster analysis (HCA), leveraging their respective strengths to improve clustering (Maugeri et al., 2021; Wisesty & Mengko, 2021).

After narrowing the scope of an analysis by outcome selection, proponents of multiple testing corrections would argue for adjusting the p-values to control for false positives. Scientific convention, somewhat arbitrarily, defines a p-value cut-off point (α) of 0.05 for statistical significance. If a single hypothesis test were conducted 20 times at $\alpha = 0.05$, there would be at least one false positive. However, if a family of 10 independent hypothesis tests were conducted, each with $\alpha = 0.05$, the probability of at least one false positive would be $1 - (1 - 0.05)^{10} = 0.40$. The desired $\alpha = 0.05$ can then be achieved by adjusting the p values for each individual test upward.

False Discovery Rate (FDR) control is often applied in the analysis of high-throughput biological data, such as immunological data (Goeman & Solari, 2014). It is less conservative than Family-Wise Error Rate (FWER) control, balancing the need to detect true positives with the need to minimise false positives. FDR controls the expected proportion of Type I errors within the set of rejected hypotheses, while FWER controls the probability of obtaining at least one false positive in a family of tests (Goeman & Solari, 2014). This difference gives FDR-based methods more power than FWER methods to detect statistically significant relationships between exposures of interest and clinical outcomes when they are truly present.

The Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) procedures are two examples of FDR control for multiple testing. These procedures make different assumptions about how hypotheses are correlated (Benjamini & Yekutieli, 2001). The BH procedure assumes independent hypotheses or positively dependent hypotheses, while the BY procedure accommodates , including negatively dependent hypotheses (Stevens et al., 2017). In immunological research, clinical outcomes may be strongly correlated (Goeman & Solari, 2014). Negatively correlated hypotheses are plausible in immunology, as the expression of one immune response can down-regulate the expression of another immune response (see Vigano et al. (2012)). The results obtained following the BH and BY procedures are compared in Chapter 7.

However, there are a number of reasons that a statistician would want to apply FWER control methods to their analyses of immunological data and other biomedical research areas. These methods are preferred for analyses where each result is required to be individually reliable - for example, in an end-stage analysis where the results will not be independently validated (Goeman & Solari, 2014). The Holm correction is an example of an FWER control method that makes no assumptions about the p-value dependence structure (Goeman & Solari, 2014). However, assuming a particular dependence structure provides more power to detect statistically significant relationships between an exposure of interest and a clinical outcome when they are truly present (Goeman & Solari, 2014).

Like the definition of statistical significance ($\alpha = 0.05$), multiple testing corrections have become a scientific convention. Critics argue that these adjustments are often arbitrarily defined and applied in inconsistent ways. Translating an example from Bender and Lange (2001) into an immunological context, consider a study that evaluates the immunogenicity of three new vaccines (T1, T2, and T3) compared to an established vaccine (C). There are two objectives: first, to determine whether T1, T2, and T3 differ from C; second, to determine whether T1, T2, and T3 differ from each other. Assuming that immunogenicity is defined by a single immune outcome, the analysis makes the following pairwise comparisons:

- T1 vs C
- T2 vs C
- T3 vs C
- T1 vs T2
- T2 vs T3
- T3 vs T1

These six comparisons may represent a family of six tests. The probability of a Type I error is then controlled across all comparisons. Alternatively, two families of three tests may be described: T1 vs C, T2 vs C, and T3 vs C; T1 vs T2, T2 vs T3, and T3 vs T1. The probability of a Type I error is then controlled within these two smaller families.

This example shows that there is some subjectivity in selecting which comparisons are included in the same family of hypothesis tests ((Bender & Lange, 2001), (Feise, 2002)) observe that the interpretation of the statistical significance of individual tests depends on how the family of tests is defined. Smaller families of tests apply less stringent corrections. Thus, there is a clear incentive to investigate fewer statistical hypotheses in an analysis, even if many potentially relevant hypotheses are left out in doing so ((Feise, 2002)).

Controlling the probability of a Type I error within a family of tests corresponds to a higher probability of a Type II error. Representing false negatives, Type II errors fail to detect statistically significant relationships between an exposure of interest and clinical outcomes when they are truly present. The consequences of both Type I and Type II errors should be considered when deciding on a strategy to correct for multiple testing (Feise, 2002).

In addition, many advocate that multiple testing corrections should not be applied in exploratory research (Bender & Lange, 2001). In these settings, hypothesis tests are often unstructured and descriptive of the data. There is little justification to control the probability of a false positive for an individual test for the probability of false positives in all tests conducted within the analysis. It is sufficient to report the observed p-values for the individual tests.

However, for a confirmatory study in which the family of tests contributes to a single conclusion, there is little debate about the need to adjust for multiple comparisons. As an example, return to the study that evaluates the immunogenicity of three new vaccines (T1, T2, and T3) compared to an established vaccine (C) for a single immune outcome. To investigate whether all three new vaccines (T1, T2, and T3) differ from C, one must control the Type I error rate within the following family of tests: T1 vs C, T2 vs C, and T3 vs C.

Chapter 3

Data Preparation

This chapter discusses the data preparation process in the outcome selection workflow (Step 1 in Figure 1.2) with application to the dissertation data set. This includes checking the quality of the data, defining the comparisons of interest for outcome selection, and preparing the data for longitudinal modelling (Step 2 in Figure 1.2).

The structure of the dissertation data set for Group A is summarised below.

Table 3.1: Structure of the dissertation data set for Group A.

ID	Time	Sex	MVA85A	QFT	Feeding	Antibiotic	$y_{1A} \cdots y_{33A}$
$i_{1A} \cdots i_{65A}$	1, 2, 3	M, F	MVA85A, Control	Positive, Negative	Breast, Formula	Yes, No	...

A total of 65 infants were enrolled in Group A and assigned a unique study identification number (ID). Thirty-three immune outcomes ($y_{1A} \cdots y_{33A}$) were recorded at three different time points ($Time$). Four exposures of interest were recorded at baseline for every infant ($MVA85A$ and QFT at birth; $Feeding$ and $Antibiotic$ at the time of BCG vaccination). These exposures of interest describe patient subgroups that may differ with respect to their immune outcome profiles ($y_{1A} \cdots y_{33A}$) over time: (1) infants receiving MVA85A priming at birth compared to a control, (2) infants born to mothers with a positive QFT test, (3) infants fed breast milk compared to formula, and (4) infants treated with cotrimoxazole. The sex of the infant (Sex) was measured as an additional covariate.

3.1 Data Quality

The completeness of immunological data is an important component of data quality. Patients with missing values for exposures of interest should be removed from the data set prior to modelling. This ensures that the model coefficient estimates extracted from the longitudinal models for dimension reduction are based on a complete and consistent data set.

- Maternal QFT (QFT) was not measured for four infants in Group A. Since this is one of the exposures of interest, these infants were removed from the analysis.
  - The analysis of Group A is therefore limited to 61 infants out of the 65 infants who were enrolled in the group.

The appropriate approach for patients with missing values for immune outcomes depends on the chosen modelling framework and the type of missing data. This dissertation applies a mixed-effect modelling framework that can estimate model parameters without explicitly imputing the missing outcome values. However, these estimates are made under the assumption of missing at random (MAR) (Fitzmaurice et al., 2012). This assumes that the probability of missing data depends on the observed data, not on the missing values themselves.

Several incomplete records are included in the modelling data set under a MAR assumption.

- In Group A, some missingness in immune outcomes is created by early termination and loss to follow-up.
  - One infant with immune outcome measurements at the time of BCG vaccination and 8 weeks after BCG vaccination ($Time = 1$ and $Time = 2$) was lost to follow-up at one year after birth ($Time = 3$).
  - Three infants with immune outcome measurements at the time of BCG vaccination and 8 weeks after BCG vaccination ($Time = 1$ and $Time = 2$) had no measurements at $Time = 3$ due to early termination.
  - Two more infants were lost to the study after $Time = 1$ due to early termination.
- At any time point ($Time \in 1, 2, 3$), no single immune outcome is observed in all Group A infants.
  - The highest proportion of incomplete trajectories is observed for the proportions of BCG-reactive proliferating $CD4^+$ T cells.
  - Excluding missing values from early termination and loss to follow up, 21 out of 61 Group A infants had at least one missing measurement of $\%GrA^+$ $Ki67^+$ $CD4^+$, $\%GrB^+$ $Ki67^+$ $CD4^+$, $\%GrK^+$ $Ki67^+$ $CD4^+$, and/or $\%Perf^+$ $Ki67^+$ $CD4^+$.

3.2 Exposure Variables

Longitudinal relationships between immune outcomes and exposures of interest are examined individually, with each model focusing on one outcome and one exposure of interest. If there is a known association between one or more exposures of interest, it should be addressed prior to modelling - for example, by creating composite exposure variables.

In the dissertation data set, *FA* was created as a composite exposure variable of two binary indicators for feeding practices (*Feeding*) and cotrimoxazole treatment (*Antibiotic*). As prophylactic cotrimoxazole treatment was strongly recommended in HIV-exposed infants fed breast milk (Bourke et al., 2019), an association was anticipated between the number of infants fed breast milk and the number of infants treated with cotrimoxazole. A two-sided Fisher's exact test was performed to examine the association between *Feeding* and *Antibiotic* in Group A:

H_0: The true odds ratio of the counts is equal to 1.
H_1: The true odds ratio of the counts is not equal to 1.

Table 3.2: Results of two-sided Fisher's Exact Test to examine the association between *Feeding* and *Cotrimoxazole* in Group A.

	Group A
Odds Ratio (OR)	0.3354
95% Confidence Interval for OR (*lower*)	0.1671
95% Confidence Interval for OR (*upper*)	0.6574
p-value	0.0001

There was strong evidence for an association between *Feeding* and *Antibiotic* in Group A ($p = 0.0001$). As *Feeding* and *Antibiotic* are binary indicators, the composite exposure variable *FA* was created from combinations of their levels.

Table 3.3: Summary of the levels of *FA*, a composite exposure variable created from *Feeding* and *Antibiotic*.

FA Level	*Feeding* Level	*Antibiotic* Level
Formula & Cotrimoxazole (*FA = 0*)	Formula	Cotrimoxazole
Breastfeeding & No Cotrimoxazole (*FA = 2*)	Breastfeeding	No Cotrimoxazole
Formula & No Cotrimoxazole (*FA = 1*)	Formula	No Cotrimoxazole
Breastfeeding & Cotrimoxazole (*FA = 3*)	Breastfeeding	Cotrimoxazole

3.3 Outcomes

Immune outcomes tend to have skewed distributions (Genser et al., 2007). Several longitudinal modelling techniques, including the mixed-effect modelling framework investigated in the dissertation, require a specified parametric form for the conditional distribution of the outcome. Few parametric distributions accommodate left-skewed outcome distributions. Hence, it is important to identify immune outcomes with left-skewed distributions and transform them into right-skewed distributions. This can be achieved through a reflection $[(x,\, y) \rightarrow (c - x,\, y)]$ where c is a value equal to (or slightly greater than) the maximum value of the immune outcome.

As an example from the dissertation data set, consider the percentage of Bulk $CD8^+$ T cells with $R7^+RA^+$ phenotype measured in Group A at three time points.

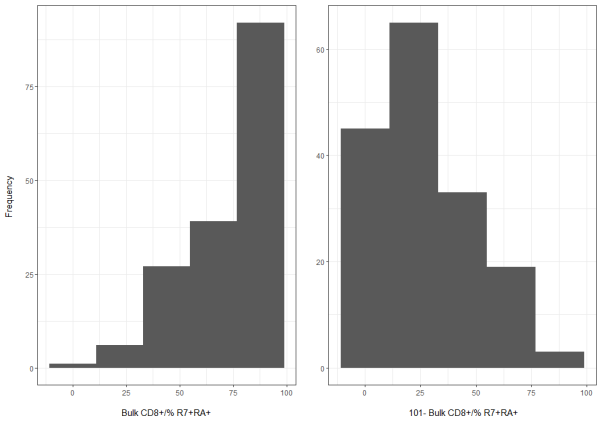


Figure 3.1: Histograms of observed percentages of bulk profiled $CD8^+$ T cells with $R7^+RA^+$ phenotype before and after reflection.

This phenotype is observed in a large percentage of cells, creating a left-skewed distribution of its measurements (*left* in Figure 3.1). Measurements of %Bulk $CD8^+$ T cells with $R7^+RA^+$ phenotype are subtracted from $c = 101$ (*right* in Figure 3.1), and left-skewness becomes right-skewness. After performing this reflection, the conditional distributions for these outcomes can be accommodated by a wider range of parametric forms.

Histograms were plotted for the measurements of each immune outcome for all patients at all visits (Appendix C).

Five left-skewed immune outcomes, all measured in Group A, were reflected when preparing the data for modelling:

- *Bulk %CD4$^+$ R7$^+$ RA$^+$*
- *Bulk %CD8$^+$ R7$^+$ RA$^+$*
- *%GrA$^+$ Ki67$^+$ NK*
- *%GrB$^+$ Ki67$^+$ NK*
- *%Perf$^+$ Ki67$^+$ NK*

3.4 Comparisons of Interest

Following data preparation, longitudinal models are formulated to estimate subgroup differences in immune outcomes over time. These subgroups are described by different levels of the exposures of interest (*EOI*). For a specific time point (described by the levels of *Time*), model coefficients compare how different exposure levels modify the expected value of the outcome. These comparisons of interest, defined for the dissertation data set in Chapter 3, are translated into a mixed-effect model structure for outcome selection in this chapter.

Within the RM-ASCA+ framework, on which the dissertation outcome selection technique is based, the comparisons of interest incorporate some (or all) of the following:

- Comparison of the effects of different *EOI* levels, holding *Time* levels constant
- Comparison of the effects of different *Time* levels, holding *EOI* levels constant
- Comparison of the effect of different *EOI* levels on different *Time* levels (*EOI:Time*)

RM-ASCA+ then uses these comparisons of interest to decompose and visualise the separate effects of experimental factors over time (Madssen et al., 2021). However, this dissertation argues that these comparisons of interest can inform outcome selection in longitudinal immunological data, in particular:

- Comparison of the effects of different *EOI* levels, holding *Time* levels constant
- Comparison of the effect of different *EOI* levels on different *Time* levels (*EOI:Time*)

By comparing how different exposure levels modify the expected value of outcomes, one can identify immune outcomes with the most evidence for subgroup differences and select them for further analysis. This restricts the scope of the analysis to a smaller number of statistical hypotheses, preserving statistical power when applying corrections for multiple hypothesis testing.

The comparisons of interest, described in collaboration with the South African Tuberculosis Vaccine Initiative (SATVI), are summarised below.

3.4.1 EOI: *MVA85A*

Group A

- How does MVA85A priming modify the effect of time at BCG vaccination (56 days after birth)? (*EOI:Time* where *MVA85A = 1* and *Time = 1*)

- What is the effect of MVA85A priming 8 weeks after BCG vaccination (112 days after birth)? (*EOI* where *MVA85A = 1* and *Time = 2*)
- How does MVA85A priming modify the effect of time $\sim$ 44 weeks after BCG vaccination (365 days after birth)? (*EOI:Time* where *MVA85A = 1* and *Time = 3*)

3.4.2 EOI: *QFT*

Group A

- How does a positive maternal QFT modify the effect of time at BCG vaccination (56 days after birth)? (*EOI:Time* where *QFT = 1* and *Time = 1*)
- What is the effect of a positive maternal QFT 8 weeks after BCG vaccination (112 days after birth)? (*EOI* where *QFT = 1* and *Time = 2*)
- How does a positive maternal QFT modify the effect of time $\sim$ 44 weeks after BCG vaccination (365 days after birth)? (*EOI:Time* where *QFT = 1* and *Time = 3*)

3.4.3 EOI: *FA*

Group A

- In infants treated with cotrimoxazole, how does feeding practice modify the effect of time at BCG vaccination (56 days after birth)? (*EOI:Time* where *FA = 0* or *FA = 3* and *Time = 1*)
- In infants treated with cotrimoxazole, what is the effect of feeding practice 8 weeks after BCG vaccination (112 days after birth)? (*EOI* where *FA = 0* or *FA = 3* and *Time = 2*)
- In infants treated with cotrimoxazole, how does feeding practice modify the effect of time $\sim$ 44 weeks after BCG vaccination (365 days after birth) for Group A? (*EOI:Time* where *FA = 0* or *FA = 3* and *Time = 3*)
- In infants fed with breast milk, how does cotrimoxazole treatment modify the effect of time at BCG vaccination (56 days after birth)? (*EOI:Time* where *FA = 2* and *Time = 1*)
- In infants fed with breast milk, what is the effect of cotrimoxazole treatment 8 weeks after BCG vaccination (112 days after birth)? (*EOI* where *FA = 2* and *Time = 2*)
- In infants fed with breast milk, how does cotrimoxazole treatment modify the effect of time $\sim$ 44 weeks after BCG vaccination (365 days after birth)? (*EOI:Time* where *FA = 2* and *Time = 3*)
- In infants fed with formula, how does cotrimoxazole treatment modify the effect of time at BCG vaccination (56 days after birth)? (*EOI:Time* where *FA = 1* and *Time = 1*)
- In infants fed with formula, what is the effect of cotrimoxazole treatment 8 weeks after BCG vaccination (112 days after birth)? (*EOI* where *FA = 1* and *Time = 2*)

– In infants fed with formula, how does cotrimoxazole treatment modify the effect of time $\sim$ 44 weeks after BCG vaccination (365 days after birth)? (*EOI:Time* where *FA = 1* and *Time = 3*)

3.4.4 Additional Covariates

The comparisons of interest should be adjusted for additional variables are known to influence the immune outcomes, including pre-randomisation covariates such as age or sex (Madssen et al., 2021). In earlier efforts to analyse the dissertation data set, SATVI identified that male and female infants often had strikingly different longitudinal response profiles. Hence, the linear predictor of the longitudinal model should estimate the comparisons of interest while holding sex differences in responses constant ($EOI * Time + Sex * Time$):

– $MVA85A * Time + Sex * Time$
– $QFT * Time + Sex * Time$
– $FA * Time + Sex * Time$

3.4.5 Reference Coding

Time

The first measurement of immune outcomes after BCG vaccination ($Time = 2$) was specified as the reference level for study time. This is when the peak immune response to BCG is anticipated.

Exposures of Interest

Infants randomised to receive MVA85A priming (*MVA85A = 1*) at birth were compared to infants randomised to receive placebo (*MVA85A = 0*). Infants born to mothers with a positive QFT (*QFT = 1*) were compared to infants born to mothers with a negative QFT (*QFT = 0*).

Different reference levels are specified for FA depending on the comparison of interest.

Table 3.4: Summary of reference levels for different comparisons of interest described by the *FA* variable.

Comparison of Interest	Reference Level
Formula & Cotrimoxazole (*FA = 0*)	Breastfeeding & Cotrimoxazole (*FA = 3*)
Breastfeeding & No Cotrimoxazole (*FA = 2*)	Breastfeeding & Cotrimoxazole (*FA = 3*)
Formula & No Cotrimoxazole (*FA = 1*)	Formula & Cotrimoxazole (*FA = 0*)
Breastfeeding & Cotrimoxazole (*FA = 3*)	Formula & Cotrimoxazole (*FA = 0*)

Additional Covariates

Female infants are the reference level for *Sex*.

3.5 Data Structure

Patients with missing values for maternal QFT are removed from the data sets, and *Feeding* and *Antibiotic* are replaced by the composite exposure FA. The structure of the prepared data sets is presented below. Reference levels are highlighted in red. Each longitudinal model estimates the comparisons of interest for one exposure variable and a single outcome. Hence, a separate data set was prepared for each exposure of interest ($MVA85A$, QFT, FA).

Table 3.5: Structure of the data sets prepared for modelling in Group A.

ID	**Time**	**Sex**	**MVA85A**	$y_{1A} \cdots y_{33A}$
$i_{1A} \ldots i_{61A}$	1, 2, 3	M, F	MVA85A, Control	...

ID	**Time**	**Sex**	**QFT**	$y_{1A} \cdots y_{33A}$
$i_{1A} \ldots i_{61A}$	1, 2, 3	M, F	Positive, Negative	...

ID	**Time**	**Sex**	**FA**	$y_{1A} \cdots y_{33A}$
$i_{1A} \ldots i_{61A}$	1, 2, 3	M, F	0, 1, 2, 3	...

Chapter 4

Modelling

This chapter discusses the longitudinal modelling step in the outcome selection workflow (Figure 1.2) in more detail. Longitudinal regression models are a key component of both the outcome selection framework and the RM-ASCA+ framework, (Madssen et al., 2021) on which it is based. An overview of regression modelling theory is presented in this chapter, including extensions to the basic regression model for the analysis of longitudinal immunological data.

In RM-ASCA+ and the outcome selection technique presented in this dissertation, longitudinal relationships between study outcomes and exposures of interest are estimated with mixed-effect models. The coefficients of these mixed-effect models describe how the levels of an exposure of interest modify the expected value of the outcome over time. These comparisons of interest, defined for the dissertation data set in Chapter 3, are translated into a mixed-effect model structure for outcome selection in this chapter.

RM-ASCA+ implements the linear mixed-effect modelling (LMM) framework to estimate how the levels of an exposure of interest modify the expected value of the outcome over time. However, the generalised linear mixed-effect (GLMM) modelling framework may be a more appropriate modelling framework as it offers greater flexibility to accommodate skew, long-tailed immune outcomes. This chapter compares these modelling frameworks through application to three immune outcomes that exemplify the characteristics of immunological data described in Chapter 2.

4.1 An Overview of Regression Modelling

Regression modelling techniques are frequently implemented to investigate hypotheses in biomedical research. As biomedical research often aims to explain and interpret complex biological processes, regression models are favoured for their ability to break down how different exposure variables contribute to observed variability in clinical measurements (Vach, 2012). Chapters 2 and 9 of Harrell et al. (2001) describes regression model structure, interpretation of model parameters, and regression model assumptions. The theory presented in these chapters is summarised in this section.

A regression model describes a relationship between the expected value of an outcome Y and a pre-specified set of independent variables ($\mathbf{X}$). The expected value of Y is modelled as conditional on a weighted sum of these independent variables in $\mathbf{X}$. The general regression model has the following form:

$$Y = g^{-1}\left(\mathbf{X}\boldsymbol{\beta} + \epsilon\right)$$

A parametric form is specified for the conditional distribution of Y which defines a particular mean-variance relationship through some link function (g).

- The conditional mean is $\mathbf{E}\left(Y \mid \mathbf{X}\right) = \boldsymbol{\mu} = g^{-1}\left(\mathbf{X}\boldsymbol{\beta}\right)$.
- The variance is a function of $\boldsymbol{\mu}$ with $\text{Var}\left(Y \mid \mathbf{X}\right) = \text{V}\left(g^{-1}\left(\mathbf{X}\boldsymbol{\beta}\right)\right) = \text{V}\left(\boldsymbol{\mu}\right)$.

If Y conditional on $\mathbf{X}$ is normally distributed, g is the identity function. This describes a linear relationship between $\boldsymbol{\mu}$ and $\mathbf{X}\boldsymbol{\beta}$. For non-normal conditional distributions of Y, g describes a generalized linear relationship between $\boldsymbol{\mu}$ and $\mathbf{X}\boldsymbol{\beta}$.

The independent variables' respective contributions to $\boldsymbol{\mu}$ are weighted with coefficients ($\boldsymbol{\beta}$). These unknown model parameters are estimated from the observed data for both Y and $\mathbf{X}$. Each coefficient represents the average change in the outcome Y corresponding with changes in an independent variable in $\mathbf{X}$. The coefficient is a conditional estimate holding all other independent variables constant.

- If the independent variable is continuous, the model coefficient estimates the average change in Y per unit change in the independent variable.
- If the independent variable is categorical, such as an exposure variable, this model coefficient compares the average effect of different levels.

When the regression model is applied to the observed data to estimate model parameters, variability in Y that is unexplained by the regression model is contained in a residual vector ($\boldsymbol{\epsilon}$).

- The distribution of $\boldsymbol{\epsilon}$ indicates whether the assumed mean-variance relationship is appropriate.
- The residuals are assumed to be independent and should have a constant variance across all observations and values or levels of the independent variables in $\mathbf{X}$.

If residual analysis demonstrates that the assumed mean-variance relationship is appropriate, the estimated model coefficients ($\hat{\boldsymbol{\beta}}$) can be interpreted to understand the relationship between an outcome Y and one or more of the independent variables in $\mathbf{X}$. A model coefficient describes the magnitude and direction of association between Y and an independent variable in $\mathbf{X}$ from the observed data, holding all other independent variables constant (Harrell et al., 2001).

If researchers wish to use the regression model to understand this relationship in broader populations beyond their observed data, probability statements - for example, constructing confidence intervals for $\boldsymbol{\beta}$ - can be estimated for $\hat{\boldsymbol{\beta}}$ (Harrell et al., 2001). However, causal interpretations require strict assumptions related to the randomness of the experimental design, measurement error, model goodness of fit, and the variables included in the regression model (Harrell et al., 2001).

As an example of $\hat{\boldsymbol{\beta}}$ interpretation from the dissertation data set, consider the following regression model:

$$\mathbf{Y} = g^{-1}\left(\beta_0 + \beta_1\,\mathbf{MVA85A} + \beta_2\,\mathbf{Sex} + \boldsymbol{\epsilon}\right),$$

where Y is an immune outcome measured in the dissertation data set, **MVA85A** is a binary variable describing whether a patient is randomly allocated to receive MVA85A priming or control, and **Sex** is a binary variable describing whether a patient is male or female.

The following expression can be obtained for $\boldsymbol{\mu}$:

$$\mathbf{E}\left(Y \mid \mathbf{MVA85A};\, \mathbf{Sex}\right) = \boldsymbol{\mu} = g^{-1}\left(\beta_0 + \beta_1\,\mathbf{MVA85A} + \beta_2\,\mathbf{Sex}\right).$$

Assume the reference category for β_1 is the control group and that β_2 is estimated with reference to female patients. Below is an exploratory interpretation of the estimated model coefficients ($\hat{\boldsymbol{\beta}}$):

- The magnitude and direction of $\hat{\beta}_0$ estimates $\boldsymbol{\mu}$ for female patients belonging to the control group, i.e., $\mathbf{E}(Y \mid \mathbf{MVA85A} = \text{Control};\ \mathbf{Sex} = \text{Female}) = \hat{\beta}_0$.
- The magnitude and direction of $\hat{\beta}_1$ estimates how MVA85A priming affects $\boldsymbol{\mu}$ relative to the control group when patient sex is held constant, i.e., $\mathbf{E}(Y|\mathbf{MVA85A} = \text{MVA85A};\ \mathbf{Sex} = \text{Female}) = \hat{\beta}_0 + \hat{\beta}_1$.
- The magnitude and direction of $\hat{\beta}_2$ estimates how being male affects $\boldsymbol{\mu}$ relative to being female when MVA85A priming is held constant, i.e., $\mathbf{E}(Y \mid \mathbf{MVA85A} = \text{Control};\ \mathbf{Sex} = \text{Male}) = \hat{\beta}_0 + \hat{\beta}_2$.

If a causal interpretation is required and assumptions are met, confidence intervals for $\boldsymbol{\beta}$ are constructed around estimated coefficients ($\hat{\boldsymbol{\beta}}$). Using β_1 as an example, a 95% confidence interval for β_1 would describe the set of all values that β_1 could assume at the 5% significance level ($\alpha = 0.5$) when holding patient sex constant.

From a regression model and its parameter estimates, one can calculate contrasts to describe or make inferences about subgroup differences that are not directly estimated by the model's coefficients. These differences are described by a contrast matrix designed to retrieve selected linear combinations of exposure variables in $\mathbf{X}$ from the fitted regression model.

For example, in the dissertation data set, one may wish to estimate pairwise contrasts for the expected value of the outcome Y for different levels of the exposure variable FA. Consider the following regression model:

$$\mathbf{Y} = g^{-1}\left(\beta_0 + \beta_1\,\mathbf{FA1} + \beta_2\,\mathbf{FA2} + \beta_3\,\mathbf{FA3} + \beta_4\,\mathbf{Sex} + \boldsymbol{\epsilon}\right),$$

where Y is an immune outcome measured in the dissertation data set. **FA** is a four-level exposure variable describing different combinations of feeding practices and cotrimoxazole treatment. $FA = 0$ is specified as the reference level and dummy variables compare all other levels of FA to $FA = 0$:

1. $FA = 0$ (reference level)
2. $FA = 1$ (dummy variable: *FA1*)
3. $FA = 2$ (dummy variable: *FA2*)
4. $FA = 3$ (dummy variable: *FA3*)

The regression model does not directly estimate all pairwise comparisons between the levels of FA. For example, one might wish to estimate the difference between $FA = 1$ and $FA = 2$ ($\hat{C}$). The matrix of model coefficient estimates $\hat{\boldsymbol{\beta}}$ is then multiplied by a corresponding contrast vector C to estimate this difference (holding *Sex* constant):

$$\begin{aligned}
\hat{C} &= C \cdot \hat{\boldsymbol{\beta}} \\
&= \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \end{bmatrix} \cdot \begin{pmatrix} \hat{\beta}_0 & 0 & 0 & 0 & 0 \\ 0 & \hat{\beta}_1 & 0 & 0 & 0 \\ 0 & 0 & \hat{\beta}_2 & 0 & 0 \\ 0 & 0 & 0 & \hat{\beta}_3 & 0 \\ 0 & 0 & 0 & 0 & \hat{\beta}_4 \end{pmatrix} \\
&= \hat{\beta}_2 - \hat{\beta}_3
\end{aligned}$$

The standard error of $\hat{C}$ is computed by considering the covariance matrix $\mathbf{V}$ of the regression model (Harrell et al., 2001):

$$SE = \sqrt{Var(C)} = \sqrt{C^T \cdot \mathbf{V} \cdot C}$$

However, in many cases, a contrast is more meaningful if it describes population-level differences rather than differences in conditional estimates. This requires formulating contrasts as differences in marginal means averaged over the distributions of all exposure variables in $\mathbf{X}$ that are not considered in the contrast. For example, suppose one wishes to estimate the population-level difference between $FA = 1$ and $FA = 2$ (C). This requires computing the marginal means μ_{FA1} and μ_{FA2}, averaged over the distribution of patient sex.

The marginal mean for $FA1$ is:

$$\mu_{FA1} = p(\text{Sex} = \text{Female}) \cdot g^{-1}(\beta_0 + \beta_1) + p(\text{Sex} = \text{Male}) \cdot g^{-1}(\beta_0 + \beta_1 + \beta_4),$$

with the probabilities $p(\text{Sex} = \text{Female})$ and $p(\text{Sex} = \text{Male})$ representing the overall proportions of male and female patients in the input data set. Similarly, the marginal mean for $FA2$ is:

$$\mu_{FA2} = p(\text{Sex} = \text{Female}) \cdot g^{-1}(\beta_0 + \beta_2) + p(\text{Sex} = \text{Male}) \cdot g^{-1}(\beta_0 + \beta_2 + \beta_4),$$

The marginal contrast C is the difference between these two marginal means (μ_{FA1} and μ_{FA2}). Once again, the standard error of C is computed by considering the variance-covariance matrix $\mathbf{V}$ of the regression model (Harrell et al., 2001):

$$SE = \sqrt{Var(C)} = \sqrt{C^T \cdot \mathbf{V} \cdot C}$$

However, the variance of C now reflects both the uncertainty in the coefficient estimates ($\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_4$) and how averaging over the distribution of patient sex affects μ_{FA1} and μ_{FA2}. Each marginal mean is weighted by the distribution of male and female patients in the input data set, and this is considered in the standard error of C. For a conditional contrast, the variance of C only reflects uncertainty in the coefficient estimates.

The *R* package *emmeans* (Lenth, 2024) provides functions to calculate and visualise these estimated marginal means, their contrasts, and the standard errors of their contrasts for a variety of regression models.

4.2 Extensions for Longitudinal Immunological Data

The basic regression model assumptions are violated when the data is produced by longitudinal study designs that take repeated measurements of an immune outcome from the same patient at different points in time (Genser et al., 2007). Repeated measurements are clustered within patients, and regression assumptions of residual independence are inappropriate (Fitzmaurice et al., 2012).

The characteristics of immunological data and their implications for regression modelling are reviewed in Chapter 2. In summary, immunological outcomes tend to have skew, long-tailed distributions with many null measurements. These characteristics make it challenging to specify an appropriate parametric form for the distribution of the expected value of an immune outcome Y conditional on a weighted sum of

independent variables in $\mathbf{X}$. However, a number of parametric forms that may accommodate these characteristics are implemented in statistical modelling packages in *R*:

- *Log-normal distributions* are frequently implemented to describe long-tailed, positively skewed continuous data generated by biological processes, including the immune system (Uh et al., 2008).
  - The immune outcome must be strictly positive, as true zeroes cannot be generated by this distribution.
  - A hurdle log-normal model allows the zero-generating process to be modelled separately from the process for non-zero values.
- *Gamma distributions* are also applied to model immunological processes (see Martin-Escolano et al. (2023) and Huang et al. (2017)).
  - The immune outcome must be strictly positive, as true zeroes cannot be generated by this distribution.
  - A hurdle Gamma model allows the zero-generating process to be modelled separately from the process for non-zero values.
- *Skew-normal and skew-T distributions* may also be applied to model immunological data (see Dias-Domingues et al. (2024) and Lin and Wang (2013)).
  - The skew-normal distribution features a shape parameter to accommodate skewness while the skew-T distribution introduces parameters to account for both skewness and long tails (Dias-Domingues et al., 2024).
  - These distributions span the real line and, in theory, the modelled immune outcome should then be able to take on any real value.
- The *compound Poisson-Gamma distribution* can accommodate continuous, positively skewed, long-tailed data with a large proportion of exact zeroes.
  - A power parameter (p) describes the respective weightings of the discrete Poisson component creating mass at zero and the continuous Gamma component (Zhang, 2013).
  - This distribution is most commonly implemented in the insurance industry but it has biomedical applications in modelling immunological data (Lapham, 2020).

Immune outcomes observed in immunological research are known to be highly variable (Walzl et al., 2011). Even in randomised controlled trial (RCT) settings where unmeasured confounding is minimised by randomisation, substantial heterogeneity is often observed in patient outcomes (Gewandter et al., 2019). Modelling immunological data therefore requires careful consideration of sources of heterogeneity that may be observed at the patient level, including (see Gewandter et al. (2019)):

1. *Between-patient variability:* There is immune outcome variability for different patients, but this difference exists irrespective of MVA85A priming.
2. *Within-patient variability:* There is immune outcome variability over time for individual patients given MVA85A priming.
3. *Treatment-by-patient variability:* There is immune outcome variability for different patients receiving MVA85A priming, and this is due to inherent individual differences in patients' responses to MVA85A priming.

4.3 (Generalised) Linear Mixed-Effect Models

(Generalised) linear mixed-effect models - (G)LMMs - are multi-level models that extend the basic regression model to longitudinal data with repeated measures. (G)LMMs are appropriate for immunological research where patient-level heterogeneity is anticipated and should be modelled explicitly (Lapham, 2020).

Chapters 1 to 5 of Pinheiro and Bates (2006) describe linear mixed-effect model structure, assumptions, and estimation and interpretation of model parameters. Chapter 7 of Demidenko (2013) describes the generalised linear mixed-effect model structure and parameter estimation. The theory presented by these authors is summarised below.

Mixed-effect models partition the outcome (dependent variable) variance into fixed and random components (Pinheiro & Bates, 2006):

- *Fixed effects* are included as model parameters to capture fixed characteristics of the experimental design or sample - for example, a patient's allocation to receive MVA85A priming or a patient's sex.
- *Random effects* are included as latent variables to permit model parameters to vary within different levels of the analysis - for example, on the level of individual patients.

By permitting model parameters to vary within different levels of the analysis, random effects capture the effects of random sampling from a population, unobserved sources of patient-level differences in the outcome, and within-patient correlations between measurements (Lapham, 2020; Pinheiro & Bates, 2006).

The random-effect structure may include a random intercept and/or a random slope. Within the RM-ASCA+ framework, the longitudinal models typically include a normally distributed random intercept only (Madssen et al., 2021). More complex random-effect structures may be motivated by the study design.

The assumption of normally distributed random effects can be met fairly loosely.

- Simulation studies have shown that linear mixed-effect model parameter estimates are fairly robust to many different violations of random-effect distributional assumptions (Schielzeth et al., 2020).
- Model coefficient estimates were strongly affected by correlated fixed effects but not skewness, bimodality or heteroscedasticity of random effects (Schielzeth et al., 2020).

Both the conditional population-average outcome profile and individual patient outcome profiles can be extracted. However, the population average is estimated using a subject-specific approach rather than a truly marginal approach (Pinheiro & Bates, 2006). These attributes position (G)LMMs as an appropriate regression model class for longitudinal immunological data.

Using an example from the dissertation data set, one can formulate a (G)LMM to describe variability of an immune outcome Y over the three levels of study time ($Time$) in terms of exposure to MVA85A priming ($MVA85A$). $Time$ is encoded using dummy variables with $Time2$ as the reference level. $Time1_i = 1$ at the time of

BCG vaccination (56 days after birth); $Time3_i = 1 \approx$44 weeks after BCG vaccination (365 days after birth).

The conditional expectation of the immune outcome for a single patient ($i \in 1 : n$) over time is modelled as follows:

$$\mathbf{y}_{ij} = g^{-1}\left((\boldsymbol{\beta}_0 + u_i) + \boldsymbol{\beta}_1 \text{ MVA85A}_{ij} + \boldsymbol{\beta}_3 \text{ Time1}_{ij} + \boldsymbol{\beta}_4 \text{ Time3}_{ij} + \epsilon_{ij}\right)$$

where g is a link function that specifies a (generalised) linear relationship between the conditional distribution of $\mathbf{Y}$ and $\mathbf{X}\boldsymbol{\beta}$.

The patient-specific random intercept u_i accounts for between-patient random variation around the population mean of Y. An immune outcome profile over time, specific to patient i, can then be modelled as a perturbation of the average immune outcome profile given by $\boldsymbol{\beta}_0$.

The model coefficients are interpreted as follows:

- $\boldsymbol{\beta}_0$ is the expectation of $\mathbf{Y}$ for a patient, holding MVA85A priming and study time constant, conditional on the unobserved random effects ($\mathbf{u}$).
- $\boldsymbol{\beta}_1$ describes the expected change in $\mathbf{Y}$ when a patient receives MVA85A priming, holding $Time$ constant, conditional on $\mathbf{u}$.
- $\boldsymbol{\beta}_3$ describes the expected change in $\mathbf{Y}$ for a patient at $Time1$ relative to $Time2$, holding MVA85A priming constant and conditional on $\mathbf{u}$.
- $\boldsymbol{\beta}_4$ describes the expected change in $\mathbf{Y}$ for a patient at $Time3$ relative to $Time2$, holding MVA85A priming constant and conditional on $\mathbf{u}$.

In matrix notation, the model for all patient outcome profiles is as follows:

$$\mathbf{Y} = g^{-1}\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}\right)$$

where $\mathbf{Z}$ is a design matrix relating $\mathbf{Y}$ to the random effects ($\mathbf{u}$). In this random intercept model, $\mathbf{Z}$ is $\mathbf{I}$.

(G)LMMs require a predetermined mean-variance relationship for the conditional distribution of $\mathbf{Y}$. A Normal distribution is typically assumed for random effects ($\mathbf{u} \sim \text{N}(\mathbf{0},\ \sigma_u^2\,\mathbf{R})$), where $\mathbf{R}$ is a variance-covariance matrix for the random effects. The residual distribution depends on the conditional distribution of $\mathbf{Y}$ given the observed data and $\mathbf{u}$.

- In a linear relationship between $\mathbf{X}\boldsymbol{\beta}$ and the conditional distribution of $\mathbf{Y}$, residuals follow a Normal distribution ($\boldsymbol{\epsilon} \sim \text{N}(\mathbf{0},\ \sigma_e^2\,\mathbf{G})$) where $\mathbf{G}$ is a variance-covariance matrix for the residuals (Pinheiro & Bates, 2006).
  - These assumptions are made by the LMM framework utilised within RM-ASCA+ and implemented in the *ALASCA* package for *R*.
- For a generalised linear relationship between $\mathbf{X}\boldsymbol{\beta}$ and the conditional distribution of $\mathbf{Y}$, the scaled residuals follow a standard Normal distribution ($s(\boldsymbol{\epsilon}) \sim \text{N}(\mathbf{0},\ \mathbf{I})$) (Hartig, 2018). These scaled residuals may be obtained using a simulation-based approach (see Appendix D).
  - Statistical methods are being explored to extend the RM-ASCA+ framework to generalised linear relationships (Madssen et al., 2021). For example, a recent Master's dissertation considered how the RM-ASCA+ method might be extended to zero-inflated count data by implementing negative binomial mixed-effect models (Haver, 2021). However, there is currently no established *R* package utilising GLMMs as the modelling framework within RM-ASCA+.

Returning to the model for all patient outcome profiles:

$$\mathbf{Y} = g^{-1}\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}\right)$$

The conditional distribution of $\mathbf{Y}$ given $\mathbf{u}$ then has the following mean and variance:

$$\mathbf{E}(\mathbf{Y} \mid \mathbf{u}) = g^{-1}\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}\right)$$

$$\text{Var}\left(\mathbf{Y} \mid \mathbf{u}\right) = \sigma_e^2\, \mathbf{G}$$

The marginal distribution of $\mathbf{Y}$ has the following mean and variance:

$$\mathbf{E}\left(\mathbf{Y} \mid \mathbf{X};\ \mathbf{u}\right) = g^{-1}\left(\mathbf{X}\boldsymbol{\beta}\right)$$

$$\text{Var}\left(\mathbf{Y} \mid \mathbf{X};\ \mathbf{u}\right) = \mathbf{Z}^T\ \left(\sigma_u^2\, \mathbf{R}\right)\mathbf{Z} + \left(\sigma_e^2\, \mathbf{G}\right)$$

Parameter estimation is likelihood-based. However, it is often impossible to find a closed form solution for the marginal likelihood - unless the mixed-effect model is linear without generalisation through the link function g (Pinheiro & Bates, 2006). This linear relationship is implemented in the RM-ASCA+ framework and *ALASCA* package in *R* (Jarmund et al., 2022).

In a generalised linear relationship, the likelihood function involves a multi-dimensional integral with respect to the random effects (Demidenko, 2013). This integral is typically approximated using quadrature methods, Monte Carlo methods, and related techniques (Demidenko, 2013).

Contrasts can be computed to describe and make inference on population-level differences in expected outcomes within or between groups described by exposure variables. These differences are averaged over the random effects and over the distributions of all fixed effects not considered in the contrast. The *R* package *emmeans* provides methods for computing contrasts of estimated marginal means for different levels of exposure variables in (G)LMMs (Lenth, 2024).

Both LMMs and GLMMs were investigated as modelling frameworks for outcome selection. The RM-ASCA+ framework, on which the dissertation's outcome selection technique is loosely based, implements LMMs to describe the relationships between immune outcomes and different levels of an exposure variables over time. However, this modelling approach is applied universally, even when the characteristics of the outcomes may not align with the assumptions underlying linear mixed-effect models.

Within the *ALASCA* package (Jarmund et al., 2022) where RM-ASCA+ is implemented in *R*, transformations can be applied to make immune outcomes more suitable for the LMM framework. However, there is no functionality for examining the distributions of random effects, a critical step in verifying model assumptions.

Through an application to immune outcomes from the dissertation data set, the validity of a LMM approach is briefly investigated and compared to a GLMM approach. It is hypothesised that the additional flexibility of GLMMs may be better suited to the characteristics of immunological data.

The key differences between these frameworks are summarised below:

- GLMMs may offer increased flexibility over LMMs for modelling immune outcomes.

    - GLMMs can specify parametric forms for the conditional distribution that accommodate skew, long-tailed immune outcomes.
    - LMMs assume a Normal conditional distribution, which cannot adequately account for these characteristics of immunological data.
- GLMMs may be better suited to immune outcomes with a high frequency of null observations than LMMs.
    - GLMMs can specify compound conditional distributions, allowing for greater mass at zero than a truly continuous distribution would afford.
    - LMMs may be unsuitable for immune outcomes with a high frequency of null observations as the conditional distribution of the outcome itself is required to be continuous.

4.4 Model Structure & Data

The comparisons of interest described in Chapter 3 inform the model structure. For an exposure of interest, each model compares how different exposure levels modify the expected value of an immune outcome. These comparisons of interest are estimated while holding sex differences constant.

A patient-specific random intercept ($\mathbf{u}$) is specified to capture the effects of random sampling, unobserved sources of patient-level differences in immune outcomes, and within-patient correlations between measurements (Lapham, 2020; Pinheiro & Bates, 2006).

Both LMMs and GLMMs were investigated as modelling frameworks for outcome selection. The fixed- and random-effect structure is identical for both frameworks, but there are differences in the specification of the relationship between $\mathbf{X}\boldsymbol{\beta}$ and the conditional distribution of $\mathbf{Y}$.

- Implementing an LMM specifies a linear relationship between $\mathbf{X}\boldsymbol{\beta}$ and the conditional distribution of $\mathbf{Y}$. The link function (g) is the identity function. The residuals follow a Normal distribution (Pinheiro & Bates, 2006). LMM coefficients have an additive interpretation.
- Implementing a GLMM specifies a generalised linear relationship between $\mathbf{X}\boldsymbol{\beta}$ and the conditional distribution of $\mathbf{Y}$ through the link function (g). The log link function is selected for this application. This gives the GLMM coefficients a multiplicative interpretation.
- A number of different parametric forms can be specified for the conditional distribution of the outcome in a GLMM. This specification should be informed by the characteristics of the observed data. Residual diagnostics can be performed on scaled residuals obtained using a simulation-based approach in the *DHARMa* package for *R* (Hartig, 2018) (see Appendix D).

For the m^{th} immune outcome in $y_{1A}...y_{33A}$ measured in Group A, (G)LMMs modelling the relationship between the EOI for patient i have the following form:

MVA85A priming

$$\mathbf{y}_{mi}^{(1)} \mid \mathbf{x}_i,\ \mathbf{u}_i \ = \ \mathrm{g}^{-1}\left[\left(\boldsymbol{\beta}_0^{(1)} + \mathbf{u}_i^{(1)}\right) + \boldsymbol{\beta}_1^{(1)}\mathrm{MVA85A}_i + \boldsymbol{\beta}_2^{(1)}\mathrm{Time3}_i + \boldsymbol{\beta}_3^{(1)}\mathrm{Time1}_i + \boldsymbol{\beta}_4^{(1)}\mathrm{MVA85A{:}Time3}_i + \boldsymbol{\beta}_5^{(1)}\mathrm{MVA85A{:}Time1}_i + \boldsymbol{\beta}_6^{(1)}\mathrm{Sex}_i + \boldsymbol{\beta}_7^{(1)}\mathrm{Sex{:}Time3}_i + \boldsymbol{\beta}_8^{(1)}\mathrm{Sex{:}Time1}_i + \mathbf{e}_i^{(1)}\right] \quad (4.1)$$

This model is fit to the following data set prepared in Section 5 of Chapter 3.

Table 4.1: Structure of the data set prepared for modelling MVA85A priming as the *EOI* in Group A.

ID	**Time**	**Sex**	**MVA85A**	$y_{1A}\cdots\ y_{33A}$
$i_{1A}...i_{61A}$	1, 2, 3	M, F	MVA85A, Control	...

Maternal Mtb exposure

$$\mathbf{y}_{mi}^{(2)} \mid \mathbf{x}_i,\ \mathbf{u}_i \ = \ \mathrm{g}^{-1}\left[\left(\boldsymbol{\beta}_0^{(2)} + \mathbf{u}_i^{(2)}\right) + \boldsymbol{\beta}_1^{(2)}\mathrm{QFT}_i + \boldsymbol{\beta}_2^{(2)}\mathrm{Time3}_i + \boldsymbol{\beta}_3^{(2)}\mathrm{Time1}_i + \boldsymbol{\beta}_4^{(2)}\mathrm{QFT{:}Time3}_i + \boldsymbol{\beta}_5^{(2)}\mathrm{QFT{:}Time1}_i + \boldsymbol{\beta}_6^{(2)}\mathrm{Sex}_i + \boldsymbol{\beta}_7^{(2)}\mathrm{Sex{:}Time3}_i + \boldsymbol{\beta}_8^{(2)}\mathrm{Sex{:}Time1}_i + \mathbf{e}_i^{(2)}\right] \quad (4.2)$$

This model is fit to the following data set prepared in Section 5 of Chapter 3.

Table 4.2: Structure of the data set prepared for modelling maternal *Mtb.* exposure as the *EOI* in Group A.

ID	**Time**	**Sex**	**QFT**	$y_{1A}\cdots\ y_{33A}$
$i_{1A}...i_{61A}$	1, 2, 3	M, F	Positive, Negative	...

Feeding practices & cotrimoxazole treatment

$$\mathbf{y}_{mi}^{(3)} \mid \mathbf{x}_i,\ \mathbf{u}_i \ = \ \mathrm{g}^{-1}\left[\left(\boldsymbol{\beta}_0^{(3)} + \mathbf{u}_i^{(3)}\right) + \boldsymbol{\beta}_1^{(3)}\mathrm{FA3}_i + \boldsymbol{\beta}_2^{(3)}\mathrm{FA2}_i + \boldsymbol{\beta}_3^{(3)}\mathrm{FA1}_i + \boldsymbol{\beta}_4^{(3)}\mathrm{Time3}_i + \boldsymbol{\beta}_5^{(3)}\mathrm{Visit1}_i + \boldsymbol{\beta}_6^{(3)}\mathrm{FA3{:}Time3}_i + \boldsymbol{\beta}_7^{(3)}\mathrm{FA2{:}Time3}_i + \boldsymbol{\beta}_8^{(3)}\mathrm{FA1{:}Time3}_i + \boldsymbol{\beta}_9^{(3)}\mathrm{FA3{:}Time1}_i + \boldsymbol{\beta}_{10}^{(3)}\mathrm{FA2{:}Time1}_i + \boldsymbol{\beta}_{11}^{(3)}\mathrm{FA1{:}Time1}_i + \boldsymbol{\beta}_{12}^{(3)}\mathrm{Sex}_i + \boldsymbol{\beta}_{13}^{(3)}\mathrm{Sex{:}Time3}_i + \boldsymbol{\beta}_{14}^{(3)}\mathrm{Sex{:}Time1}_i + \mathbf{e}_i^{(3)}\right] \quad (4.3)$$

This model is fit to the following data set prepared in Section 5 of Chapter 3.

Table 4.3: Structure of the data set prepared for modelling feeding practices and cotrimoxazole treatment as the *EOI* in Group A, with formula feeding and cotrimoxazole treatment as the reference group.

ID	**Time**	**Sex**	**FA**	$y_{1A}\cdots\ y_{33A}$
$i_{1A}...i_{61A}$	1, 2, 3	M, F	0, 1, 2, 3	...

$$\mathbf{y}_{mi}^{(4)} \mid \mathbf{x}_i,\ \mathbf{u}_i \ = \ \mathrm{g}^{-1}\Big[\Big(\boldsymbol{\beta}_0^{(4)} + \mathbf{u}_i^{(4)}\Big) + \boldsymbol{\beta}_1^{(4)}\mathrm{FA2}_i + \boldsymbol{\beta}_2^{(4)}\mathrm{FA1}_i + \boldsymbol{\beta}_3^{(4)}\mathrm{FA0}_i + \boldsymbol{\beta}_4^{(4)}\mathrm{Time3}_i + \boldsymbol{\beta}_5^{(4)}\mathrm{Visit1}_i + \boldsymbol{\beta}_6^{(4)}\mathrm{FA2{:}Time3}_i + \boldsymbol{\beta}_7^{(4)}\mathrm{FA1{:}Time3}_i + \boldsymbol{\beta}_8^{(4)}\mathrm{FA0{:}Time3}_i + \boldsymbol{\beta}_9^{(4)}\mathrm{FA2{:}Time1}_i + \boldsymbol{\beta}_{10}^{(4)}\mathrm{FA1{:}Time1}_i + \boldsymbol{\beta}_{11}^{(4)}\mathrm{FA0{:}Time1}_i + \boldsymbol{\beta}_{12}^{(4)}\mathrm{Sex}_i + \boldsymbol{\beta}_{13}^{(4)}\mathrm{Sex{:}Time3}_i + \boldsymbol{\beta}_{14}^{(4)}\mathrm{Sex{:}Time1}_i + \mathbf{e}_i^{(4)}\Big] \quad (4.4)$$

This model is fit to the following data set prepared in Section 5 of Chapter 3.

Table 4.4: Structure of the data set prepared for modelling feeding practices and cotrimoxazole treatment as the *EOI* in Group A, with breastfeeding and cotrimoxazole treatment as the reference group.

ID	**Time**	**Sex**	**FA**	y_{1A}... y_{33A}
i_{1A}...i_{61A}	1, 2, 3	M, F	0, 1, 2, 3	...

For the LMMs, the link function g is the identity link. The GLMMs make use of a log link function.

If the conditional distribution and link function are appropriately specified, the modelling framework will meet a number of distributional assumptions for the random effects and residuals. These assumptions are summarised in the tables below.

Table 4.5: Summary of distributional assumptions and model evaluation for LMMs.

Assumption	Method of Evaluation
$\mathbf{u}^{(n)} \sim \mathrm{N}(\mathbf{0},\ \sigma^2_{(n)}\mathbf{R})$	Normal Q-Q plots
$\mathbf{e}^{(n)} \sim \mathrm{N}(\mathbf{0},\ 1)$	Normal Q-Q plots

Table 4.6: Summary of distributional assumptions and model evaluation for GLMMs.

Assumption	Method of Evaluation
$\mathbf{u}^{(n)} \sim \mathrm{N}(\mathbf{0},\ \sigma^2_{(n)}\mathbf{R})$	Normal Q-Q plots
$s(\mathbf{e}^{(n)}) \sim \mathrm{N}(\mathbf{0},\ 1)$	Normal Q-Q plots
$s(\mathbf{e}^{(n)})$ quantiles $\approx$ rank-transformed model predictions	Residual vs. Predicted plots
$s(\mathbf{e}^{(n)})$ dispersion $\approx$ specified dispersion	A non-parametric dispersion test

Note that $s(\mathbf{e}^{(n)})$ represents scaled residuals obtained using a simulation-based approach in the *DHARMa* package (Hartig, 2018) (see Appendix D).

The appropriateness of the model specification was also visually inspected by comparing the predicted conditional mean longitudinal trajectory with the observed mean trajectory.

- The shapes of the predicted and observed trajectories should correspond.
- The predicted and observed mean trajectories should have similar ranges.

If all model evaluations were satisfactory, model coefficients (and their standard errors) corresponding to the comparisons of interest were extracted from the fitted

models. These model coefficients estimate the expected change in an immune outcome for different levels of *MVA85A*, *QFT*, and *FA*. Estimates hold *Sex* constant and are conditional on the random effects.

Instead of specifying different reference levels for *FA* to reflect the comparison of interest and extracting model coefficient estimates, pairwise contrasts could be calculated to estimate differences in *EOI* levels that are not directly estimated by the regression model.

However, this changes the nature of the input to outcome selection:

- A longitudinal regression model coefficient describes the relationship between levels of the *EOI*, *Time* or *EOI:Time* and the outcome, conditional on the random effects and holding additional covariates constant.
- A contrast describes the expected difference in the outcome for different levels of the *EOI*, *Time* or *EOI:Time*, averaged over the random effects and covariates not considered in the contrast.

4.5 Comparison of Modelling Frameworks

The linear mixed-effect modelling (LMM) framework and generalised linear mixed-effect modelling (GLMM) framework were applied to three immune outcomes measured in Group A: %CD4+ T cells expressing IL-22, %Bulk CD4+ T cells with R7+RA+ phenotype, and %GrK$^+$ Ki67$^+$ NK. Comparing their goodness of fit provides a rough assessment of the validity of the LMM approach, applied universally by the *ALASCA* package's implementation of RM-ASCA+, for longitudinal modelling of immunological data.

The histogram of %CD4$^+$ T cells expressing IL-22 is right-skewed, long-tailed, and has a fairly high frequency of zero observations.

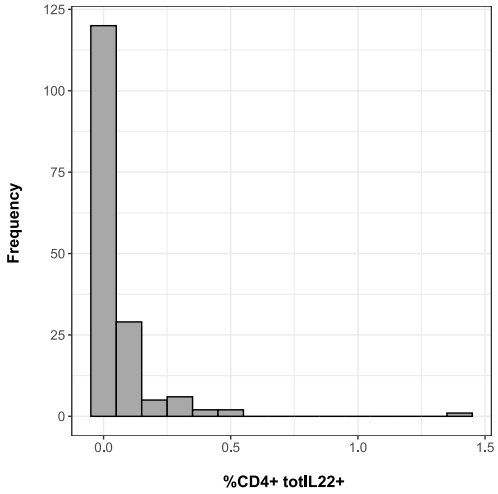


Figure 4.1: Histogram of %CD4$^+$ T cells expressing IL-22.

The histogram of %Bulk CD4+ T cells with R7+RA+ phenotype is left skewed. After applying a transformation, the histogram is right skewed. However, this transformed histogram has lighter tails and lower frequencies of zeroes than those observed in the histogram of %CD4$^+$ T cells expressing IL-22.

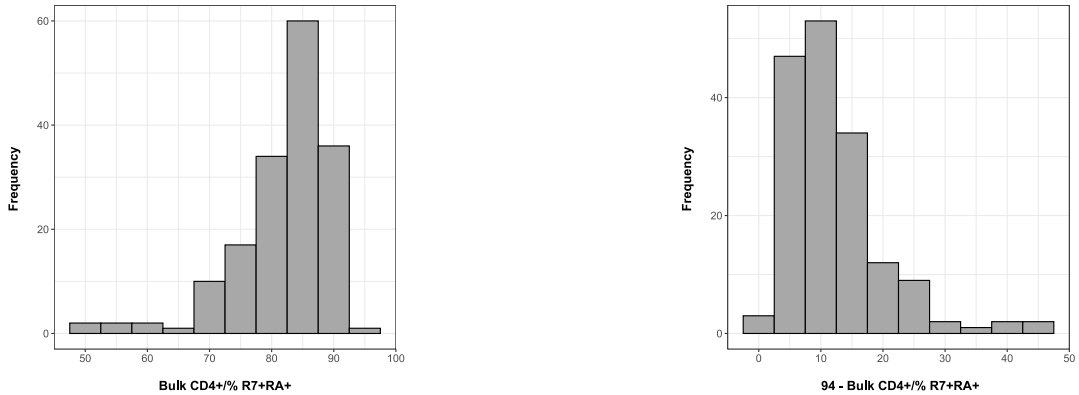


Figure 4.2: Histograms of %Bulk CD4+ T cells with R7+RA+ phenotype over its original domain (*left*) and corrected for left skewness (*right*) prior to modelling.

The histogram of %GrK$^+$ Ki67$^+$ NK cells resembles a folded normal distribution. If its observations were reflected ($(x,\ y) \to (-x,\ y)$), the histogram would resemble a Normal distribution. Compared to immune outcomes %CD4$^+$ T cells expressing IL-22 and %Bulk CD4+ T cells with R7$^+$RA$^+$ phenotype, this histogram is reasonably symmetric with light tails.

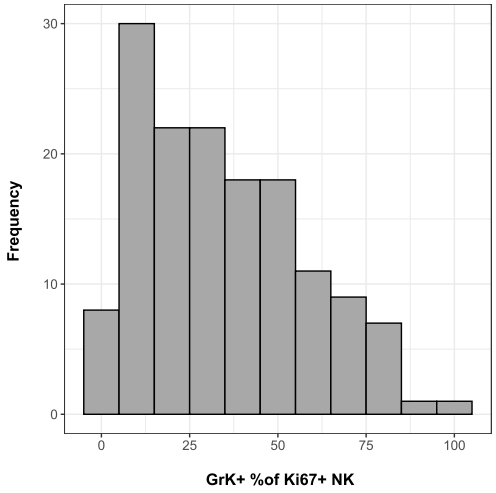


Figure 4.3: Histogram of %GrK$^+$ Ki67$^+$ NK cells.

The dissertation data set contains a number of immunological outcomes with similar distributions (see Appendix C).

The complete modelling procedure applied to %CD4+ T cells expressing IL-22, %Bulk CD4+ T cells with R7+RA+ phenotype, and %GrK$^+$ Ki67$^+$ NK within the LMM and GLMM frameworks is contained in the code repository. The GLMM modelling framework appears to be the most suitable for modelling immune response profiles within the dissertation's outcome selection techniques.

4.5.1 Random Effects

The distributional assumptions for the random intercepts in models of CD4+ T cells expressing IL-22 are best met for the GLMM framework.

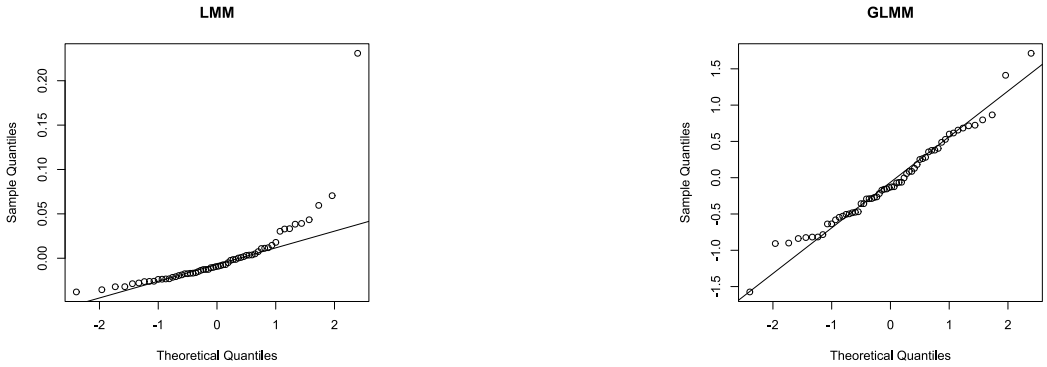


Figure 4.4: Normal Q-Q plots for the random intercept in the LMM and GLMM of CD4+ T cells expressing IL-22 considering MVA85A priming.

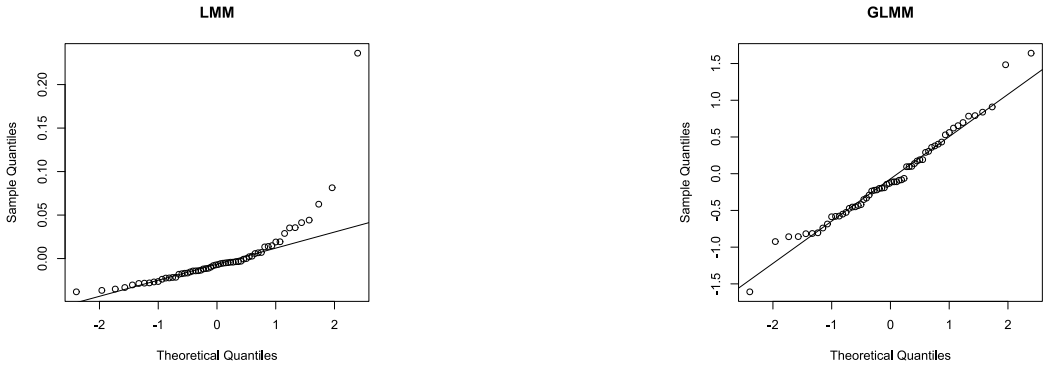


Figure 4.5: Normal Q-Q plots for the random intercept in the LMM and GLMM models of CD4+ T cells expressing IL-22 considering maternal QFT.

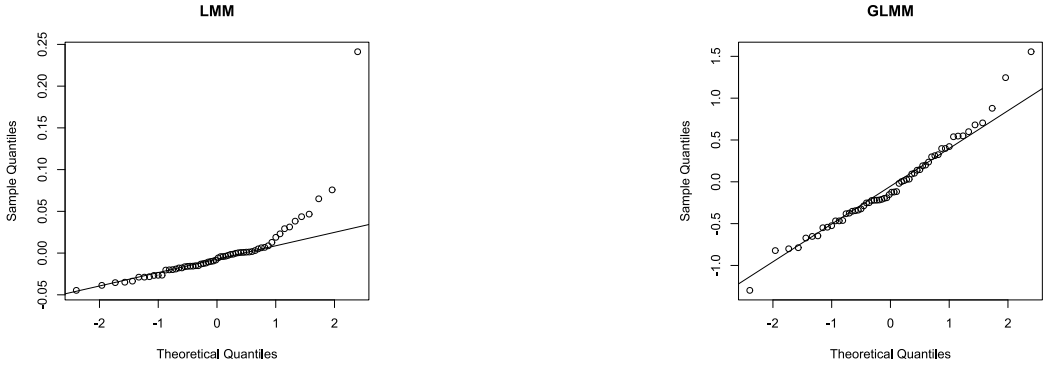


Figure 4.6: Normal Q-Q plots for the random intercept in the LMM and GLMM of CD4+ T cells expressing IL-22 considering combinations of feeding practice and cotrimoxazole treatment.

The LMM random intercept distributions deviate from normality at the upper tails. A very large random intercept is estimated for a patient within the LMM, but not the GLMM. The largest GLMM estimates still lie fairly close to the Normal Q-Q line.

Considering the models of %Bulk CD4+ T cells with R7+RA+ phenotype, the distributional assumptions for the random intercepts are also best met for the GLMM framework.

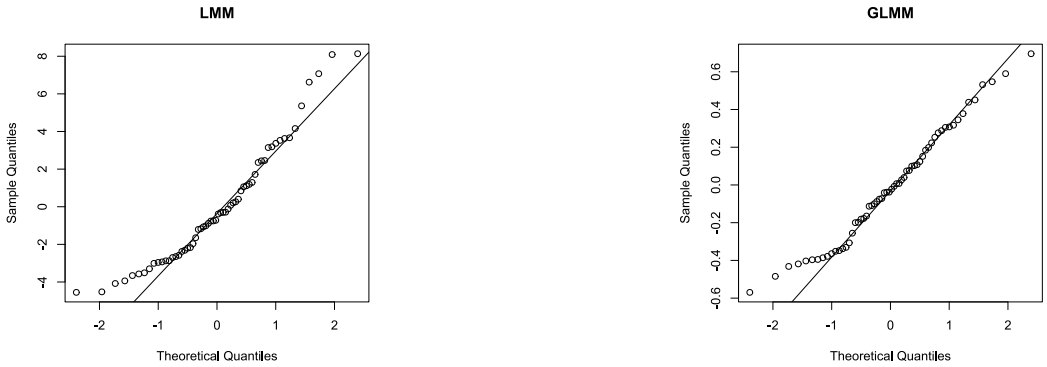


Figure 4.7: Normal Q-Q plots for the random intercept in the LMM and GLMM of %Bulk CD4+ T cells with R7+RA+ phenotype, considering MVA85A priming.

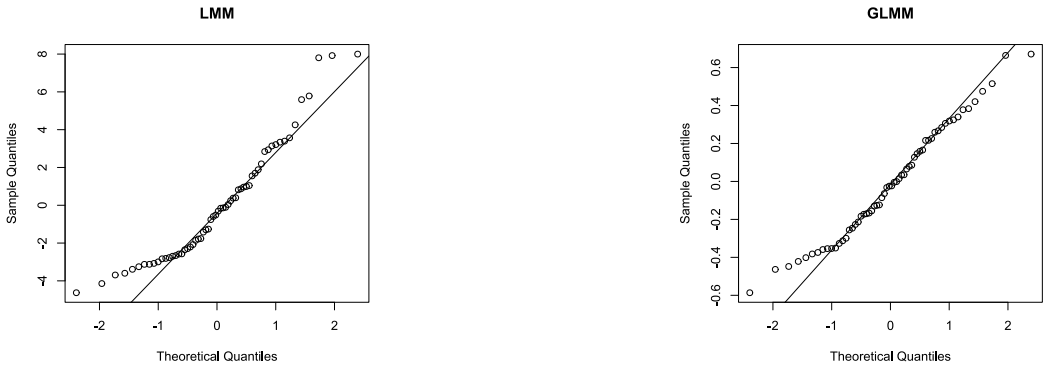


Figure 4.8: Normal Q-Q plots for the random intercept in the LMM and GLMM of %Bulk CD4+ T cells with R7+RA+ phenotype, considering maternal QFT.

The LMM random intercept distributions deviate from normality at both the lower and upper tails. The GLMM estimates lie closer to the Normal Q-Q line.

The distributional assumptions for the random intercepts are met for both LMM and GLMM models of %GrK$^+$ Ki67$^+$ NK.

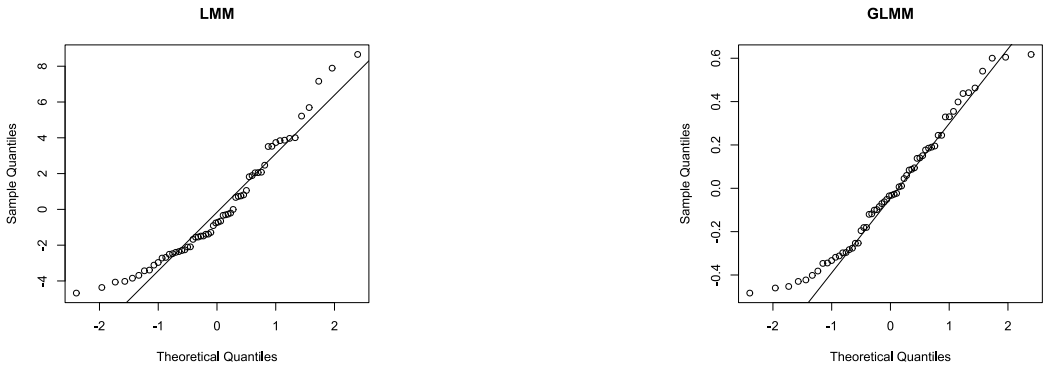


Figure 4.9: Normal Q-Q plots for the random intercept in the LMM and GLMM of %Bulk CD4+ T cells with R7+RA+ phenotype, considering combinations of feeding practice and cotrimoxazole treatment.

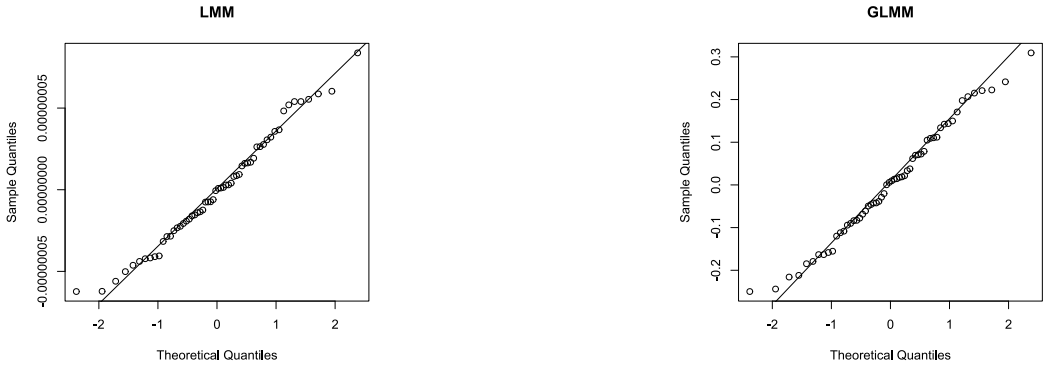


Figure 4.10: Normal Q-Q plots for the random intercept in the LMM and GLMM of %GrK$^+$ Ki67$^+$ NK considering MVA85A priming.

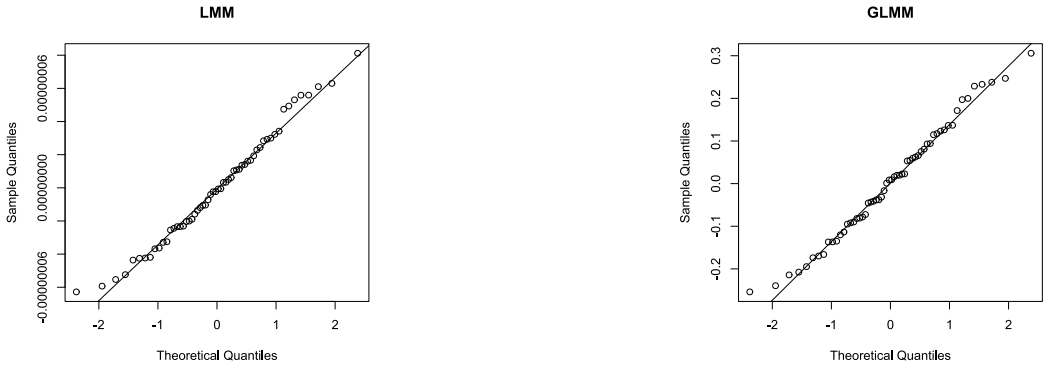


Figure 4.11: Normal Q-Q plots for the random intercept in the LMM and GLMM models of %GrK$^+$ Ki67$^+$ NK considering maternal QFT.

The LMM random intercept estimates are very small, and there seems to be little variability in the random intercepts estimated for different patients. This is not the case for the GLMM or LQMM models and could indicate a singular model fit. The random intercept distribution for the LQMM model considering combinations of feeding and cotrimoxazole treatment deviates from normality at the upper and lower tails.

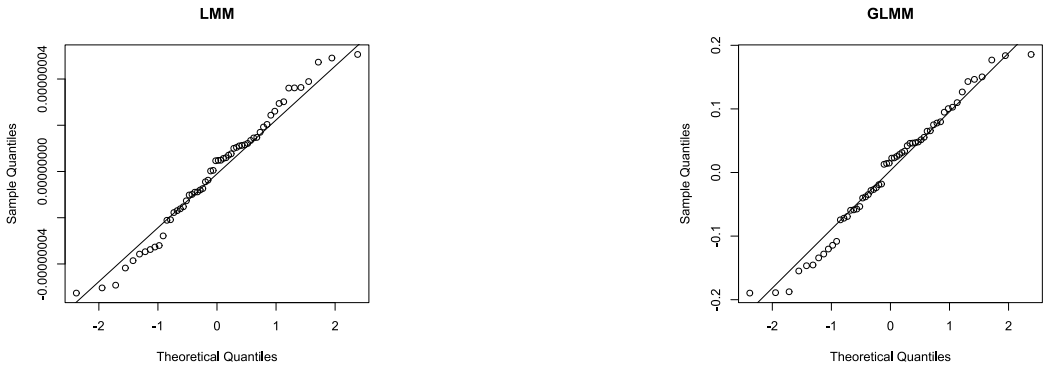


Figure 4.12: Normal Q-Q plots for the random intercept in the LMM and GLMM models of %GrK$^+$ Ki67$^+$ NK considering combinations of feeding practice and cotrimoxazole treatment.

4.5.2 Residuals

The distributional assumptions for the model residuals are well met within the GLMM framework for CD4+ T cells expressing IL-22, %Bulk CD4+ T cells with R7+RA+ phenotype, and %GrK$^+$ Ki67$^+$ NK cells. Scaled residuals were obtained using a simulation-based approach in the *DHARMa* package (Hartig, 2018) (see Appendix D).

MVA85A Priming

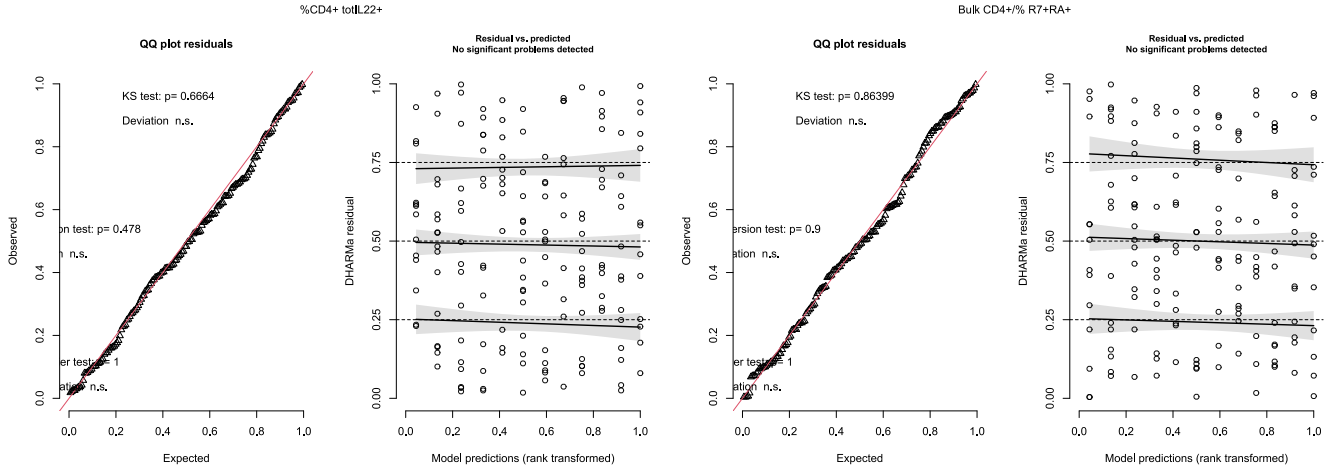


Figure 4.13: Residual diagnostics in the GLMM models of %CD4+ T cells expressing IL-22 and %Bulk CD4+ T cells with R7+RA+ phenotype considering MVA85A priming.

The Normal Q-Q plots show that the scaled residuals are approximately normally distributed for CD4+ T cells expressing IL-22, %Bulk CD4+ T cells with R7+RA+ phenotype, and %GrK$^+$ Ki67$^+$ NK cells in the GLMM models describing MVA85A priming. There is no evidence of under- or overdispersion of scaled residuals, and the scaled residual quantiles approximate the rank-transformed model predictions well.

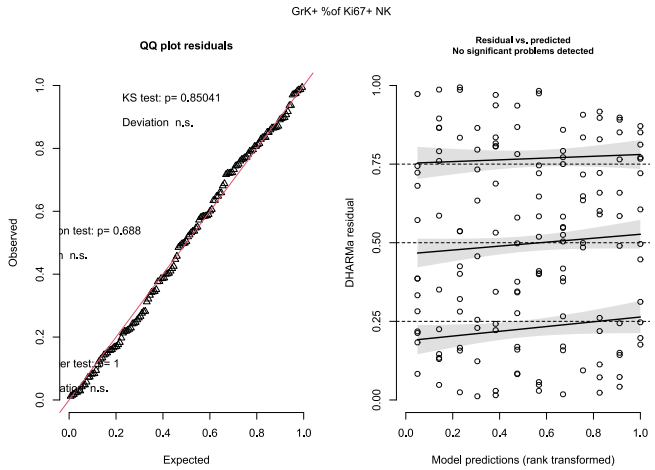


Figure 4.14: Residual diagnostics in the GLMM model of %GrK$^+$ Ki67$^+$ NK cells considering MVA85A priming.

Maternal QFT

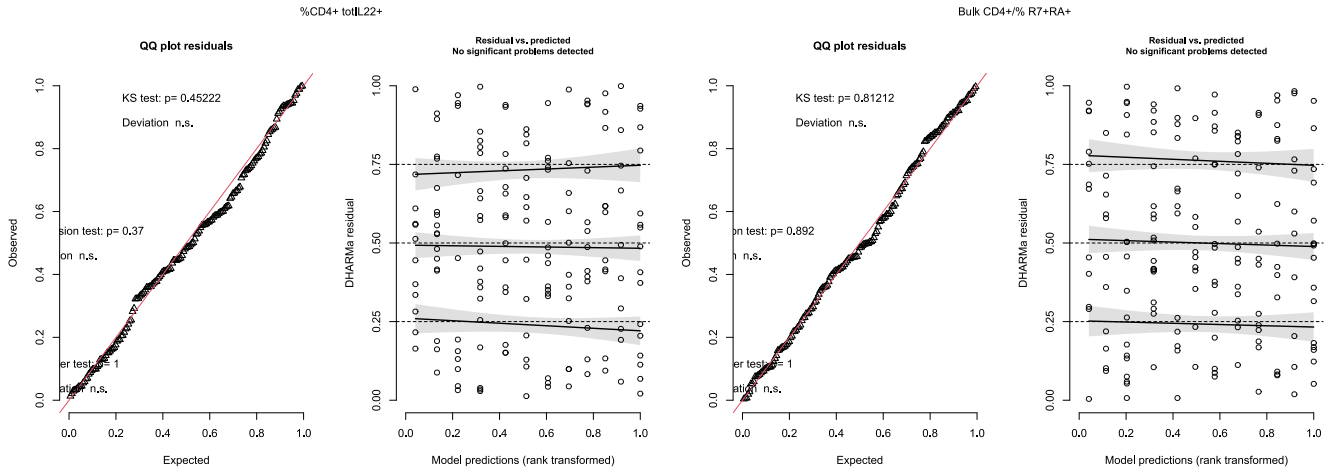


Figure 4.15: Residual diagnostics in the GLMM models considering maternal Mtb exposure, indicated by a positive QFT, for %CD4+ T cells expressing IL-22 and %Bulk CD4+ T cells with R7+RA+ phenotype.

The Normal Q-Q plots show that the scaled residuals are approximately normally distributed for %CD4+ T cells expressing IL-22, %Bulk CD4+ T cells with R7+RA+ phenotype, and %GrK$^+$ Ki67$^+$ NK cells in the GLMM models describing maternal QFT. There is no evidence of under- or overdispersion of scaled residuals, and the scaled residual quantiles approximate the rank-transformed model predictions well.

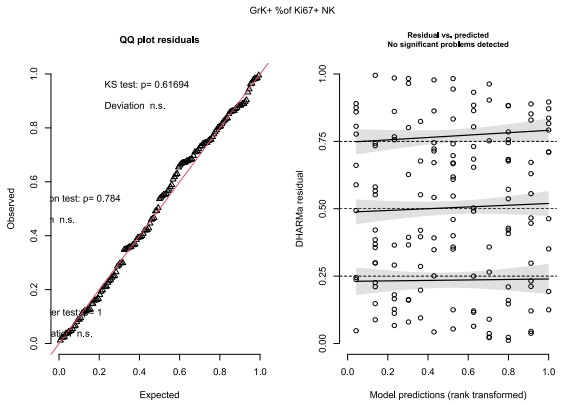


Figure 4.16: Residual diagnostics in the GLMM model considering maternal Mtb exposure, indicated by a positive QFT, for %GrK$^+$ Ki67$^+$ NK cells.

Feeding & Cotrimoxazole

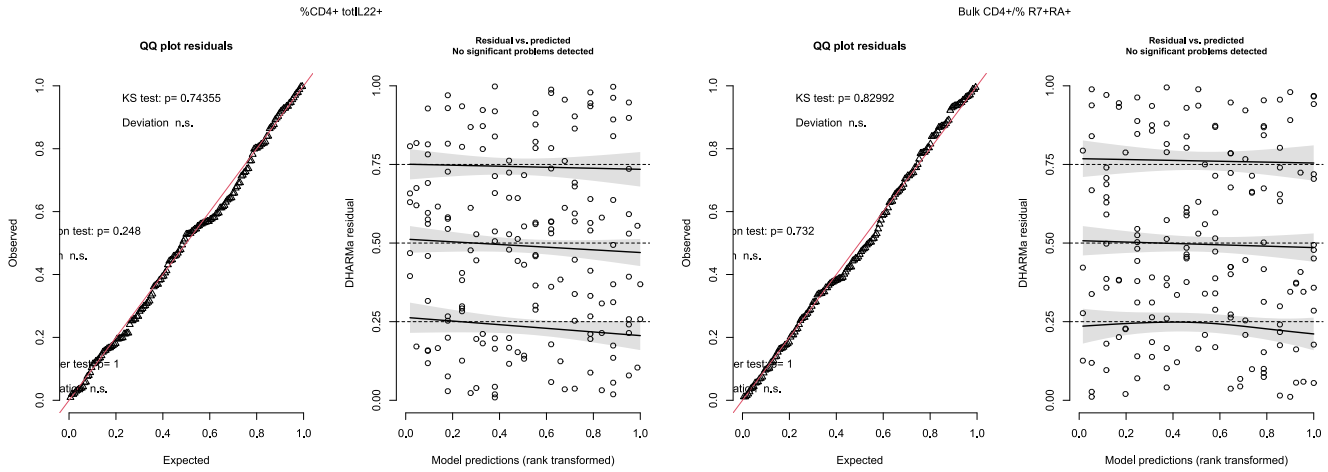


Figure 4.17: Residual diagnostics in the GLMM models considering combinations of feeding practice and cotrimoxazole treatment for %CD4+ T cells expressing IL-22 and %Bulk CD4+ T cells with R7+RA+ phenotype.

The Normal Q-Q plots show that the scaled residuals are approximately normally distributed for %CD4+ T cells expressing IL-22, %Bulk CD4+ T cells with R7+RA+ phenotype, and %GrK$^+$ Ki67$^+$ NK cells in the GLMM models describing combinations of feeding practice and cotrimoxazole treatment. There is no evidence of under- or overdispersion of scaled residuals, and the scaled residual quantiles approximate the rank-transformed model predictions well.

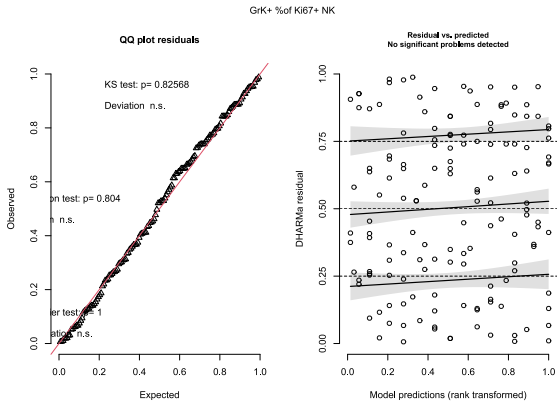


Figure 4.18: Residual diagnostics in the GLMM model considering combinations of feeding practice and cotrimoxazole treatment for %GrK$^+$ Ki67$^+$ NK cells.

The predicted conditional subgroup trajectories and empirical mean trajectories correspond fairly well within the GLMM framework. The GLMMs of %CD4+ T cells expressing IL-22 considering MVA85A priming and maternal QFT underpredict the empirical trajectories. However, the MVA85A priming, maternal QFT and feeding-cotrimoxazole combination GLMMs correspond closely with the empirical mean trajectories for %Bulk CD4+ T cells with R7+RA+ phenotype and %GrK$^+$ Ki67$^+$ NK cells in all subgroups.

%CD4+ T cells expressing IL-22

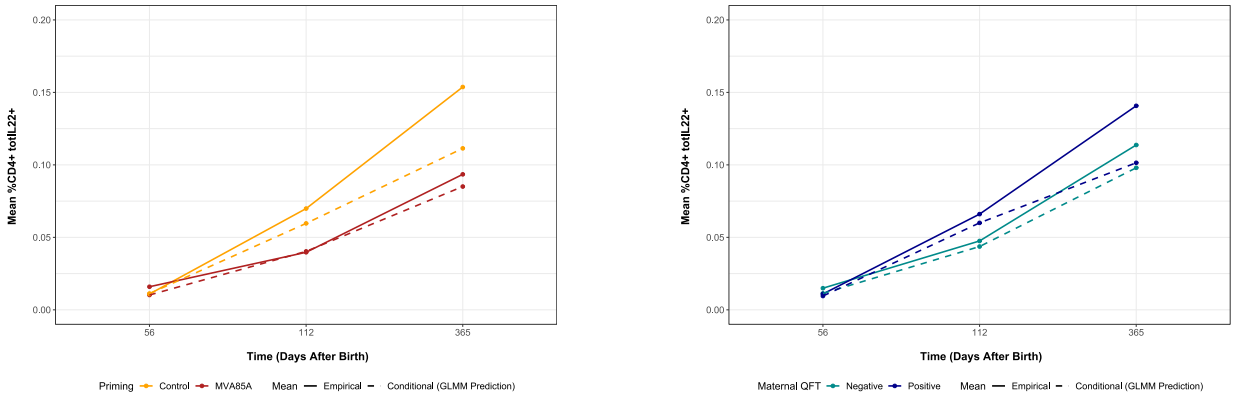


Figure 4.19: Comparison of empirical group mean trajectories (*solid lines*) to conditional group mean trajectories estimated by the GLMM models (*dotted lines*) of %CD4+ T cells expressing IL-22 for MVA85A priming and maternal QFT.

The MVA85A priming GLMM underpredicts the empirical mean trajectory for the control group, but not the group receiving MVA85A priming. The maternal QFT GLMM underpredicts the empirical mean trajectory of CD4+ T cells expressing IL-22 for both groups, but the underprediction is most striking for infants born to mothers with a positive maternal QFT.

There is a slight underprediction of the empirical mean trajectory for formula-fed infants treated with cotrimoxazole.

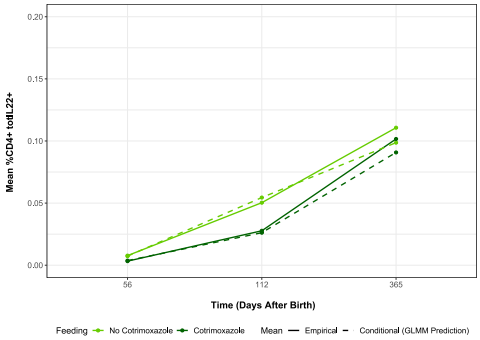


Figure 4.20: Comparison of empirical group mean trajectories (*solid lines*) to conditional group mean trajectories estimated by the GLMM models (*dotted lines*) of %CD4+ T cells expressing IL-22 for feeding-cotrimoxazole treatment.

%Bulk CD4+ T cells with R7+RA+ phenotype

The MVA85A priming and maternal QFT GLMMs correspond closely to the empirical mean trajectory for %Bulk CD4+ T cells with R7+RA+ phenotype in all subgroups, as well as the feeding-cotrimoxazole combination GLMM.

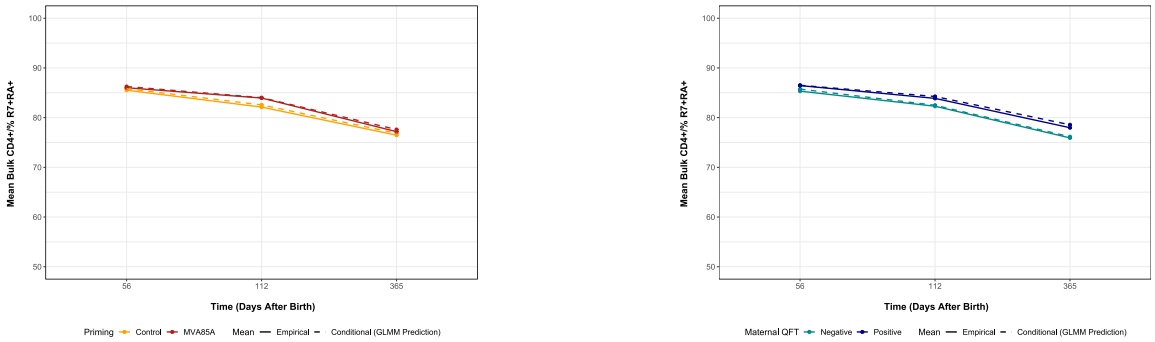


Figure 4.21: Comparison of empirical group mean trajectories (*solid lines*) to conditional group mean trajectories estimated by the GLMM models (*dotted lines*) of Bulk CD4+ T cells with R7+RA+ phenotype for MVA85A priming and maternal QFT.

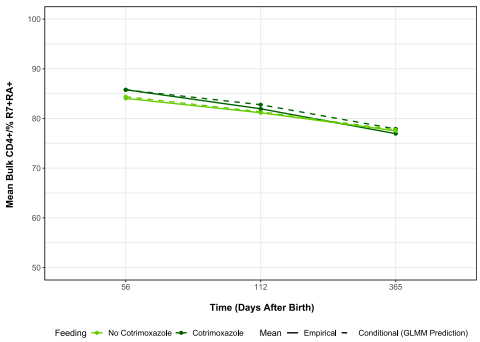


Figure 4.22: Comparison of empirical group mean trajectories (*solid lines*) to conditional group mean trajectories estimated by the GLMM models (*dotted lines*) of %Bulk CD4+ T cells with R7+RA+ phenotype.

%GrK$^+$ Ki67$^+$ NK cells

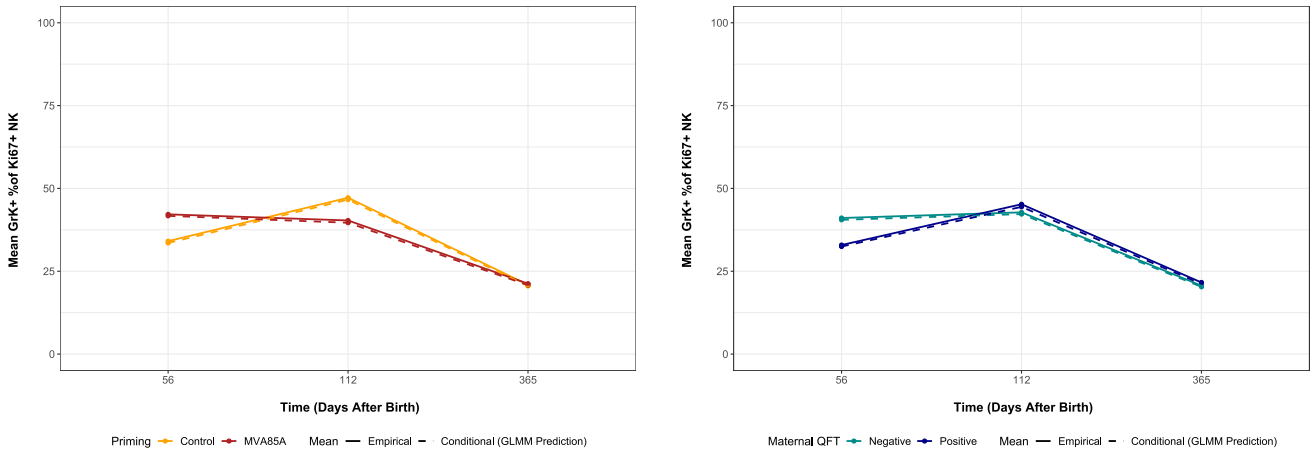


Figure 4.23: Comparison of empirical group mean trajectories (*solid lines*) to conditional group mean trajectories estimated by the GLMM models (*dotted lines*) of %GrK$^+$ Ki67$^+$ NK cells for MVA85A priming and maternal QFT.

Similarly, the MVA85A priming and maternal QFT GLMMs correspond closely to the empirical mean trajectory for %GrK$^+$ Ki67$^+$ NK cells in all subgroups. The feeding-cotrimoxazole combination GLMM also corresponds closely to the empirical mean trajectory for %GrK$^+$ Ki67$^+$ NK cells in all subgroups.

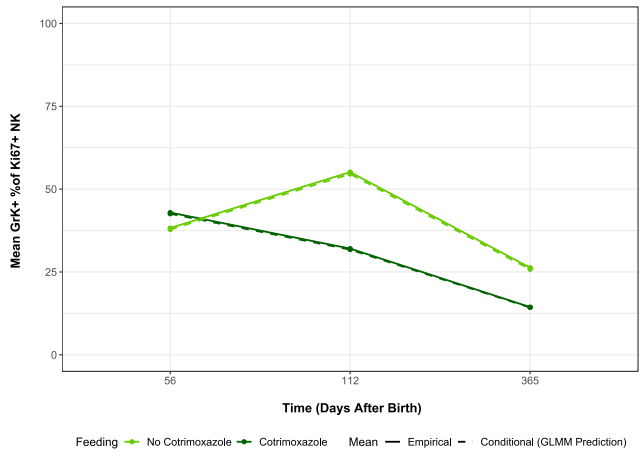


Figure 4.24: Comparison of empirical group mean trajectories (*solid lines*) to conditional group mean trajectories estimated by the GLMM models (*dotted lines*) of %GrK$^+$ Ki67$^+$ NK cells for feeding-cotrimoxazole combination.

%CD4+ T cells expressing IL-22, %Bulk CD4+ T cells with R7+RA+ phenotype, and %GrK$^+$ Ki67$^+$ NK cells exemplify the distributional characteristics of immunological data. The random effect and residual distributional assumptions do not deviate enormously from assumptions for the LMMs (as implemented in RM-ASCA+). However, it is clear that the GLMM modelling framework is more suitable for these immune outcomes. Hence, the GLMM framework was selected as the longitudinal modelling framework for the dissertation's outcome selection technique.

A summary of the GLMM conditional distributions is included in Appendix E. Parametric forms were selected by plotting histograms to identify features of the observed outcome distributions (see Appendix C). A compound Poisson-Gamma conditional distribution was specified for most immune outcomes (see Appendix E). Model evaluation for Group A immune outcomes is included as Appendix F. GLMMs did not fully meet assumptions for a few immune outcomes. These cases are discussed in detail in Appendix G.

4.6 Extracting Model Coefficient Estimates

As the GLMM framework's assumptions are met and model predictions correspond well with the observed data, model coefficient estimates are extracted as input data sets for dimension reduction. Every input data set contains model coefficients estimating comparisons of interest for a single exposure variable. These estimates are made for every immune outcome, comparing how different levels of the exposure modify their expected values over time. As all GLMMs make use of the log link function, the extracted model coefficient estimates are comparable.

This application produces three data sets as inputs to dimension reduction. Each data set corresponds to a single exposure variable (*MVA85A* or *QFT* or *FA*).

The structure of these data sets is summarised below:

1. EOI: *MVA85A*

Group A Input Data Set: $MVA85A{:}Time1^{(1)}$, $MVA85A^{(2)}$, $MVA85A{:}Time3^{(3)}$

Number of standardised model coefficients (q): 3
Number of immune outcomes (m): 33

2. EOI: *QFT*

Group A Input Data Set:$QFT{:}Time1^{(1)}$, $QFT^{(2)}$, $QFT{:}Time3^{(3)}$

Number of standardised model coefficients (q): 3
Number of immune outcomes (m): 33

3. EOI: *FA*

Group A Input Data Set

- $FA3{:}Time1^{(1)}$
- $FA0{:}Time1^{(2)}$
- $FA3^{(3)}$
- $FA0^{(4)}$
- $FA3{:}Time3^{(5)}$
- $FA0{:}Time3^{(6)}$
- $FA2{:}Time1^{(7)}$
- $FA2^{(8)}$)
- $FA2{:}Time3^{(9)}$
- $FA1{:}Time1^{(10)}$
- $FA1^{(11)}$
- $FA1{:}Time3^{(12)}$

Number of standardised model coefficients (q): 12
Number of immune outcomes (m): 33

Chapter 5

Dimension Reduction Techniques

This chapter discusses the *Dimension Reduction* step in the outcome selection workflow (Figure 1.2) in more detail. Dimension reduction techniques identify a lower-dimensional representation of a input data set's most meaningful properties. In this application, the input data sets consist of coefficient estimates extracted from longitudinal models (see Chapter 4). These model coefficient estimates compare how different exposure variable levels modify the expected values of immune outcomes over time. After preprocessing these input data sets for dimension reduction, lower-dimensional representations of subgroup differences are obtained for each exposure variable. From this lower-dimensional representation, subset(s) of immune outcomes with the most evidence of subgroup differences over time (in the complete set of immune outcomes or in clusters of immune outcomes) can be selected for further analysis.

Two dimension reduction techniques are considered: first, PCA (as implemented in RM-ASCA+), and second, agglomerative hierarchical cluster analysis (HCA) followed by PCA. A brief theoretical overview is presented for each technique, and the workflows described in Chapter 1 (Figures 1.3 and 1.4) are described in more detail. To demonstrate the techniques, dimension reduction is applied to obtain a lower-dimensional representation of the relationships between the exposure of interest *FA* and the Group A immune outcomes in the dissertation data set.

5.1 Principal Component Analysis

Principal Component Analysis (PCA) is a dimension reduction technique that maps an input data set to a new coordinate system. This coordinate system is defined by orthogonal axes called principal components (PCs), which are obtained from the eigendecomposition of a covariance or correlation matrix of the input data:

$$C = V\Lambda V^\top$$

where:

- C is the covariance or correlation matrix ($q \times q$),
- V is the matrix of eigenvectors (*loadings*, $q \times q$),
- Λ is the diagonal matrix of eigenvalues ($q \times q$).

If C is the covariance matrix, the q inputs are centred but unscaled. If C is the correlation matrix, the q inputs are standardised to have a mean of 0 and a variance of 1.

As C is symmetric, its eigendecomposition identifies orthogonal eigenvectors V. The eigenvectors define the directions of the PCs and describe each input variable's contribution to the orthogonal axes.

- In this application, the input variables to the PCA are the q standardised coefficients extracted from longitudinal models.

- These coefficients represent comparisons of interest for MVA85A priming, maternal QFT, and combinations of feeding practices and cotrimoxazole treatment.
- Model coefficients are extracted and scaled by their standard errors to create the q standardised coefficients. This ensures that the inputs have similar scales.

Every observation contained in the input data ($\mathbf{X}$) is projected onto the PCs (defined by the eigenvectors) to obtain its **scores** on each PC:

$$X' = X_{\text{centred}} \times V$$

where X' $(m \times q)$ are the scores representing the coordinates of each observation in the new PC space. V contains the loadings.

- In this application, every immune outcome considered in the analysis $(y_1 ... y_m)$ corresponds to an observation with a score on the PCs.

PCA is performed with the *princomp* function in the *R stats* package on correlation matrices of the input data sets described in the previous chapter. Using the correlation matrix ensures that all variables contribute equally because they are standardised to unit variance.

- Although the coefficient estimates are scaled by their standard errors (and, therefore, measured on similar scales), working with the correlation matrix is necessary when PCA is performed on clusters identified by HCA. This is discussed in more detail in the next section of this chapter.
- Using the correlation matrix when applying PCA only, although not strictly necessary, ensures consistency between the two approaches.

Each PC explains a portion of the total variance in the input data, proportional to its eigenvalue in Λ. The first $q^* \leq q$ PCs that explain a sufficient cumulative proportion of this variance may be selected as latent variables to replace the original input variables in a further analysis. However, for the purpose of outcome selection, the latent variables themselves are not of interest. Instead, the focus is on how much the respective immune outcomes contribute to the first q^* PCs. These contributions are reflected in their scores.

All observations are assumed to contribute uniformly to each PC. For example, Group A considers 33 unique immune outcomes. Every immune outcome is therefore expected to contribute $(100 \times \frac{1}{33})\%$ of the variance explained by each PC. For a given PC, this contribution is then adjusted by the ratio of the immune outcome's squared coordinate position to the squared standard deviation of the PC.

As PCs are ranked by the proportion of input variance that they explain, immune outcomes with contributions exceeding this uniform expectation on one or more of the first $q^* \leq q$ PCs explain the majority of the variance in the input data. Therefore, these immune outcomes can be selected as a subset of the complete set of immune outcomes for further analysis of the comparisons of interest.

The basic workflow for applying this dimension reduction technique for outcome selection is included below. For each of these p exposures of interest, a single immune outcome subset containing $m^* < m$ outcomes (where m refers to the complete set of m outcomes) is selected for further analysis.

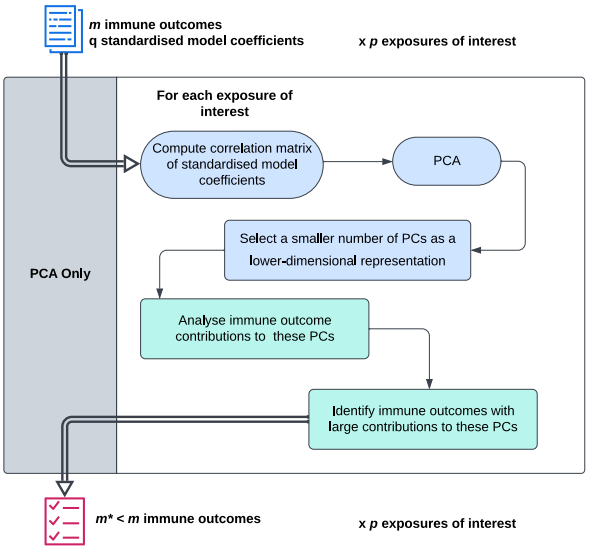


Figure 5.1: An overview of the outcome selection workflow when applying PCA only as a dimension reduction technique.

PCA can only be performed when the number of immune outcomes (m) in the input data set equals or exceeds the number of standardised model coefficients (q). Referring back to the input data sets described in Chapter 4, all input data sets contain at least as many immune outcomes as standardised model coefficients ($m \geq q$).

5.2 Hierarchical Cluster Analysis Then PCA

Clusters are groups of proximal objects identified with respect to some chosen distance metric and grouping rule (or linkage method). Objects that are 'close' to one another are 'similar' to one another, and they are assumed to represent the intrinsic structure of the input data.

- In this application, immune outcomes are cluster objects. In Group A, 33 immune outcomes are clustered with respect to the chosen distance metric and linkage.
- Immune outcomes are clustered by their similarity with respect to the standardised model coefficients estimating the comparisons of interest for MVA85A priming, maternal QFT, combinations of feeding practices, and cotrimoxazole treatment.
- This similarity is quantified using the Pearson correlation distance. The Pearson correlation coefficient (r) quantifies the linear relationship between two observations, ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation).

To convert Pearson correlation into a distance, the distance between two observations i and j is defined as:

$$d_{ij} = 1 - r_{ij}$$

where r_{ij} is the Pearson correlation coefficient between observations i and j. Higher correlations correspond to smaller distances, aligning similarity with proximity in clustering.

- Average linkage is selected as the grouping rule for cluster formation. The distance between two clusters is defined as the average of all pairwise distances between observations in the two clusters. For two clusters C_1 and C_2, the distance is given by:

$$d(C_1, C_2) = \frac{1}{|C_1| \cdot |C_2|} \sum_{i \in C_1} \sum_{j \in C_2} d_{ij}$$

where $|C_1|$ and $|C_2|$ are the sizes of the clusters and d_{ij} is the distance between observations i and j. In this application, d_{ij} is $1 - r_{ij}$ where r_{ij} is the Pearson correlation coefficient between observations i and j.

Hierarchical Cluster Analysis (HCA) considers clusters to have a hierarchical structure. In other words, there may be smaller groups of objects nested within larger groups of objects. Different clusterings of objects can be extracted depending on the amount of consideration given to the granular, local relationships versus broadly applicable global relationships between objects in the input data set.

- The number of clusters extracted from this hierarchical structure does not need to be specified before performing the clustering.
- As there is limited prior knowledge of which immune outcomes are 'similar' with respect to the exposures of interest, HCA is better suited to this application than other clustering methods (for example, k-means clustering) that require the number of clusters to be specified upfront.

This dissertation considers agglomerative hierarchical clustering, a bottom-up clustering method that begins with each observation as its own cluster and iteratively merges clusters until all objects are contained in a single cluster. The algorithm is as follows:

1. Calculate the distance (in this application, the Pearson correlation distance $(1 - r)$) for all pairs of objects.
2. Treat each object as its own initial cluster.
3. Iteratively merge the two closest clusters based on the linkage criterion (in this application, average linkage).
4. Recalculate the distances (in this application, the Pearson correlation distance $(1 - r)$) between the new cluster and all remaining clusters.
5. Continue merging clusters until all immune outcomes form a single cluster.

The distance at which clusters merge is known as the link height.

- In this application, when two clusters merge, the link height describes the average pairwise Pearson distance between the immune outcomes in the two clusters.
- Objects with high correlations (i.e., $r_{ij} = 0.95$ or $d_{ij} = 0.05$) are locally similar. Retaining more clusters formed at lower link heights emphasises these relationships.
- Objects with moderate correlations (i.e., $r_{ij} = 0.5$ or $d_{ij} = 0.5$) may share a broad and consistent relationship across the input data set. Extracting clusters formed at higher link heights emphasises these relationships and produces a smaller number of clusters.

The HCA process and nested cluster structure is visualised using a dendogram. When considering Pearson correlation distances between clusters, clusters that merge near the maximum link height ($d_{ij} \approx 2$) are maximally dissimilar based on this distance metric. As an example from the dissertation data set, consider the dendogram produced by the hierarchical clustering of the standardised model coefficients corresponding to subgroup differences in Group A immune outcomes based on maternal QFT.

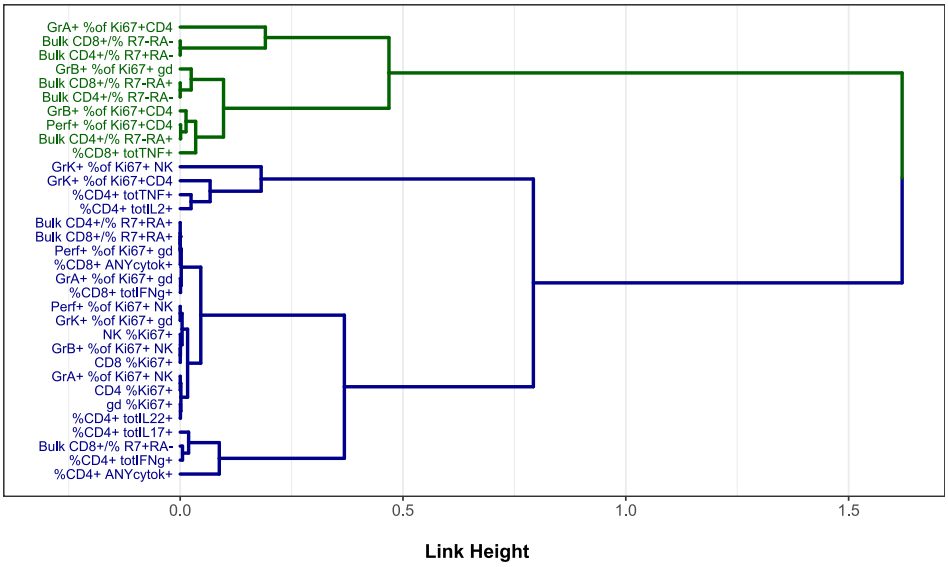


Figure 5.2: Dendogram of hierarchical clustering of standardised model coefficients estimating comparisons of interest for maternal QFT in Group A.

With application to the dissertation data set, the immune outcomes contained in Cluster 1 (*green*) and Cluster 2 (*blue*) that merge at $d_{12} \approx 2$ display inverse patterns in their relationships with the standardised model coefficients (*QFT:Time1*, *QFT:Time2* and *QFT:Time3*). A single cluster of immune outcomes containing both the Cluster 1 and Cluster 2 outcomes would likely not represent any intrinsic structure of the input data. Hence, clusters are typically extracted at lower link heights to ensure that they represent interpretable groups of similar objects.

If a minimum of two clusters ($k \geq 2$) are extracted from the hierarchy of groupings at an appropriate link height, every cluster will contain fewer immune outcomes than the complete set ($m_k^* < m$). The intrinsic structure of the input data is then represented by two or more groups ($k \geq 2$) of 'similar' objects. No dimension reduction has been

performed; the m input objects have merely been grouped into k clusters of different sizes ($\sum_{i=1}^{k} m_i^* = m$). However, these k clusters can then become units of analysis.

The workflow for dimension reduction by HCA followed by PCA is summarised in the diagram below.

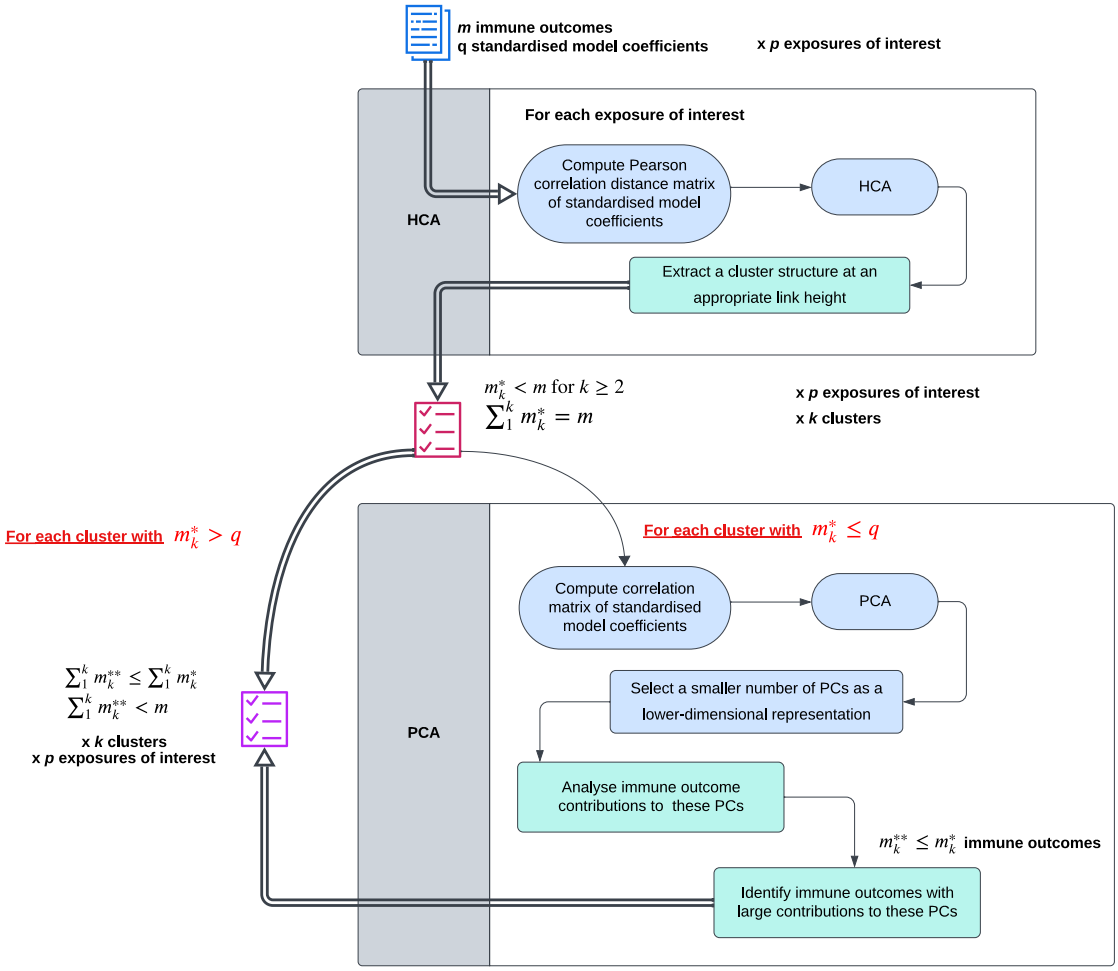


Figure 5.3: An overview of the outcome selection workflow when applying HCA followed by PCA as a dimension reduction technique.

For an exposure of interest, HCA organises m immune outcomes into k nested groups that are similar with respect to the q standardised model coefficient estimates. These coefficients estimate subgroup differences over time in all immune outcomes $(y_1 ... y_m)$ for different levels of the exposure variable. After identifying these k clusters of immune outcomes, it is convenient to select k single objects as their representatives. These representative objects then become units of analysis.

However, this application considers a wide range of immune outcomes with diverse immunological functions. Clusters may contain immune outcomes that share weak or partial similarities in their standardised model coefficients. If further analysis was restricted to a single representative object for each cluster, the results would probably not describe the behaviour shared by all cluster members.

Instead of selecting a single representative object, dimension reduction (for example, PCA) can be performed on standardised model coefficients at the cluster level. The number of immune outcomes in a cluster may equal or exceed the number of standardised model coefficients ($m_k^* \geq q$). In these cases, a PCA is performed on

the correlation matrix of the model coefficient estimates for the immune outcomes contained within the cluster. A number of PCs ($q* \leq q$) are then selected to provide a lower-dimensional representation of the cluster variance in standardised model coefficient estimates (where possible). By analysing immune outcome contributions to these PCs, the immune outcomes that contribute the most to the cluster variance of model coefficient estimates can be selected as a subset for further analysis.

Alternatively, there may be fewer immune outcomes than coefficients ($m_k^* < q$). In these cases, a PCA cannot be applied. Without performing dimension reduction, all immune outcomes contained in the cluster are selected for further analysis.

Unlike the PCA-only approach, multiple immune outcome subsets can be identified per exposure of interest. For each of the p exposures of interest, k immune outcome subsets (one for each cluster) are selected for further analysis. Each subset contains $m_k^{**} \leq m_k^*$ cluster outcomes. Depending on the size of the clusters (m_k^{**}), this approach may investigate fewer statistical hypotheses simultaneously, compared to dimension reduction by PCA only.

5.3 Application

The two dimension reduction techniques, PCA only and HCA followed by PCA, were then applied to the input data sets for Group A. To demonstrate the techniques, dimension reduction is presented for the exposure of interest *FA*. Appendix I contains the application of the techniques to *MVA85A* and *QFT*.

5.3.1 Data Preprocessing

Before performing dimension reduction, these model coefficient estimates are preprocessed to account for reflected immune outcomes and the precision of the model coefficient estimates.

First, left skewness should be restored for immune outcomes reflected during data preparation $[(x,\ y) \rightarrow (c - x,\ y)]$. To restore left skewness, the model coefficient estimates are multiplied by a factor of -1 $[\beta \rightarrow -\beta]$. This reflection was applied to model coefficient estimates for the following immune outcomes, all measured in Group A:

- *Bulk %CD4$^+$ R7$^+$ RA$^+$*
- *Bulk %CD8$^+$ R7$^+$ RA$^+$*
- *%GrA$^+$ Ki67$^+$ NK*
- *%GrB$^+$ Ki67$^+$ NK*
- *%Perf$^+$ Ki67$^+$ NK*

Second, model coefficient estimates should be divided by their standard errors. This ensures that the precision of the coefficient estimates is taken into account when comparing how different exposure levels modify the expected values of the immune outcomes. For the left-skewed outcomes listed above, the standard errors were estimated for their reflections $[(x,\ y) \rightarrow (c - x,\ y)]$. However, unlike transformations that stretch or shrink measurements about an axis, a reflection about an axis maintains

the observed distances between measurements. Reflecting the model coefficient estimates to restore left skewness $[\beta \to -\beta]$, and then scaling these coefficient estimates by standard errors estimated on the reflected scale, does not introduce bias.

PCA only and HCA followed by PCA are then applied to the input data sets described in Chapter 3. For each exposure of interest, these techniques identify the subset(s) of immune outcomes with the most evidence for subgroup differences. These immune outcomes are then selected for further analysis.

5.3.2 PCA Only

Dimension reduction by PCA only is demonstrated with application to Group A immune outcomes and the exposure of interest *FA* A PCA was performed on the correlation matrix of the following standardised model coefficients for the Group A outcomes, adjusted for the average sex effect:

- $FA3{:}Time1^{(1)}$
- $FA0{:}Time1^{(2)}$
- $FA3^{(3)}$
- $FA0^{(4)}$
- $FA3{:}Time3^{(5)}$
- $FA0{:}Time3^{(6)}$
- $FA2{:}Time1^{(7)}$
- $FA2^{(8)}$
- $FA2{:}Time3^{(9)}$
- $FA1{:}Time1^{(10)}$
- $FA1^{(11)}$
- $FA1{:}Time3^{(12)}$

As 33 immune outcomes are measured in Group A, PCA is performed on a 33×12 input data set. This produces 12 principal components (PCs). The scree plot of the variance in the standardised model coefficients explained by each component is included below.

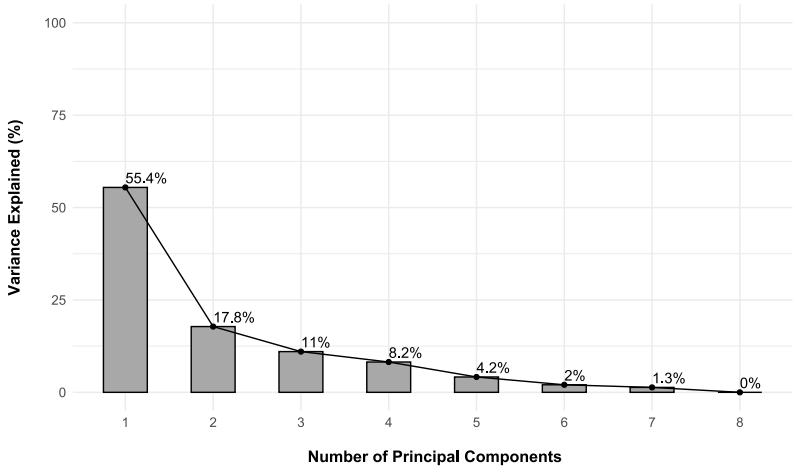


Figure 5.4: Scree plot of the percentage variance in standardised model coefficients for *FA* explained by each principal component across Group A immune outcomes.

Contributions to the first $q^* < q$ total PCs explaining at least 60% of the variance in the standardised model coefficients were investigated.

- The first two PCs explain 73.2% of the variance in the standardised model coefficients.
- In other words, for the 33 immune outcomes in Group A, PC1 and PC2 capture most of the variance in the longitudinal effects of combinations of feeding practices and cotrimoxazole treatment. PC1 and PC2 can act as a lower-dimensional representation of these effects in Group A.

The contributions of standardised model coefficients and immune outcomes were then investigated for PC1 and PC2. The scree plot shows which standardised model coefficients make large contributions to the variance explained by PC1 and PC2.

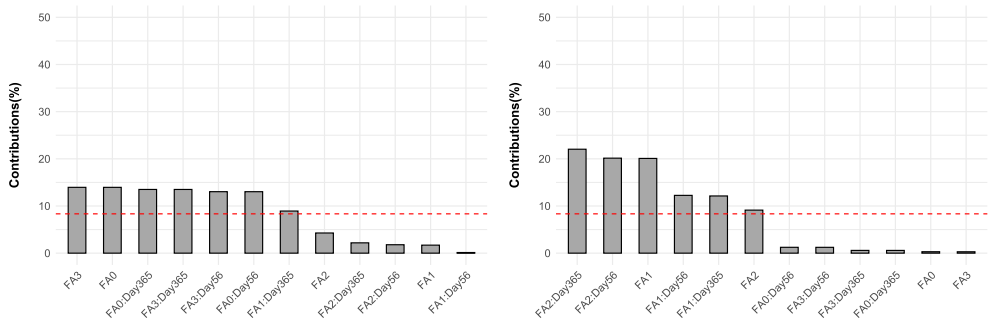


Figure 5.5: Scree plots of the percentage variance contributed by standardised model coefficients for *FA* to PC1 (*left*) and PC2 (*right*) across Group A immune outcomes.

PC1 primarily describes the variance in the standardised coefficients for the effect of breastfeeding ($FA3$, $FA3 : Time1$ and $FA3 : Time3$) and formula feeding ($FA0$, $FA0 : Time1$ and $FA0 : Time3$) in infants treated with cotrimoxazole. This indicates that longitudinal immune outcome profiles may vary considerably for different feeding practices.

PC2 primarily describes variance in the standardised coefficients for no cotrimoxazole treatment in formula-fed infants ($FA1$, $FA1 : Time1$ and $FA1 : Time3$) and breastfed infants ($FA2$, $FA2 : Time1$ and $FA2 : Time3$). This indicates that longitudinal immune outcome profiles can vary considerably with cotrimoxazole treatment within different feeding practices.

The scree plot below identifies six immune outcomes that make larger contributions to the variance explained by PC1 than would be expected under the assumption of uniform contributions.

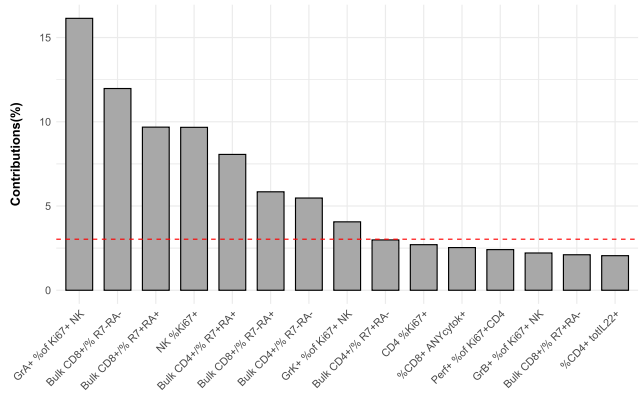


Figure 5.6: Scree plot of the percentage variance in *FA* contributed by the Group A immune outcomes to PC1.

Similarly, the scree plot below identifies ten immune outcomes that make large contributions to the variance explained by PC2.

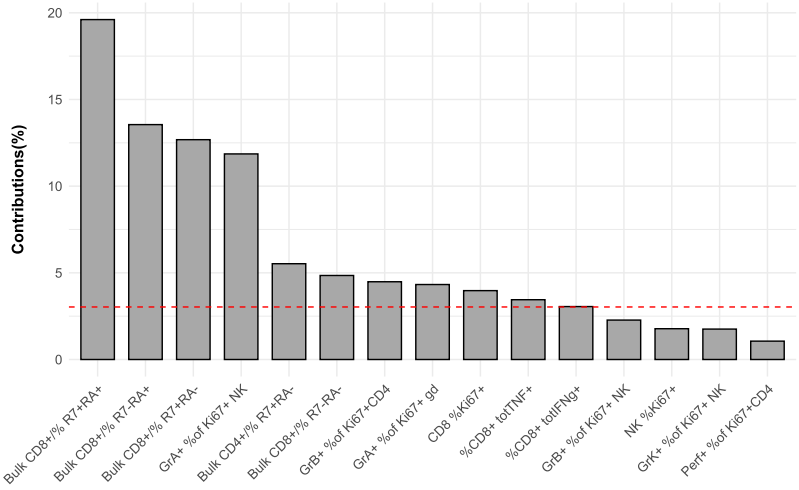


Figure 5.7: Scree plot of the percentage variance in *FA* contributed by the Group A immune outcomes to PC2.

Both PC1 and PC2 are required for a lower-dimensional representation that sufficiently captures the variance in the longitudinal effects of combinations of feeding practices and cotrimoxazole treatment in Group A.

Hence, the 15 unique immune outcomes making large contributions to PC1 and PC2 are combined to form an immune outcome subset for further analysis:

- $\%GrA^{+}\ Ki67^{+}\ NK^{(1)}$
- $Bulk\ \%CD8^{+}\ R7^{-}\ RA^{-(2)}$
- $Bulk\ \%CD8^{+}\ R7^{+}\ RA^{+(3)}$
- $\%NK^{+}\ Ki67^{+(4)}$
- $Bulk\ \%CD4^{+}\ R7^{+}\ RA^{+(5)}$
- $Bulk\ \%CD8^{+}\ R7^{-}\ RA^{+}\ ^{(6)}$
- $Bulk\ \%CD4^{+}\ R7^{-}\ RA^{-(7)}$
- $\%GrK^{+}\ Ki67^{+}\ NK^{(8)}$

- $Bulk\ \%CD8^{+}\ R7^{+}\ RA^{-}$(9)
- $Bulk\ \%CD4^{+}\ R7^{+}\ RA^{-}$(10)
- $\%GrB^{+}\ Ki67^{+}\ CD4^{+}$(11)
- $\%GrA^{+}\ Ki67^{+}\ gd$(12)
- $\%CD8^{+}\ Ki67^{+}$(13)
- $\%CD8^{+}\ totTNF^{+}$(14)
- $\%CD8^{+}\ totIFNg^{+}$(15)

Instead of analysing the longitudinal effects of combinations of feeding practices and cotrimoxazole treatment for all 33 outcomes, the analysis can proceed with a subset of 15 immune outcomes describing the majority of the variance in the effects.

5.3.3 HCA Followed By PCA

Dimension reduction by HCA followed by PCA is then demonstrated by application to Group A immune outcomes for exposure of interest *FA*. A hierarchical cluster analysis with average linkage was performed on the Pearson correlation distance matrix of the following standardised model coefficients for the Group A outcomes, adjusted for the average sex effect:

- $FA3{:}Time1$(1)
- $FA0{:}Time1$(2)
- $FA3$(3)
- $FA0$(4)
- $FA3{:}Time3$(5)
- $FA0{:}Time3$(6)
- $FA2{:}Time1$(7)
- $FA2$(8)
- $FA2{:}Time3$(9)
- $FA1{:}Time1$(10)
- $FA1$(11)
- $FA1{:}Time3$(12)

As 33 immune outcomes are measured in Group A, HCA is performed on a 33×12 input data set. As the input data set contains 12 standardised model coefficients, it is important to extract a cluster structure in which at least one cluster contains twelve or more immune outcomes. If all clusters contain less than 12 immune outcomes, all 33 immune outcomes will be analysed further.

A cluster structure ($k = 7$) is extracted at a link height of $d_{ij} = 0.15$, corresponding to an average pairwise Pearson correlation of at least 0.85 between all immune outcomes contained in a cluster. The dendogram is included below.

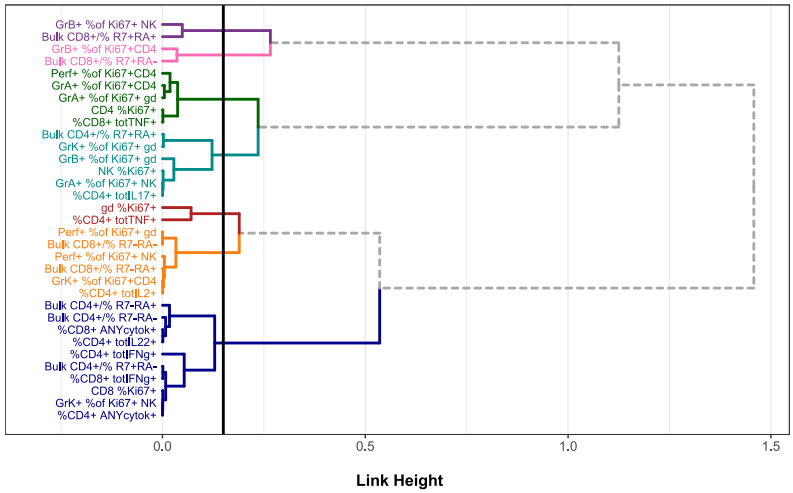


Figure 5.8: Dendogram of hierarchical clustering of standardised model coefficients describing comparisons of interest for *FA* in Group A, extracted at a link height of 0.15.

At this link height, all clusters contain fewer immune outcomes than the number of standardised model coefficients ($n_k^* < q$ for all $k \in 1, ..., 7$).

Extracting a second cluster structure ($k = 5$) at a link height of $d_{ij} = 0.25$ creates larger clusters. These clusters still contain highly similar immune outcomes, corresponding to an average pairwise Pearson correlation of at least 0.75 between all immune outcomes contained in a cluster.

The dendogram is included below.

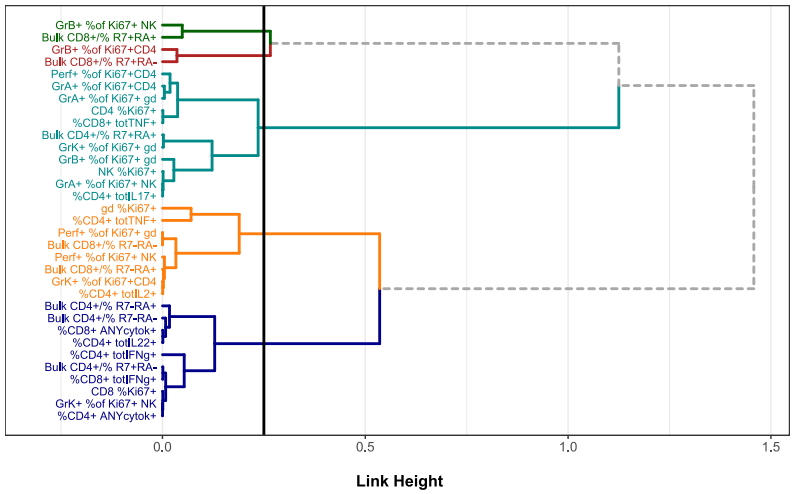


Figure 5.9: Dendogram of hierarchical clustering of standardised model coefficients describing comparisons of interest for *FA* in Group A, extracted at a link height of 0.25.

All clusters still contain fewer immune outcomes than the number of standardised model coefficients ($n_k^* < q$ for all $k \in 1, ..., 5$).

Extracting a third cluster structure ($k = 3$) at a link height of $d_{ij} = 0.55$ creates larger clusters. These clusters contain very broadly similar immune outcomes, corresponding to an average pairwise Pearson correlation of at least 0.45 between all immune outcomes contained in a cluster. The dendogram is included below.

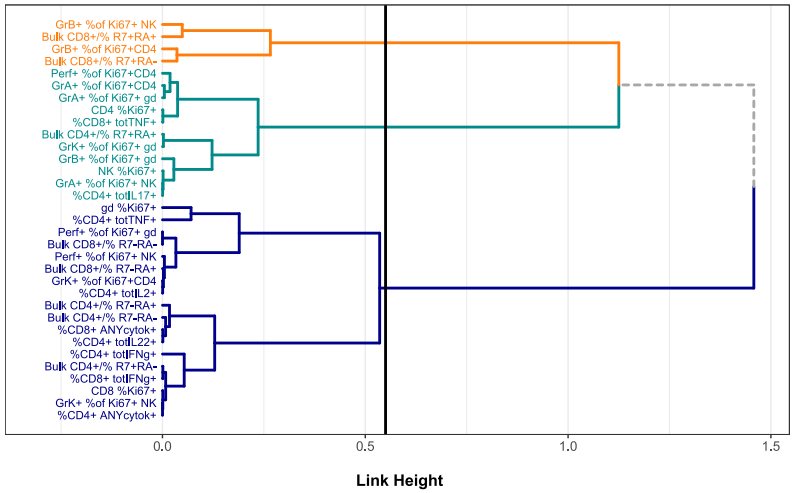


Figure 5.10: Dendogram of hierarchical clustering of standardised model coefficients describing comparisons of interest for *FA* in Group A, extracted at a link height of 0.55.

At this link height, two of the three clusters contain fewer immune outcomes than the number of standardised model coefficients ($n_1^* = 4 < q$ (*orange*); $n_2^* = 11 < q$ (*turquoise*)). The immune outcomes contained in each of these clusters are then selected for further analysis of subgroup differences described by levels of *FA*.

A PCA can be performed on the standardised model coefficients for the members of Cluster 3 (*blue*). As 18 immune outcomes are measured in this cluster, PCA is performed on a 18×12 input data set. The scree plot of the variance in standardised model coefficients explained by each of the 12 PCs is included below.

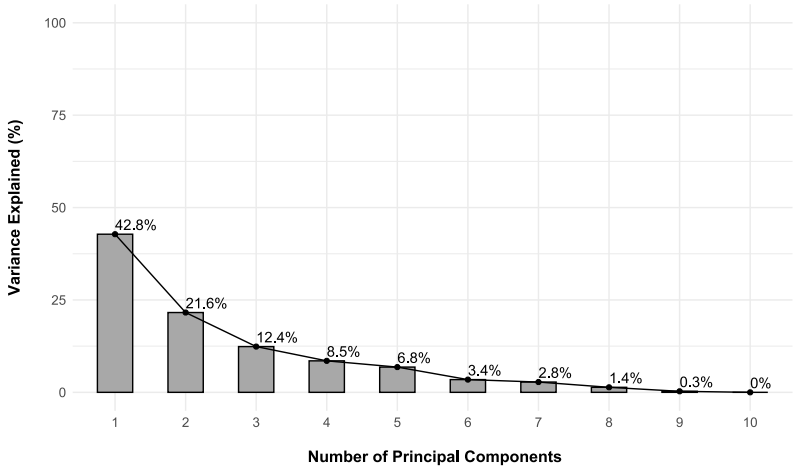


Figure 5.11: Scree plot of the percentage variance in standardised model coefficients for *FA* explained by each principal component across Group A immune outcomes.

The first two PCs explain 64.4% of the variance in the standardised model coefficients for the cluster outcomes. In other words, for the 18 immune outcomes in this cluster, PC1 and PC2 capture most of the cluster variance in the longitudinal effects of combinations of feeding practices and cotrimoxazole treatment. PC1 and PC2 can act as a lower-dimensional representation of these effects in this cluster.

Contributions to the first two PCs were investigated to identify a subset of the cluster

outcomes for further analysis. The scree plots below show which standardised model coefficients make large contributions to the variance explained by PC1 and PC2.

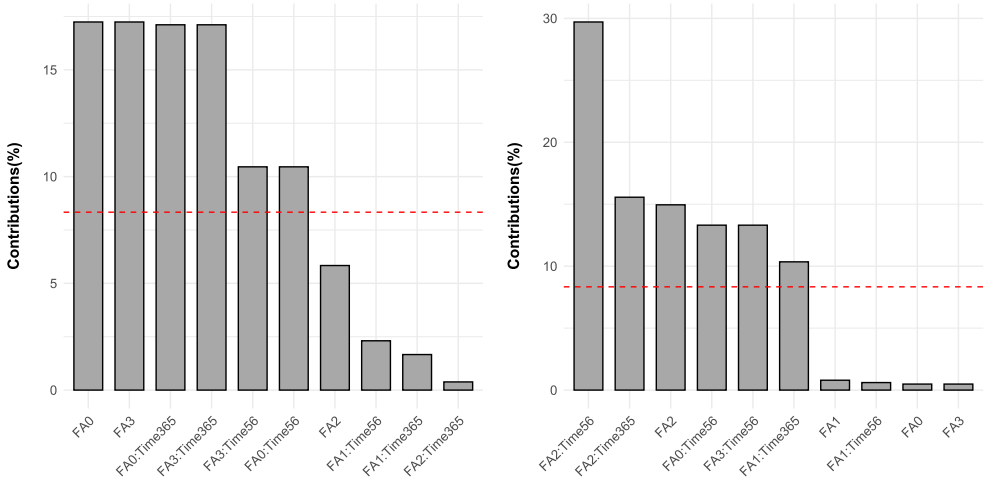


Figure 5.12: Scree plots of the contributions of standardised FA model coefficients to the Cluster 3 variance explained by PC1 and PC2.

PC1 primarily describes variance in the standardised coefficients for the effect of breastfeeding ($FA3$, $FA3 : Time1$ and $FA3 : Time3$) and formula feeding ($FA0$, $FA0 : Time1$ and $FA0 : Time3$) in infants treated with cotrimoxazole. PC2 primarily describes the effects of no cotrimoxazole treatment in breastfed infants ($FA2$, $FA2 : Time1$ and $FA2 : Time3$).

The scree plot below identifies three immune outcomes that make larger contributions to the variance explained by PC1 than would be expected under the assumption of uniform contributions.

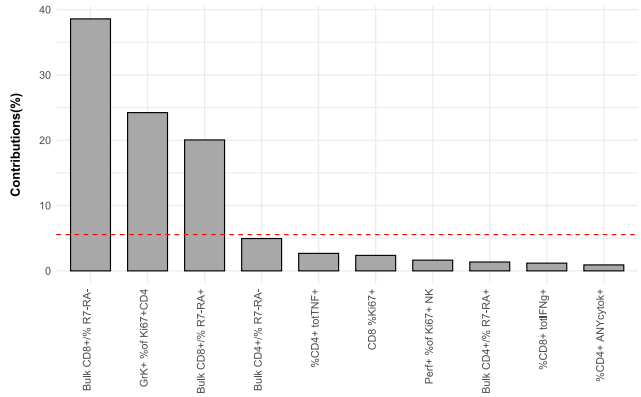


Figure 5.13: Scree plot of the contributions of Group A immune outcomes to the Cluster 3 variance explained by PC1.

Similarly, the scree plot below identifies four immune outcomes that make large contributions to the variance explained by PC2.

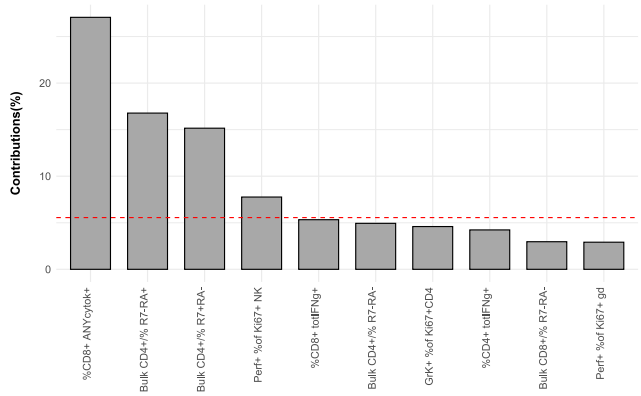


Figure 5.14: Scree plot of the contributions of Group A immune outcomes to the Cluster 3 variance explained by PC2.

Both PC1 and PC2 are required for a lower-dimensional representation that sufficiently captures the variance in the longitudinal effects of combinations of feeding practices and cotrimoxazole treatment in this cluster. Hence, the seven unique immune outcomes making large contributions to PC1 and PC2 are combined to form an immune outcome subset for further analysis:

- *Bulk* $\%CD8^{+}$ $R7^{-}$ $RA^{-(1)}$
- $\%GrK^{+}$ $Ki67^{+}$ $CD4^{+(2)}$
- *Bulk* $\%CD8^{+}$ $R7^{-}$ $RA^{+(3)}$
- $\%CD8^{+}$ $ANYcytok^{+(4)}$
- *Bulk* $\%CD4^{+}$ $R7^{-}$ $RA^{+(5)}$
- *Bulk* $\%CD4^{+}$ $R7^{+}$ $RA^{-(6)}$
- $\%Perf^{+}$ $Ki67^{+}$ $NK^{(7)}$

The immune outcomes contained in the smaller clusters each constitute a subset for further analysis. The smallest cluster (*yellow*) consists of four immune outcomes:

- $\%GrB^{+}$ $Ki67^{+}$ $NK^{(1)}$
- *Bulk* $\%CD8^{+}$ $R7^{+}$ $RA^{+(2)}$
- $\%GrB^{+}$ $Ki67^{+}$ $CD4^{+(3)}$
- *Bulk* $\%CD8^{+}$ $R7^{+}$ $RA^{-(4)}$

The second-largest cluster (*green*) consists of 11 immune outcomes:

- $\%Perf^{+}$ $Ki67^{+}$ $CD4^{+(1)}$
- $\%GrA^{+}$ $Ki67^{+}$ $CD4^{+(2)}$
- $\%GrA^{+}$ $Ki67^{+}$ $gd^{(3)}$
- $\%CD4$ $Ki67^{+(4)}$
- $\%CD8^{+}$ $totTNF^{+(5)}$
- *Bulk* $\%CD4^{+}$ $R7^{+}$ $RA^{+(6)}$
- $\%GrK^{+}$ $Ki67^{+}$ $gd^{(7)}$
- $\%GrB^{+}$ $Ki67^{+}$ $gd^{(8)}$
- $\%NK$ $Ki67^{+(9)}$
- $\%GrA^{+}$ $Ki67^{+}$ $NK^{(10)}$
- $\%CD4$ $totIL7^{+(11)}$

Instead of analysing the subgroup differences described by levels of *FA* for all 33 outcomes simultaneously, the analysis is performed in three subsets of immune outcomes identified from clusters describing immune outcomes with similar behaviour over time.

Chapter 6

Outcome Selection

This chapter presents the outcomes selected for each exposure of interest by PCA only and by HCA followed by PCA. These dimension reduction techniques represent different approaches to outcome selection. For HCA followed by PCA, groups of similar immune outcomes (clusters) become the units of analysis. For every cluster, an immune outcome subset is selected for further analysis of subgroup differences per exposure of interest. The PCA-only approach, on the other hand, selects a single immune outcome subset per exposure of interest from lower-dimensional representations of subgroup differences.

The size and contents of the selected immune outcome subset(s) are compared for each exposure of interest (*EOI*) and across dimension reduction techniques. *MVA85A*, *QFT*, and *FA* are initially assumed to have independent effects on the immune outcomes in Group A. However, this assumption is challenging to uphold for immune outcomes selected for multiple subsets corresponding to different *EOI*.

6.0.1 PCA Only

The heat map below compares the contents of Group A immune outcome subsets selected for further analysis into subgroup differences based on the levels of *MVA85A*, *QFT* and *FA*.

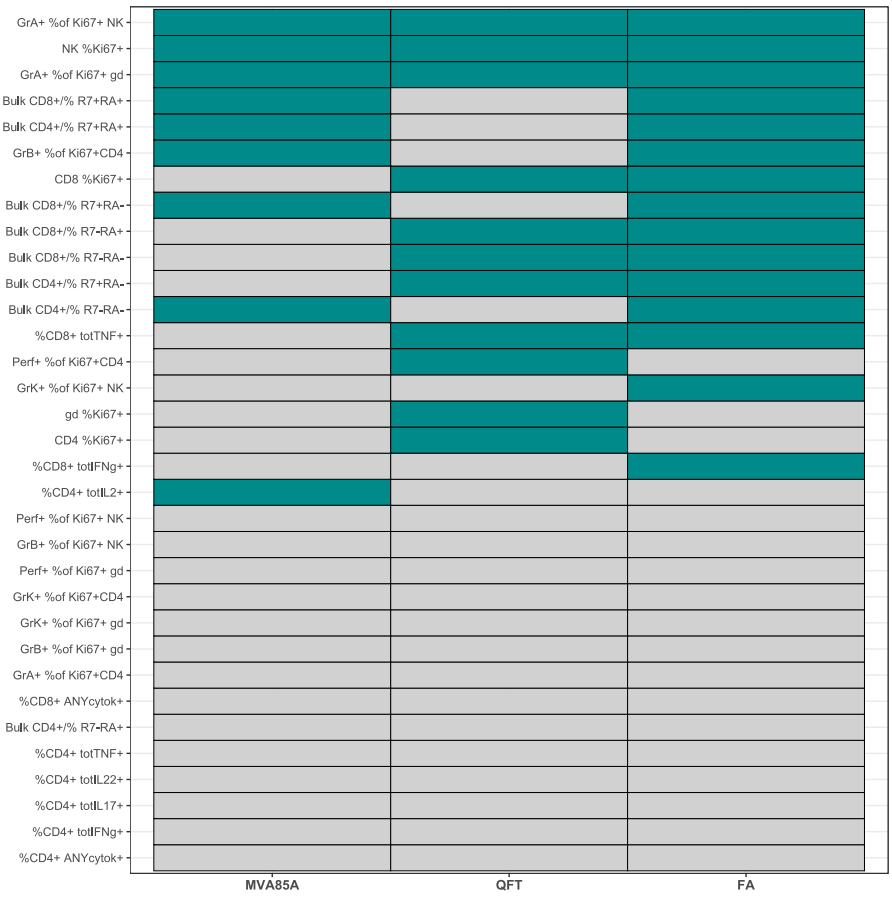


Figure 6.1: Outcome selection by exposure of interest for Group A immune outcomes after dimension reduction by PCA only. Selected outcomes are indicated by green cells.

For an exposure of interest (*EOI*), an immune outcome is selected (*green*) if it makes a large contribution to the lower-dimensional representation of the correlation matrix of standardised model coefficient estimates. These model coefficient estimates compare how different levels of the *EOI* modify the expected value of the immune outcome over time. Hence, the selected immune outcomes explain most of the variance between subgroups described by levels of the *EOI*.

Nineteen unique immune outcomes are selected for further analysis into differences between subgroups defined by different levels of *MVA85A*, *QFT* and *FA*.

- Contrasts computing subgroup differences based on MVA85A priming can be estimated and corrected for multiple hypothesis testing within the *MVA85A* subset of immune outcomes ($m_1^* = 9$).
- Contrasts computing subgroup differences based on maternal *Mtb* sensitisation (measured by a positive maternal QFT) can be estimated and corrected for multiple hypothesis testing within the *QFT* subset of immune outcomes ($m_2^* = 11$).
- Contrasts computing subgroup differences based on combinations of feeding and cotrimoxazole treatment can be estimated and corrected for multiple hypothesis testing within the *FA* subset of immune outcomes ($m_3^* = 15$).

A number of immune outcomes are selected for further analysis for a single *EOI*:

- *%CD4$^+$ totIL2$^+$* was selected for further analysis for *MVA85A* subgroup differences only.
- Three immune outcomes (*%Perf$^+$ Ki67$^+$ CD4$^+$*, *%CD4$^+$ Ki67$^+$*, *%gd Ki67$^+$ T*) were selected for further analysis for *QFT* only.
- Two immune outcomes (*%GrK$^+$ Ki67$^+$ NK* and *%CD8$^+$ totIFNg$^+$*) were selected for further analysis for *FA* only.

It is reasonable to assume that the *EOI* have independent effects on these immune outcomes. Contrasts estimating the corresponding subgroup differences can be computed from the respective longitudinal models without considering model adjustments for additional exposures.

However, many immune outcomes in Group A are selected for further analysis for more than one *EOI*.

- Three immune outcomes (*%GrA$^+$ Ki67$^+$ NK*, *%NK$^+$ Ki67$^+$* and *%GrA$^+$ Ki67$^+$ gd*) were selected for further analysis for all three exposures of interest.
- *Bulk %CD8$^+$ R7$^+$ RA$^+$*, *Bulk %CD4$^+$ R7$^+$ RA$^+$*, *%GrB$^+$ Ki67$^+$ CD4$^+$*, *Bulk %CD8$^+$ R7$^+$ RA$^-$* and *Bulk %CD4$^+$ R7$^-$ RA$^-$* were selected for further analysis for both *MVA85A* and *FA* but not *QFT*.
- Five immune outcomes (*%CD4$^+$ Ki67$^+$*, *Bulk %CD8$^+$ R7$^-$ RA$^+$*, *Bulk %CD8$^+$ R7$^-$ RA$^-$*, *Bulk %CD4$^+$ R7$^+$ RA$^-$* and *%CD8$^+$ totTNF$^+$*) were selected for further analysis for both *QFT* and *FA* but not *MVA85A*.

For these immune outcomes, it is less reasonable to assume that the effects estimated for each *EOI* are independent. The effects of the levels of multiple *EOI* should be disentangled before computing contrasts to estimate subgroup differences. When

computing contrasts estimating the corresponding subgroup differences from the respective longitudinal models, model adjustments should be considered for additional *EOI*. For example, when computing contrasts for the effect of a positive maternal QFT on $\%CD4^{+}$ $Ki67^{+}$, *Bulk* $\%CD8^{+}$ $R7^{-}$ RA^{+}, *Bulk* $\%CD8^{+}$ $R7^{-}$ RA^{-}, *Bulk* $\%CD4^{+}$ $R7^{+}$ RA^{-} or $\%CD8^{+}$ $totTNF^{+}$, the results should be compared to those from a longitudinal model adjusted for feeding-cotrimoxazole combinations (*FA*). Similarly, when computing contrasts for the effects of feeding-cotrimoxazole combinations on these immune outcomes, the results should be compared to those from a longitudinal model adjusted for the effect of a positive maternal QFT.

6.0.2 HCA then PCA

The heat map above compares the contents of Group A immune outcome subsets selected for further analysis into subgroup differences based on the levels of *MVA85A*, *QFT* and *FA*.

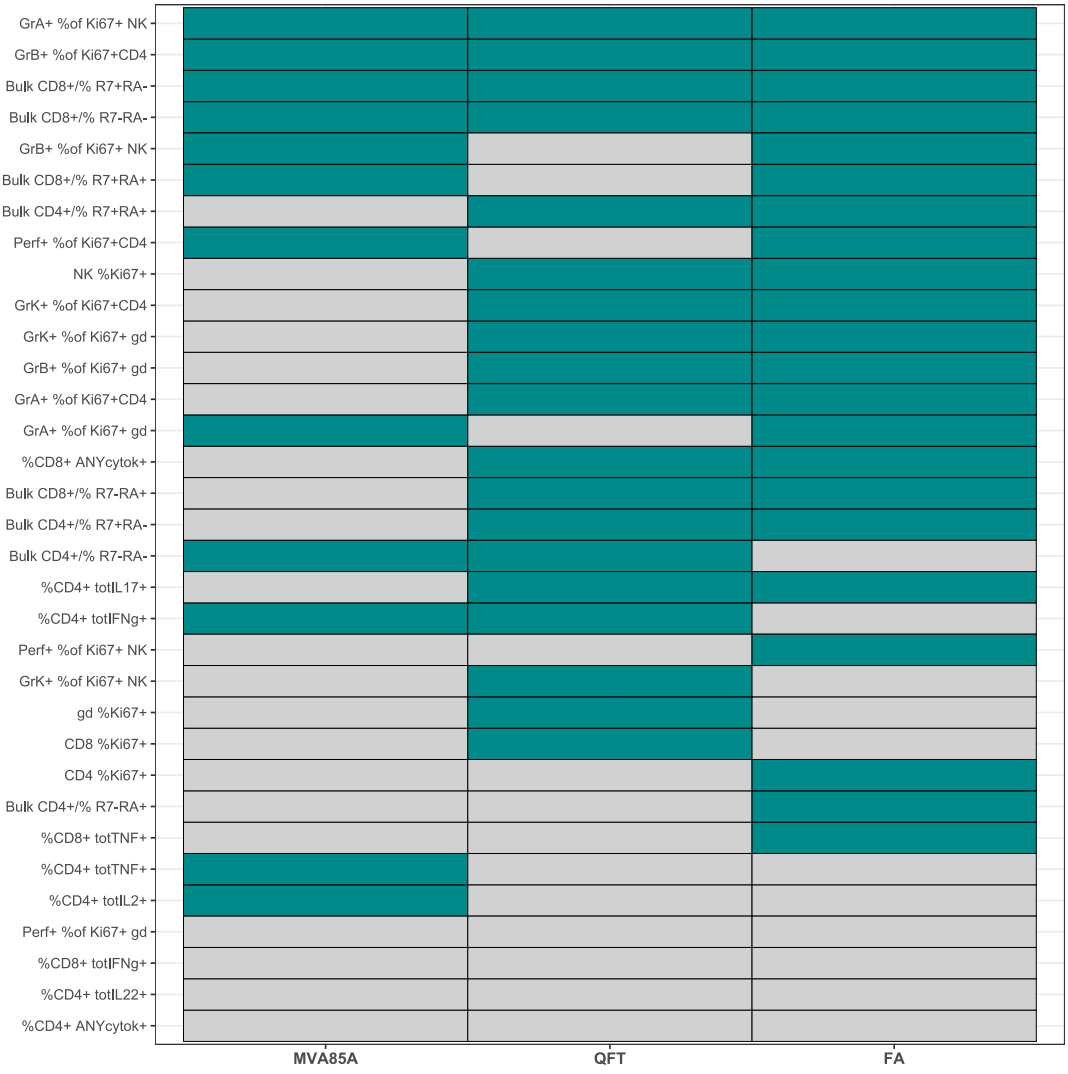


Figure 6.2: Outcome selection by exposure of interest for Group A immune outcomes by HCA followed by PCA. Selected outcomes are indicated by green cells.

HCA defines clusters of immune outcomes that are similar with respect to the standardised model coefficients for an exposure of interest (*EOI*). These clusters become

the units of the analysis. An immune outcome is selected (*blue*) under one of two conditions:

1. The cluster contains fewer immune outcomes than there are standardised model coefficients. In this case, all immune outcomes contained in the cluster are selected for further analysis.
2. The cluster contains at least as many immune outcomes as there are standardised model coefficients. In this case, a PCA is applied to the correlation matrix of the standardised regression coefficients estimated for the cluster members. An immune outcome is selected for further analysis if it makes a large contribution to the cluster variance.

Twenty-seven unique immune outcomes are selected for further analysis into differences between subgroups defined by different levels of *MVA85A*, *QFT* and *FA*.

- Twelve immune outcomes in six cluster subsets are selected for further analysis into *MVA85A* subgroup differences.
- Nineteen immune outcomes in seven cluster subsets are selected for further analysis into *QFT* subgroup differences.
- Twenty-two immune outcomes in three cluster subsets are selected for further analysis into *FA* subgroup differences.

The heat map below describes the structure of the cluster subsets for the exposure of interest *MVA85A*.

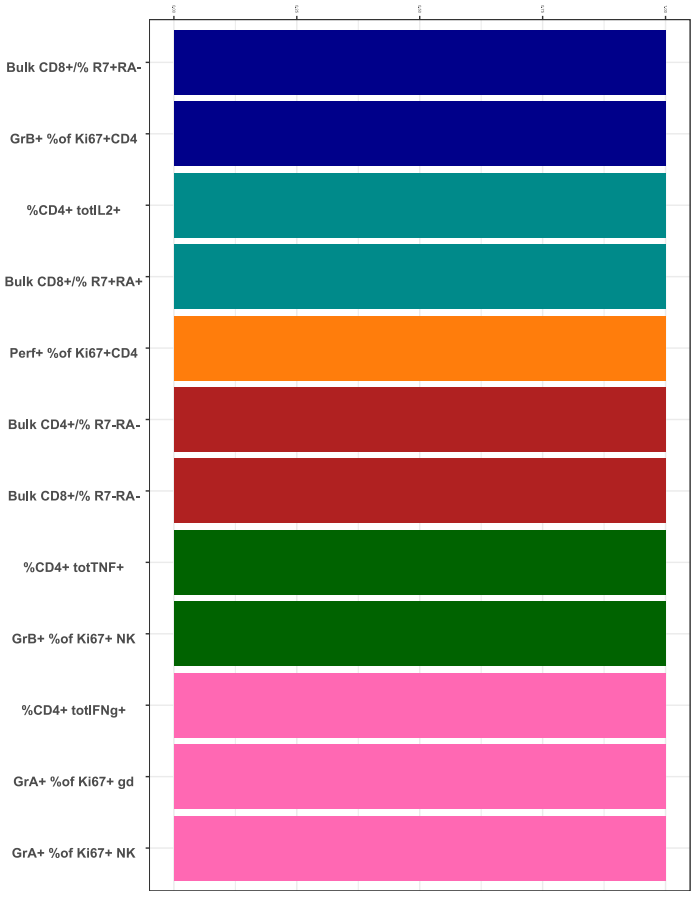


Figure 6.3: Cluster subset structure for Group A immune outcome selection for further analysis into subgroup differences described by *MVA85A*.

These cluster subsets become the units of further analysis:

– Contrasts computing subgroup differences based on MVA85A priming can be estimated and corrected for multiple hypothesis testing within these six subsets ($m_1^{**} = 2$, $m_2^{**} = 2$, $m_3^{**} = 1$, $m_4^{**} = 2$, $m_5^{**} = 2$, $m_6^{**} = 3$).

The heat map below describes the structure of the cluster subsets for the exposure of interest *QFT*.

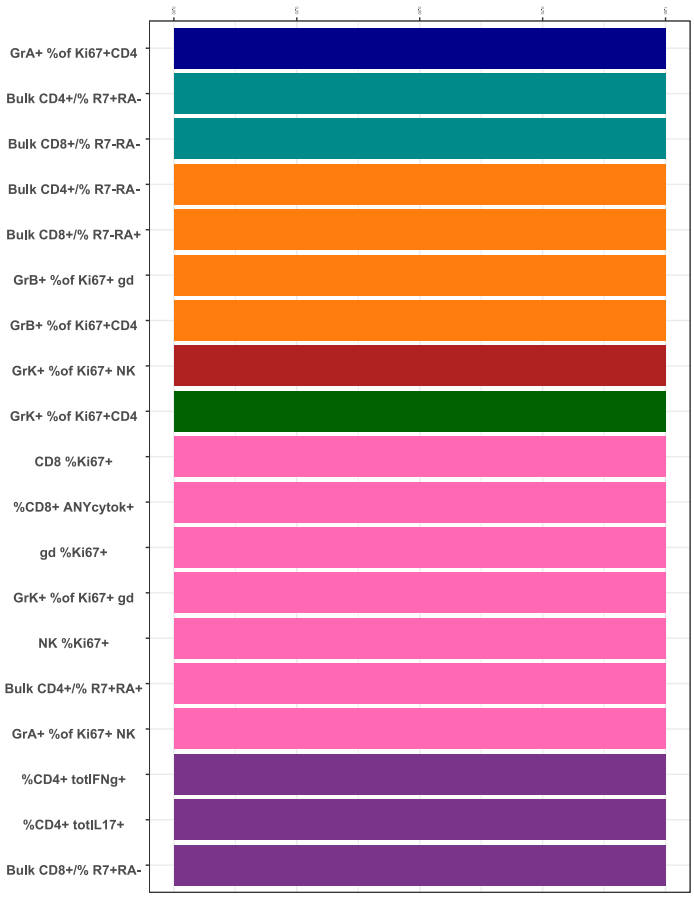


Figure 6.4: Cluster subset structure for Group A immune outcome selection for further analysis into subgroup differences described by *QFT*.

These cluster subsets become the units of further analysis:

– Contrasts computing subgroup differences based on maternal QFT can be estimated and corrected for multiple hypothesis testing within the seven subsets corresponding to the seven clusters ($m_1^{**} = 1$, $m_2^{**} = 2$, $m_3^{**} = 4$, $m_4^{**} = 1$, $m_5^{**} = 1$, $m_6^{**} = 7$, $m_7^{**} = 3$).

The heat map below describes the structure of the cluster subsets for the exposure of interest *FA*.

Figure 6.5: Cluster subset structure for Group A immune outcome selection for further analysis into subgroup differences described by *FA*.

These cluster subsets become the units of further analysis:

- Contrasts computing subgroup differences based on combinations of feeding and cotrimoxazole treatment can be estimated and corrected for multiple hypothesis testing within the three subsets corresponding to the three clusters ($m_1^{**} = 4$, $m_2^{**} = 11$, $m_3^{**} = 5$)

Referring back to Figure 6.2, a number of immune outcomes were selected for further analysis for a single *EOI*:

- $\%CD4^+$ $totIL2^+$ and $\%CD4^+$ $totTNF^+$ were selected for further analysis for *MVA85A* subgroup differences only.
- $\%GrK^+$ $Ki67^+$ *NK*, $\%gd$ $Ki67^+$, and $\%CD8^+$ $Ki67^+$ were selected for further analysis for *QFT* subgroup differences only.
- $\%Perf^+$ $Ki67^+$ *NK*, $\%CD4^+$ $Ki67^+$, *Bulk* $\%CD4^+$ $R7^+$ RA^-, $\%CD8^+$ $totTNF^+$ were selected for further analysis for *FA* subgroup differences only.

However, many immune outcomes are selected for further analysis for more than one *EOI*.

- Four immune outcomes (*%GrA$^+$ Ki67$^+$ NK*, *%GrB$^+$ Ki67$^+$ CD4$^+$*, *Bulk %CD8$^+$ R7$^+$ RA$^-$* and *Bulk %CD8$^+$ R7$^-$ RA$^-$*) were selected for further analysis for all three exposures of interest.
- Four immune outcomes (*%GrB$^+$ Ki67$^+$ NK*, *Bulk %CD8$^+$ R7$^+$ RA$^+$*, *%Perf$^+$ Ki67$^+$ CD4$^+$* and *%GrA$^+$ Ki67$^+$ gd*) were selected for further analysis for both *MVA85A* and *FA* but not *QFT*.
- Ten immune outcomes were selected for further analysis for both *QFT* and *FA*, but not *MVA85A*.
  - *Bulk %CD4$^+$ R7$^+$ RA$^+$*
  - *%NK Ki67$^+$*
  - *%GrK$^+$ Ki67$^+$ CD4$^+$*
  - *%GrK$^+$ Ki67$^+$ gd*
  - *%GrB$^+$ Ki67$^+$ gd*
  - *%GrA$^+$ Ki67$^+$ CD4$^+$*
  - *%CD8$^+$ ANYcytok$^+$*
  - *Bulk %CD8$^+$ R7$^-$ RA$^+$*
  - *Bulk %CD4$^+$ R7$^+$ RA$^-$*
  - *%CD4$^+$ totIL17$^+$*
- Two immune outcomes (*Bulk %CD4$^+$ R7$^-$ RA$^-$* and *%CD4$^+$ totIFNg$^+$*) were selected for further analysis for both *MVA85A* and *QFT*, but not *FA*.

For these immune outcomes, as described in outcome selection by PCA only, the effects of the levels of multiple *EOI* must be disentangled by adjusting the corresponding longitudinal models for additional *EOI*, as required, before calculating contrasts to estimate subgroup differences.

Chapter 7 investigates how these different approaches to selecting outcomes and computing contrasts may affect the number of discoveries and q-values after applying FDR control. Following both approaches, the contrasts will be computed and adjusted for multiple hypothesis testing.

Chapter 7

Corrections for Multiple Testing

This chapter describes and demonstrates the *Corrections for Multiple Testing* step in the outcome selection workflow (Figure 1.2). The preceding step, *Dimension Reduction*, identified immune outcome subset(s) with the most evidence of subgroup differences with respect to the exposure variables of interest (*MVA85A*, *QFT* and *FA*). Contrasts estimating these subgroup differences can be computed from the corresponding longitudinal models of the relationships between the expectations of the immune outcomes and the levels of the exposure variables. The contrast p-values quantify the evidence for these subgroup differences in wider populations beyond the observed data.

This chapter demonstrates how corrections for multiple hypothesis testing should be applied to contrasts after outcome selection and before reporting results. Different dimension reduction techniques may select different outcomes and lead to different numbers of comparisons. Although the number of contrasts is predetermined, these contrasts may be estimated for different numbers of immune outcomes. Hence, this chapter compares the discoveries (for $q < 0.15$) and q-values for outcomes selected by PCA only and by HCA followed by PCA.

Two False Discovery Rate (FDR) control methods are compared in this chapter: the Benjamini-Hochberg (BH) procedure (Benjamini & Hochberg, 1995) and the more conservative Benjamini-Yekutieli (BY) procedure (Benjamini & Yekutieli, 2001). Brief theoretical overviews are provided before applying these methods to outcomes selected by PCA only and by HCA followed by PCA.

MVA85A, *QFT*, and *FA* are initially assumed to have independent effects on immune outcomes. However, as demonstrated in Chapter 6, a number of immune outcomes were selected for further analysis for at least two of these *EOI*.

The contrasts estimated under the assumption of independence are therefore compared to contrasts estimated from the same longitudinal models adjusted for the effects of additional *EOI*, where indicated by overlapping outcome selections in Chapter 6.

For example, *Bulk %CD4$^+$ R7$^-$ RA$^-$* was selected for further analysis for both *MVA85A* and *QFT*, but not *FA*. First, *MVA85A* and *QFT* were assumed to have independent effects on *Bulk %CD4$^+$ R7$^-$ RA$^-$*. Under this assumption, contrasts computing subgroup differences based on MVA85A priming were estimated from the following longitudinal model.

$$\begin{aligned}\textbf{Bulk \%CD4}^+\ \textbf{R7}^-\ \textbf{RA}^-{}_i^{(1)} | \mathbf{x}_i, \mathbf{u}_i = \mathrm{g}^{-1}\Big[\Big(\boldsymbol{\beta}_0^{(1)} + \mathbf{u}_i^{(1)}\Big) + \boldsymbol{\beta}_1^{(1)}\mathrm{MVA85A}_i + \boldsymbol{\beta}_2^{(1)}\mathrm{Time3}_i + \boldsymbol{\beta}_3^{(1)}\mathrm{Time1}_i + \\ \boldsymbol{\beta}_4^{(1)}\mathrm{MVA85A{:}Time3}_i + \boldsymbol{\beta}_5^{(1)}\mathrm{MVA85A{:}Time1}_i + \\ \boldsymbol{\beta}_6^{(1)}\mathrm{Sex}_i + \boldsymbol{\beta}_7^{(1)}\mathrm{Sex{:}Time3}_i + \boldsymbol{\beta}_8^{(1)}\mathrm{Sex{:}Time1}_i + \mathbf{e}_i^{(1)}\end{aligned}$$

Second, because *Bulk %CD4$^+$ R7$^-$ RA$^-$* was also selected for further analysis for *QFT*, contrasts computing subgroup differences based on MVA85A priming were estimated from the following adjusted longitudinal model.

$$\textbf{Bulk \%CD4}^{+}\ \textbf{R7}^{-}\ \textbf{RA}^{-}{}_{i}^{(1)}|\mathbf{x}_i, \mathbf{u}_i = \mathrm{g}^{-1}\left[\left(\boldsymbol{\beta}_0^{(1)} + \mathbf{u}_i^{(1)}\right) + \boldsymbol{\beta}_1^{(1)}\mathrm{MVA85A}_i + \boldsymbol{\beta}_2^{(1)}\mathrm{Time3}_i + \boldsymbol{\beta}_3^{(1)}\mathrm{Time1}_i + \boldsymbol{\beta}_4^{(1)}\mathrm{MVA85A{:}Time3}_i + \boldsymbol{\beta}_5^{(1)}\mathrm{MVA85A{:}Time1}_i + \boldsymbol{\beta}_6^{(1)}\mathrm{Sex}_i + \boldsymbol{\beta}_7^{(1)}\mathrm{Sex{:}Time3}_i + \boldsymbol{\beta}_8^{(1)}\mathrm{Sex{:}Time1}_i + \boldsymbol{\beta}_9^{(1)}\mathrm{QFT}_i + \mathbf{e}_i^{(1)}$$

Both sets of contrasts were corrected for multiple testing. The number and nature of discoveries (for $q < 0.15$) could then be compared when assuming accounting for additional exposures.

7.1 An Overview of False Discovery Rate Control

When statistical analyses are performed for multiple outcomes, multiple comparisons are made simultaneously (Goeman & Solari, 2014). Each hypothesis test has some bounded probability α of making a Type I error - in other words, falsely identifying a statistically significant relationship between an exposure variable and an outcome where there is truly none (Goeman & Solari, 2014). Given this, if a family of hypotheses are tested simultaneously, the probability of at least one false positive is greater than α. False Discovery Rate (FDR) control methods maintain a proportion of false positives (q) within the family of tests, corresponding to the desired α for an individual test. The Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) procedures are two well-known approaches for FDR control.

7.1.1 Benjamini-Hochberg Procedure

This procedure controls the FDR under the assumption of independent or positively dependent hypotheses. The procedure is implemented as follows:

1. Sort the p-values in ascending order:

$$p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(k)},$$

   where $p_{(k)}$ is the kth smallest p-value ($k \in 1 \ldots m$) of m comparisons in the family of tests.

2. Compute the corresponding q-values:

$$q_{(1)} \leq q_{(2)} \leq \ldots \leq q_{(k)},$$

   where $q_{(k)}$ is the kth smallest q-value ($q \in 1 \ldots m$) of m comparisons in the family of tests. The kth smallest q-value is defined as:

$$q_{(k)} = \min_{j \geq k}\left(\frac{m p_{(j)}}{j}\right),$$

   with the condition that the q-values are non-decreasing:

$$q_{(k)} = \min\left(q_{(k)},\ q_{(k+1)}\right).$$

   where m is the total number of hypotheses.

3. Reject all hypotheses where the q-value is less than or equal to the desired FDR level α:

$$q_{(k)} \leq \alpha.$$

The q-value represents the minimum FDR at which a given hypothesis test can be declared significant within a family of tests. In other words, it is an adjusted measure of significance that accounts for the expected proportion of false discoveries among the rejected hypotheses.

7.1.2 Benjamini-Yekutieli Procedure

The BY procedure extends the BH approach to control the FDR under arbitrary dependence, including positive, negative and complex dependence among test statistics.

To achieve this, it introduces a correction factor, $c(m)$, defined as:

$$c(m) = \sum_{i=1}^{m} \frac{1}{i}.$$

The inclusion of $c(m)$ makes the BY procedure more conservative than BH, particularly for large families of tests, because $c(m)$ grows logarithmically with the number of tests.

The BY procedure modifies the BH procedure as follows:

1. Sort the p-values in ascending order:

$$p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(k)},$$

   where $p_{(k)}$ is the kth smallest p-value ($k \in 1 \ldots m$) of m comparisons in the family of tests.

2. Compute the corresponding q-values:

$$q_{(1)} \leq q_{(2)} \leq \ldots \leq q_{(k)},$$

   where $q_{(k)}$ is the kth smallest q-value ($q \in 1 \ldots m$) of m comparisons in the family of tests. The kth smallest q-value is defined as:

$$q_{(k)} = \min_{j \geq k} \left(\frac{m p_{(j)}}{j \times c(m)} \right),$$

   with the condition that the q-values are non-decreasing:

$$q_{(k)} = \min \left(q_{(k)},\ q_{(k+1)} \right).$$

   where m is the total number of hypotheses.

3. Reject all hypotheses where the q-value is less than or equal to the desired FDR level α:

$$q_{(k)} \leq \alpha.$$

7.2 Application

Contrasts estimating subgroup differences for the exposures of interest were computed from longitudinal models corresponding to outcomes selected by PCA only, and HCA followed by PCA. The exposures of interest were initially assumed to have independent effects on the immune outcomes. However, these contrasts were compared to contrasts estimated from longitudinal models adjusted for the effects of additional exposures.

Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) procedures were applied to the contrast p-values to control the False Discovery Rate (FDR). The family of tests includes all pairwise comparisons: (a) within-subgroup differences between successive time points (for example, the difference between $\%GrA^+$ $Ki67^+$ NK at $Time1$ and $Time2$ in infants primed with MVA85A), and (b) between-subgroup differences at the same time point (for example, the difference between $\%GrA^+$ $Ki67^+$ NK at $Time1$ in infants primed with MVA85A and infants primed with the control). Although only the comparisons in (b) are of primary interest for outcome selection and are therefore displayed in the tables, those in (a) are biologically relevant and must be included in the correction procedure. Q values were ranked from smallest to largest, and comparisons were selected for display if they met the threshold of q ¡ 0.15 within category (b). Comparisons in (a) that met the same threshold were retained in the correction but are not shown, as they are not directly relevant to the primary analytic question. Notably, many of the smallest q values corresponded to within-subgroup time point comparisons (category a), which is why only comparisons with larger rank positions are visible in the displayed tables.

The discoveries (for $q < 0.15$ and $q < 0.05$) and q-values are compared for:

- Outcomes selected by PCA only and by HCA followed by PCA
- Contrasts assuming independent exposures of interest and contrasts adjusted for additional exposures
- The BH and BY procedures for FDR control

Multiple testing corrections are presented for two exposures of interest (*MVA85A* and *QFT*). Appendix I contains the multiple testing corrections for the selected *FA* outcomes.

7.2.1 MVA85A priming

Contrasts

The matrix of model coefficient estimates $\hat{\boldsymbol{\beta}}$ from the corresponding longitudinal models (Equation 4.1) is multiplied by the following contrast matrix (C) to estimate subgroup differences ($\hat{C}$) for this exposure of interest.

Contrast	**Control** *Time*1	**Control** *Time*2	**Control** *Time*3	**MVA85A** *Time*1	**MVA85A** *Time*2	**MVA85A** *Time*3
Control *Time*2 vs *Time*1	-1	0	1	0	0	0
Control *Time*2 vs *Time*3	-1	0	0	0	1	0
Control *Time*1 vs *Time*3	0	0	-1	0	1	0
MVA85A *Time*2 vs *Time*1	0	-1	0	-1	0	0
MVA85A *Time*2 vs *Time*3	0	-1	0	0	0	1
MVA85A *Time*1 vs *Time*3	0	0	0	-1	0	1
*Time*2 MVA85A vs Control	-1	1	0	0	0	0
*Time*1 MVA85A vs Control	0	0	-1	1	0	0
*Time*3 MVA85A vs Control	0	0	0	0	-1	1

PCA Only

Contrasts computing subgroup differences based on MVA85A priming were estimated on the log response scale from longitudinal models for a subset of Group A immune outcomes ($m_1^* = 9$).

- There are nine contrasts and nine immune outcomes, meaning that the contrast p-values are adjusted for false discoveries in $9 \times 9 = 81$ comparisons.
- Fifty contrasts are significantly different from zero for a Benjamini-Hochberg FDR of 15%.
- Three left-skewed immune outcomes selected for this subset were reflected prior to longitudinal modelling (*Bulk %CD4$^+$ R7$^+$ RA$^+$*, *Bulk %CD8$^+$ R7$^+$ RA$^+$* and *%GrA$^+$ Ki67$^+$ NK*). Hence, their estimated subgroup differences were multiplied by -1 before exponentiating ($\exp(\hat{C}) = \exp(\text{-}\hat{C})$). These outcomes are marked with an asterisk when presenting the results.

As the GLMMs make use of a log link function, the contrast estimates are exponentiated for interpretation. The key results are summarised below.

Table 7.1: Summary of differences in MVA85A priming subgroups ($q < 0.15$) after correcting for multiple hypothesis testing with the Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) procedures.

Rank	Outcome	Contrast	exp($\hat{C}$)	p-value	BH(q)	BY(q)
47/50	*%NK$^+$ Ki67$^+$*	*Time*1: MVA85A vs Control	1.555	0.074	0.127	0.631
48/50	*Bulk %CD4$^+$ R7$^-$ RA$^-$*	*Time*2: MVA85A vs Control	0.768	0.076	0.129	0.640
49/50	*%CD4$^+$ totIL2$^+$*	*Time*3: MVA85A vs Control	0.660	0.078	0.130	0.645
50/50	*%GrA$^+$ Ki67$^+$ NK**	*Time*2: MVA85A vs Control	0.707	0.088	0.143	0.712

Four immune outcomes are significantly different in the MVA85A priming subgroup when contrasts are corrected for multiple hypothesis testing with the BH procedure ($q < 0.15$). However, these discoveries are not retained when the more conservative BY procedure is applied ($q > 0.6$). If the analysis followed scientific convention for statistical significance ($q < 0.05$), there would be no discoveries of significant subgroup differences for *MVA85A*.

However, these contrasts assume that *MVA85A*, *QFT*, and *FA* have independent effects on immune outcomes. Referring back to the immune outcome subsets selected for the exposures of interest in Chapter 6, the longitudinal models can be adjusted for the effects of multiple exposures. The adjustment was applied to immune outcomes selected for *MVA85A* and at least one of *QFT* and *FA*.

For example, *%NK$^+$ Ki67$^+$* is selected for further analysis with respect to *MVA85A*, *QFT* and *FA*. The corresponding longitudinal model in Equation 4.1 (Chapter 4) is adjusted as follows. The changes are printed in red:

$$\begin{aligned}\textbf{\%NK}^+\ \textbf{Ki67}^+{}_i^{(1)} \mid \mathbf{x_i},\ \mathbf{u_i} \ = \ \mathrm{g}^{-1}\Big[&\left(\boldsymbol{\beta}_0^{(1)} + \mathbf{u}_i^{(1)}\right) + \boldsymbol{\beta}_1^{(1)}\mathrm{MVA85A}_i + \boldsymbol{\beta}_2^{(1)}\mathrm{Time3}_i + \boldsymbol{\beta}_3^{(1)}\mathrm{Time1}_i + \\ &\boldsymbol{\beta}_4^{(1)}\mathrm{MVA85A{:}Time3}_i + \boldsymbol{\beta}_5^{(1)}\mathrm{MVA85A{:}Time1}_i + \\ &\boldsymbol{\beta}_6^{(1)}\mathrm{Sex}_i + \boldsymbol{\beta}_7^{(1)}\mathrm{Sex{:}Time3}_i + \boldsymbol{\beta}_8^{(1)}\mathrm{Sex{:}Time1}_i + \\ &\boldsymbol{\beta}_9^{(1)}\mathrm{QFT}_i + \boldsymbol{\beta}_{10}^{(1)}\mathrm{FA1}_i + \boldsymbol{\beta}_{11}^{(1)}\mathrm{FA2}_i + \boldsymbol{\beta}_{12}^{(1)}\mathrm{FA3}_i + \mathbf{e}_i^{(1)}\Big]\end{aligned}$$

Contrasts computing subgroup differences based on MVA85A priming were then estimated from these adjusted longitudinal models. The number of comparisons

does not change, as the contrast is computed by averaging over the levels of the additional covariates ($m_1^* \times 9 = 9 \times 9 = 81$ comparisons).

Table 7.2: Summary of differences in MVA85A priming subgroups ($q < 0.15$), adjusted for additional exposures, and corrected for multiple hypothesis testing with the Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) procedures.

Rank	Outcome	Contrast	Adjustment	exp($\hat{C}$)	p-value	BH(q)	BY(q)
44/51	*Bulk %CD4$^+$ R7$^-$ RA$^-$*	*Time*2: MVA85A vs Control	$+FA$	1.377	0.028	0.052	0.257
48/51	*Bulk %CD8$^+$ R7$^+$ RA^{+*}*	*Time*1: MVA85A vs Control	$+FA$	1.323	0.049	0.083	0.412
49/51	*%NK$^+$ Ki67$^+$*	*Time*1: MVA85A vs Control	$+QFT + FA$	1.555	0.074	0.123	0.613
50/51	*%CD4$^+$ totIL2$^+$*	*Time*3: MVA85A vs Control	None	0.660	0.078	0.125	0.624
51/51	*%GrB$^+$ Ki67$^+$ CD4$^+$*	*Time*2: MVA85A vs Control	$+FA$	1.356	0.079	0.125	0.624

After adjusting the longitudinal models for additional exposures, five differences in MVA85A priming subgroups are identified after following the BH procedure ($q < 0.15$). The difference observed for *%GrA$^+$ Ki67$^+$ NK* at *Time*2 is no longer detected when the models are adjusted for the average effects of the other exposures of interest (*QFT* and *FA*), as required.

A difference in *Bulk %CD8$^+$ R7$^+$ RA$^+$* at *Time*1 was not discovered before making these adjustments.

- The difference in *Bulk %CD8$^+$ R7$^+$ RA$^+$* between MVA85A priming and control subgroups at the time of BCG vaccination (*Time*1), averaged over the distributions of *Sex* and *FA*, is $\exp(\hat{C}) = 1.323$.
- In other words, mean *Bulk %CD8$^+$ R7$^+$ RA$^+$* at *Time*1 is estimated to be $\approx 32\%$ higher in infants who received MVA85A priming. A statistically significant subgroup difference ($q < 0.15$) does not persist after BCG vaccination (*Time*2 and *Time*3).
- If the analysis follows scientific convention for statistical significance ($q < 0.05$), this discovery would not be significant, whether assuming independent or positive dependence of tests (BH(q) = 0.052) or arbitrary dependence (BY).

Similarly, a difference in *%GrB$^+$ Ki67$^+$ CD4$^+$* at *Time*2 was not discovered for $q < 0.15$ before making these adjustments.

- The difference in *%GrB$^+$ Ki67$^+$ CD4$^+$* between MVA85A priming and control subgroups 8 weeks after BCG (*Time*2), averaged over the distributions of *Sex* and *FA*, is $\exp(\hat{C}) = 1.356$.
- In other words, mean *%GrB$^+$ Ki67$^+$ CD4$^+$* at *Time*2 is estimated to be $\approx 36\%$ higher in infants who received MVA85A priming.
- A statistically significant subgroup difference ($q < 0.15$) is not detected at the time of BCG vaccination (*Time*1) or one year after BCG vaccination (*Time*3).
- Once again, if the analysis followed scientific convention for statistical significance ($q < 0.05$), this discovery would not be significant, whether assuming independent or positive dependence of tests (BH(q) = 0.083) or arbitrary dependence (BY).

Estimated subgroup differences in *%NK$^+$ Ki67$^+$* and *%CD4$^+$ totIL2$^+$* are identical after the adjustments. This is expected for *%CD4$^+$ totIL2$^+$* as this immune outcome was uniquely selected to the *MVA85A* outcome subset and no adjustments were applied to its longitudinal model. However, the *%NK$^+$ Ki67$^+$* model was adjusted for both *QFT* and *FA*.

- The estimated subgroup difference in $\%NK^+$ $Ki67^+$ at $Time1$ is identical ($\exp(\hat{C}) = 1.555$).
- In other words, mean $\%NK^+$ $Ki67^+$ at the time of BCG ($Time1$) is $\approx 56\%$ higher in MVA85A primed infants, whether averaged over the distribution of Sex only or over the distributions of all other exposure variables.
- This is a far stronger result than estimating subgroup differences without consideration of other exposure variables.

HCA Then PCA

Contrasts computing subgroup differences based on MVA85A priming were estimated from longitudinal models for six subsets of Group A immune outcomes (Figure 6.3: $m_1^{**} = 2$, $m_2^{**} = 2$, $m_3^{**} = 1$, $m_4^{**} = 2$, $m_5^{**} = 2$, $m_6^{**} = 3$). Once again, nine contrasts are estimated for each subset. For each subset, contrast p-values are adjusted for false discoveries. These adjustments consider different numbers of comparisons:

- Cluster 1: $m_1^{**} \times 9 = 2 \times 9 = 18$
- Cluster 2: $m_2^{**} \times 9 = 2 \times 9 = 18$
- Cluster 3: $m_3^{**} \times 9 = 1 \times 9 = 9$
- Cluster 4: $m_4^{**} \times 9 = 2 \times 9 = 18$
- Cluster 5: $m_5^{**} \times 9 = 2 \times 9 = 18$
- Cluster 6: $m_6^{**} \times 9 = 3 \times 9 = 27$

The estimated subgroup differences for left-skewed outcomes reflected prior to longitudinal modelling were multiplied by -1 before exponentiating ($\exp(\hat{C}) = \exp(\text{-}\hat{C})$). As in the previous section, these outcomes are marked with an asterisk when presenting the results.

- One immune outcome selected for Cluster 1 was reflected prior to longitudinal modelling ($\%GrA^+$ $Ki67^+$ NK).
- One immune outcome selected for Cluster 5 was reflected prior to longitudinal modelling ($\%GrB^+$ $Ki67^+$ NK).
- One immune outcome selected for Cluster 4 was reflected prior to longitudinal modelling ($Bulk$ $\%CD4^+$ $R7^+$ RA^+).

Multiple testing corrections were applied to the contrast p-values for each of the six clusters. The key results for a Benjamini-Hochberg FDR of 15% are summarised below:

Table 7.3: Summary of differences in MVA85A priming subgroups ($q < 0.15$) after correcting for multiple hypothesis testing by cluster with the Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) procedures.

Cluster	Rank	Outcome	Contrast	$\exp(\hat{C})$	p-value	BH(q)	BY(q)
2	13/13	$\%CD4^+$ $totIL2^+$	$Time3$: MVA85A vs Control	0.660	0.078	0.109	0.379
4	13/13	$Bulk$ $\%CD4^+$ $R7^-$ RA^-	$Time2$: MVA85A vs Control	0.768	0.076	0.106	0.369
5	11/13	$\%GrB^+$ $Ki67^+$ NK^*	$Time1$: MVA85A vs Control	1.538	0.010	0.017	0.058
5	13/13	$\%CD4^+$ $totTNF^+$	$Time3$: MVA85A vs Control	0.636	0.043	0.059	0.208
6	17/17	$\%GrA^+$ $Ki67^+$ NK^*	$Time2$: MVA85A vs Control	0.707	0.088	0.140	0.546

Five immune outcomes are significantly different in the MVA85A priming subgroup when contrasts are corrected for multiple hypothesis testing with the BH procedure ($q < 0.15$). Significant differences in $\%CD4^+$ $totIL2^+$ at $Time3$, $Bulk$ $\%CD4^+$ $R7^-$ RA^- at $Time2$, and $\%GrA^+$ $Ki67^+$ NK at $Time2$ are also identified after corrections

in the PCA-only approach. However, only one discovery is retained when the more conservative BY procedure is applied ($\%GrB^+$ $Ki67^+$ NK at $Time1$).

If the analysis followed scientific convention for statistical significance ($q < 0.05$), there would be a single discovery of significant subgroup differences for *MVA85A*.

- The difference in $\%GrB^+$ $Ki67^+$ NK between MVA85A priming and control subgroups at the time of BCG vaccination ($Time1$), averaged over the distribution of *Sex*, is $\exp(\hat{C}) = 1.538$.
- In other words, mean $\%GrB^+$ $Ki67^+$ NK at $Time1$ is estimated to be $\approx 54\%$ higher in infants who received MVA85A priming. A statistically significant subgroup difference ($q < 0.15$) does not persist after BCG vaccination ($Time2$ and $Time3$).
- This discovery would not be significant ($q < 0.05$) assuming arbitrary dependence (BY) among the family of tests.

However, these contrasts assume that *MVA85A*, *QFT*, and *FA* have independent effects on the immune outcomes. Referring back to the immune outcome subsets selected for the exposures of interest in Chapter 7 (p. 64), the longitudinal models can be adjusted for the effects of multiple exposures. The adjustment was applied to immune outcomes selected for *MVA85A* and at least one of *QFT* and *FA*.

Contrasts computing subgroup differences based on MVA85A priming were then estimated from these adjusted longitudinal models. The number of comparisons does not change, as the contrasts are computed by averaging over the levels of the additional exposure variables:

- Cluster 1: $m_1^{**} \times 9 = 2 \times 9 = 18$
- Cluster 2: $m_2^{**} \times 9 = 2 \times 9 = 18$
- Cluster 3: $m_3^{**} \times 9 = 1 \times 9 = 9$
- Cluster 4: $m_4^{**} \times 9 = 2 \times 9 = 18$
- Cluster 5: $m_5^{**} \times 9 = 2 \times 9 = 18$
- Cluster 6: $m_6^{**} \times 9 = 3 \times 9 = 27$

After adjusting the longitudinal models for additional exposures (as required), multiple testing corrections were applied to the contrast p-values for each of the six clusters. The key results for a Benjamini-Hochberg FDR of 15% are summarised below:

Table 7.4: Summary of differences in MVA85A priming subgroups ($q < 0.15$), adjusted for additional exposures, and corrected for multiple hypothesis testing by cluster with the Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) procedures.

Cluster	Rank	Outcome	Contrast	Adjustments	$\exp(\hat{C})$	p-value	BH(q)	BY(q)
1	10/12	$\%GrB^+$ $Ki67^+$ $CD4^+$	$Time2$: MVA85A vs Control	$+QFT + FA$	1.422	0.043	0.072	0.250
2	13/14	Bulk $\%CD8^+$ $R7^+$ RA^+	$Time1$: MVA85A vs Control	$+FA$	1.323	0.049	0.068	0.237
2	14/14	$\%CD4^+$ $totIL2^+$	$Time3$: MVA85A vs Control	None	0.660	0.078	0.101	0.352
4	13/13	Bulk $\%CD4^+$ $R7^-$ RA^-	$Time2$: MVA85A vs Control	$+QFT$	0.758	0.062	0.086	0.302
5	11/13	$\%GrB^+$ $Ki67^+$ NK^*	$Time1$: MVA85A vs Control	$+FA$	1.594	0.005	0.009	0.031
5	13/13	$\%CD4^+$ $totTNF^+$	$Time3$: MVA85A vs Control	None	0.636	0.043	0.059	0.207

After adjusting the longitudinal models for additional exposures, six differences in MVA85A priming subgroups are identified after following the BH procedure ($q < 0.15$). If the analysis follows scientific convention for statistical significance ($q < 0.05$), one of these contrasts would be a significant discovery, assuming independence or positive dependence among the family of tests (BH), as well as arbitrary dependence (BY). This is a stronger result than any of the discoveries obtained without adjusting for additional exposures.

This difference in $\%GrB^+$ $Ki67^+$ NK at $Time1$ was not significant before making these adjustments.

- The difference in $\%GrB^+$ $Ki67^+$ NK between MVA85A priming and control subgroups at the time of BCG vaccination ($Time1$), averaged over the distributions of Sex and FA, is $\exp(\hat{C}) = 1.594$.
- In other words, mean $\%GrB^+$ $Ki67^+$ NK at $Time1$ is estimated to be $\approx 59\%$ higher in infants who received MVA85A priming.
- This estimate is smaller than the contrast obtained from the unadjusted model ($\exp(\hat{C}) = 1.538$), averaged only over the distribution of Sex.
- The q-values are smaller after adjusting the longitudinal model for additional exposure variables (BH: $q = 0.017$; $q = 0.009$; BY: $q = 0.058$; $q = 0.031$).
- A statistically significant subgroup difference ($q < 0.15$) does not persist after BCG vaccination ($Time2$ and $Time3$).

There were two new discoveries ($q < 0.15$) after adjusting the longitudinal models for additional exposures. First, a difference n $\%GrB^+$ $Ki67^+$ $CD4^+$ at $Time1$ was discovered after corrections for multiple hypothesis testing within adjusted models for Cluster 1 immune outcomes.

- The difference in $\%GrB^+$ $Ki67^+$ $CD4^+$ between MVA85A priming and control subgroups at the time of BCG vaccination ($Time1$), averaged over the distributions of Sex, QFT and FA, is $\exp(\hat{C}) = 1.422$.
- In other words, mean $\%GrB^+$ $Ki67^+$ $CD4^+$ at $Time1$ is estimated to be $\approx 42\%$ higher in infants who received MVA85A priming.
- If the analysis follows scientific convention for statistical significance ($q < 0.05$), this discovery would not be significant assuming independence or positive dependence (BH) or arbitrary dependence (BY) within the family of tests.

Second, a difference in $Bulk$ $\%CD8^+$ $R7^+$ RA^+ at $Time1$ was not discovered before making these adjustments.

- The difference in $Bulk$ $\%CD8^+$ $R7^+$ RA^+ between MVA85A priming and control subgroups at the time of BCG vaccination ($Time1$), averaged over the distributions of Sex and FA, is $\exp(\hat{C}) = 1.323$.
- In other words, mean $Bulk$ $\%CD8^+$ $R7^+$ RA^+ at $Time1$ is estimated to be $\approx 32\%$ higher in infants who received MVA85A priming. A statistically significant subgroup difference ($q < 0.15$) does not persist after BCG vaccination ($Time2$ and $Time3$).
- If the analysis follows scientific convention for statistical significance ($q < 0.05$), this discovery would not be significant assuming independence or positive dependence (BH) or arbitrary dependence (BY) within the family of tests.

The contrast q-values for $\%GrA^+$ $Ki67^+$ NK at $Time2$ are larger after adjusting for additional exposure variables. This difference is no longer detected ($q > 0.15$), as was also observed when applying corrections to outcomes selected by PCA only.

The subgroup difference in $\%CD4^+$ $totIL2^+$ has a smaller q-value for outcome selection by HCA followed by PCA:

- When FDR control is limited to the two outcomes in Cluster 2, the BH q-value assuming independent or positive dependence of tests is 0.101. When assuming arbitrary dependence, the BY q-value is 0.352.
- When FDR control considers nine outcomes selected for further analysis by the PCA only approach, the BH q-value assuming independent or positive dependence of tests is 0.125. When assuming arbitrary dependence, the BY q-value is 0.624.

The subgroup difference in *Bulk* $\%CD4^+$ $R7^-$ RA^- has a larger q-value for outcome selection by HCA followed by PCA:

- When FDR control is limited to the two outcomes in Cluster 4, the BH q-value assuming independent or positive dependence of tests is 0.086. When assuming arbitrary dependence, the BY q-value is 0.302.
- When FDR control considers nine outcomes selected for further analysis by the PCA only approach, the BH q-value assuming independent or positive dependence of tests is 0.052. When assuming arbitrary dependence, the BY q-value is 0.257.

7.2.2 Maternal QFT

Contrasts

The matrix of model coefficient estimates $\hat{\boldsymbol{\beta}}$ from the corresponding longitudinal models (Equation 4.2, p. 36) is multiplied by the following contrast matrix (C) to estimate subgroup differences ($\hat{C}$) for this exposure of interest.

Contrast	**Negative** *Time*1	**Negative** *Time*2	**Negative** *Time*3	**Positive** *Time*1	**Positive** *Time*2	**Positive** *Time*3
Negative *Time*2 vs *Time*1	-1	0	1	0	0	0
Negative *Time*2 vs *Time*3	-1	0	0	0	1	0
Negative *Time*1 vs *Time*3	0	0	-1	0	1	0
Positive *Time*2 vs *Time*1	0	-1	0	1	0	0
Positive *Time*2 vs *Time*3	0	-1	0	0	0	1
Positive *Time*1 vs *Time*3	0	0	0	-1	0	1
*Time*2 Positive vs Negative	-1	1	0	0	0	0
*Time*1 Positive vs Negative	0	0	-1	1	0	0
*Time*3 Positive vs Negative	0	0	0	0	-1	1

PCA Only

Contrasts computing subgroup differences based on maternal *Mtb* sensitisation (measured by a positive maternal QFT) were estimated from longitudinal models for a subset of Group A immune outcomes ($m_2^* = 11$, p. 64).

- There are nine contrasts and 11 immune outcomes, meaning that the contrast p-values are adjusted for false discoveries in $9 \times 11 = 99$ comparisons.
- Fifty-one contrasts are significantly different from zero for a Benjamini-Hochberg FDR of 15%.

– One immune outcome selected for this subset was reflected prior to longitudinal modelling ($\%GrA^+$ $Ki67^+$ NK). Hence, its estimated subgroup differences were multiplied by -1 before exponentiating ($\exp(\hat{C}) = \exp(\text{-}\hat{C})$). This outcome was marked with an asterisk when presenting the results.

The key results are summarised below after exponentiating contrast estimates for interpretation:

Table 7.5: Summary of differences in maternal QFT subgroups ($q < 0.15$) after correcting for multiple hypothesis testing with the Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) procedures.

Rank	Outcome	Contrast	exp(Est)	p-value	BH(q)	BY(q)
39/51	$\%NK^+$ $Ki67^+$	$Time1$: Positive vs Negative	0.472	0.005	0.012	0.061
47/51	$\%GrA^+$ $Ki67^+$ NK^*	$Time2$: Positive vs Negative	1.530	0.037	0.077	0.399
49/51	$\%NK^+$ $Ki67^+$	$Time2$: Positive vs Negative	1.478	0.065	0.130	0.671
50/51	$\%CD8^+$ $Ki67^+$	$Time2$: Positive vs Negative	1.933	0.065	0.130	0.671
51/51	$\%CD8^+$ $totTNF^+$	$Time3$: Positive vs Negative	1.479	0.077	0.149	0.771

There are five discoveries of differences in maternal QFT subgroups after following the BH procedure ($q < 0.15$). However, four of these discoveries are not retained when the more conservative BY procedure is applied ($q > 0.3$). If the analysis followed scientific convention for statistical significance ($q < 0.05$), a significant difference would be detected for $\%NK^+$ $Ki67^+$ at $Time1$ by maternal QFT, assuming independence or positive dependence among the family of tests.

These estimates assume that *QFT*, *MVA85A*, and *FA* have independent effects on Group A immune outcomes. As demonstrated for *MVA85A*, the longitudinal models can be adjusted for the effects of multiple exposures. The adjustment was applied to immune outcomes selected for *QFT* and at least one of *MVA85A* and *FA*.

Contrasts computing subgroup differences based on maternal QFT were then estimated from these adjusted longitudinal models. The number of comparisons remains ($m_2^* \times 9 = 11 \times 9 = 99$ comparisons).

Table 7.6: Summary of differences in maternal QFT subgroups ($q < 0.15$), adjusted for additional exposures, and corrected for multiple hypothesis testing with the Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) procedures.

Rank	Outcome	Contrast	Adjustment	exp(Est)	p-value	BH(q)	BY(q)
37/50	$\%NK^+$ $Ki67^+$	$Time1$: Positive vs Negative	$+MVA85A + FA$	0.457	0.003	0.009	0.047
48/50	$\%CD8^+$ $Ki67^+$	$Time2$: Positive vs Negative	$+FA$	2.055	0.044	0.090	0.387
49/50	$\%GrA^+$ $Ki67^+$ NK^*	$Time2$: Positive vs Negative	$+MVA85A + FA$	1.505	0.047	0.094	0.487
50/50	$\%NK^+$ $Ki67^+$	$Time2$: Positive vs Negative	$+MVA85A + FA$	1.488	0.060	0.119	0.615

After adjusting the longitudinal models for additional exposures, four differences in maternal QFT subgroups are identified after following the BH procedure ($q < 0.15$). The difference observed for $\%CD8^+$ $totTNF^+$ at $Time3$ is no longer detected when the models are adjusted for the average effects of the other exposures of interest (*MVA85A* and *FA*), as required.

After adjusting for the average effects of *MVA85A* and *FA*, the q-values for the difference between subgroups defined by maternal QFT are smaller for $\%NK^+$ $Ki67^+$ at $Time1$ (BH: $q = 0.012$; $q = 0.009$; BY: $q = 0.062$; $q = 0.047$). Even under the assumption of arbitrary dependence within the family of tests, statistical convention would consider this to be a significant result ($q < 0.05$).

- The difference in *%NK$^+$ Ki67$^+$* between positive and negative maternal QFT subgroups at the time of BCG vaccination ($Time1$), averaged over the distributions of *Sex*, *MVA85A*, and *FA*, is $\exp(\hat{C}) = 0.457$.
- In other words, mean *%NK$^+$ Ki67$^+$* at $Time1$ is estimated to be $\approx 54\%$ lower in infants born to *Mtb*-sensitised mothers with a positive QFT test.
- When averaged over the distribution of *Sex* only, the estimated subgroup difference is similar ($\exp(\hat{C}) = 0.472$) but the q-values are larger.

The q-values are also smaller for *%CD8$^+$ Ki67$^+$* at $Time2$ (BH: $q = 0.130$; $q = 0.090$; BY: $q = 0.671$; $q = 0.387$), and *%NK$^+$ Ki67$^+$* at $Time2$ (BH: $q = 0.130$; $q = 0.119$; BY: $q = 0.671$; $q = 0.615$) after adjusting the longitudinal models. However, this does not lead to additional discoveries at $q < 0.05$.

On the other hand, the q-values are larger for *%GrA$^+$ Ki67$^+$ NK* at $Time2$ (BH: $q = 0.077$; $q = 0.094$; BY: $q = 0.399$; $q = 0.487$) after adjusting the longitudinal models.

- The difference in *%GrA$^+$ Ki67$^+$ NK* between positive and negative maternal QFT subgroups 8 weeks after BCG vaccination ($Time2$), averaged over the distributions of *Sex*, *MVA85A* and *FA*, is $\exp(\hat{C}) = 1.505$.
- In other words, mean *%GrA$^+$ Ki67$^+$ NK* at $Time2$ is estimated to be $\approx 51\%$ higher in infants born to *Mtb*-sensitised mothers with a positive QFT test.
- This is a far stronger result than estimating subgroup differences without considering other exposure variables.
- When averaged over the distribution of *Sex* only, the estimated subgroup difference is slightly smaller ($\exp(\hat{C}) = 1.530$).

HCA Then PCA

Contrasts computing subgroup differences based on maternal QFT were estimated from longitudinal models for seven subsets of Group A immune outcomes ($m_1^{**} = 1$, $m_2^{**} = 2$, $m_3^{**} = 4$, $m_4^{**} = 1$, $m_5^{**} = 1$, $m_6^{**} = 7$, $m_7^{**} = 3$). Once again, nine contrasts are estimated for each subset. For each subset, contrast p-values are adjusted for false discoveries. These adjustments consider different numbers of comparisons:

- Cluster 1: $m_1^{**} \times 9 = 1 \times 9 = 9$
- Cluster 2: $m_2^{**} \times 9 = 2 \times 9 = 18$
- Cluster 3: $m_3^{**} \times 9 = 4 \times 9 = 36$
- Cluster 4: $m_4^{**} \times 9 = 1 \times 9 = 9$
- Cluster 5: $m_5^{**} \times 9 = 1 \times 9 = 9$
- Cluster 6: $m_6^{**} \times 9 = 7 \times 9 = 63$
- Cluster 7: $m_7^{**} \times 9 = 3 \times 9 = 27$

The estimated subgroup differences for outcomes reflected prior to longitudinal modelling were multiplied by -1 before exponentiating ($\exp(\hat{C}) = \exp(\text{-}\hat{C})$). As in the previous section, these outcomes are marked with an asterisk when presenting the results. Two immune outcomes selected for Cluster 6 were reflected prior to longitudinal modelling (*%GrA$^+$ Ki67$^+$ NK*; *Bulk %CD4$^+$ R7$^+$ RA$^+$*).

Multiple testing corrections were applied to the contrast p-values for each of the seven clusters. The key results for a Benjamini-Hochberg FDR of 15% are summarised below:

Table 7.7: Summary of differences in maternal QFT subgroups ($q < 0.15$) after correcting for multiple hypothesis testing by cluster with the Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) procedures.

Cluster	Rank	Outcome	Contrast	exp($\hat{C}$)	p-value	BH(q)	BY(q)
5	6/7	$\%GrK^+$ $Ki67^+$ $CD4^+$	$Time3$: Positive vs Negative	2.109	0.031	0.047	0.132
6	18/25	$\%NK^+$ $Ki67^+$	$Time1$: Positive vs Negative	0.472	0.005	0.017	0.079
6	25/25	$\%GrA^+$ $Ki67^+$ NK^*	$Time2$: Positive vs Negative	1.530	0.037	0.092	0.435

There are three discoveries of differences in maternal QFT subgroups after following the BH procedure ($q < 0.15$) across seven clusters. Differences in $\%NK^+$ $Ki67^+$ at $Time1$ and $\%GrA^+$ $Ki67^+$ NK at $Time2$ are also identified when outcome selection is performed by PCA only. However, no discoveries are retained when the more conservative BY procedure is applied.

However, these contrasts assume that *MVA85A*, *QFT*, and *FA* have independent effects on the immune outcomes. Referring back to the immune outcome subsets selected for the exposures of interest in Chapter 7 (p. 64), the longitudinal models can be adjusted for the effects of multiple exposures. The adjustment was applied to immune outcomes selected for *QFT* and at least one of *MVA85A* and *FA*.

Contrasts computing subgroup differences based on maternal QFT were then estimated from these adjusted longitudinal models. The number of comparisons does not change, as the contrast is computed by averaging over the levels of the additional covariates.

After adjusting the longitudinal models for additional exposures (as required), multiple testing corrections were applied to the contrast p-values for each of the six clusters. The key results for a Benjamini-Hochberg FDR of 15% are summarised below:

Table 7.8: Summary of differences in maternal QFT subgroups ($q < 0.15$), adjusted for additional exposures, and corrected for multiple hypothesis testing by cluster with the Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) procedures.

Cluster	Rank	Outcome	Contrast	Adjustments	exp($\hat{C}$)	p-value	BH(q)	BY(q)
5	6/7	$\%GrK^+$ $Ki67^+$ $CD4^+$	$Time3$: Positive vs Negative	$+FA$	2.344	0.011	0.016	0.046
6	16/25	$\%NK^+$ $Ki67^+$	$Time1$: Positive vs Negative	$+FA$	0.455	0.003	0.012	0.057
6	25/25	$\%GrA^+$ $Ki67^+$ NK^*	$Time2$: Positive vs Negative	$+MVA85A + FA$	1.505	0.047	0.117	0.555
7	15/17	$\%CD4^+$ $totIL17^+$	$Time1$: Positive vs Negative	$+FA$	1.751	0.080	0.143	0.555
7	16/17	$\%CD4^+$ $totIL17^+$	$Time2$: Positive vs Negative	$+FA$	0.725	0.085	0.143	0.555

After adjusting the longitudinal models for additional exposures, five differences in maternal QFT subgroups are identified after following the BH procedure ($q < 0.15$). If the analysis follows scientific convention for statistical significance ($q < 0.05$), two of these contrasts would be significant discoveries, assuming independence or positive dependence between the family of tests (BH).

Estimated subgroup differences in $\%GrK^+$ $Ki67^+$ $CD4^+$ at $Time3$ are larger after adjusting the longitudinal models for the immune outcomes in Cluster 5.

- The estimated subgroup difference in $\%GrK^+$ $Ki67^+$ $CD4^+$ at $Time3$ is considerably larger ($\exp(\hat{C}) = 2.344$; $\exp(\hat{C}) = 2.109$).
- In other words, mean $\%GrK^+$ $Ki67^+$ $CD4^+ \approx 44$ weeks after BCG vaccination ($Time3$) is $\approx 134\%$ higher in infants born to *Mtb*-sensitised mothers, averaged over the distributions of *Sex* and *FA*.
- When averaged over the distribution of *Sex* only, mean $\%GrK^+$ $Ki67^+$ $CD4^+$ at $Time3$ is $\approx 111\%$ higher in infants born to *Mtb*-sensitised mothers.

- Without adjusting the longitudinal models for additional exposures, the estimated subgroup difference is $\approx 23\%$ lower.
- After adjustments, the q-values for this contrast are much smaller (BH: $q = 0.016$; $q = 0.047$; BY: $q = 0.046$; $q = 0.132$).

Estimated subgroup differences in $\%NK^+$ $Ki67^+$ at the time of BCG vaccination ($Time1$) are similar after adjusting the longitudinal models, but the contrast q-values are slightly smaller.

- The estimated subgroup difference in $\%NK^+$ $Ki67^+$ at $Time1$ is slightly larger ($\exp(\hat{C}) = 0.472$; $\exp(\hat{C}) = 0.455$).
- In other words, mean $\%NK^+$ $Ki67^+$ at $Time1$ is $\approx 55\%$ lower in infants born to *Mtb*-sensitised mothers, averaged over the distributions of *Sex*, *MVA85A*, and *FA*.
- Without adjusting the longitudinal models for additional exposures, the estimated subgroup difference is $\approx 2\%$ smaller.
- After adjustments, the q-values for this contrast are slightly smaller (BH: $q = 0.017$; $q = 0.012$; BY: $q = 0.079$; $q = 0.057$).

If the analysis follows scientific convention for statistical significance ($q < 0.05$), only the subgroup difference for $\%GrK^+$ $Ki67^+$ $CD4^+$ at $Time3$ would be considered significant assuming arbitrary dependence (BY) within the family of tests. This is a far stronger result than when subgroup differences are estimated without considering other exposure variables.

Differences in $\%CD4^+$ $totIL17^+$ at $Time1$ and $Time2$ were not discovered before making these adjustments.

- $\%CD4^+$ $totIL17^+$ in infants born to *Mtb*-sensitised mothers, averaged over the distributions of *Sex* and *FA*, is $\exp(\hat{C}) = 1.751$ times higher at the time of BCG vaccination ($Time1$).
- However, 8 weeks after BCG vaccination ($Time2$), the estimated difference in $\%CD4^+$ $totIL17^+$ is $\exp(\hat{C}) = 0.725$. In other words, mean $\%CD4^+$ $totIL17^+$ is $\approx 27\%$ lower in infants born to *Mtb*-sensitised mothers at $Time2$.
- If the analysis follows scientific convention for statistical significance ($q < 0.05$), neither discovery would not be significant, whether assuming independent or positive dependence of tests (BH) or arbitrary dependence (BY).

Chapter 8

Overview of Implementation

This dissertation implemented an outcome selection technique that combines longitudinal modelling with dimension reduction to identify immune outcome subset(s) with the most evidence of subgroup differences. Contrasts estimating these subgroup differences were computed from the corresponding longitudinal models and corrected for multiple hypothesis testing. Several different methodologies were considered:

- *Modelling Frameworks:* Linear mixed-effect models (LMM); generalised linear mixed-effect models (GLMM).
- *Dimension Reduction Techniques:* PCA only, Hierarchical Cluster Analysis followed by PCA.
- *Corrections for Multiple Hypothesis Testing:* Benjamini-Hochberg procedure; Benjamini-Yekutieli procedure.

This chapter summarises the insights obtained for outcome selection from implementing these different modelling frameworks, dimension reduction techniques, and corrections for multiple hypothesis testing. The code supporting this dissertation is available on GitHub: https://github.com/ShannonHolcroft/MScDissertation.

8.1 Modelling

Frameworks

In RM-ASCA+ and the outcome selection technique presented in this dissertation, mixed-effect models describe the longitudinal relationships between study outcomes and exposures of interest. Model coefficients estimate how the levels of an exposure of interest modify the expected value of an outcome over time. After extracting coefficient estimates from fitted models, a dimension reduction technique is applied to identify their most meaningful properties. In doing so, RM-ASCA+ merely aims to decompose and visualise the separate effects of exposures over time (Madssen et al., 2021). However, when outcome selection is the objective, these properties can be used to select subset(s) of immune outcomes with the most evidence of subgroup differences over time for further analysis.

RM-ASCA+ implements the linear mixed-effect modelling (LMM) framework, which specifies a linear relationship between the conditional distribution of an immune outcome $\mathbf{Y}$ and $\mathbf{X}\boldsymbol{\beta}$. However, this is often an unrealistic assumption for immunological data in which outcomes tend to have skew, long-tailed distributions. Chapter 4 compares the linear mixed-effect modelling (LMM) and generalised linear mixed-effect (GLMM) modelling frameworks through application to three immune outcomes that exemplify these characteristics of immunological data. Model assumptions were better met within the GLMM framework. In particular, the LMM random-effect distributions deviated more strongly from normality assumptions. This is an important consideration: mixed-effect model estimates, which are the inputs to dimension reduction, are conditional on the random effects.

If model coefficient estimates were extracted from LMMs, the estimated subgroup differences may have been quite different. As an example, consider the LMM and GLMM contrasts ($\hat{C}$) comparing *Bulk %CD4$^+$ R7$^+$ RA$^+$* by maternal QFT. As this left-skewed immune outcome was reflected prior to modelling, $-\hat{C}$ is presented below. The LMM contrasts describe the expected additive difference in *Bulk %CD4$^+$ R7$^+$ RA$^+$* for a positive maternal QFT. The GLMM contrasts describe the expected multiplicative difference in *Bulk %CD4$^+$ R7$^+$ RA$^+$* for a positive maternal QFT.

Table 8.1: Comparison of LMM and GLMM contrasts estimating maternal QFT subgroup differences in *Bulk %CD4$^+$ R7$^+$ RA$^+$*.

Contrast	$(-\hat{C})$**(LMM)**	$\exp(-\hat{C})$**(GLMM)**
$Time2$: Positive vs Negative	1.363	1.323
$Time1$: Positive vs Negative	1.271	1.115
$Time3$: Positive vs Negative	1.77	1.131

When an infant is born to a mother with a positive QFT test, both the LMM and GLMM estimate slightly larger proportions of CD4$^+$ T cells with R7$^+$ RA$^+$ phenotype. The LMM estimates the largest difference (1.77%) at $Time3$ ($\approx$ 44 weeks after BCG vaccination), averaged over the distribution of patient *Sex*. The GLMM, on the other hand, estimates the largest difference (by a multiplicative factor of 1.323) at $Time2$ (8 weeks after BCG vaccination) when the peak immune response is anticipated. This difference between modelling frameworks could be a consequence of violated LMM assumptions.

By specifying a log link function, GLMM coefficients have a multiplicative interpretation. LMM coefficients, on the other hand, have an additive interpretation. As an example of the interpretation differences, consider the LMM and GLMM of the relationship between *Bulk %CD4$^+$ R7$^+$ RA$^+$* and maternal QFT over time. Recall that this left-skewed immune outcome is transformed to be right-skewed prior to modelling (see Figure 4.2).

In the LMM below, the model coefficients have an additive interpretation and describe an absolute subgroup difference. When an infant is exposed to a positive maternal QFT, an absolute increase of 1.363% is expected in Bulk %CD4$^+$ R7$^+$ RA$^+$ 8 weeks after BCG (QFT$_i$), holding *Sex* constant.

$$\begin{aligned}\mathbf{94}-\text{Bulk \%CD4}^+\ \text{R7}^+\ \text{RA}^+{}_i^{(2)} \mid \mathbf{x}_i,\ \mathbf{u}_i\ =\ \exp\Big[\Big(11.322+\mathbf{u}_i^{(2)}\Big)-1.363\ \text{QFT}_i+5.347\ \text{Time3}_i- \\ 2.507\ \text{Time1}_i\ 0.414\ \text{QFT:Time3}_i+0.092\ \text{QFT:Time1}_i+ \\ 0.367\ \text{Sex}_i+2.276\ \text{Sex:Time3}_i-0.631\ \text{Sex:Time1}_i+\mathbf{e}_i^{(2)}\Big] \quad (8.1)\end{aligned}$$

On the other hand, in the GLMM below (Tweedie; $p = 0.148$), the model coefficients have a multiplicative interpretation. When an infant is exposed to a positive maternal QFT, a multiplicative increase of $\exp(-(-0.136)) = 1.146$ is expected in Bulk %CD4$^+$ R7$^+$ RA$^+$ 8 weeks after BCG (QFT$_i$), holding *Sex* constant.

$$\begin{aligned}\mathbf{94}-\text{Bulk \%CD4}^+\ \text{R7}^+\ \text{RA}^+{}_i^{(2)} \mid \mathbf{x}_i,\ \mathbf{u}_i\ =\ \exp\Big[\Big(2.239+\mathbf{u}_i^{(2)}\Big)-0.136\ \text{QFT}_i+0.431\ \text{Time3}_i- \\ 0.279\ \text{Time1}_i\ +0.012\ \text{QFT:Time3}_i+0.027\ \text{QFT:Time1}_i+ \\ 0.107\ \text{Sex}_i+0.055\ \text{Sex:Time3}_i-0.055\ \text{Sex:Time1}_i+\mathbf{e}_i^{(2)}\Big] \quad (8.2)\end{aligned}$$

The LMM framework estimates an absolute subgroup difference. However, the GLMM framework estimates a relative subgroup difference. This will be illustrated through the following examples:

- Referring to Equations 8.1 and 8.2, suppose bulk profiled T cells are expected to have a 1% $R7^+$ RA^+ phenotype at $Time2$ in infants born to mothers with a negative QFT test. According to the GLMM coefficients in Equation 8.1, a $(1\% \times 1.146) = 1.146\%$ $R7^+$ RA^+ phenotype is expected in bulk profiled T cells of infants born to mothers with a positive QFT test. There is a 0.146% difference between subgroups. Referring to the LMM in Equation 8.1, the expected proportion of $Bulk\ \%CD4^+\ R7^+\ RA^+$ T cells in infants born to QFT positive mothers would be $(1 + 1.363)\% = 2.363\%$. There is a 1.363% difference between subgroups.
- Still referring to Equations 8.1 and 8.2, suppose a 10% $R7^+$ RA^+ phenotype is expected at $Time2$ instead. According to the GLMM coefficients in Equation 8.2, for infants born to mothers with a positive QFT test, $(10\% \times 1.146) = 11.46\%$ of bulk profiled $CD4^+$ T cells would have the $R7^+$ RA^+ phenotype. There is now a 1.46% difference between subgroups. Referring to the LMM in Equation 8.1, the expected proportion of $Bulk\ \%CD4^+\ R7^+\ RA^+$ T cells in infants born to QFT positive mothers would be $(10 + 1.363)\% = 11.363\%$. There is still a 1.363% difference between subgroups.

Structure

The outcome selection technique presented in this dissertation utilises model coefficient estimates. These coefficient estimates are a function of the model structure. Initially, the exposures of interest were assumed to have independent effects on immune outcomes. Every longitudinal model described the relationship between the levels of a single exposure variable and a single immune outcome. For every exposure of interest, model coefficient estimates were extracted, a dimension reduction technique was applied, and outcomes were selected for further analysis. After outcome selection, Chapter 7 demonstrates how contrasts estimating subgroup differences can be computed from the corresponding longitudinal models and corrected for multiple hypothesis testing.

However, Chapter 6 shows that the selected immune outcome subset(s) overlap for different exposures of interest. For example, three immune outcomes ($\%GrA^+\ Ki67^+\ NK$, $\%NK^+\ Ki67^+$ and $\%GrA^+\ Ki67^+\ gd$) were selected for further analysis for $MVA85A$, QFT and FA. A contrast estimating subgroup differences based on MVA85A priming could therefore indirectly include subgroup differences based on maternal QFT and combinations of feeding practices and cotrimoxazole treatment. Hence, for selected outcomes, Chapter 7 compares multiplicity-corrected contrasts from the original longitudinal models to estimates from longitudinal models adjusted for additional exposure variables.

Differences were observed when estimating contrasts from longitudinal models adjusted for additional exposures. These differences were often small. For example, consider the longitudinal model describing the relationship between $\%CD8^+\ totTNF^+$ and QFT over time. When dimension reduction is performed by PCA only, this immune outcome is also selected for further analysis in FA over time.

– Assuming that the longitudinal model only describes the relationship between $\%CD8^+$ $totTNF^+$ and QFT over time, $\%CD8^+$ $totTNF^+$ at $Time3$ is estimated to be 1.479 times higher in infants born to mothers with a positive QFT, averaged over the distribution of *Sex*.

– After adjusting the longitudinal model for an additional exposure variable (*FA*), $\%CD8^+$ $totTNF^+$ at $Time3$ is estimated to be 1.450 times higher in infants born to mothers with a positive QFT. There is a ≈2% (multiplicative) difference between subgroups when the estimate is averaged over the distributions of *Sex* and *FA*.

However, occasionally, the differences between contrasts estimated from longitudinal models adjusted for additional exposures were fairly large. As an example, consider the longitudinal model describing the relationship between $\%CD8^+$ $Ki67^+$ and QFT over time. When dimension reduction is performed by PCA only, this immune outcome is also selected for further analysis in *FA* over time.

– Assuming that the longitudinal model only describes the relationship between $\%CD8^+$ $Ki67^+$ and QFT over time, $\%CD8^+$ $Ki67^+$ at $Time2$ is estimated to be 1.933 times higher in infants born to mothers with a positive QFT, averaged over the distribution of *Sex*.

– After adjusting the longitudinal model for an additional exposure variable (*FA*), $\%CD8^+$ $Ki67^+$ at $Time2$ is estimated to be 2.055 times higher in infants born to mothers with a positive QFT. There is a ≈12% (multiplicative) difference between subgroups when the estimate is averaged over the distributions of *Sex* and *FA*.

When performing dimension reduction by HCA followed by PCA, there was no evidence for adjusting the longitudinal model of the relationship between $\%CD8^+$ $Ki67^+$ and QFT for *FA*. As $\%CD8^+$ $Ki67^+$ was only selected for further analysis in QFT over time, $\%CD8^+$ $Ki67^+$ is estimated to be 1.933 times higher (and not 2.055 times higher) at $Time2$ in infants born to mothers with a positive QFT. If longitudinal models are only retrospectively adjusted for additional exposure variables, important adjustments to subgroup differences may be missed and estimates could be misleading.

This is further supported by differences in the number of discoveries ($q < 0.15$ and $q < 0.05$) after correcting contrasts estimated from longitudinal models adjusted for additional exposure variables for multiple hypothesis testing. Several discoveries made when estimating contrasts from unadjusted longitudinal models are no longer detected after adjusting the models for additional exposure variables. For example, consider the contrasts estimated from the longitudinal model describing the relationship between $\%GrA^+$ $Ki67^+$ NK and $MVA85A$ over time. When dimension reduction is performed by PCA only, this immune outcome is also selected for further analysis in QFT and *FA*.

– Assuming that the longitudinal model only describes the relationship between $\%GrA^+$ $Ki67^+$ NK and $MVA85A$ over time, differences in $\%GrA^+$ $Ki67^+$ NK based on MVA85A priming were detected at $Time2$ (BH: $q < 0.15$). $\%GrA^+$ $Ki67^+$ NK is estimated to be 1.414 times higher at $Time2$ in MVA85A-primed infants (BH: $q = 0.143$), averaged over the distribution of *Sex*.

– This difference was no longer detected (BH: $q \geq 0.15$) after the longitudinal models corresponding to the outcome subset were adjusted for *QFT* and *FA*.

– Importantly, the direction of the estimated subgroup difference has also changed after adjusting for additional exposure variables. *%GrA$^+$ Ki67$^+$ NK* is estimated to be 0.719 times higher (or $\approx 28\%$ lower) at *Time*2 in MVA85A primed infants (BH: $q = 0.158$), averaged over the distributions of *Sex*, *QFT* and *FA*.

However, adjusting for additional exposure variables did not consistently reduce the number of discoveries and increase contrast q-values. Two discoveries (BH: $q < 0.15$) were only made after adjusting their respective longitudinal models for additional exposure variables: *Bulk %CD8$^+$ R7$^+$ RA$^+$* at *Time*1 and *%GrB$^+$ Ki67$^+$ CD4$^+$* at *Time*2.

It is also important to note that retrospectively adjusting the original longitudinal models for additional exposure variables creates new models with different coefficient estimates. If outcome selection was based on coefficient estimates extracted from these adjusted longitudinal models, different immune outcomes may have been selected for further analysis.

Finally, the model coefficient estimates and contrasts adjust the exposure's effects for the *average* effect of *Sex* over time. However, the model structure could instead allow *Sex* to modify the exposure's effects over time by specifying a three-way interaction. For example, consider the longitudinal model of the relationship between *QFT* and *%NK$^+$ Ki67$^+$* over time (Equation 4.2).

$$\begin{aligned}\textbf{\%NK}^+\ \textbf{Ki67}^{+(1)}_i \mid \mathbf{x}_i,\ \mathbf{u}_i\ =\ \mathrm{g}^{-1}\Big[&\left(\boldsymbol{\beta}_0^{(1)} + \mathbf{u}_i^{(1)}\right) + \boldsymbol{\beta}_1^{(1)}\mathrm{QFT}_i + \boldsymbol{\beta}_2^{(1)}\mathrm{Time3}_i + \boldsymbol{\beta}_3^{(1)}\mathrm{Time1}_i + \\ &\boldsymbol{\beta}_4^{(1)}\mathrm{QFT{:}Time3}_i + \boldsymbol{\beta}_5^{(1)}\mathrm{QFT{:}Time1}_i + \boldsymbol{\beta}_6^{(1)}\mathrm{Sex}_i + \boldsymbol{\beta}_7^{(1)}\mathrm{Sex{:}Time3}_i + \boldsymbol{\beta}_8^{(1)}\mathrm{Sex{:}Time1}_i + \mathbf{e}_i^{(1)}\Big]\end{aligned}$$

This model adjusts the effects of a positive maternal QFT on **%NK$^+$ Ki67$^+$** for the average effect of *Sex* over time. However, the model could be modified as follows.

$$\begin{aligned}\textbf{\%NK}^+\ \textbf{Ki67}^{+(1)}_i \mid \mathbf{x}_i,\ \mathbf{u}_i\ =\ \mathrm{g}^{-1}\Big[&\left(\boldsymbol{\beta}_0^{(1)} + \mathbf{u}_i^{(1)}\right) + \boldsymbol{\beta}_1^{(1)}\mathrm{QFT}_i + \boldsymbol{\beta}_2^{(1)}\mathrm{Time3}_i + \boldsymbol{\beta}_3^{(1)}\mathrm{Time1}_i + \\ &\boldsymbol{\beta}_4^{(1)}\mathrm{QFT{:}Time3}_i + \boldsymbol{\beta}_5^{(1)}\mathrm{QFT{:}Time1}_i + \boldsymbol{\beta}_6^{(1)}\mathrm{Sex}_i + \boldsymbol{\beta}_7^{(1)}\mathrm{Sex{:}Time3}_i + \boldsymbol{\beta}_8^{(1)}\mathrm{Sex{:}Time1}_i + \\ &\boldsymbol{\beta}_9^{(1)}\mathrm{QFT{:}Sex}_i + \boldsymbol{\beta}_{10}^{(1)}\mathrm{QFT{:}Sex{:}Time3}_i + \boldsymbol{\beta}_{11}^{(1)}\mathrm{QFT{:}Sex{:}Time1}_i + \mathbf{e}_i^{(1)}\Big]\end{aligned}$$

Considering the relationship between *QFT* and *%NK$^+$ Ki67$^+$* over time for fixed *Sex*, there are clear differences in the estimated subgroup differences when the model structure allows *Sex* to modify the effects of the levels of *QFT* on an immune outcome.

Contrast	$\exp(\hat{C})_{(\textbf{Sex*Time})}$	$\exp(\hat{C})_{(\textbf{QFT*Sex*Time})}$
F *Time*2: Positive vs Negative	1.478	1.215
F *Time*1: Positive vs Negative	0.489	0.664
F *Time*3: Positive vs Negative	1.090	0.900

Table 8.2: Summary of differences in estimated contrasts when *Sex* is permitted to modify the effects of a positive maternal QFT on *%NK$^+$ Ki67$^+$*.

If *Sex* is permitted to modify the effects of a positive maternal QFT, there is a smaller estimated difference in *%NK$^+$ Ki67$^+$* for the positive maternal QFT subgroup of female patients.

– There is a $\approx 25\%$ (multiplicative) difference between subgroups at *Time*2.

- There is a $\approx 18\%$ (multiplicative) difference between subgroups at $Time1$.
- There is a $\approx 19\%$ (multiplicative) difference between subgroups at $Time3$.

Hence, if a different relationship was specified between the exposure(s), time, and patient sex, outcome selection would be based on different model coefficient estimates. These examples from the implementation of the outcome selection techniques highlight the importance of determining an appropriate model structure when following a model-based outcome selection approach.

8.2 Dimension Reduction

RM-ASCA+ implements Principal Component Analysis (PCA) as a dimension reduction technique to decompose and visualise the separate effects of exposures over time (Madssen et al., 2021). PCA is applied to identify the most meaningful properties from an input data set of extracted model coefficient estimates describing how the levels of an exposure of interest modify the expected value of an outcome over time. However, when outcome selection is the objective, these properties can be used to select subset(s) of immune outcomes with the most evidence of subgroup differences over time for further analysis. The outcome selection technique presented in this dissertation considers two dimension reduction techniques: (1) PCA only, as implemented in RM-ASCA+, and (2) Hierarchical Cluster Analysis (HCA) followed by PCA.

These techniques take different approaches to dimension reduction:

- The PCA-only approach considers the complete set of immune outcomes. It takes a global approach to dimension reduction that captures the overall variance across all outcomes.
- HCA followed by PCA narrows the focus to subsets of similar outcomes contained within clusters. Dimension reduction, if applied to model coefficient estimates for cluster members, captures localized variance within these groups.

Due to these differences, these techniques offer distinct strategies to identify the subset(s) of immune outcomes with the most evidence of subgroup differences.

- The PCA-only approach treats all immune outcomes equally in outcome selection. For an exposure of interest, a subset of outcomes is selected based on their contributions to the total variance of all standardised model coefficient estimates.
- HCA followed by PCA selects outcomes within clusters. Where the PCA-only approach considers all immune outcomes simultaneously, these groups of similar immune outcomes become units of analysis for an exposure of interest.
    * Clusters containing a few immune outcomes are selected as subsets for further analysis of subgroup differences.
    * PCA was applied to reduce the dimensionality of larger clusters. From this lower-dimensional representation, a subset of immune outcomes is selected based on their contributions to the group-specific variance of standardised model coefficient estimates.

Although these techniques identify many of the same immune outcomes for further analysis, there are some differences in the size and contents of their selected outcome subsets. A larger number of unique immune outcomes are selected for further analysis when outcome selection is performed by HCA followed by PCA. This difference is probably due to its localised approach to dimension reduction.

- Twenty-seven unique immune outcomes are identified for further analysis after performing HCA followed by PCA on the standardised regression coefficients comparing subgroups described by different levels of *MVA85A*, *QFT* and *FA*.
- In comparison, PCA only selects 19 unique immune outcomes.

Considering 33 immune outcomes are measured in the complete set, outcome selection by HCA followed by PCA does not substantially reduce the number of immune outcomes considered in the analysis. Subgroup differences are investigated for 27 immune outcomes. However, these immune outcomes are never compared simultaneously. At most, 11 immune outcomes will be compared simultaneously when computing contrasts comparing different combinations of feeding practices and antibiotic treatment (*FA*). This is less than the maximum number of contrasts ($n_3 = 15$; *FA*) computed simultaneously after outcome selection by PCA only. Fewer simultaneous statistical comparisons may result in less stringent corrections for multiple hypothesis testing. Neither dimension reduction technique selects *%CD4*$^+$ *totIL22*$^+$ or *%CD4*$^+$ *ANYcytok*$^+$ for further analysis. Some immune outcomes selected by HCA followed by PCA are not selected by the PCA-only approach:

- *%Perf*$^+$ *Ki67*$^+$ *NK* for *FA*
- *%GrB*$^+$ *Ki67*$^+$ *NK* for *MVA85A* and *FA*
- *%GrK*$^+$ *Ki67*$^+$ *CD4*$^+$ for *QFT* and *FA*
- *%GrK*$^+$ *Ki67*$^+$ *gd* for *QFT* and *FA*
- *%GrB*$^+$ *Ki67*$^+$ *gd* for *QFT* and *FA*
- *%GrA*$^+$ *Ki67*$^+$ *CD4*$^+$ for *QFT* and *FA*
- *%CD8*$^+$ *ANYcytok*$^+$ for *QFT* and *FA*
- *%CD4*$^+$ *totTNF*$^+$ for *MVA85A*
- *%CD4*$^+$ *totIL17*$^+$ for *QFT* and *FA*
- *%CD4*$^+$ *totIFNg*$^+$ for *MVA85A* and *QFT*

The heat maps summarising the content of the immune outcome subsets identified for further analysis by PCA only and HCA followed by PCA are placed side by side for comparison on the next page. Immune outcomes not selected by the PCA-only approach do not contribute substantially to the total variance in standardised model coefficient estimates. However, if these immune outcomes are selected by HCA followed by PCA, they play an important role in explaining the variance within specific clusters of similar immune outcomes.

Figure 8.1: Comparison of outcome selection by applying PCA only (*left*) and by applying HCA followed by PCA (*right*) for Group A immune outcomes.

Clusters may represent specific biological processes or interactions that are obscured when all outcomes are analysed simultaneously in the PCA-only approach.

- $\%GrB^+$ $Ki67^+$ NK and $\%CD4^+$ $totTNF^+$ belong to the same cluster. These immune outcomes are selected for further analysis with respect to *MVA85A* after performing a PCA on the correlation matrix of standardised model coefficient estimates for the cluster members.
  - The cluster also contains $\%gd$ $Ki67^+$ and $\%CD8^+$ $Ki67^+$.
  - These immune outcomes may represent a coordinated immune response to MVA85A priming, involving an interaction between innate ($\%GrB^+$ $Ki67^+$ NK; $\%gd$ $Ki67^+$) and adaptive ($\%CD4^+$ $totTNF^+$; $\%CD8^+$ $Ki67^+$) components of immunity.
  - Within this coordinated immune response (but not within all outcomes), there are clear subgroup differences in $\%GrB^+$ $Ki67^+$ NK and $\%CD4^+$ $totTNF^+$.
- $\%CD4^+$ $totIL17^+$ and $\%CD4^+$ $totIFNg^+$ belong to the same cluster. These immune outcomes are selected for further analysis with respect to *QFT* after

performing a PCA on the correlation matrix of standardised model coefficient estimates for the cluster members.

  - The cluster also contains *%CD4$^+$ ANYcytok$^+$* and *Bulk %CD8$^+$ R7$^+$ RA$^-$*.
  - These immune outcomes may represent a coordinated immune response to a positive maternal QFT involving an interaction between adaptive components of immunity (*%CD4$^+$ totIL17$^+$*;*%CD4$^+$ totIFNg$^+$*; *%CD4$^+$ ANYcytok$^+$*) and the development of CD8 memory T cells (*Bulk %CD8$^+$ R7$^+$ RA$^-$*).
  - Within this coordinated immune response (but not within all outcomes), there are clear subgroup differences in *%CD4$^+$ totIL17$^+$* and *%CD4$^+$ totIFNg$^+$*.

Across both techniques, *%CD4$^+$ totIL2$^+$*, *Bulk %CD8$^+$ R7$^-$ RA$^+$*, *%gd Ki67$^+$*, *Bulk %CD8$^+$ R7$^+$ RA$^+$* and *%GrA$^+$ Ki67$^+$ NK* are selected for further analysis of the same subgroup differences.

- *%CD4$^+$ totIL2$^+$* is selected for further analysis for *MVA85A* only.
- *Bulk %CD8$^+$ R7$^-$ RA$^+$* is selected for further analysis for *QFT* and *FA*.
- *%gd Ki67$^+$* is selected for further analysis for *QFT* only.
- *Bulk %CD8$^+$ R7$^+$ RA$^+$* is selected for further analysis for *MVA85A* and *FA*.
- *%GrA$^+$ Ki67$^+$ NK* is selected for further analysis for *MVA85A*, *QFT* and *FA*.

These immune outcomes therefore make large contributions to (i) the total variance when all immune outcomes are analysed simultaneously in the PCA-only approach, and (ii) the cluster variance for groups of similar outcomes when applying HCA followed by PCA. There is therefore especially strong evidence for the subgroup differences described above.

In other cases, both techniques select the same immune outcomes for further analysis - but for different exposure variables, for example:

- *%Perf$^+$ Ki67$^+$ CD4$^+$* is selected for further analysis for *QFT* by the PCA-only approach. HCA followed by PCA selects this outcome for further analysis of subgroup differences based on *MVA85A* and *FA*.
- *%NK Ki67$^+$* is selected for *MVA85A*, *QFT* and *FA* by the PCA-only approach and for *QFT* and *FA* only by HCA followed by PCA.
- *%CD4$^+$ Ki67$^+$* is selected for *QFT* by the PCA-only approach and for *FA* by HCA followed by PCA.
- *%GrK$^+$ Ki67$^+$ NK* is selected for *QFT* by PCA only and for *FA* by HCA followed by PCA.
- *Bulk %CD4$^+$ R7$^+$ RA$^+$* is selected for *MVA85A* and *FA* by PCA only and for *QFT* and *FA* by HCA followed by PCA.

This inconsistency reflects the different approaches to dimension reduction. Considering the complete set of outcomes and taking a global approach to dimension reduction (PCA only) will select different outcomes compared to considering groups of similar outcomes and taking a localised approach to dimension reduction (HCA followed by PCA). This does not mean that the outcomes selected are merely artefacts of the selection technique, but rather that different dimension reduction techniques prioritise different characteristics of the data.

8.3 Corrections for Multiple Hypothesis Testing

Different approaches to dimension reduction also define different families of tests to correct for multiple comparisons. In the PCA-only approach, the family of tests includes all immune outcomes that are selected for further analysis based on their contribution to the principal components (PCs). The family of tests is defined by immune outcomes' contributions to the total variance in standardised model coefficient estimates for an exposure of interest.

However, for HCA followed by PCA, the family of tests is cluster-specific. Hypotheses are tested for subgroup differences for outcomes within the same cluster, rather than across the entire data set. The inherent structure of the data, described by the cluster structure, therefore defines the families of tests. In this application, the structure of the data is described in terms of pairwise Pearson correlations. Every family of tests corresponds to a group of highly similar immune outcomes with an average pairwise Pearson correlation of at least 0.85 (*MVA85A* and *QFT*) or at least 0.45 (*FA*). Outcomes outside the cluster are not considered part of the family of tests.

When outcomes are selected by PCA-only dimension reduction, corrections account for a larger number of tests. This potentially corresponds with stricter significance thresholds than HCA followed by PCA. Therefore, it was hypothesised that there would be smaller q-values and more discoveries (for false discovery rates of 5 and 15%) after corrections for multiple hypothesis testing when outcome selection was performed by HCA followed by PCA.

In some cases, the BH q-value is considerably smaller when the FDR is controlled for fewer comparisons. For *MVA85A*, the BH q-value for the subgroup difference in $\%GrB^+$ $Ki67^+$ $CD4^+$ at $Time2$ is much smaller when controlled for 18 comparisons (HCA followed by PCA; $q = 0.072$; Cluster 1) than 81 comparisons (PCA only; $q = 0.125$).

However, a smaller number of simultaneous comparisons did not reliably correspond with smaller q-values after corrections for multiple hypothesis testing:

- *MVA85A*: The BH q-value for the subgroup difference in *Bulk* $\%CD4^+$ $R7^-$ RA^- at $Time2$ is slightly larger when controlled for 18 comparisons (HCA followed by PCA; $q = 0.086$; Cluster 4) than 81 comparisons (PCA only; $q = 0.052$).
- *QFT*: The BY q-value for the subgroup difference in $\%NK^+$ $Ki67^+$ at $Time1$ is slightly larger when controlled for 63 comparisons (HCA followed by PCA: $q = 0.057$) than 99 comparisons (PCA only: $q = 0.047$).

In addition, using a cluster-based definition of a family of tests did not reliably correspond to more discoveries (for false discovery rates of 5 and 15%). The table below summarises the number of discoveries obtained after correcting outcomes selected by PCA only for multiple comparisons.

Table 8.3: Number of discoveries ($q < 0.15$; $q < 0.05$) for outcome selection by PCA only, correcting estimated contrasts for outcomes for multiple comparisons.

Exposure Variable	$q < 0.15$ (BH)	$q < 0.05$ (BH)	$q < 0.05$ (BY)
$MVA85A$	4	0	0
QFT	5	1	1
FA	16	4	3

The table below summarises the number of discoveries obtained after correcting outcomes selected by HCA followed by PCA for multiple comparisons.

Table 8.4: Number of discoveries ($q < 0.15$; $q < 0.05$) for outcome selection by HCA followed by PCA, correcting estimated contrasts for outcomes for multiple comparisons.

Exposure Variable	$q < 0.15$ (BH)	$q < 0.05$ (BH)	$q < 0.05$ (BY)
$MVA85A$	6	2	1
QFT	3	2	0
FA	16	6	2

There is little difference in the number of *FA* subgroup differences discovered after PCA-only and HCA followed by PCA outcome selection. There are fewer discoveries of QFT subgroup differences after outcome selection by HCA followed by HCA than by PCA only. This includes discoveries at a liberal significance threshold ($q < 0.15$), assuming independence or positive dependence within the family of tests.

Assuming independence or positive dependence within the family of tests (BH), there are more discoveries of *MVA85A* subgroup differences after PCA-only outcome selection. This includes discoveries at a liberal significance threshold ($q < 0.15$) and at the conventional statistical significance threshold ($q < 0.05$). Assuming arbitrary dependence (BY), there are no statistically significant discoveries after PCA-only outcome selection, but a single discovery for outcome selection by HCA followed by PCA.

Adjusting the models to account for additional exposures reliably corresponded to an increased number of discoveries. Assuming independence or positive dependence within the family of tests, there are six more discoveries at the conventional statistical significance threshold ($q < 0.05$) after adjusting the longitudinal *FA* models for additional exposure variables after PCA-only outcome selection.

Table 8.5: Number of discoveries ($q < 0.15$; $q < 0.05$) for outcome selection by PCA only, after adjusting longitudinal models for additional exposures.

Exposure Variable	$q < 0.15$ (BH)	$q < 0.05$ (BH)	$q < 0.05$ (BY)
Exposure	$q < 0.15$ (BH)	$q < 0.05$ (BH)	$q < 0.05$ (BY)
$MVA85A$	5	0	0
QFT	4	1	1
FA	16	10	3

Similarly, assuming independence or positive dependence among the family of tests, there are 3 more discoveries at the conventional statistical significance threshold ($q < 0.05$) after adjusting the longitudinal *FA* models for additional exposure variables after HCA followed by PCA outcome selection.

Table 8.6: Number of discoveries ($q < 0.15$; $q < 0.05$) for outcome selection by HCA followed by PCA, after adjusting longitudinal models for additional exposures.

Exposure Variable	$q < 0.15$ (BH)	$q < 0.05$ (BH)	$q < 0.05$ (BY)
Exposure	$q < 0.15$ (BH)	$q < 0.05$ (BH)	$q < 0.05$ (BY)
$MVA85A$	6	0	0
QFT	4	1	1
FA	19	9	4

Once again, this highlights the importance of determining an appropriate model structure when following a model-based outcome selection approach.

Different methods selected different outcomes for further analysis. Similarly, different methods give different estimates of the statistical significance of the same discoveries.

- For example, assuming arbitrary dependence (BH) among the family of tests, the contrast q-value for the *QFT* subgroup difference in $\%NK^+$ $Ki67^+$ at the time of BCG vaccination is $q = 0.047$ (PCA only). On the other hand, the q-value is slightly larger ($q = 0.057$) when cluster-based corrections are applied to the contrast q-values (HCA followed by PCA).

The family of tests is defined differently by the two approaches to dimension reduction.

- In the PCA-only approach, the family of tests is defined by immune outcomes' contributions to the total variance in standardised model coefficient estimates for *QFT*. FDR is controlled for 99 comparisons.
- For HCA followed by PCA, the family of tests is defined by the cluster structure. $\%NK^+$ $Ki67^+$ is contained in Cluster 6, along with six other immune outcomes. FDR is controlled for 66 comparisons.

The definition of the family of tests is perhaps more important than reducing the number of simultaneous comparisons. For HCA followed by PCA, corrections are applied only to the outcomes within the cluster (or its lower-dimensional representation, if PCA is applied). Probability statements about subgroup differences are corrected *only* for immune outcomes that behave similarly with respect to the exposure of interest. This may increase the statistical power to detect subgroup differences within groups of closely related outcomes, but not the complete set of outcomes. However, this approach is reasonable only if the clusters themselves are biologically meaningful.

This application also highlights that different assumptions about the dependence within the family of hypothesis tests may lead to very different insights. Many discoveries that would be considered statistically significant at $q < 0.05$ under independence or positive dependence (Benjamini-Hochberg) would not be considered significant under arbitrary dependence (Benjamini-Yekutieli). In immunological research where negative dependencies are plausible within a family of tests, the dependence assumptions of the chosen multiple comparison correction procedure should be carefully considered.

Chapter 9

Discussion and Conclusion

Immunological research often compares subgroups that are defined by exposure variables known or hypothesised to influence continuous immune responses. As many immune responses are measured over time, effective outcome selection ensures that immunological research focuses on the immune outcomes with the strongest signals for subgroup differences. By restricting the scope of the analysis to the subset(s) of immune outcomes identified by outcome selection, a smaller number of statistical hypotheses are investigated. In doing so, the analysis preserves statistical power when stricter significance thresholds are imposed to account for type-I error inflation. Furthermore, analysing fewer immune outcomes makes it easier to interpret and discuss research findings.

As variable selection is a well-established research area, it is convenient to restructure outcome selection problems to fit within the variable selection framework. In other words, the outcomes are inputs to the selection problem, and the outcomes are treated as predictive of the levels of the exposure of interest. This structure changes the direction of causality. If inference is the objective of the analysis, it is important to develop specific outcome selection techniques that do not require this restructuring. This dissertation targeted this previously unaddressed methodological challenge.

The RM-ASCA+ framework, although not explicitly an outcome selection technique, showed promise for outcome selection in longitudinal immunological data. RM-ASCA+ combines linear mixed-effect models (LMM) with principal component analysis (PCA) to decompose and visualise the separate effects of experimental factors over time on continuous outcomes (Madssen et al., 2021).

9.1 Key Findings and Methodological Insights

This dissertation showed that RM-ASCA+, as currently implemented in *R*, was not directly applicable for outcome selection in longitudinal immunological data. Compared to the LMM framework implemented in RM-ASCA+, the generalized linear mixed-effect model (GLMM) framework was demonstrated to be more appropriate for outcomes with skew, long-tailed distributions. These characteristics are common in immunological data.

9.1.1 Modelling Frameworks

Model assumptions were better met within the GLMM framework. In particular, the LMM random-effect distributions deviated more strongly from normality assumptions. This is an important consideration: mixed-effect model estimates, which are the inputs to dimension reduction in RM-ASCA+, are conditional on the random effects. RM-ASCA+, as currently implemented in *R* (Jarmund et al., 2022), does not store the random-effect estimates for assumption checking. For a given immune outcome, model coefficient estimates extracted from the LMM may be considerably different from estimates extracted from the GLMM. As an example, Chapter

8 compares estimated *QFT* subgroup differences in *Bulk* $\%CD4^+$ $R7^+$ RA^+ from the LMM and GLMM. The largest subgroup difference is identified at different time points. This difference between modelling frameworks could be a consequence of violated LMM assumptions, but these violations would not have been visible using the current R implementation of RM-ASCA+.

Additionally, the LMM framework implemented in RM-ASCA+ specifies a linear relationship between the conditional distribution of an immune outcome $\mathbf{Y}$ and $\mathbf{X}\boldsymbol{\beta}$. The model coefficient estimates, which are the inputs to dimension reduction for outcome selection, have an additive interpretation and describe an absolute subgroup difference. Within a GLMM, one can specify a log-link function describing a generalised linear relationship between $\mathbf{Y}$ and $\mathbf{X}\boldsymbol{\beta}$. GLMM coefficients have a multiplicative interpretation. Basing outcome selection on multiplicative model coefficients is more intuitive, meaningful and possibly more biologically appropriate. Basing outcome selection on relative subgroup differences offers a number of advantages over basing outcome selection on absolute subgroup differences:

1. An absolute subgroup difference is only meaningful in context. For example, an expected subgroup difference of 1.363% has different implications for a baseline proportion of 1% or 50%. A relative subgroup difference (e.g., a multiplicative difference of 1.323) applies equally no matter the starting point.
2. Many biological processes are multiplicative. Estimating absolute subgroup differences may distort relationships between exposures of interest and immune outcomes.

RM-ASCA+ aims to decompose and visualise the separate effects of experimental factors on continuous outcomes over time (Madssen et al., 2021). In theory, this information could be used to select smaller sets of immune outcomes for which there is strong evidence of subgroup differences. However, the current *R* implementation of RM-ASCA+ is not well-suited to the characteristics of immunological data and the nature of biological processes.

9.1.2 Dimension Reduction

In both RM-ASCA+ and the outcome selection technique presented in this dissertation, model coefficients for the exposure comparisons of interest are extracted from longitudinal models of each immune outcome and supplied as inputs to a dimension reduction technique. Two dimension reduction techniques were considered in this dissertation: first, PCA (as implemented in RM-ASCA+), and second, agglomerative hierarchical cluster analysis (HCA) followed by PCA. The techniques offer distinct strategies to select the subset(s) of immune outcomes with the most evidence of subgroup differences over time for further analysis. The PCA-only approach considers the complete set of immune outcomes. It takes a global approach to dimension reduction that captures the overall variance across all outcomes. HCA followed by PCA narrows the focus to subsets of similar outcomes contained within clusters. Dimension reduction, if applied to model coefficient estimates for cluster members, captures localised variance within these groups.

Due to the localised approach to dimension reduction in HCA followed by PCA, it was hypothesised that there would be more immune outcomes selected for further analysis

when following this method. As expected, this was observed during implementation. However, the two dimension reduction techniques selected different immune outcomes for further analysis (often for different exposure variables). This does not mean that the outcomes selected are merely artefacts of the selection technique, but rather that different dimension reduction techniques prioritise different characteristics of the data. The chosen dimension reduction technique should depend on the objectives of the analysis.

PCA only, as implemented in RM-ASCA+, takes a global approach to dimension reduction. All immune outcomes are treated equally in outcome selection. For an exposure of interest, a subset of outcomes is selected based on their contributions to the total variance of all standardised model coefficient estimates. On the other hand, HCA followed by PCA takes a localised approach to dimension reduction. Where the PCA-only approach considers all immune outcomes simultaneously, outcomes are selected from groups of similar immune outcomes. This is analogous to identifying a representative object for the cluster, but the focus is identifying one or more outcomes with the strongest signal for subgroup differences within a given cluster. However, this localised approach to dimension reduction only makes sense if the biological meaning of a cluster is clear. Clusters may represent specific biological processes or interactions that are obscured when all outcomes are analysed simultaneously in the PCA-only approach. However, they may also represent spurious groupings and warrant careful interpretation to distinguish biologically meaningful patterns from artefacts of the clustering process.

For this reason, dimension reduction by PCA only (as implemented in RM-ASCA+) is more broadly applicable for outcome selection. It does not rely on prior assumptions about the underlying relationships between immune outcomes (which may be unknown or poorly defined). Instead, dimension reduction captures the largest sources of variance in standardised model coefficients across all immune outcomes. This provides a concise representation of the primary drivers of variation, albeit at the expense of capturing more granular, local relationships that could be biologically significant.

Corrections for Multiple Hypothesis Testing

In addition to providing distinct strategies for selecting the subset(s) of immune outcomes with the strongest evidence of subgroup differences over time, the two dimension reduction techniques also influence how the family of tests is defined for multiple hypothesis testing corrections. In the PCA-only approach, the family of tests is defined by immune outcomes' contributions to the total variance in standardised model coefficient estimates for an exposure of interest. For HCA followed by PCA, the family of tests is defined by the cluster structure.

Different approaches to dimension reduction also define different families of tests to correct for multiple comparisons. In the PCA-only approach, the family of tests includes all immune outcomes that are selected for further analysis based on their contribution to the principal components (PCs). The family of tests is defined by immune outcomes' contributions to the total variance in standardised model coefficient estimates for an exposure of interest. However, for HCA followed by PCA, the family of tests is cluster-specific. This approach can be advantageous if clusters reflect biologically meaningful groupings. However, if a cluster represents a spurious

grouping of outcomes, the false discovery rate is controlled for the wrong comparisons, potentially leading to misleading conclusions.

In biomedical research, it is scientific convention to apply corrections for multiple comparisons before reporting results. Hence, there is a strong emphasis on restricting the number of outcomes compared within a study. This approach is driven by the goal of increasing the ability to detect meaningful differences when they truly exist, without inflating the rate of false positives due to multiple comparisons. When outcomes are selected by PCA-only dimension reduction, corrections account for a larger number of tests. Therefore, it was hypothesised that there would be smaller q-values and more discoveries (for false discovery rates of 5 and 15%) after corrections for multiple hypothesis testing when outcome selection was performed by HCA followed by PCA. However, a smaller number of simultaneous comparisons did not reliably correspond with smaller q-values and more discoveries after corrections for multiple hypothesis testing.

This highlights the importance of defining the family of tests appropriately, rather than aiming to reducing the number of simultaneous comparisons. For HCA followed by PCA, corrections are applied only to the outcomes within the cluster (or its lower-dimensional representation, if PCA is applied). Probability statements about subgroup differences are corrected *only* for immune outcomes that behave similarly with respect to the exposure of interest. This may increase the statistical power to detect subgroup differences within groups of closely related outcomes but not the complete set of outcomes. Once again, dimension reduction by PCA only (as implemented in RM-ASCA+) is shown to be more broadly applicable for outcome selection.

9.2 Limitations

The outcome selection technique presented in this dissertation depends heavily on the model specification and the assumptions underlying the longitudinal modelling frameworks. If the model inadequately describes the relationship between the exposure of interest and the immune outcome over time, the outcome selection results may be biased or misleading. While dividing the model coefficient estimates by their standard errors before dimension reduction accounts for uncertainty in the estimates, this step cannot compensate for an incorrectly specified model. Striking a balance between a model structure that is broadly applicable to all immune outcomes and one that is not overly general, while ensuring the underlying assumptions are met, remains a significant challenge.

The methods presented here are tailored to the characteristics of this specific data set. It is important to investigate whether they can be applied to other immunological datasets produced by different experimental designs. The technique is limited to selection of continuous outcomes based on their association with one or more exposure variables (factors). It is not applicable to binary or categorical outcomes, or to continuous independent variables. Finally, outcome selection is driven by estimates extracted from many univariate models when multivariate relationships are likely to exist among immune outcomes. Although this simplification makes outcome selection more computationally feasible, it could lead to biased or incomplete interpretations.

9.3 Recommendations for Future Research

9.3.1 Outcome Selection Methodology

If a GLMM-based implementation of RM-ASCA+ were developed, the RM-ASCA+ framework could be repurposed for outcome selection in longitudinal immunological data. A compound Poisson-Gamma (Tweedie) conditional distribution was specified for most ($n = 23$; $\approx 70\%$) of the 33 immune outcomes analysed, followed by a Gamma conditional distribution ($n = 9$; $\approx 27\%$). A Negative binomial conditional distribution was estimated for a single immune outcome ($\%GrA^+$ $Ki67^+$ $CD4^+$). The conditional distributions specified for every immune outcome are summarised in Appendix E.

For convenience, a compound Poisson-Gamma conditional distribution *could* have been specified for immune outcomes with a Gamma conditional distribution. The power parameter ($0 < p < 1$) weights the contributions of the discrete Poisson component and the continuous Gamma component of the compound distribution. A power parameter very close to 1 would approximate a Gamma conditional distribution and provide a suitable fit for the generalised linear relationship between $\mathbf{Y}$ and $\mathbf{X}\boldsymbol{\beta}$. Hence, an *R* package implementation of RM-ASCA+ - specifying a compound Poisson-Gamma conditional distribution with a log-link function - would position the RM-ASCA+ framework as a tool for outcome selection in immunological data.

9.3.2 Model Structure for Outcome Selection

Above all else, this dissertation highlighted the importance of model structure. Before extracting model coefficient estimates of comparisons of interest, the longitudinal model should appropriately describe the relationship between exposure variables, time, and other covariates. Misspecification can lead to biased model coefficient estimates. As outcome selection is based on model coefficient estimates extracted from the original longitudinal models, it is not advisable to retrospectively adjust models for additional exposure variables or more complex relationships between the exposure(s), time, and other covariates.

If this analysis were redesigned, the longitudinal model structure would adjust for the average effects of the other exposure variables, even if the data itself does not suggest that the adjustment is necessary. For the m^{th} immune outcome in $y_{1A}...y_{33A}$ measured in Group A, (G)LMMs modelling the relationship between the *EOI* for patient i would have the following form:

$$\begin{aligned}\mathbf{y}_{mi}^{(1)} \mid \mathbf{x_i},\, \mathbf{u_i} \;=\; \mathrm{g}^{-1}\Big[&\Big(\boldsymbol{\beta}_0^{(1)} + \mathbf{u}_i^{(1)}\Big) + \boldsymbol{\beta}_1^{(1)}\mathrm{MVA85A}_i + \boldsymbol{\beta}_2^{(1)}\mathrm{Time3}_i + \boldsymbol{\beta}_3^{(1)}\mathrm{Time1}_i + \\ &\boldsymbol{\beta}_4^{(1)}\mathrm{MVA85A{:}Time3}_i + \boldsymbol{\beta}_5^{(1)}\mathrm{MVA85A{:}Time1}_i + \\ \boldsymbol{\beta}_6^{(1)}\mathrm{Sex}_i + \boldsymbol{\beta}_7^{(1)}&\mathrm{Sex{:}Time3}_i + \boldsymbol{\beta}_8^{(1)}\mathrm{Sex{:}Time1}_i + \boldsymbol{\beta}_9^{(1)}\mathrm{QFT}_i + \beta_{10}^{(2)}\mathrm{FA1}_i + \beta_{11}^{(2)}\mathrm{FA2}_i + \beta_{12}^{(2)}\mathrm{FA3}_i + \mathbf{e}_i^{(1)}\Big]\end{aligned} \tag{9.1}$$

$$\mathbf{y}_{mi}^{(2)} \mid \mathbf{x}_i,\ \mathbf{u}_i \ =\ \mathrm{g}^{-1}\left[\left(\boldsymbol{\beta}_0^{(2)} + \mathbf{u}_i^{(2)}\right) + \boldsymbol{\beta}_1^{(2)}\mathrm{QFT}_i + \boldsymbol{\beta}_2^{(2)}\mathrm{Time3}_i + \boldsymbol{\beta}_3^{(2)}\mathrm{Time1}_i + \boldsymbol{\beta}_4^{(2)}\mathrm{QFT{:}Time3}_i + \boldsymbol{\beta}_5^{(2)}\mathrm{QFT{:}Time1}_i + \boldsymbol{\beta}_6^{(2)}\mathrm{Sex}_i + \boldsymbol{\beta}_7^{(2)}\mathrm{Sex{:}Time3}_i + \boldsymbol{\beta}_8^{(2)}\mathrm{Sex{:}Time1}_i + \boldsymbol{\beta}_9^{(2)}\mathrm{MVA85A}_i + \boldsymbol{\beta}_{10}^{(2)}\mathrm{FA1}_i + \boldsymbol{\beta}_{11}^{(2)}\mathrm{FA2}_i + \boldsymbol{\beta}_{12}^{(2)}\mathrm{FA3}_i + \mathbf{e}_i^{(2)}\right] \tag{9.2}$$

$$\mathbf{y}_{mi}^{(3)} \mid \mathbf{x}_i,\ \mathbf{u}_i \ =\ \mathrm{g}^{-1}\left[\left(\boldsymbol{\beta}_0^{(3)} + \mathbf{u}_i^{(3)}\right) + \boldsymbol{\beta}_1^{(3)}\mathrm{FA3}_i + \boldsymbol{\beta}_2^{(3)}\mathrm{FA2}_i + \boldsymbol{\beta}_3^{(3)}\mathrm{FA1}_i + \boldsymbol{\beta}_4^{(3)}\mathrm{Time3}_i + \boldsymbol{\beta}_5^{(3)}\mathrm{Visit1}_i + \boldsymbol{\beta}_6^{(3)}\mathrm{FA3{:}Time3}_i + \boldsymbol{\beta}_7^{(3)}\mathrm{FA2{:}Time3}_i + \boldsymbol{\beta}_8^{(3)}\mathrm{FA1{:}Time3}_i + \boldsymbol{\beta}_9^{(3)}\mathrm{FA3{:}Time1}_i + \boldsymbol{\beta}_{10}^{(3)}\mathrm{FA2{:}Time1}_i + \boldsymbol{\beta}_{11}^{(3)}\mathrm{FA1{:}Time1}_i + \boldsymbol{\beta}_{12}^{(3)}\mathrm{Sex}_i + \boldsymbol{\beta}_{13}^{(3)}\mathrm{Sex{:}Time3}_i + \boldsymbol{\beta}_{14}^{(3)}\mathrm{Sex{:}Time1}_i + \boldsymbol{\beta}_{15}^{(3)}\mathrm{MVA85A}_i + \boldsymbol{\beta}_{16}^{(3)}\mathrm{QFT}_i + \mathbf{e}_i^{(3)}\right] \tag{9.3}$$

$$\mathbf{y}_{mi}^{(4)} \mid \mathbf{x}_i,\ \mathbf{u}_i \ =\ \mathrm{g}^{-1}\left[\left(\boldsymbol{\beta}_0^{(4)} + \mathbf{u}_i^{(4)}\right) + \boldsymbol{\beta}_1^{(4)}\mathrm{FA2}_i + \boldsymbol{\beta}_2^{(4)}\mathrm{FA1}_i + \boldsymbol{\beta}_3^{(4)}\mathrm{FA0}_i + \boldsymbol{\beta}_4^{(4)}\mathrm{Time3}_i + \boldsymbol{\beta}_5^{(4)}\mathrm{Visit1}_i + \boldsymbol{\beta}_6^{(4)}\mathrm{FA2{:}Time3}_i + \boldsymbol{\beta}_7^{(4)}\mathrm{FA1{:}Time3}_i + \boldsymbol{\beta}_8^{(4)}\mathrm{FA0{:}Time3}_i + \boldsymbol{\beta}_9^{(4)}\mathrm{FA2{:}Time1}_i + \boldsymbol{\beta}_{10}^{(4)}\mathrm{FA1{:}Time1}_i + \boldsymbol{\beta}_{11}^{(4)}\mathrm{FA0{:}Time1}_i + \boldsymbol{\beta}_{12}^{(4)}\mathrm{Sex}_i + \boldsymbol{\beta}_{13}^{(4)}\mathrm{Sex{:}Time3}_i + \boldsymbol{\beta}_{14}^{(4)}\mathrm{Sex{:}Time1}_i + \boldsymbol{\beta}_{15}^{(4)}\mathrm{MVA85A}_i + \boldsymbol{\beta}_{16}^{(4)}\mathrm{QFT}_i + \mathbf{e}_i^{(4)}\right] \tag{9.4}$$

As a minimum, these adjustments are necessary to achieve SATVI's objectives for outcome selection. Without adjusting the comparisons of interest for the average effects of other exposure variables, it is impossible to disentangle subgroup differences in immune outcomes for *MVA85A*, *QFT* and *FA*. After these adjustments, it will be possible to extract model coefficient estimates of comparisons of interest. By performing dimension reduction by PCA on these adjusted coefficient estimates, it will be possible to conclusively select immune outcomes for further analysis. Contrasts will be computed from the longitudinal models corresponding with these immune outcome subsets and corrected for multiple hypothesis testing using the approach demonstrated in this dissertation.

9.3.3 Defining the Family of Tests

In exploratory research, such as SATVI's intended analysis of the dissertation dataset, the concept of defining a family of tests often becomes challenging or impractical. The family of tests may not be easily or meaningfully defined, and adjustments for multiple comparisons may not offer a practical control of the false discovery rate (FDR). When the structure of the family of tests lacks clear justification, reporting exact p-values for comparisons of interest may be more prudent.

Additionally, different assumptions about the dependence within the family of hypothesis tests may lead to very different insights. Many discoveries that would be

considered statistically significant at $q < 0.05$ under independence or positive dependence (Benjamini-Hochberg) would not be considered significant under arbitrary dependence (Benjamini-Yekutieli). In immunological research where negative dependencies are plausible within a family of tests, the dependence assumptions of the chosen multiple comparison correction procedure should be carefully considered.

References

Arboretti, R., Bathke, A. C., Carrozzo, E., Pesarin, F., & Salmaso, L. (2020). Multivariate permutation tests for two sample testing in presence of nondetects with application to microarray data. *Statistical Methods in Medical Research*, *29*(1), 258–271.

Baumgartner, C., Osl, M., Netzer, M., & Baumgartner, D. (2011). Bioinformatic-driven search for metabolic biomarkers in disease. *Journal of clinical bioinformatics*, *1*, 1–10.

Bender, R., & Lange, S. (2001). Adjusting for multiple testing—when and how? *Journal of clinical epidemiology*, *54*(4), 343–349.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289–300.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, 1165–1188.

Bourke, C. D., Gough, E. K., Pimundu, G., Shonhai, A., Berejena, C., Terry, L., Baumard, L., Choudhry, N., Karmali, Y., Bwakura-Dangarembizi, M., et al. (2019). Cotrimoxazole reduces systemic inflammation in HIV infection by altering the gut microbiome and immune activation. *Science translational medicine*, *11*(486), eaav0537.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400.

Cadena, A. M., Fortune, S. M., & Flynn, J. L. (2017). Heterogeneity in tuberculosis. *Nature Reviews Immunology*, *17*(11), 691–702.

Choi, G., Buckley, J. P., Kuiper, J. R., & Keil, A. P. (2022). Log-transformation of independent variables: Must we? *Epidemiology*, *33*(6), 843–853.

Chow, S.-C., & Liu, J.-P. (2013). *Design and analysis of clinical trials: Concepts and methodologies*. John Wiley & Sons.

Demidenko, E. (2013). *Mixed models: Theory and applications with R*. John Wiley & Sons.

Dias-Domingues, T., Mouriño, H., & Sepúlveda, N. (2024). Classification methods for the serological status based on mixtures of skew-normal and skew-t distributions. *Mathematics*, *12*(2), 217.

Erdős, B., Westerhuis, J. A., Adriaens, M. E., O'Donovan, S. D., Xie, R., Singh-Povel, C. M., Smilde, A. K., & Arts, I. C. W. (2023). Analysis of high-

dimensional metabolomics data with complex temporal dynamics using RM-ASCA+. *PLOS Computational Biology*, *19*, e1011221.

Feise, R. J. (2002). Do multiple outcome measures require p-value adjustment? *BMC medical research methodology*, *2*, 1–4.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied longitudinal analysis*. John Wiley & Sons.

Forlin, R., James, A., & Brodin, P. (2023). Making human immune systems more interpretable through systems immunology. *Trends in immunology*.

Genser, B., Cooper, P. J., Yazdanbakhsh, M., Barreto, M. L., & Rodrigues, L. C. (2007). A guide to modern statistical analysis of immunological data. *BMC immunology*, *8*, 1–15.

Gewandter, J. S., McDermott, M. P., He, H., Gao, S., Cai, X., Farrar, J. T., Katz, N. P., Markman, J. D., Senn, S., Turk, D. C., et al. (2019). Demonstrating heterogeneity of treatment effects among patients: An overlooked but important step toward precision medicine. *Clinical Pharmacology & Therapeutics*, *106*(1), 204–210.

Goeman, J. J., & Solari, A. (2014). Multiple hypothesis testing in genomics. *Statistics in medicine*, *33*(11), 1946–1978.

Grissa, D., Pétéra, M., Brandolini, M., Napoli, A., Comte, B., & Pujos-Guillot, E. (2016). Feature selection methods for early predictive biomarker discovery using untargeted metabolomic data. *Frontiers in molecular biosciences*, *3*, 30.

Harrell, F. E., et al. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis* (Vol. 608). Springer.

Hartig, F. (2018). Dharma: Residual diagnostics for hierarchical (multi-level/mixed) regression models. *R Packag version 020*.

Haver, A. (2021). Simulation study for zero inflated count RM-ASCA+. https://github.com/AukeHaver/ZICRM-ASCA_plus

Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection–a review and recommendations for the practicing statistician. *Biometrical journal*, *60*(3), 431–449.

Huang, Y., Chen, J., & Yin, P. (2017). Hierarchical mixture models for longitudinal immunologic data with heterogeneity, non-normality, and missingness. *Statistical methods in medical research*, *26*(1), 223–247.

Jarmund, A. H., Madssen, T. S., & Giskeødegård, G. F. (2022). Alasca: An r package for longitudinal and cross-sectional analysis of multivariate data by ASCA-based methods. *Frontiers in Molecular Biosciences*, *9*, 962431.

Lapham, B. M. (2020). Modelling multivariate nonlinear vaccine induced immune responses.

Lenth, R. V. (2024). *Emmeans: Estimated marginal means, aka least-squares means* [R package version 1.10.5-0900001, https://rvlenth.github.io/emmeans/]. https://rvlenth.github.io/emmeans/

Li, L., & Liu, Z.-P. (2021). Detecting prognostic biomarkers of breast cancer by regularized cox proportional hazards models. *Journal of translational medicine*, *19*, 1–20.

Lin, T.-I., & Wang, W.-L. (2013). Multivariate skew-normal at linear mixed models for multi-outcome longitudinal data. *Statistical Modelling*, *13*(3), 199–221.

Liu, L., Shih, Y.-C. T., Strawderman, R. L., Zhang, D., Johnson, B. A., & Chai, H. (2019). Statistical analysis of zero-inflated nonnegative continuous data. *Statistical Science*, *34*(2), 253–279.

Lubyayi, L., Mawa, P. A., Nabakooza, G., Nakibuule, M., Tushabe, J. V., Serubanja, J., Aibo, D., Akurut, H., Tumusiime, J., Hasso-Agopsowicz, M., et al. (2020). Maternal latent mycobacterium tuberculosis does not affect the infant immune response following bcg at birth: An observational longitudinal study in uganda. *Frontiers in Immunology*, *11*, 929.

Luckett, P. H., Chen, C., Gordon, B. A., Wisch, J., Berman, S. B., Chhatwal, J. P., Cruchaga, C., Fagan, A. M., Farlow, M. R., Fox, N. C., et al. (2023). Biomarker clustering in autosomal dominant alzheimer's disease. *Alzheimer's & Dementia*, *19*(1), 274–284.

Madssen, T. S., Giskeødegård, G. F., Smilde, A. K., & Westerhuis, J. A. (2021). Repeated measures ASCA+ for analysis of longitudinal intervention studies with multivariate outcome data. *PLoS Computational Biology*, *17*(11), e1009585.

Maghsoudloo, M., Jamalkandi, S. A., Najafi, A., & Masoudi-Nejad, A. (2020). An efficient hybrid feature selection method to identify potential biomarkers in common chronic lung inflammatory diseases. *Genomics*, *112*(5), 3284–3293.

Mahanty, S., Prigent, A., & Garraud, O. (2015). Immunogenicity of infectious pathogens and vaccine antigens. *BMC immunology*, *16*, 1–6.

Martin-Escolano, R., Vidal-Alcantara, E. J., Crespo, J., Ryan, P., Real, L. M., Lazo-Alvarez, J. I., Cabezas-Gonzalez, J., Macias, J., Arias-Loste, M. T., Cuevas, G., et al. (2023). Immunological and senescence biomarker profiles in patients after spontaneous clearance of hepatitis c virus: Gender implications for long-term health risk. *Immunity & Ageing*, *20*(1), 62.

Maugeri, A., Barchitta, M., Basile, G., & Agodi, A. (2021). Applying a hierarchical clustering on principal components approach to identify different patterns of the SARS-CoV-2 epidemic across italian regions. *Scientific reports*, *11*(1), 7082.

Metwally, A. A., Zhang, T., Wu, S., Kellogg, R., Zhou, W., Contrepois, K., Tang, H., & Snyder, M. (2022). Robust identification of temporal biomarkers in longitudinal omics studies. *Bioinformatics*, *38*(15), 3802–3811.

Murtagh, F., & Contreras, P. (2017). Algorithms for hierarchical clustering: An overview, ii. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *7*(6), e1219.

Nemes, E., Hesseling, A. C., Tameris, M., Mauff, K., Downing, K., Mulenga, H., Rose, P., van der Zalm, M., Mbaba, S., Van As, D., et al. (2018). Safety and immunogenicity of newborn mva85a vaccination and selective, delayed bacille calmette-guerin for infants of human immunodeficiency virus-infected mothers: A phase 2 randomized, controlled trial. *Clinical Infectious Diseases*, *66*(4), 554–563.

Pineda, C., Castañeda Hernández, G., Jacobs, I. A., Alvarez, D. F., & Carini, C. (2016). Assessing the immunogenicity of biopharmaceuticals. *BioDrugs*, *30*, 195–206.

Pinheiro, J., & Bates, D. (2006). *Mixed-effects models in s and s-plus*. Springer science & business media.

Rahman, M. J., Degano, I. R., Singh, M., & Fernandez, C. (2010). Influence of maternal gestational treatment with mycobacterial antigens on postnatal immunity in an experimental murine model. *PLoS One*, *5*(3), e9699.

Reisetter, A. C., & Breheny, P. (2021). Penalized linear mixed models for structured genetic data. *Genetic epidemiology*, *45*(5), 427–444.

Remeseiro, B., & Bolon-Canedo, V. (2019). A review of feature selection methods in medical applications. *Computers in biology and medicine*, *112*, 103375.

Rojo, J. (2013). Heavy-tailed densities. *Wiley Interdisciplinary Reviews: Computational Statistics*, *5*(1), 30–40.

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in ecology and evolution*, *11*(9), 1141–1152.

Solorio-Fernandez, S., Carrasco-Ochoa, J. A., & Martinez-Trinidad, J. F. (2020). A review of unsupervised feature selection methods. *Artificial Intelligence Review*, *53*(2), 907–948.

St. Clair, L. A., Chaulagain, S., Klein, S. L., Benn, C. S., & Flanagan, K. L. (2023). Sex-differential and non-specific effects of vaccines over the life course. *Sex and Gender Differences in Infection and Treatments for Infectious Diseases*, 225–251.

Stevens, J. R., Al Masud, A., & Suyundikov, A. (2017). A comparison of multiple testing adjustment methods with block-correlation positively-dependent tests. *Plos one*, *12*(4), e0176124.

Uh, H.-W., Hartgers, F. C., Yazdanbakhsh, M., & Houwing-Duistermaat, J. J. (2008). Evaluation of regression methods when immunological measurements are constrained by detection limits. *BMC immunology*, *9*, 1–10.

Vach, W. (2012). *Regression models as a tool in medical research*. CRC Press.

Victora, C. G., Bahl, R., Barros, A. J., França, G. V., Horton, S., Krasevec, J., Murch, S., Sankar, M. J., Walker, N., & Rollins, N. C. (2016). Breastfeeding in the 21st century: Epidemiology, mechanisms, and lifelong effect. *The lancet*, *387*(10017), 475–490.

Vigano, S., Perreau, M., Pantaleo, G., & Harari, A. (2012). Positive and negative regulation of cellular immune responses in physiologic conditions and diseases. *Journal of Immunology Research*, *2012*(1), 485781.

Walzl, G., Ronacher, K., Hanekom, W., Scriba, T. J., & Zumla, A. (2011). Immunological biomarkers of tuberculosis. *Nature Reviews Immunology*, *11*(5), 343–354.

Weiner 3rd, J., & Kaufmann, S. H. (2014). Recent advances towards tuberculosis control: Vaccines and biomarkers. *Journal of internal medicine*, *275*(5), 467–480.

Williamson, P. R., Altman, D. G., Blazeby, J. M., Clarke, M., Devane, D., Gargon, E., & Tugwell, P. (2012). Developing core outcome sets for clinical trials: Issues to consider. *Trials*, *13*, 1–8.

Wisesty, U., & Mengko, T. (2021). Comparison of dimensionality reduction and clustering methods for SARS-CoV-2 genome. *Bulletin of Electrical Engineering and Informatics*, *10*, 2170–2180.

World Health Organisation. (2004). Joint WHO/UNAIDS/UNICEF statement on use of cotrimoxazole as prophylaxis in HIV exposed and HIV infected children.

Zhang, Y. (2013). Likelihood-based and Bayesian methods for Tweedie compound poisson linear mixed models. *Statistics and Computing*, *23*, 743–757.

Zimmermann, P., & Curtis, N. (2019). Factors that influence the immune response to vaccination. *Clinical microbiology reviews*, *32*(2), 10–1128.

Appendix A

Table 9.1: Proportions of BCG-specific $CD4^+$ and $CD8^+$ T cells expressing single cytokines. Measured in whole blood collected from Group A after 56 days, 112 days, and 365 days.

$CD4^+$ T cells	***$CD8^+$ T cells***
(1) Expressing any cytokine (*$\%CD4^+$ $ANYcytok^+$*)	**(2)** Expressing any cytokine (*$\%CD8^+$ $ANYcytok^+$*)
(3) Expressing INFg (*$\%CD4^+$ $totIFNg^+$*)	**(4)** Expressing INFg (*$\%CD8^+$ $totIFNg^+$*)
(5) Expressing IL-2 (*$\%CD4^+$ $totIL2^+$*)	
(6) Expressing IL-17 (*$\%CD4^+$ $totIL17^+$*)	
(7) Expressing IL-22 (*$\%CD4^+$ $totIL22^+$*)	
(8) Expressing TNF (*$\%CD4^+$ $totTNF^+$*)	**(9)** Expressing TNF (*$\%CD8^+$ $totTNF^+$*)

Table 9.2: Proportions of $CD4^+$ and $CD8^+$ T cells with particular phenotypes (CCR7). Quantified by bulk profile from whole blood collected in Group A after 56 days, 112 days, and 365 days.

$CD4^+$ T cells	***$CD8^+$ T cells***
(10) Naive (*Bulk $\%CD4^+$ $R7^+$ RA^+*)	**(11)** Naive (*Bulk $\%CD8^+$ $R7^+$ RA^+*)
(12) Effector Memory (*Bulk $\%CD4^+$ $R7^-$ RA^-*)	**(13)** Effector Memory (*Bulk $\%CD8^+$ $R7^-$ RA^-*)
(14) Central Memory (*Bulk $\%CD4^+$ $R7^+$ RA^-*)	**(15)** Central Memory (*Bulk $\%CD8^+$ $R7^+$ RA^-*)
(16) Effector (*Bulk $\%CD4^+$ $R7^-$ RA^+*)	**(17)** Effector (*Bulk $\%CD8^+$ $R7^-$ RA^+*)

Table 9.3: Proportions of BCG-reactive proliferating T cells and NK cells. Measured in whole blood collected from Group A after 56 days, 112 days, and 365 days.

(18) $CD4^+$ T cells expressing Ki67 (*$\%CD4^+$ $Ki67^+$*)
(19) $CD8^+$ T cells expressing Ki67 (*$\%CD8^+$ $Ki67^+$*)
(20) $\gamma\delta$ (gd) T cells expressing Ki67 (*$\%gd$ $Ki67^+$*)
(21) NK cells expressing Ki67 (*$\%NK$ $Ki67^+$*)

Table 9.4: Proportions of BCG-reactive proliferating $CD4^+$ and $\gamma\delta$ (gd) T cells expressing particular markers. Measured in whole blood collected from Group A after 56 days, 112 days, and 365 days.

$CD4^+$ T cells	***gd T cells***
(22) Expressing Ki67 & Gr.A (*$\%GrA^+$ $Ki67^+$ $CD4^+$*)	**(23)** Expressing Ki67 & Gr. A (*$\%GrA^+$ $Ki67^+$ gd*)
(24) Expressing Ki67 & Gr.B (*$\%GrB^+$ $Ki67^+$ $CD4^+$*)	**(25)** Expressing Ki67 & Gr.B (*$\%GrB^+$ $Ki67^+$ gd*)
(26) Expressing Ki67 & Gr.K (*$\%GrK^+$ $Ki67^+$ $CD4^+$*)	**(27)** Expressing Ki67 & Gr.K (*$\%GrK^+$ $Ki67^+$ gd*)
(28) Expressing Ki67 & Perforin (*$\%Perf^+$ $Ki67^+$ $CD4^+$*)	**(29)** Expressing Ki67 & Perforin (*$\%Perf^+$ $Ki67^+$ gd*)

Table 9.5: Proportions of BCG-reactive proliferating NK cells expressing particular markers. Measured in whole blood collected from Group A after 56 days, 112 days, and 365 days.

NK cells

(30) Expressing Ki67 and Gr.A ($\%GrA^{+}\ Ki67^{+}\ NK$)
(31) Expressing Ki67 and Gr.B ($\%GrB^{+}\ Ki67^{+}\ NK$)
(32) Expressing Ki67 and Gr.K ($\%GrK^{+}\ Ki67^{+}\ NK$)
(33) Expressing Ki67 and Perforin ($\%Perf^{+}\ Ki67^{+}\ NK$)

Appendix B

Table 9.6: Counts and proportions of patient sex in Group A.

Group	Female	Male	Total
A	30 (50%)	30 (50%)	60 (100%)

Table 9.7: Counts and proportions of MVA85A priming in Group A.

Group	Control	MVA85A	Total
A	31 (51.7%)	29 (48.3%)	60 (100%)

Table 9.8: Counts and proportions of maternal QFT result in Group A.

Group	Negative	Positive	Total
A	36 (60%)	24 (40%)	60 (100%)

Table 9.9: Counts and proportions of feeding practices in Group A.

Group	Breast Milk	Formula	Total
A	23 (38.3%)	37 (61.7%)	60 (100%)

Table 9.10: Counts and proportions of cotrimoxazole treatment in Group A.

Group	No Antibiotic	Antibiotic	Total
A	28 (46.7%)	32 (53.3%)	60 (100%)

Table 9.11: Results of two-sided Fisher's Exact Test to examine the association between *Feeding* and *Antibiotic* in Group A. H_1: The true odds ratio of the counts is not equal to 1.

	Group A
Odds Ratio (OR)	0.3354
95% Confidence Interval for OR (*lower*)	0.1671
95% Confidence Interval for OR (*upper*)	0.6574
p-value	0.0001

Table 9.12: Counts and proportions of combinations of feeding practices and cotrimoxazole treatment in Group A.

	Group A
Formula & Cotrimoxazole	16 (26.7%)
Formula & No Cotrimoxazole	21 (35%)
Breast Milk & No Cotrimoxazole	7 (11.7%)
Breast Milk & Cotrimoxazole	16 (26.7%)
Total	60 (100%)

Table 9.13: Counts and proportions of MVA85A priming by patient sex in Group A and Group B.

Sex	Control	MVA85A	Total
Female	29 (53.7%)	25 (46.3%)	54 (100%)
Male	29 (51.8%)	27 (48.2%)	56 (100%)

Table 9.14: Counts and proportions of maternal QFT by patient sex in Group A and Group B.

Sex	Negative QFT	Positive QFT	Total
Female	30 (55.6%)	24 (44.4%)	54 (100%)
Male	34 (60.7%)	22 (39.3%)	56 (100%)

Table 9.15: Counts and proportions of combinations of feeding practice and cotrimoxazole treatment by patient sex in Group A and Group B.

	Female	Male
Formula & Cotrimoxazole	15 (27.8%)	19 (33.9%)
Formula & No Cotrimoxazole	13 (24.1%)	16 (28.6%)
Breast Milk & No Cotrimoxazole	10 (18.5%)	6 (10.7%)
Breast Milk & Cotrimoxazole	16 (29.6%)	15 (26.8%)
Total	54 (100%)	56 (100%)

Appendix C

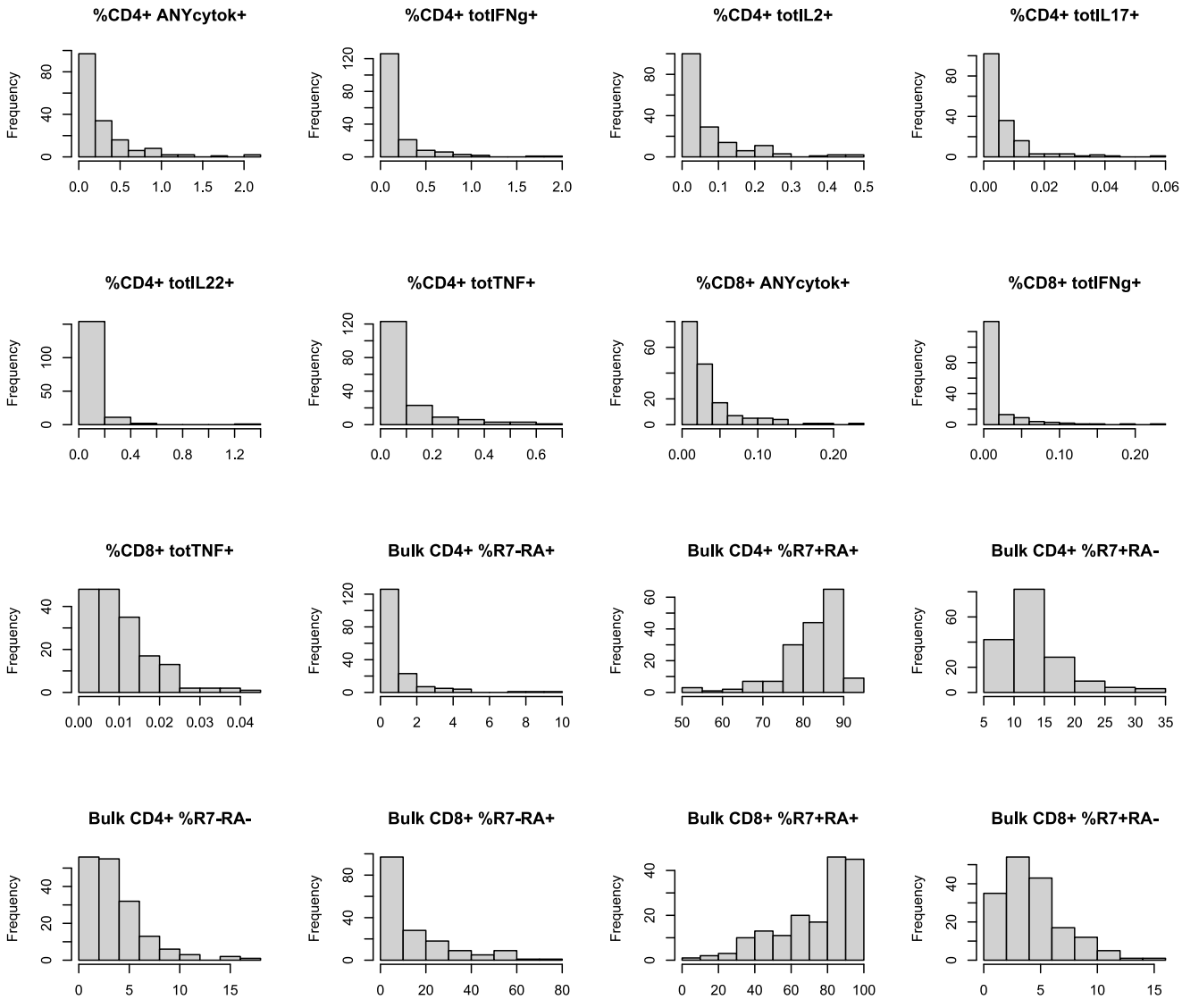


Figure 9.1: Histograms of Group A immune outcome measurements (*Part 1*). Taken at 56 days, 112 days, and 365 days after BCG vaccination.

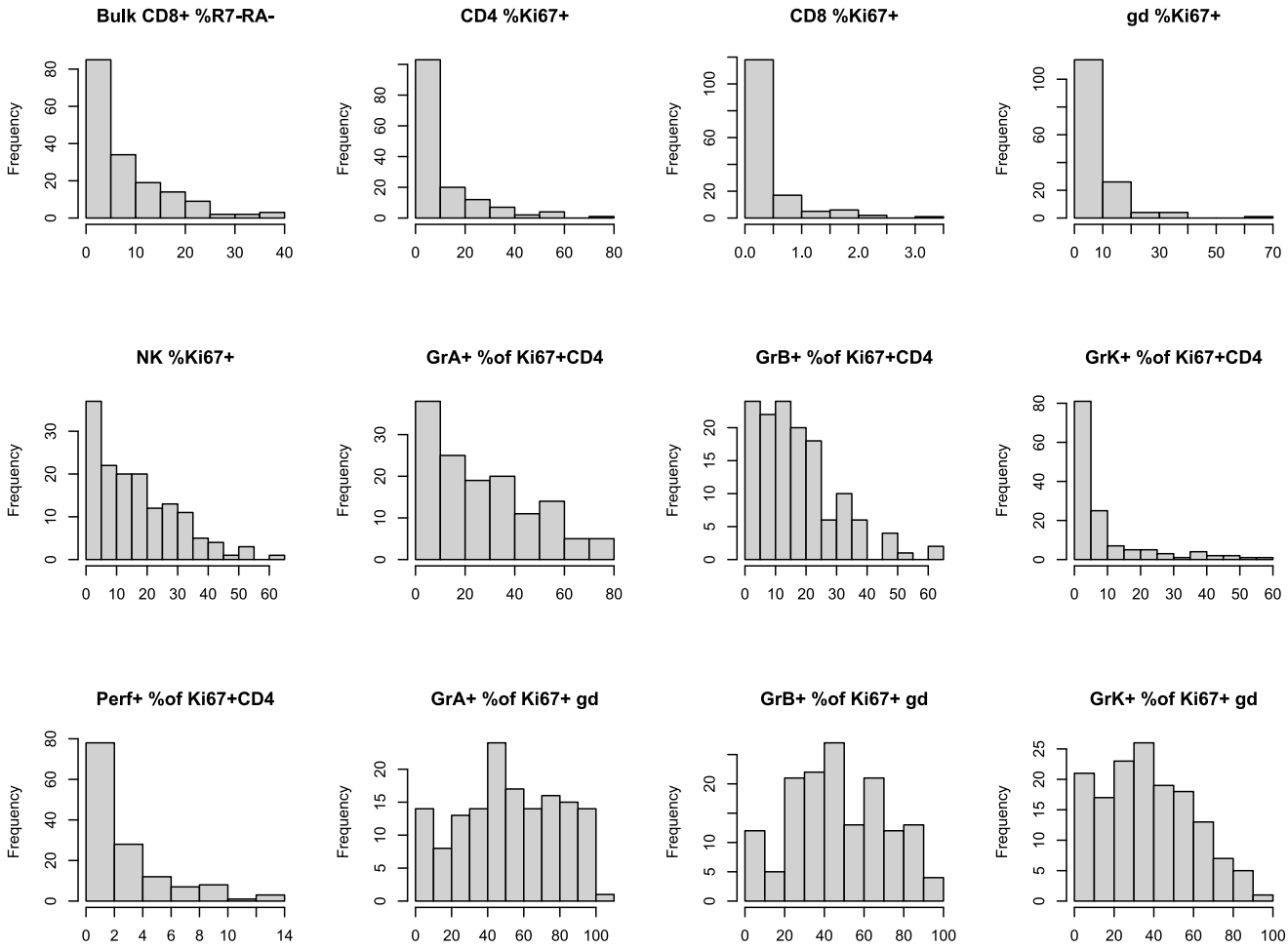


Figure 9.2: Histograms of Group A immune outcome measurements (*Part 2*). Taken at 56 days, 112 days, and 365 days after BCG vaccination.

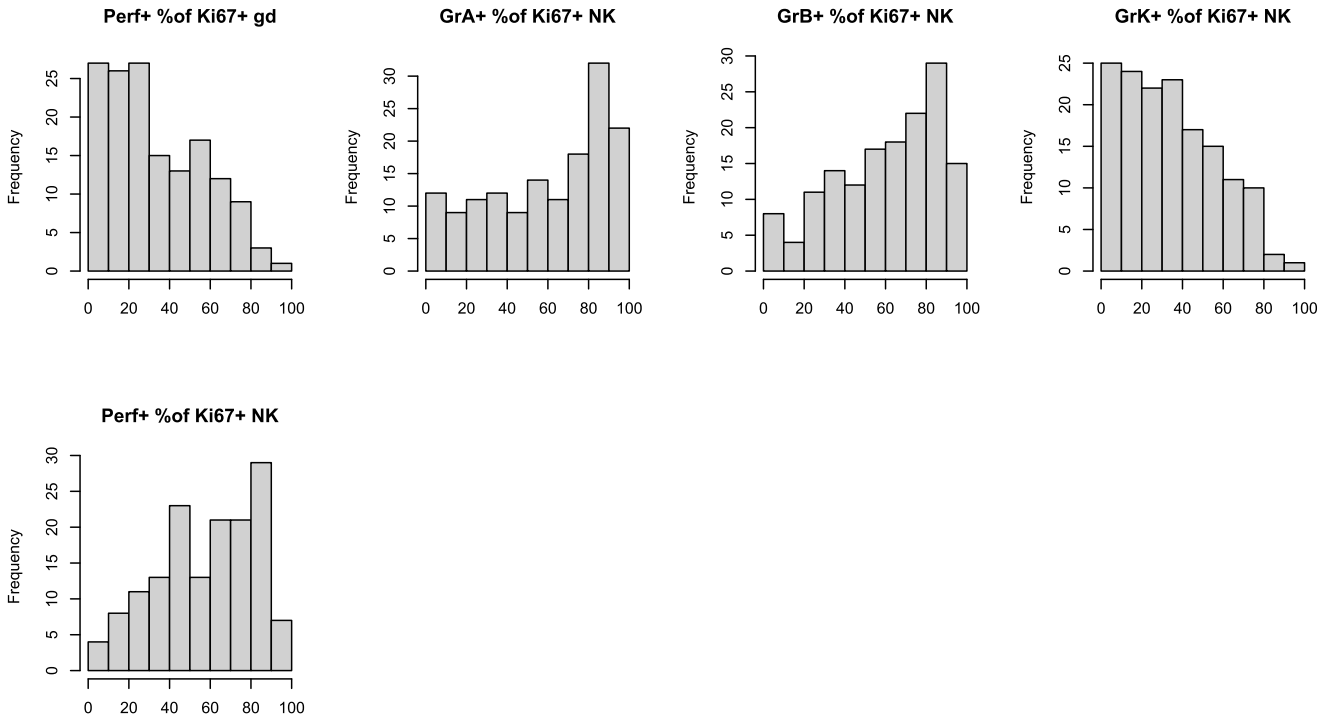


Figure 9.3: Histograms of Group A immune outcome measurements (*Part 3*). Taken at 56 days, 112 days, and 365 days after BCG vaccination.

Appendix D

Overview

Residual diagnostics for GLMMs were performed using a simulation-based approach as implemented in the *DHARMa* package (Hartig, 2018). As linear models, linear mixed-effect models (LMMs) produce readily interpretable residuals. This is not the case for generalized linear mixed-effect models. Simulation studies have shown that heteroscedasticity and non-normality of residuals may appear to be present, even when the model is correctly specified (Hartig, 2018). This is because GLMM residuals are not on a comparable scale, as the expected distribution of the data, i.e., the predictive distribution, changes with the fitted values for each observation's covariate pattern.

In the *DHARMa* package, the GLMM is assumed to represent the true data generating process (Hartig, 2018). New response data is then simulated from the model for each observation. Empirical cumulative density functions are calculated for the simulated observations. The probability associated with the simulated observation is taken as a scaled residual bounded between 0 and 1. If the scaled residual is 0, all simulated observations exceed the true observation. If the scaled residual is 0.5, half of the simulated observations are larger than what was observed. Using this definition of residuals (assuming that the model is correctly specified), all models are expected to have the same (flat) residual distribution.

Residual diagnostics compare these scaled residuals to rank-transformed model predictions. Any visible quantile deviations indicate that the model is likely to be incorrectly specified. Additionally, a Normal quantile-quantile (QQ) plot considers whether the scaled residuals meet distributional assumptions. A dispersion test is also performed on the scaled residuals to consider whether the residual dispersion corresponds to what would be expected under the model specification.

Simulation may be unconditional *or* conditional on the random effects. However, conditional simulations are not possible from *glmmTMB* model objects (Brooks et al., 2017). Hartig (2018) notes that re-simulating random effects may reduce power to detect over- or underdispersion. However, by simulating from all levels in the hierarchical model, the full model specification is investigated (Hartig, 2018). Unconditional simulations are therefore the *DHARMa* package default (Hartig, 2018).

Quantile Deviations

When a model contains continuous predictors only or categorical predictors with more than two levels, the *DHARMa* package compares quantiles of these scaled residuals to what would be expected if the GLMM represents the true data generating process: a uniform distribution (Hartig, 2018). The R package *qgam* is implemented within *DHARMa* to fit spline-based quantile non-parametric additive models to the scaled residuals. If the splines significantly differ from the expected flat distribution at the lower quartile, median or upper quartile, *DHARMa* will report a significant deviation at that quantile. A combined p-value is reported that considers both the intercept and the spline by applying a Benjamini-Hochberg correction to control the false discovery rate.

A plot is generated to represent this comparison. Quantile lines that deviate from model assumptions are highlighted in red. *DHARMa* also reports a total combined

p-value for quantile deviations at the lower quartile, median and upper quartile. Once again, a Benjamini-Hochberg correction is applied.

Categorical Dependencies

When a model contains binary predictors or a combination of continuous and binary predictors, the *DHARMa* package performs two tests (Hartig, 2018). First, within-group uniformity of scaled residuals is assessed by performing multiple Kolmogorov-Smirnov tests of uniformity. The p-value reported for this result is adjusted for multiple testing with Holm's correction. *DHARMa* provides a box plot of scaled residual distributions by group. If scaled residuals are not distributed according to model assumptions for a given group, the corresponding bar is highlighted in red. Second, a Levene test is performed to assess whether between-group variance of scaled residuals is constant. The p-value is reported for this result.

Dispersion Test

The *DHARMa* package performs tests for overdispersion and/or underdispersion on the scaled residuals (Hartig, 2018). A significant p-value indicates that there is evidence for the residual dispersion differing from what would be expected under the model specification. The default is a simulation-based dispersion test, where a ratio is constructed comparing the variance of the observed raw residuals with the variance of simulated residuals. The variances are scaled in terms of the mean simulated variance. A ratio greater than $|1|$ then reflects overdispersion or underdispersion relative to the model specification. A ratio of 1 is expected under the null hypothesis, and a significant p-value identifies a significant difference from this expectation. This, in turn, reflects possible model misspecification.

Outlier Test

When all simulated observations are greater or smaller than the true observation, the *DHARMa* package considers this observation to be an outlier. In the former case, the scaled residual is 0. In the latter case, the scaled residual is 1. Within the *DHARMa* package, an outlying scaled residual does not describe the magnitude of the residual deviation. An outlier indicates that the range of simulated observations does not correspond well to the true observation. This, in turn, reflects possible model misspecification.

The strength of the evidence for an excess or lack of simulation outliers is tested using either a binomial test or a bootstrap test. The binomial test is the *DHARMa* package default, and the bootstrap test is recommended for integer-valued distributions. No outliers are expected, and a significant p-value identifies a significant difference from this null expectation. This difference could be caused by either overdispersion or underdispersion as the hypothesis is two-sided.

Appendix E

Table 9.16: Summary of (G)LMM conditional distributions for BCG-specific $CD4^+$ and $CD8^+$ T cells expressing single cytokines. Measured in whole blood collected from Group A after 56 days, 112 days, and 365 days.

$CD4^+$ T cells	***$CD8^+$ T cells***
(1) *$\%CD4^+$ $ANYcytok^+$*: Gamma	**(2)** *$\%CD8^+$ $ANYcytok^+$*: Tweedie ($p = 1.93$)
(3) *$\%CD4^+$ $totIFNg^+$*: Tweedie ($p = 1.79$)	**(4)** *$\%CD8^+$ $totIFNg^+$*: Tweedie ($p = 1.64$)
(5) *$\%CD4^+$ $totIL2^+$*: Tweedie ($p = 1.99$)	
(6) *$\%CD4^+$ $totIL17^+$*: Tweedie ($p = 1.56$)	
(7) *$\%CD4^+$ $totIL22^+$*: Gamma	
(8) *$\%CD4^+$ $totTNF^+$*: Gamma	**(9)** *$\%CD8^+$ $totTNF^+$*: Tweedie ($p = 1.46$)

Table 9.17: Summary of (G)LMM conditional distributions for BCG-specific cytokine-expressing $CD4^+$ and $CD8^+$ T cells with particular phenotypes. Quantified by bulk profile from whole blood collected in Group A after 56 days, 112 days, and 365 days.

$CD4^+$ T cells	***$CD8^+$ T cells***
(10) *Bulk $\%CD4^+$ $R7^+$ RA^+*: Tweedie ($p = 1.89$)	**(11)** *Bulk $\%CD8^+$ $R7^+$ RA^+*: Gamma
(12) *Bulk $\%CD4^+$ $R7^-$ RA^-*: Gamma	**(13)***Bulk $\%CD8^+$ $R7^-$ RA^-*: Gamma
(14) *Bulk $\%CD4^+$ $R7^+$ RA^-*: Gamma	**(15)** *Bulk $\%CD8^+$ $R7^+$ RA^-*: Tweedie ($p = 1.07$)
(16) *Bulk $\%CD4^+$ $R7^-$ RA^+*: Gamma	**(17)** *Bulk $\%CD8^+$ $R7^-$ RA^+*: Gamma

Table 9.18: Summary of (G)LMM conditional distributions for BCG-reactive proliferating T cells and NK cells. Measured in whole blood collected from Group A after 56 days, 112 days, and 365 days.

(18) *$\%CD4^+$ $Ki67^+$*: Tweedie ($p = 1.75$)
(19) *$\%CD8^+$ $Ki67^+$*: Tweedie ($p = 1.53$)
(20) *$\%gd$ $Ki67^+$*: Tweedie ($p = 1.70$)
(21) *$\%NK$ $Ki67^+$*: Tweedie ($p = 1.39$)

Table 9.19: Summary of (G)LMM conditional distributions for BCG-reactive proliferating $CD4^+$ and $\gamma\delta$ (gd) T cells expressing particular functional markers. Measured in whole blood collected from Group A after 56 days, 112 days, and 365 days.

$CD4^+$ T cells	***gd T cells***
(22) *$\%GrA^+$ $Ki67^+$ $CD4^+$*: Negative binomial	**(23)** *$\%GrA^+$ $Ki67^+$ gd*: Tweedie ($p = 1.07$)
(24) *$\%GrB^+$ $Ki67^+$ $CD4^+$*: Tweedie ($p = 1.38$)	**(25)** *$\%GrB^+$ $Ki67^+$ gd*: Tweedie ($p = 1.04$)
(26) *$\%GrK^+$ $Ki67^+$ $CD4^+$*: Tweedie ($p = 1.56$)	**(27)** *$\%GrK^+$ $Ki67^+$ gd*: Tweedie ($p = 1.17$)
(28) *$\%Perf^+$ $Ki67^+$ $CD4^+$*: Tweedie ($p = 1.37$)	**(29)** *$\%Perf^+$ $Ki67^+$ gd*: Tweedie ($p = 1.14$)

Table 9.20: Summary of (G)LMM conditional distributions for BCG-reactive proliferating NK cells expressing particular functional markers. Measured in whole blood collected from Group A after 56 days, 112 days, and 365 days.

NK cells

(30) $\%GrA^{+}$ $Ki67^{+}$ NK: Tweedie ($p = 1.85$)
(31) $\%GrB^{+}$ $Ki67^{+}$ NK: Tweedie ($p = 1.46$)
(32) $\%GrK^{+}$ $Ki67^{+}$ NK: Tweedie ($p = 1.18$)
(33) $\%Perf^{+}$ $Ki67^{+}$ NK: Tweedie ($p = 1.33$)

Appendix F

The GLMM modelling framework was demonstrated to be a suitable framework for CD4+ T cells expressing IL-22, %Bulk CD4+ T cells with R7+RA+ phenotype, and %GrK$^+$ Ki67$^+$ NK cells. Although this basic screening identified the GLMM modelling framework as best suited to modelling immune response profiles for outcome selection, certain immune outcome models did not fully meet assumptions. These cases are discussed in detail below.

Random Effect Assumptions

Random intercept distributions show noticeable deviations from normality assumptions at the tails across all models for the following immune outcomes:

Group A

- The percentage of CD4$^+$ T cells expressing any cytokine (Figure 9.4)
- The percentage of CD4$^+$ T cells expressing IL-2 (Figure 9.6)
- The bulk profiled percentage of CD4$^+$ T cells with phenotype R7$^-$RA$^+$ (Figure 9.16)
- The percentage of proliferating (Ki67$^+$) Natural Killer (NK) cells expressing Granzyme A (Figure 9.36)
- The percentage of proliferating (Ki67$^+$) Natural Killer (NK) cells expressing Granzyme B (Figure 9.37)

These immune outcomes have long right tails with outliers that are many standard deviations away from the centre of the distribution (see Appendix D). This characteristic may be difficult to capture, even by specifying a Tweedie conditional distribution. However, in all other respects, the model specifications appear appropriate.

Scaled Residual Assumptions

For the following immune outcomes in Group A, the scaled residual quantiles deviate from the quantiles of rank-transformed model predictions:

- The percentage of CD4$^+$ T cells expressing TNF (*lower quantile deviation*: MVA85A priming, Feeding & cotrimoxazole) (Figure 9.9)
- The bulk profiled percentage of CD4$^+$ T cells with phenotype R7$^-$RA$^+$ (*lower quantile deviation*: MVA85A priming, Maternal QFT status, Feeding & cotrimoxazole) (Figure 9.16)

For the percentage of CD4$^+$ T cells expressing TNF, the combined p-value for quantile deviations is fairly small for the model concerned with MVA85A priming ($p = 0.091$). Although there is evidence for a significant lower quartile deviation ($p = 0.031$), the combined p-value is large enough to not discredit this model specification entirely. For the model concerned with feeding and cotrimoxazole, the combined p-value for quantile deviations is large ($p = 0.1868$). There is some evidence of a significant lower quantile deviation ($p = 0.06228$), but the quantiles of the scaled residuals correspond well with what would be expected if the model specification was correct.

However, for the bulk profiled percentage of $CD4^+$ T cells with phenotype $R7^-RA^+$, the combined p-value for quantile deviations is small in all models (MVA85A priming: $p = 0.043$, Maternal QFT status: $p = 0.059$, Feeding & cotrimoxazole: $p = 0.070$). There is evidence for a significant lower quartile deviation throughout (MVA85A priming: $p = 0.014$, Maternal QFT status: $p = 0.020$, Feeding & cotrimoxazole: $p = 0.023$). A Gamma conditional distribution was specified for these models. Looking at the histogram of observed bulk profiled percentage of $CD4^+$ T cells with phenotype $R7^-RA^+$, a Tweedie conditional distribution might appear to be more appropriate. However, it was not possible to simulate residuals from a Tweedie GLMM modelling this immune outcome - possibly due to the additional complexity of estimating a power parameter. By specifying a Gamma conditional distribution, adequate model evaluation could be performed. The random intercept meets distributional assumptions and the predictions correspond well with observed data - despite the quantile deviations.

Underdispersion or overdispersion of scaled residuals was identified for the following immune outcomes:

- The percentage of proliferating $Ki67^+$ $\gamma\delta$ (gd) T cells expressing Granzyme A (Figure 9.32)
- The bulk profiled percentage of $CD8^+$ T cells with phenotype $R7^-RA^+$ (Figure 9.20)

For the percentage of proliferating $Ki67^+$ $\gamma\delta$ (gd) T cells expressing Granzyme, a two-sided non-parametric dispersion test identified a significant difference between the standard deviation of fitted and scaled residuals for the model considering maternal QFT ($p = 0.046$). A small p-value was also observed for the same test performed for the model considering feeding and cotrimoxazole ($p = 0.162$). One-sided dispersion tests suggest underdispersion ($p = 0.023$ for QFT; $p = 0.081$ for feeding and cotrimoxazole) relative to the specified model. As there was stronger evidence for underdispersion when specifying a Poisson conditional distribution for this immune outcome, the analysis proceeded with the Tweedie GLMM.

For the bulk profiled percentage of $CD8^+$ T cells with phenotype $R7^-RA^+$, a two-sided non-parametric dispersion test identified a significant difference for the model considering maternal QFT status ($p = 0.040$). One-sided dispersion tests suggest underdispersion ($p = 0.020$). As it was not possible to simulate residuals from a Tweedie GLMM modelling this immune outcome, the analysis proceeded with the Gamma GLMM.

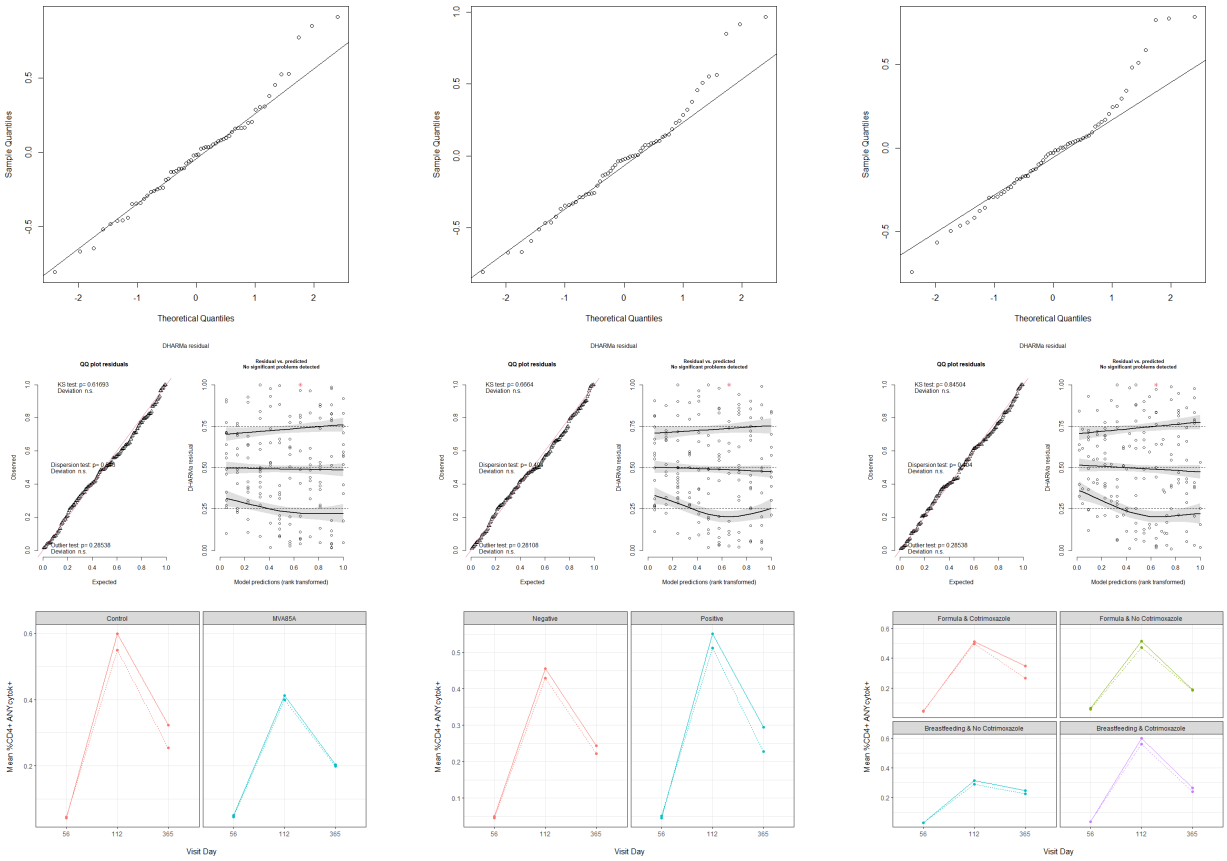
Excess outliers were detected in scaled residuals for the following immune outcomes by an exact binomial test, but not a more robust bootstrapped test:

- The percentage of proliferating $Ki67^+$ $\gamma\delta$ (gd) T cells expressing Granzyme B (Maternal QFT status: $p = 0.035$ and $p = 1$; Feeding & cotrimoxazole: $p = 0.036$ and $p = 0.440$) (*Figure 9.33*)
- The percentage of proliferating $Ki67^+$ $\gamma\delta$ (gd) T cells (Feeding & cotrimoxazole: $p = 0.036$ and $p = 0.420$) (*Figure 9.26*)

Predicted Trajectories

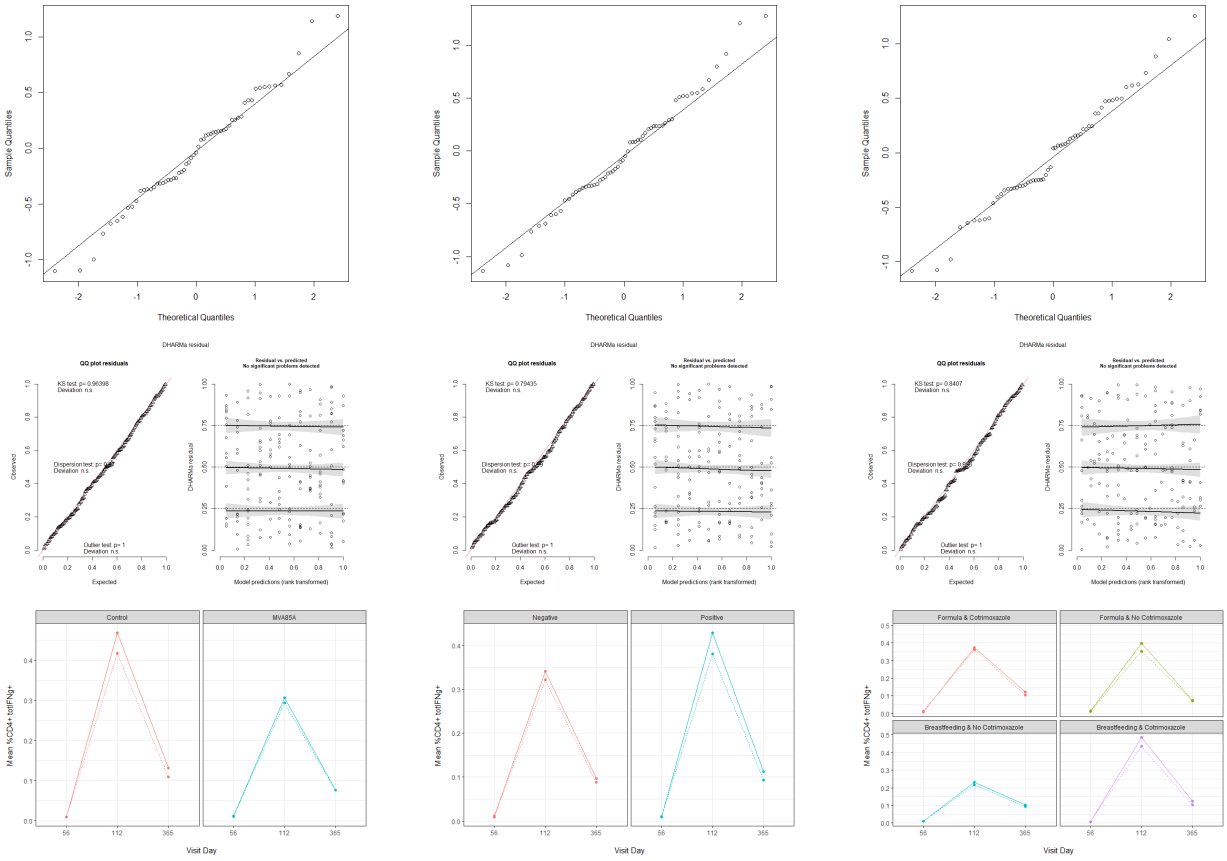
The shape and/or magnitude of GLMM predicted trajectories correspond well with what was observed over time. However, more underpredictions are observed for GLMMs of immune outcomes measured in Group B (see Appendix VII) than in Group A. This could be the consequence of fewer data points ($n = 2$) for Group B immune outcomes compared to Group A ($n = 3$). However, for the percentage of $CD8^+$ T cells expressing IL-2 measured in Group A, the shape and magnitude of the predicted trajectory from the maternal QFT status model differs considerably from what was observed (Figure 9.12).

Appendix G



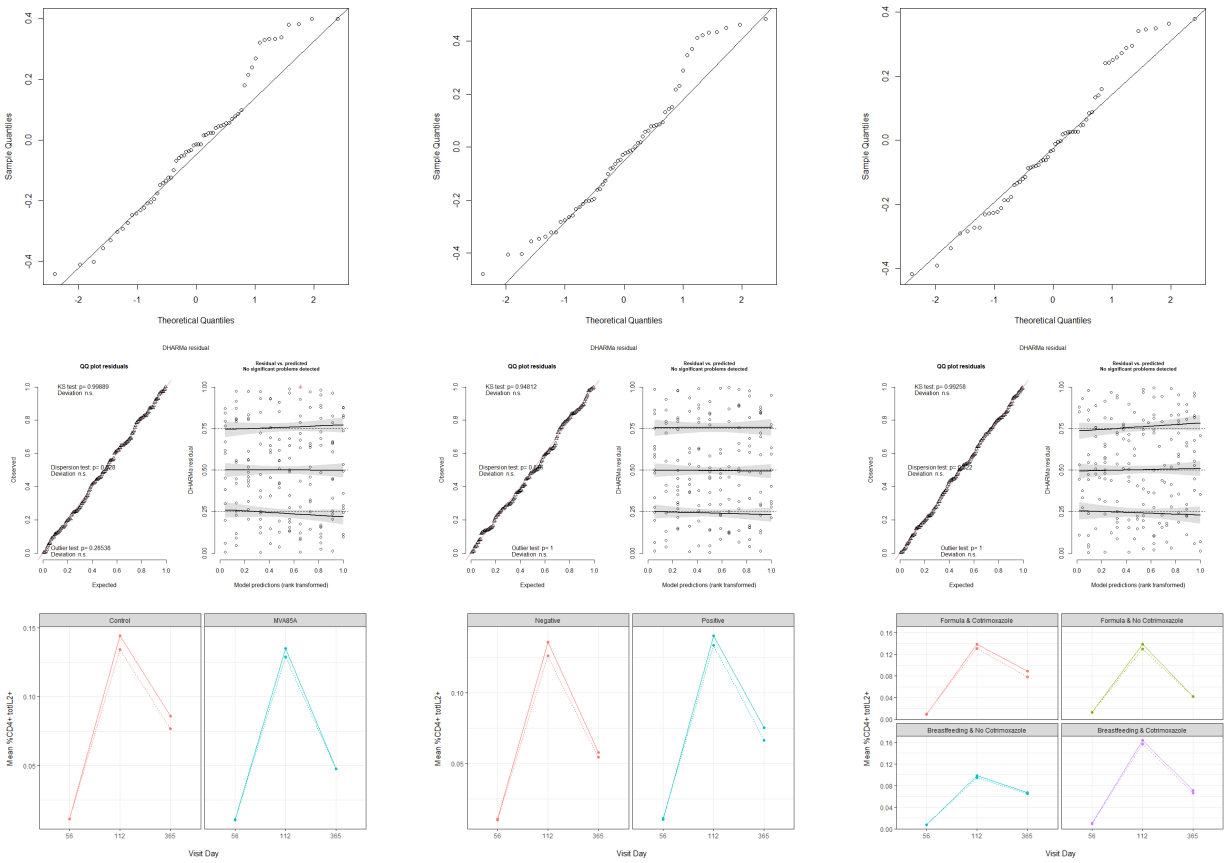
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.4: The percentage of $CD4^+$ T cells expressing any cytokine.



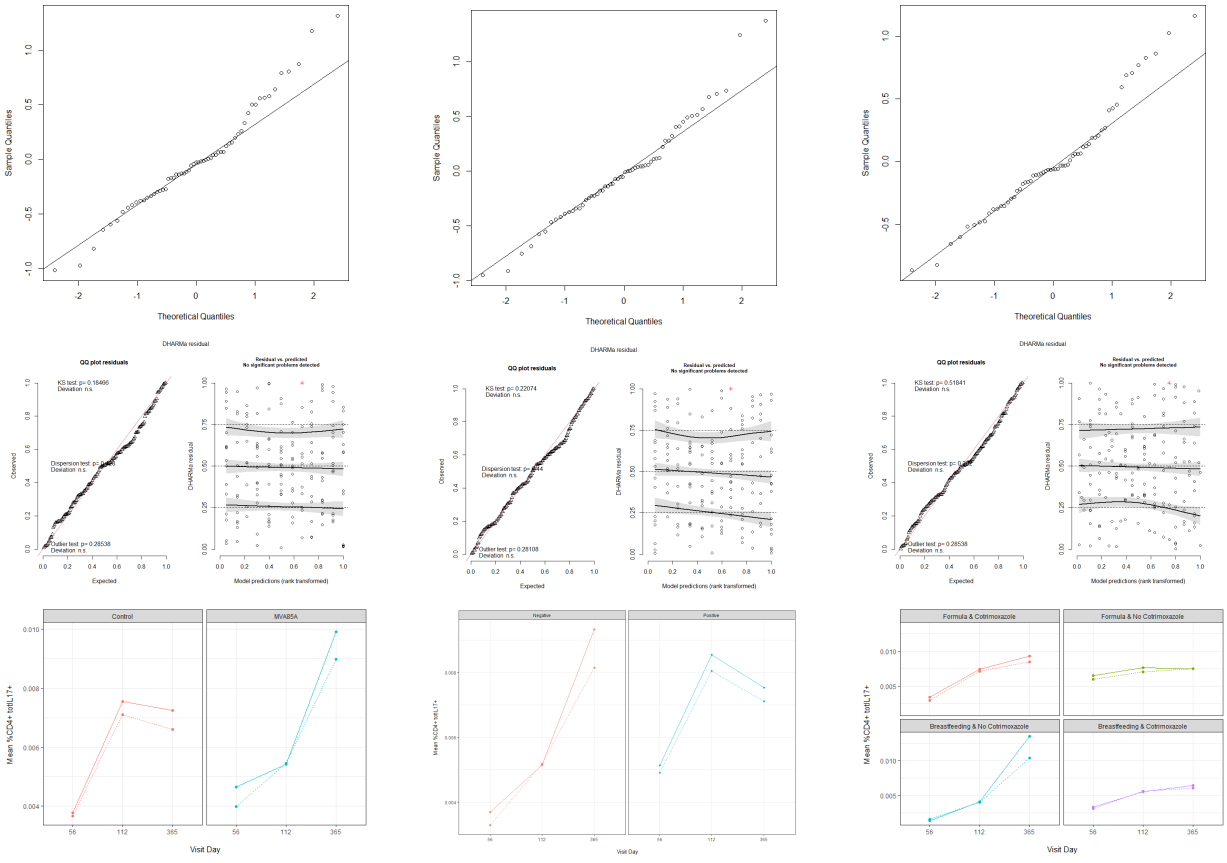
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.5: The percentage of $CD4^+$ T cells expressing IFNg.



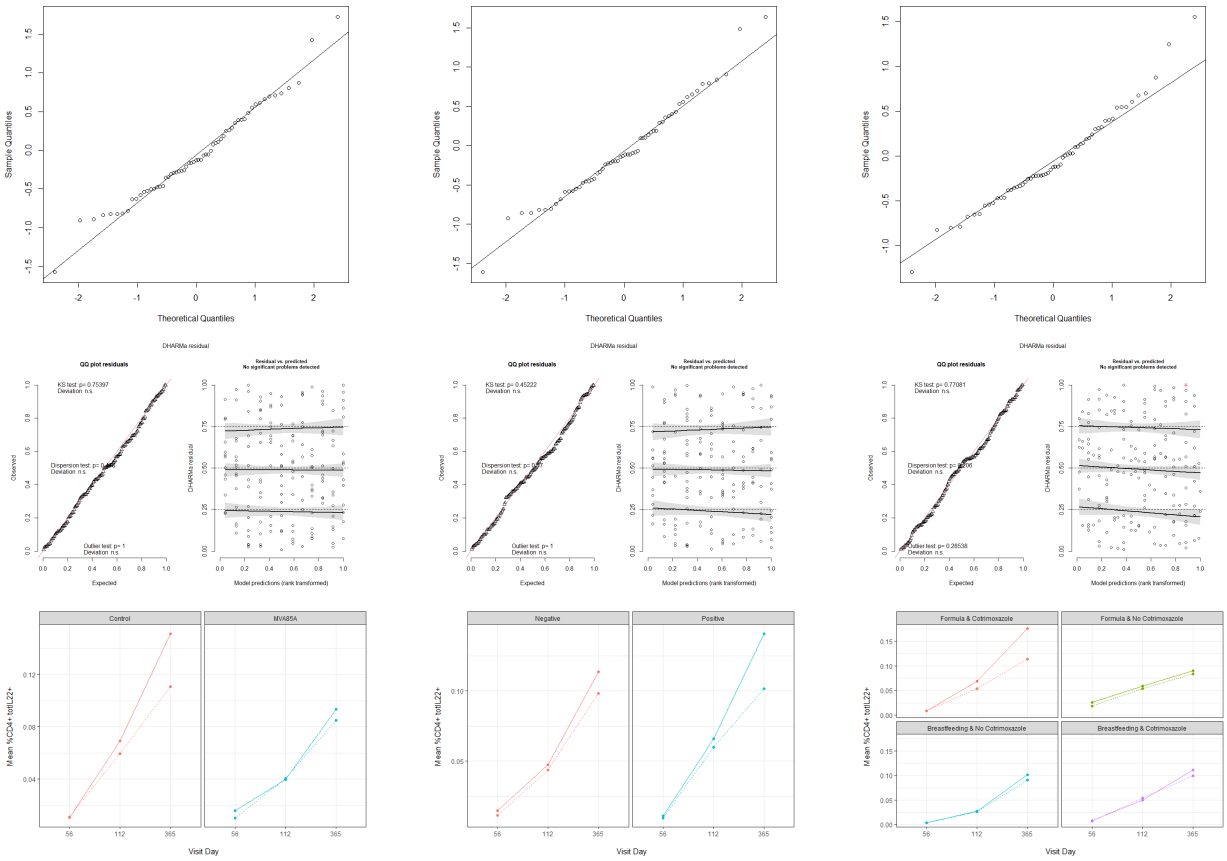
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.6: The percentage of $CD4^+$ T cells expressing IL-2.



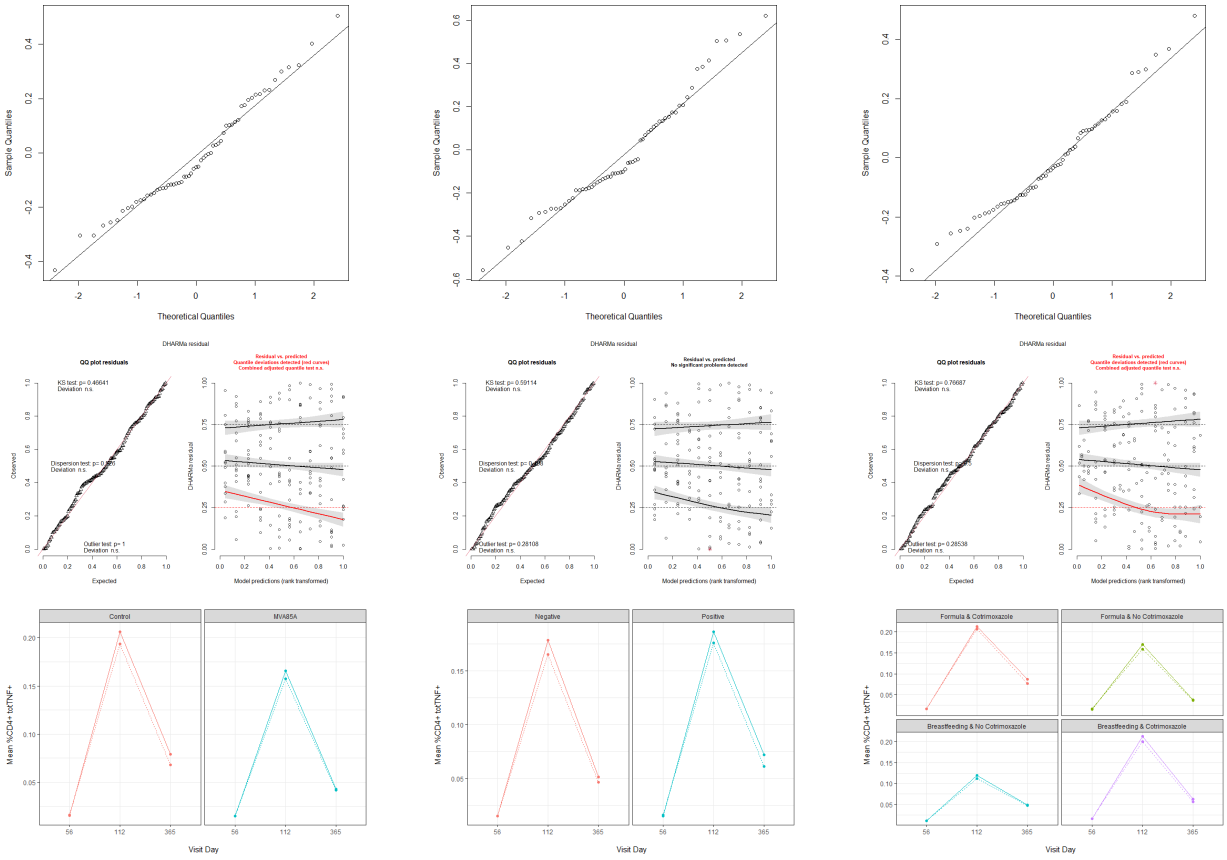
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.7: The percentage of $CD4^+$ T cells expressing IL-17.



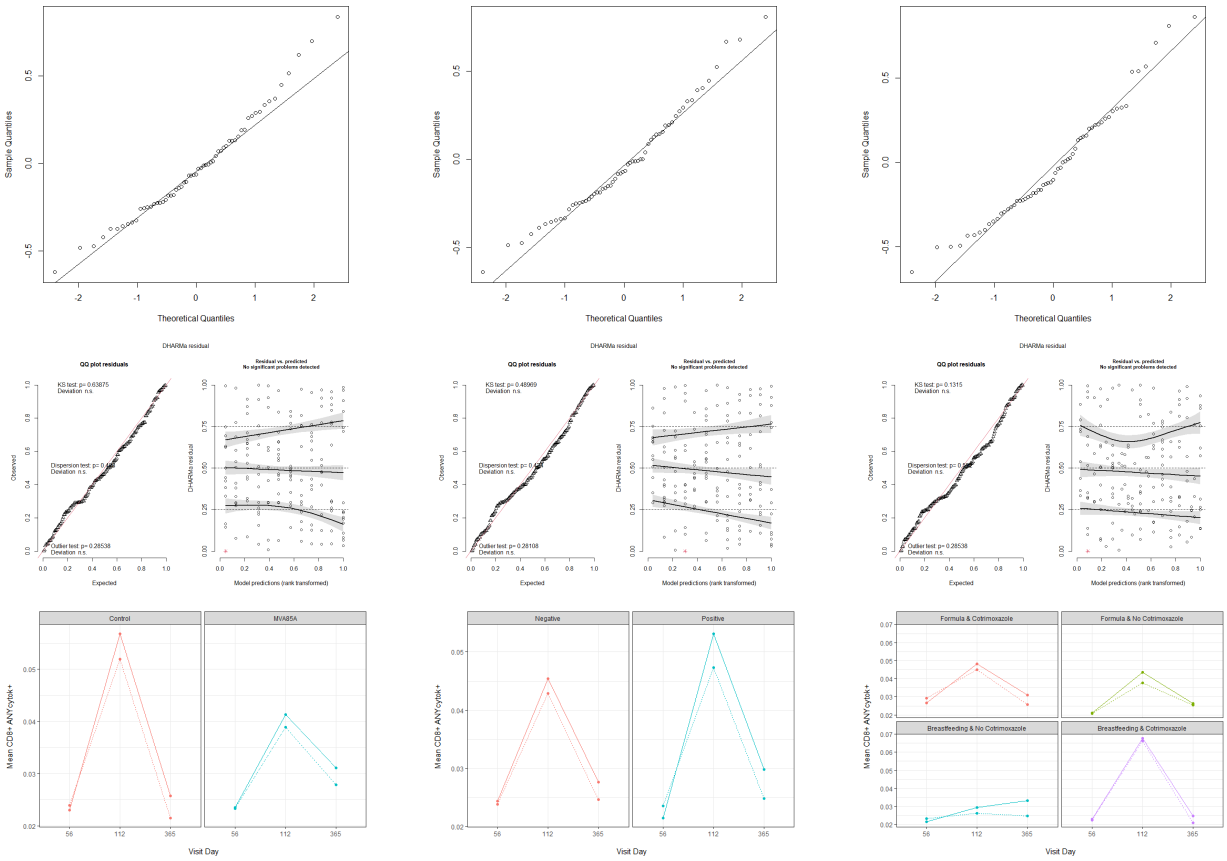
MVA85A priming Maternal QFT status Feeding & cotrimoxazole

Figure 9.8: The percentage of CD4$^+$ T cells expressing IL-22.



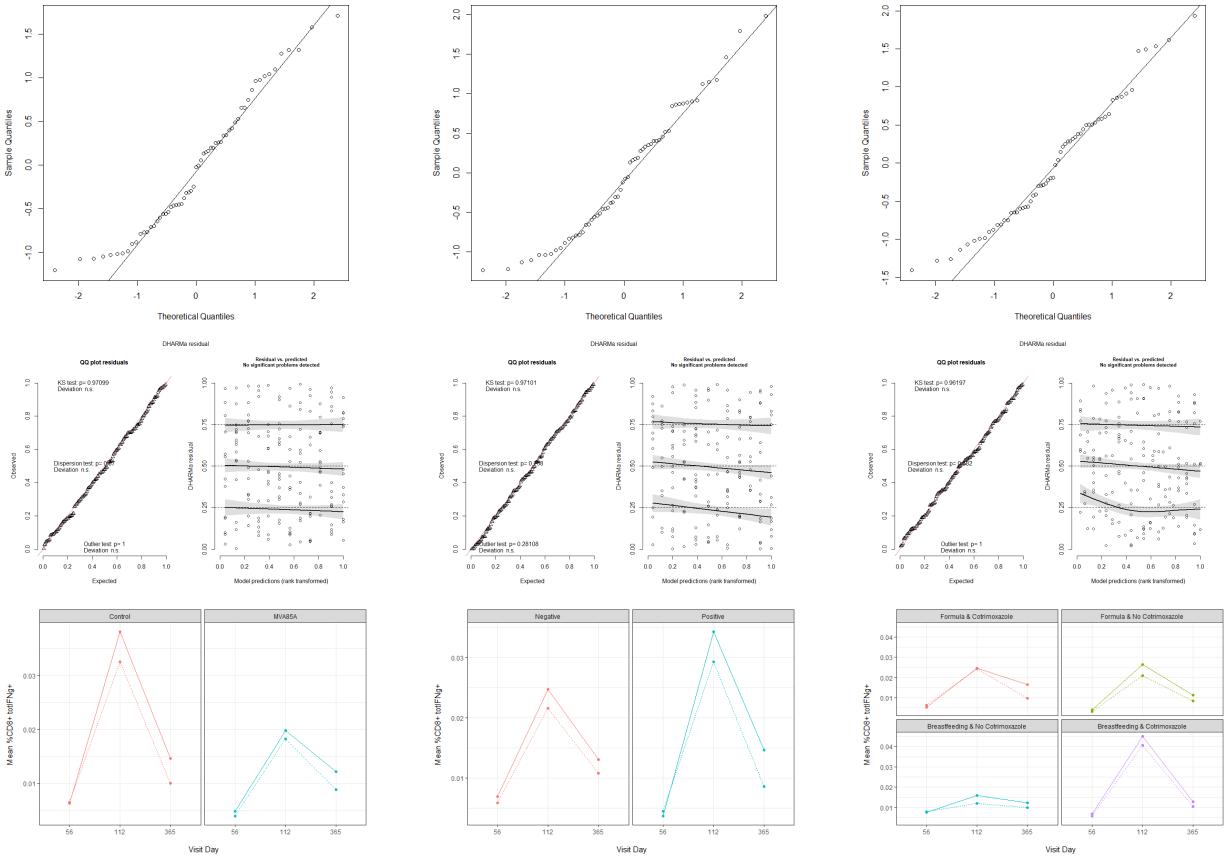
MVA85A priming Maternal QFT status Feeding & cotrimoxazole

Figure 9.9: The percentage of CD4$^+$ T cells expressing TNF.



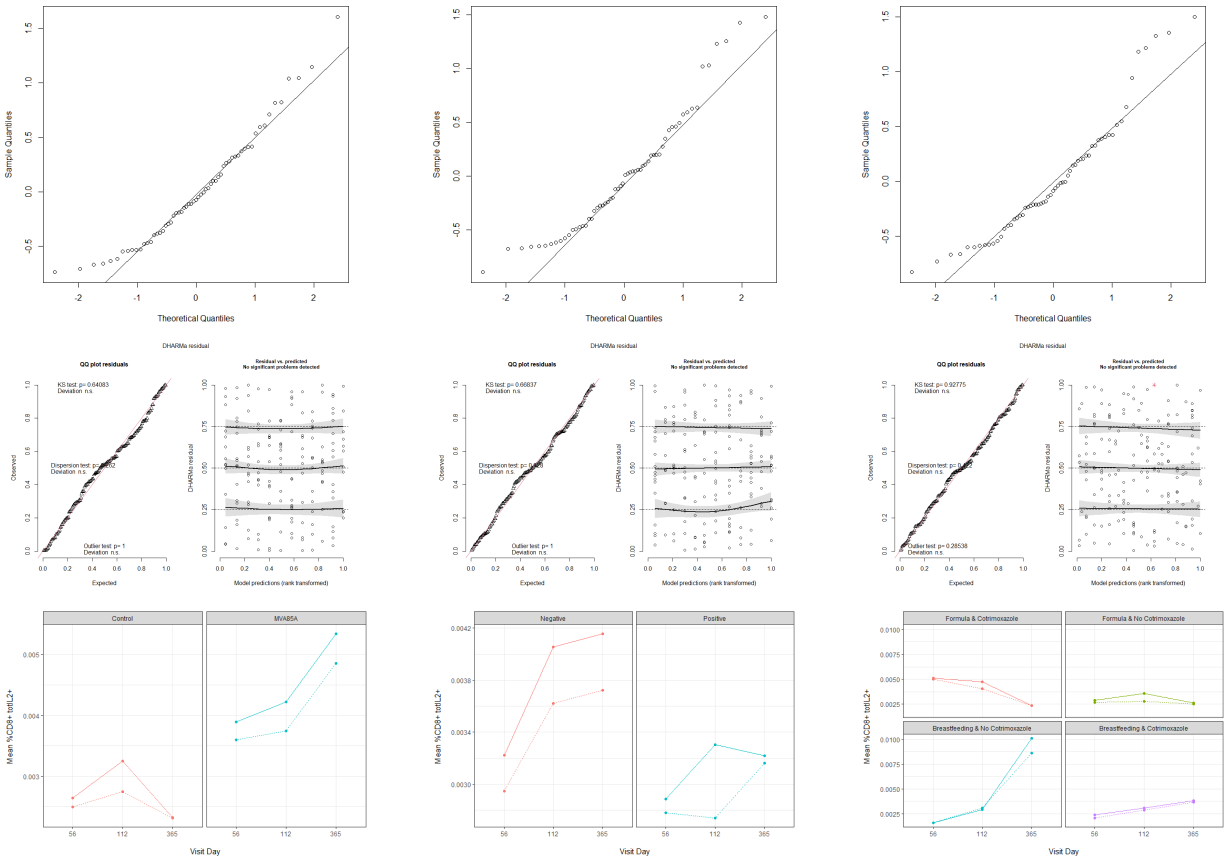
MVA85A priming Maternal QFT status Feeding & cotrimoxazole

Figure 9.10: The percentage of $CD8^+$ T cells expressing any cytokine.



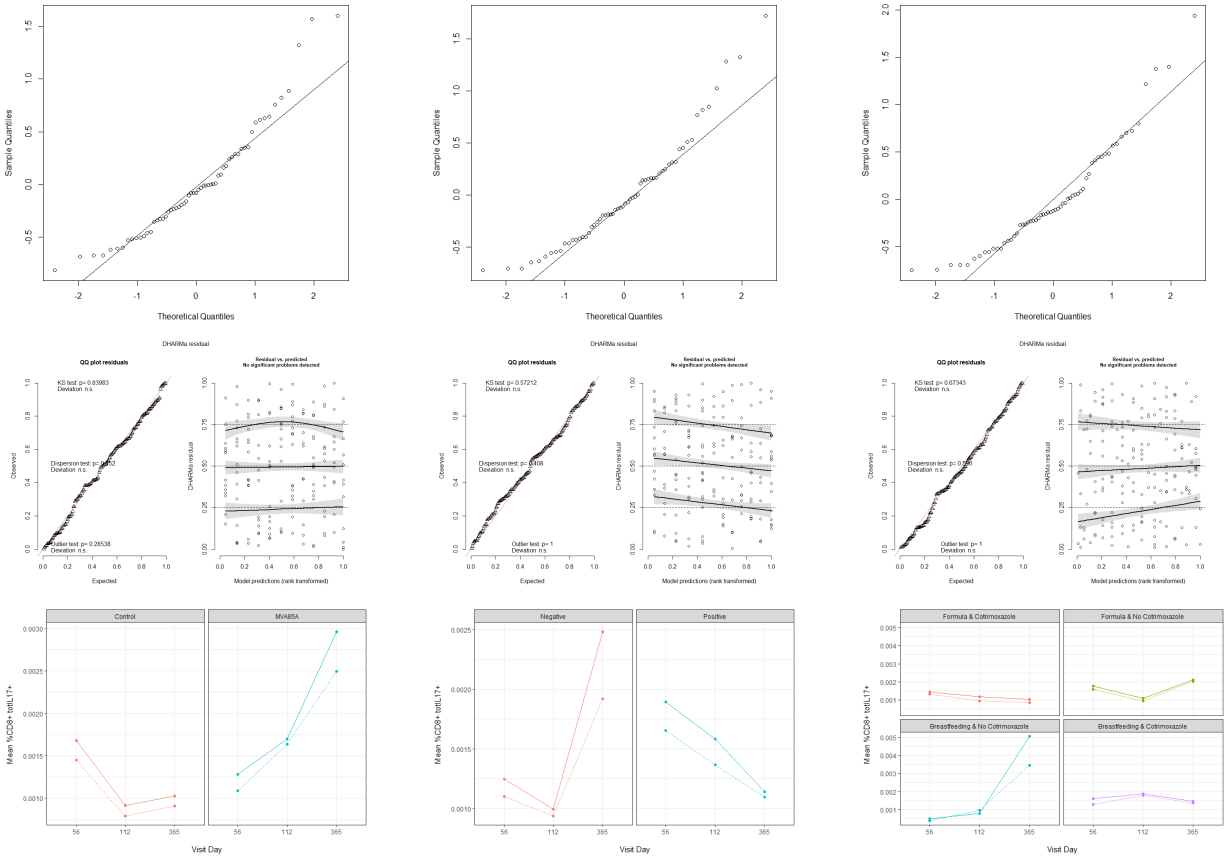
MVA85A priming Maternal QFT status Feeding & cotrimoxazole

Figure 9.11: The percentage of $CD8^+$ T cells expressing IFNg.



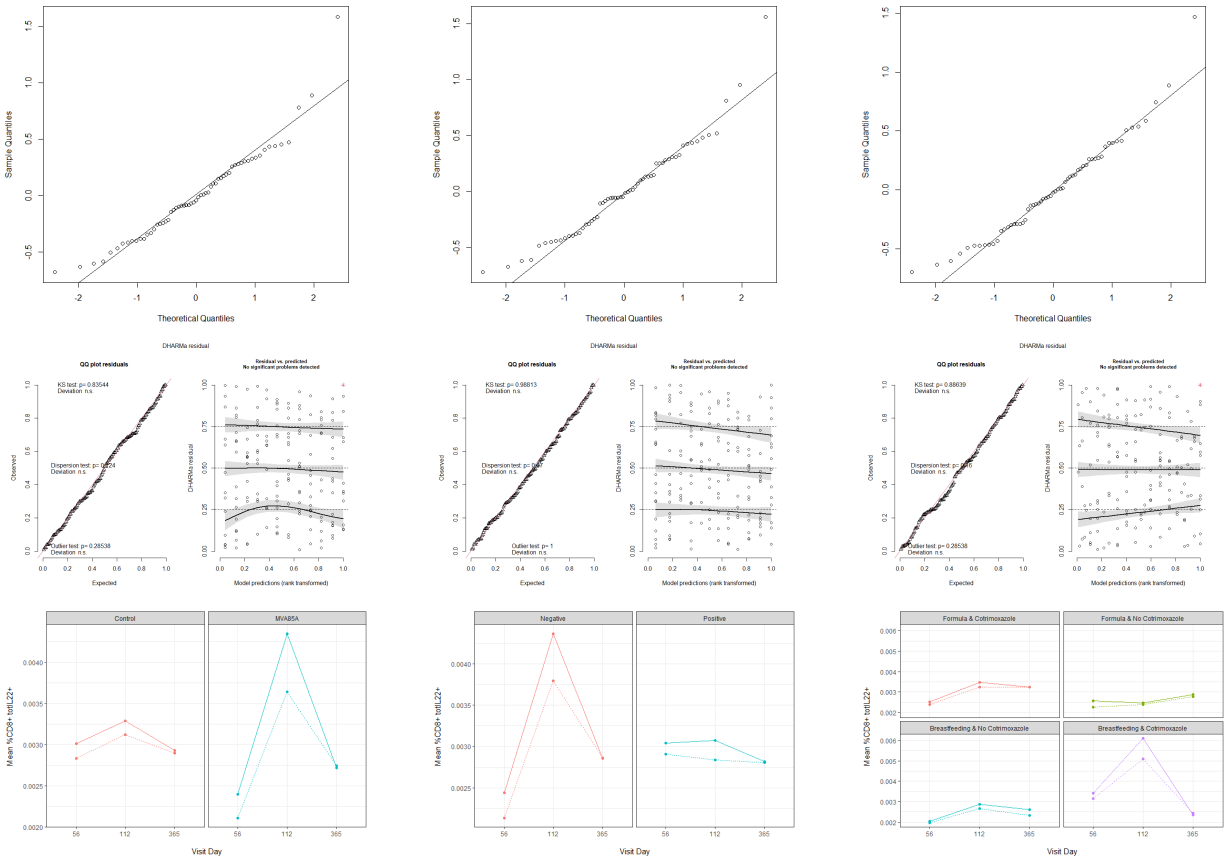
MVA85A priming Maternal QFT status Feeding & cotrimoxazole

Figure 9.12: The percentage of CD8$^+$ T cells expressing IL-2.



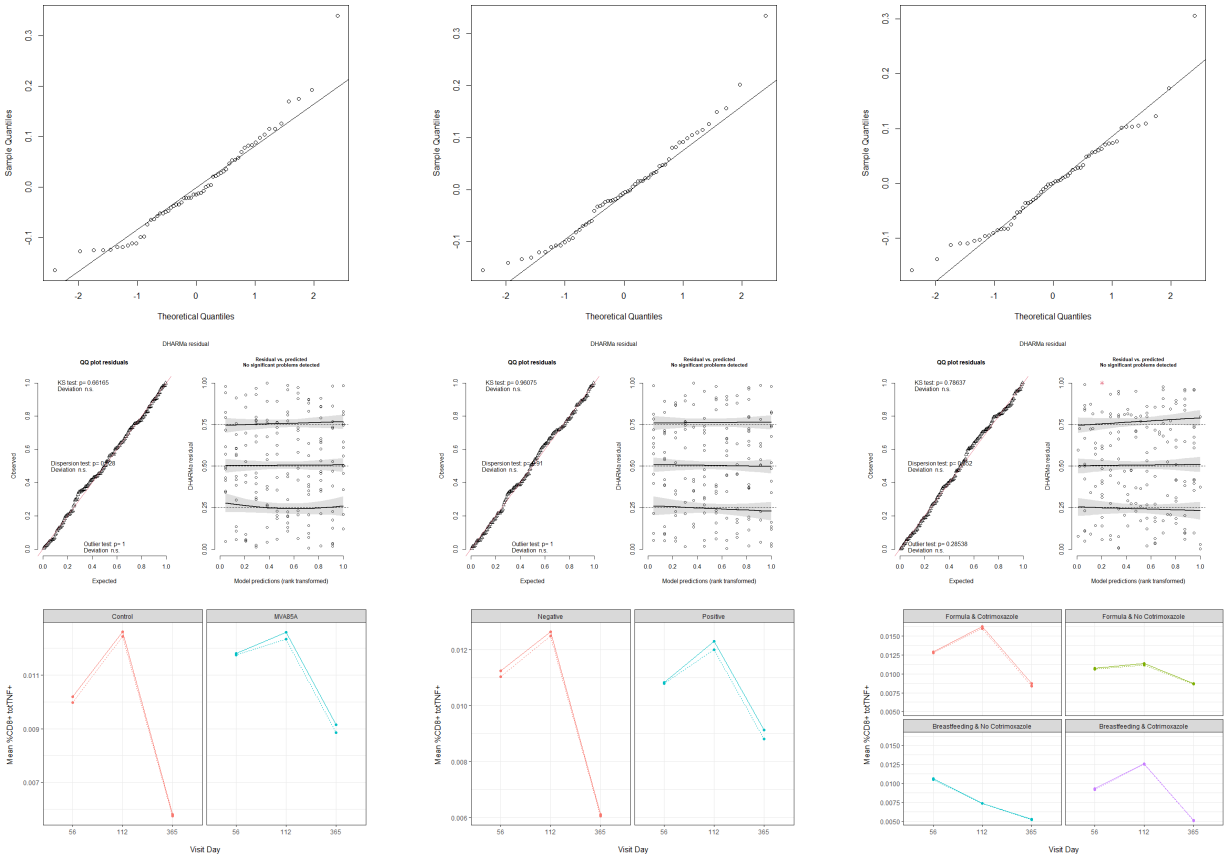
MVA85A priming Maternal QFT status Feeding & cotrimoxazole

Figure 9.13: The percentage of CD8$^+$ T cells expressing IL-17.



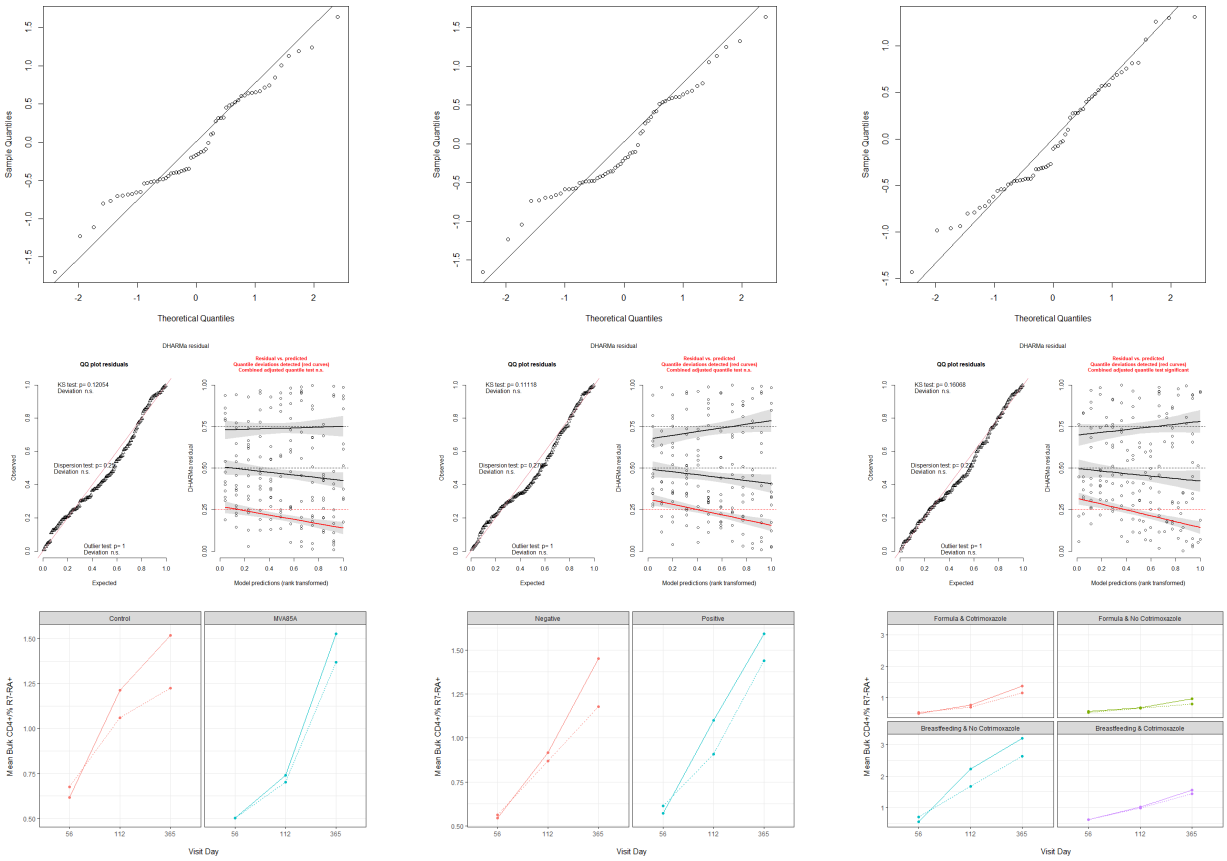
MVA85A priming Maternal QFT status Feeding & cotrimoxazole

Figure 9.14: The percentage of CD8$^+$ T cells expressing IL-22.



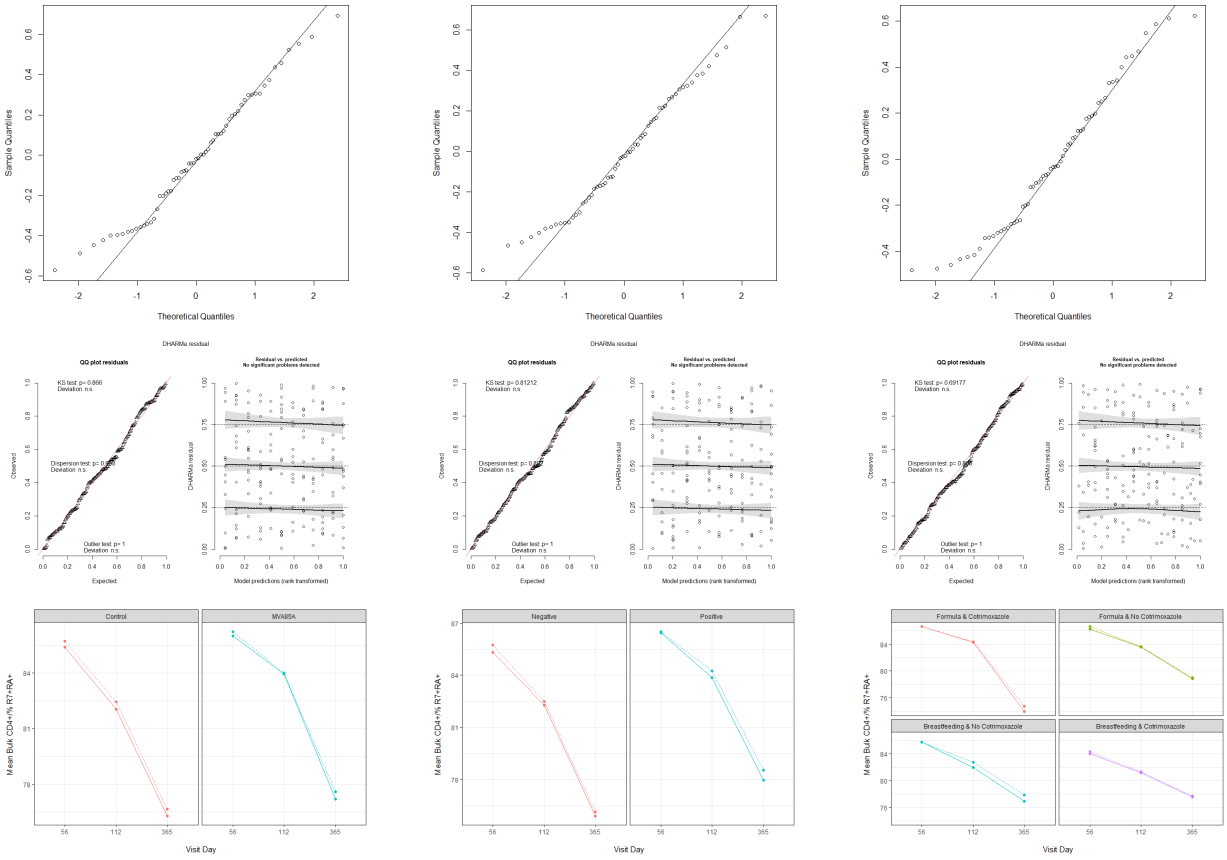
MVA85A priming Maternal QFT status Feeding & cotrimoxazole

Figure 9.15: The percentage of CD8$^+$ T cells expressing TNF.



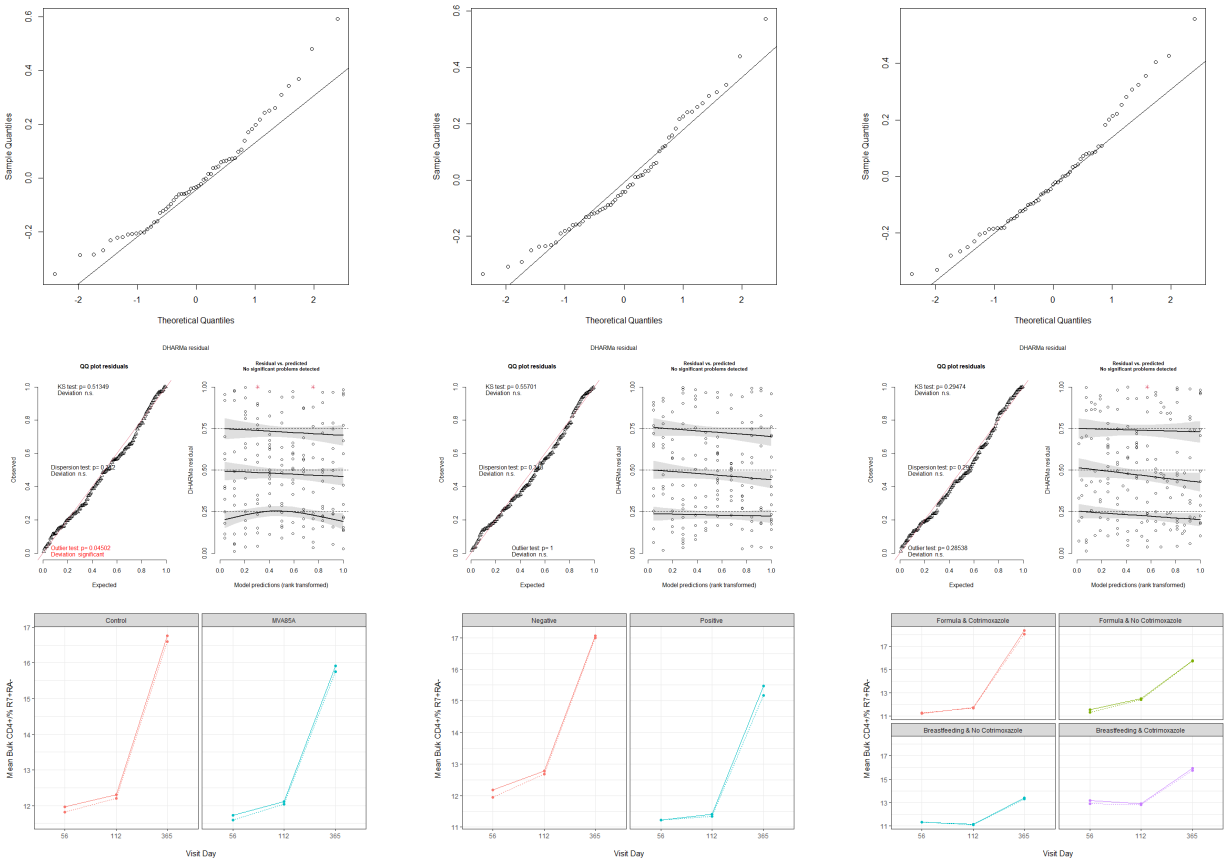
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.16: The percentage of bulk profiled $CD4^+$ T cells with $R7^-RA^+$ phenotype.



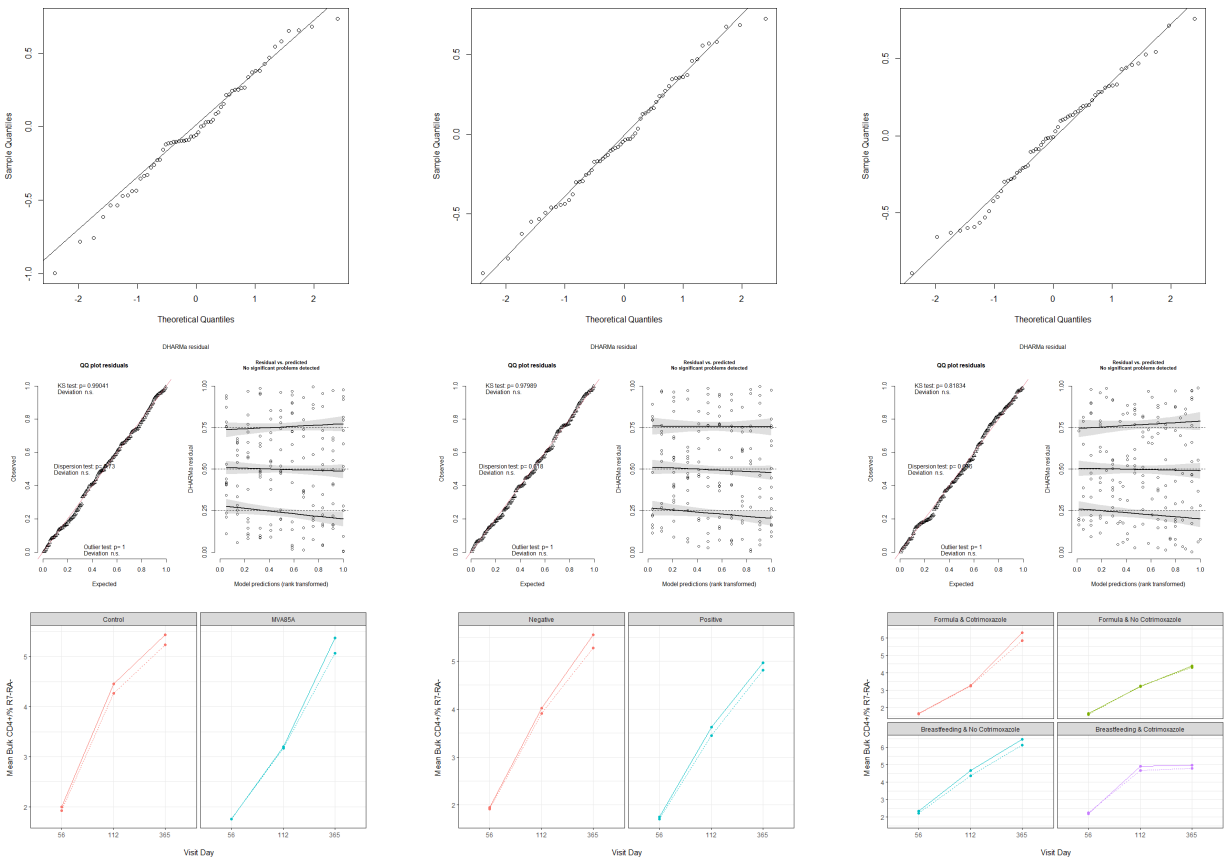
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.17: The percentage of bulk profiled $CD4^+$ T cells with $R7^+RA^+$ phenotype.



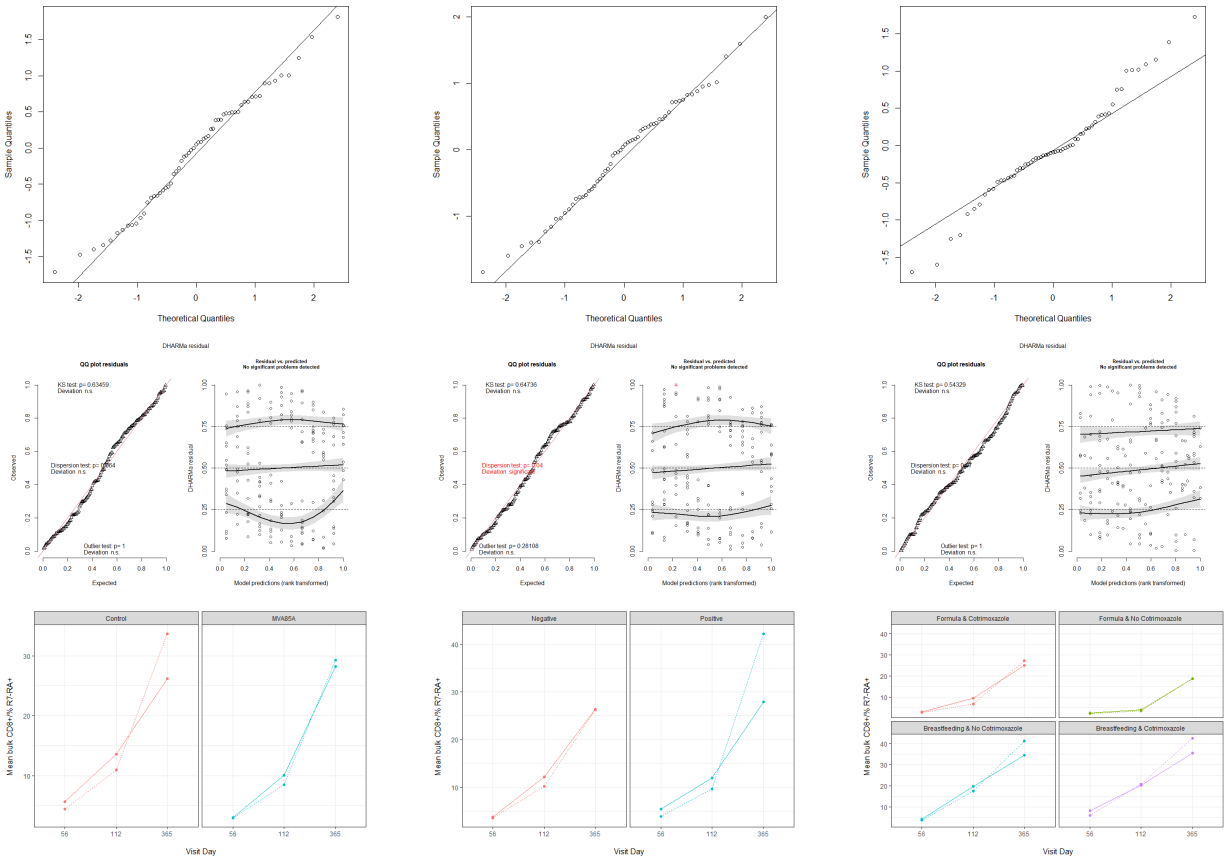
MVA85A priming Maternal QFT status Feeding & cotrimoxazole

Figure 9.18: The percentage of bulk profiled $CD4^+$ T cells with $R7^+RA^-$ phenotype.



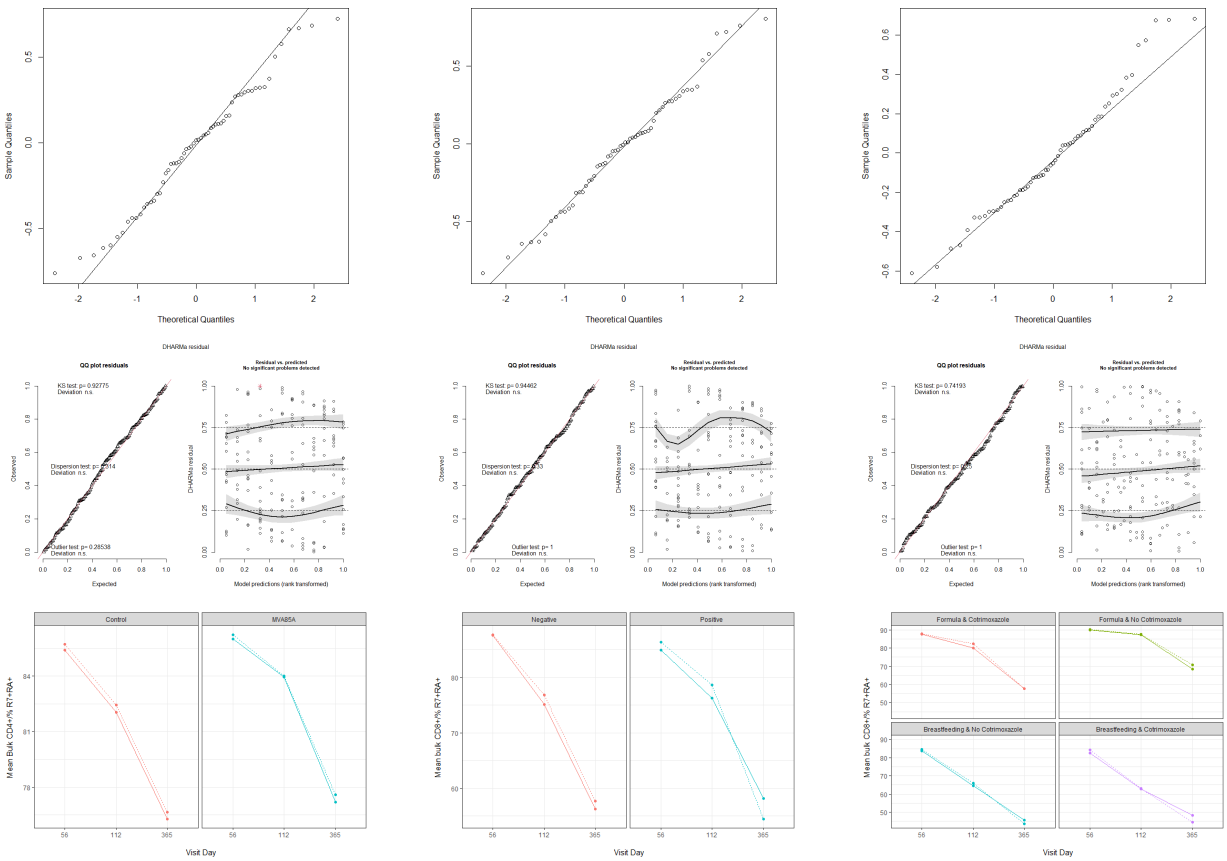
MVA85A priming Maternal QFT status Feeding & cotrimoxazole

Figure 9.19: The percentage of bulk profiled $CD4^+$ T cells with $R7^-RA^-$ phenotype.



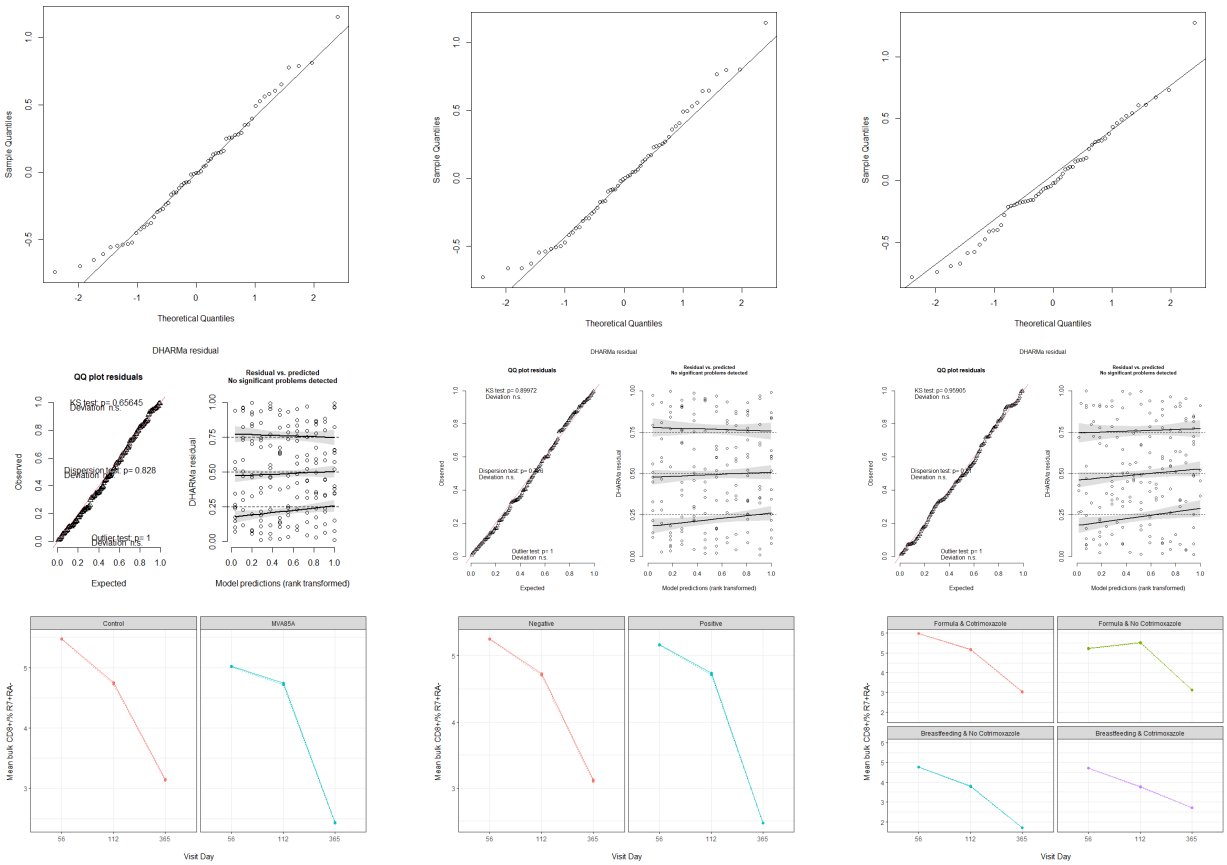
MVA85A priming Maternal QFT status Feeding & cotrimoxazole

Figure 9.20: The percentage of bulk profiled $CD8^+$ T cells with $R7^-RA^+$ phenotype.



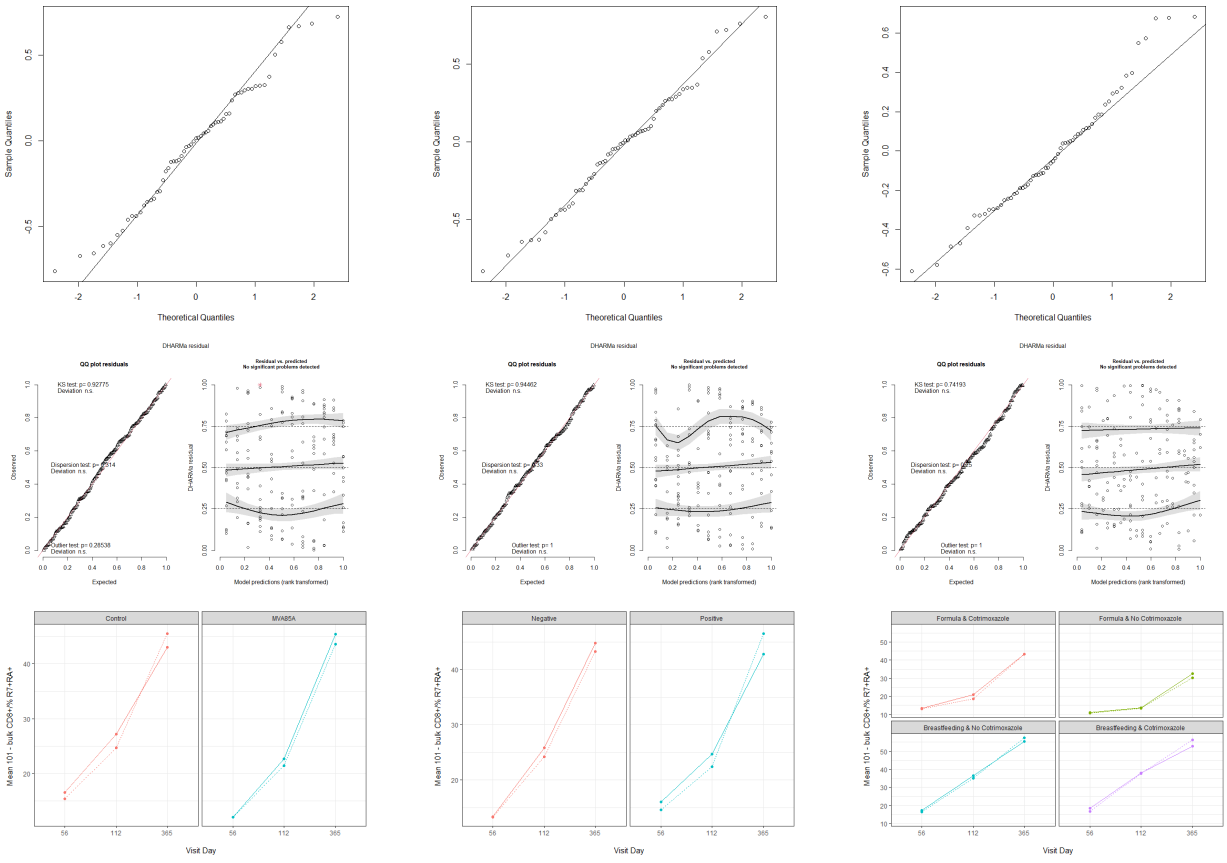
MVA85A priming Maternal QFT status Feeding & cotrimoxazole

Figure 9.21: The percentage of bulk profiled $CD8^+$ T cells with $R7^+RA^+$ phenotype.



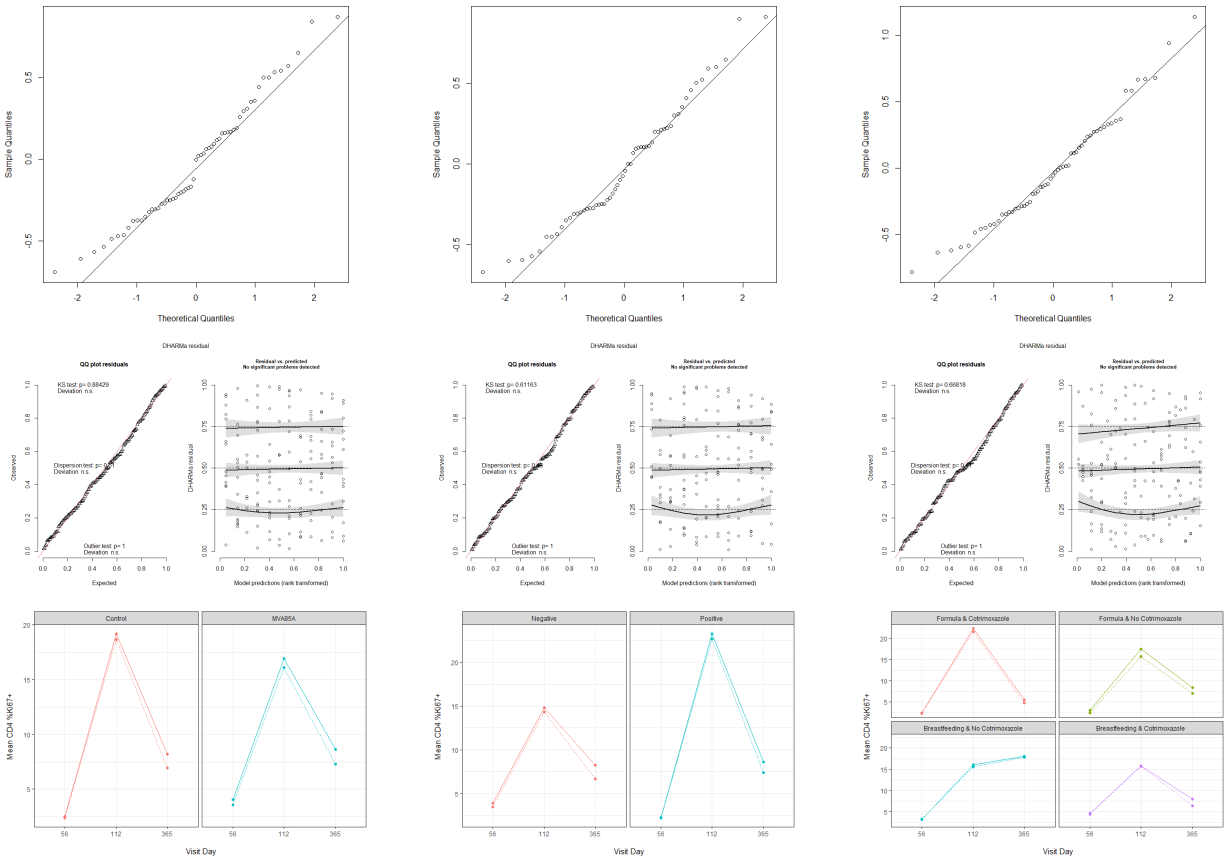
MVA85A priming Maternal QFT status Feeding & cotrimoxazole

Figure 9.22: The percentage of bulk profiled $CD8^+$ T cells with $R7^+RA^-$ phenotype.



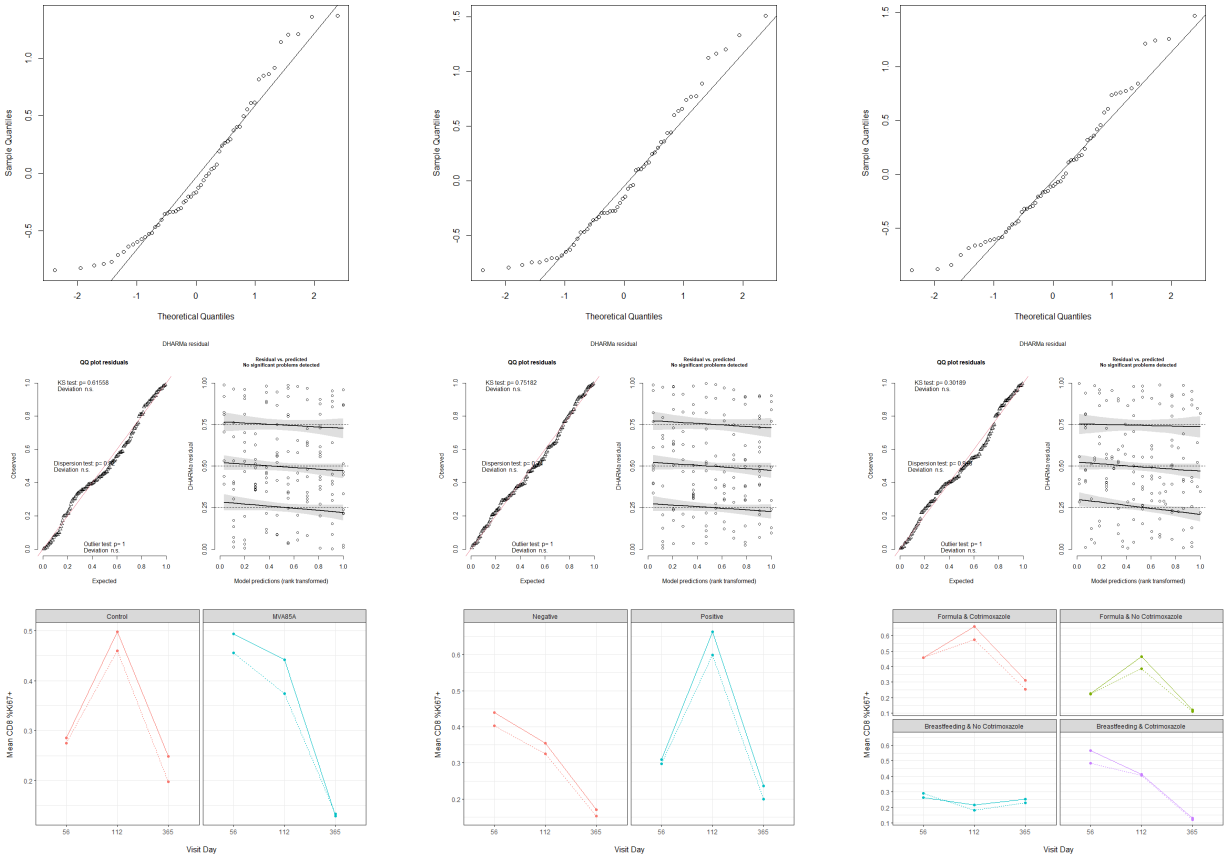
MVA85A priming Maternal QFT status Feeding & cotrimoxazole

Figure 9.23: The percentage of bulk profiled $CD8^+$ T cells with $R7^-RA^-$ phenotype.



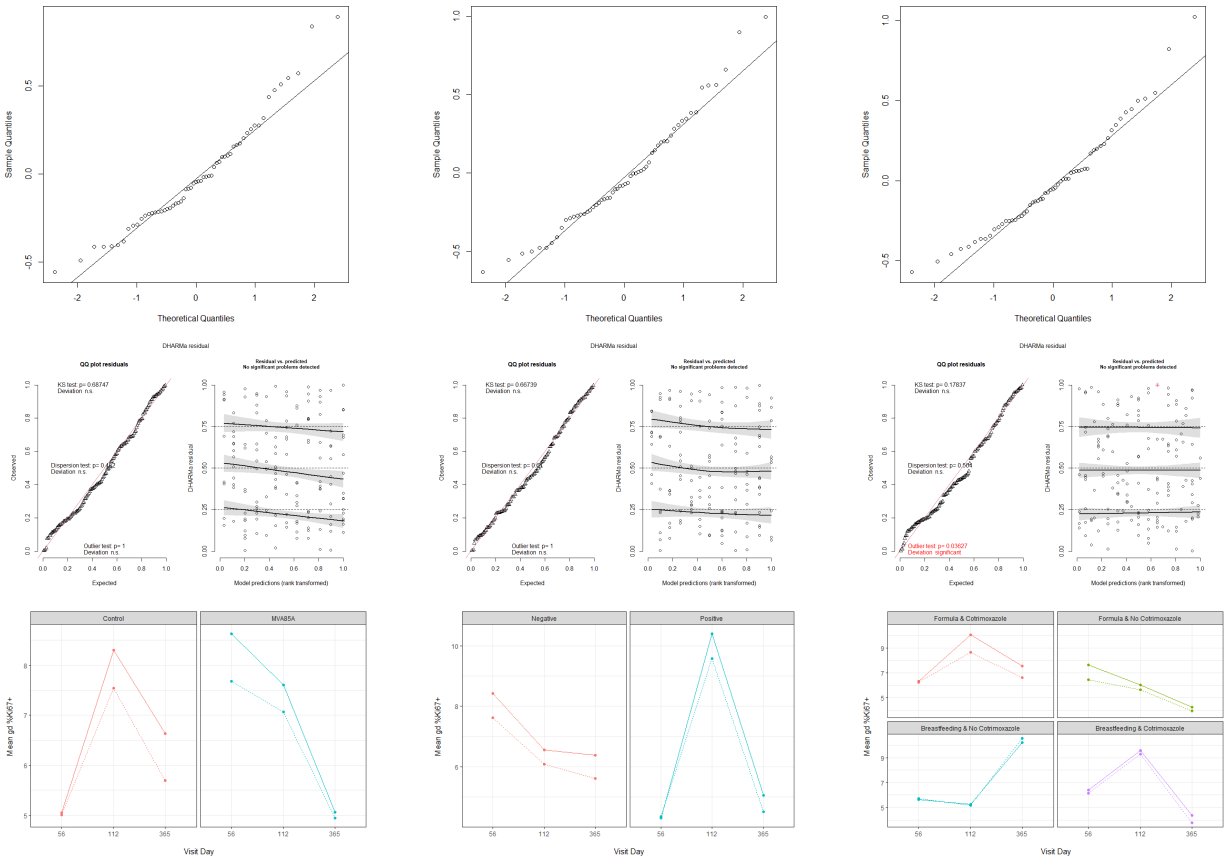
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.24: The percentage of proliferating Ki67$^+$ CD4$^+$ T cells.



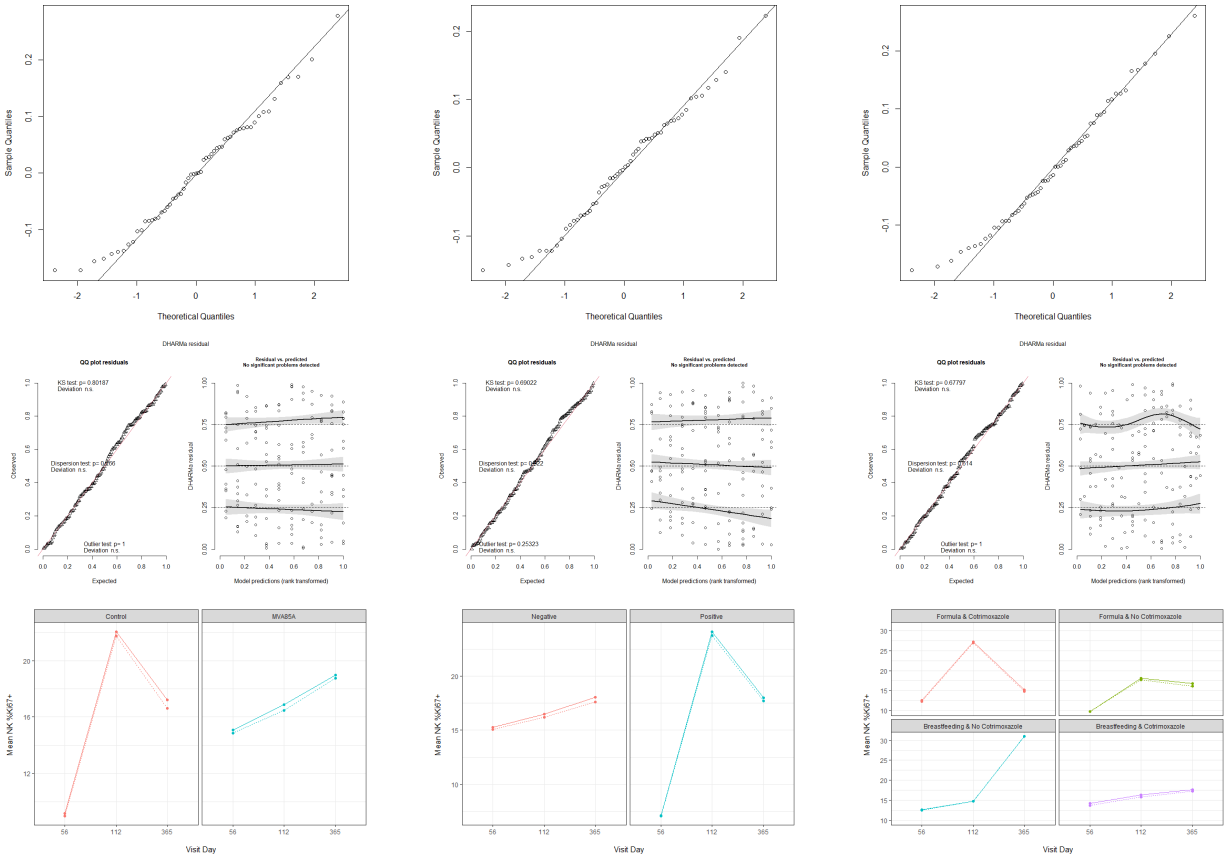
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.25: The percentage of proliferating Ki67$^+$ CD8$^+$ T cells.



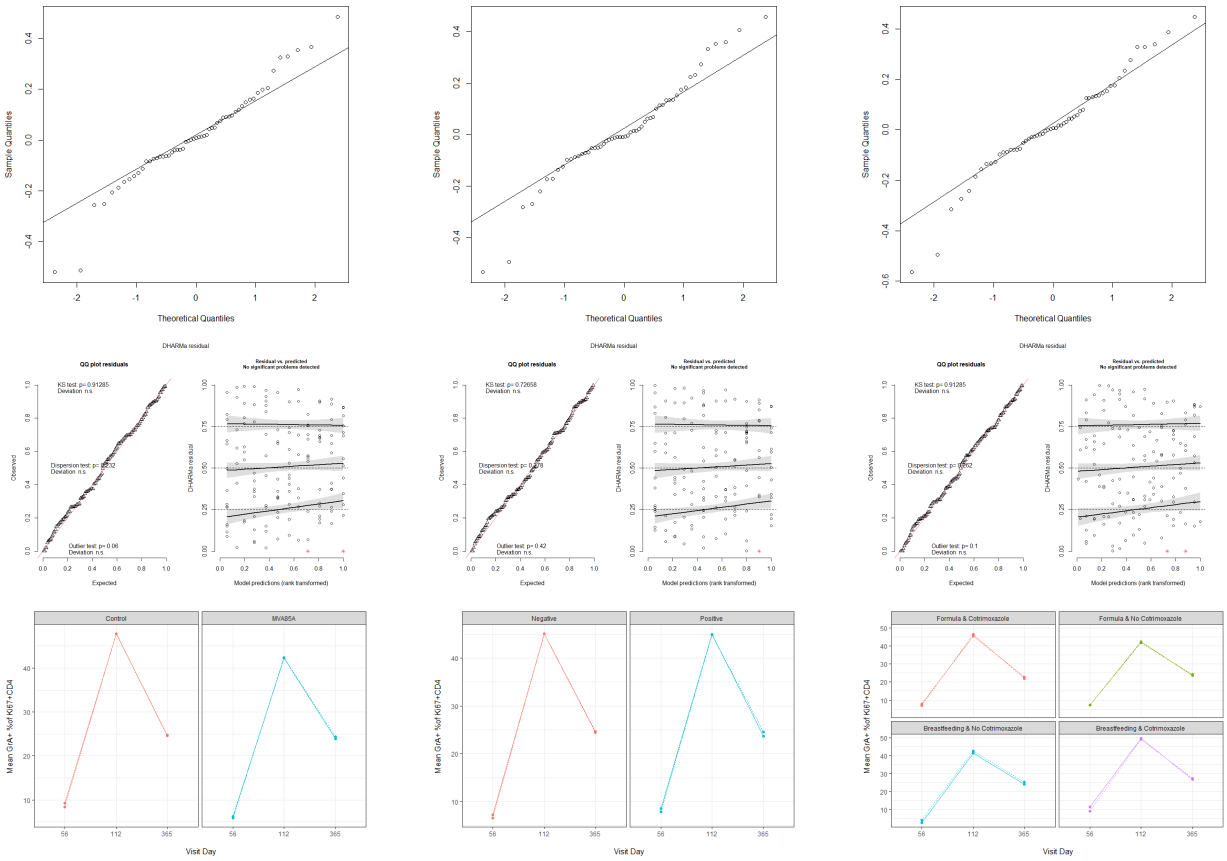
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.26: The percentage of proliferating Ki67$^+$ $\gamma\delta$ (gd) T cells.



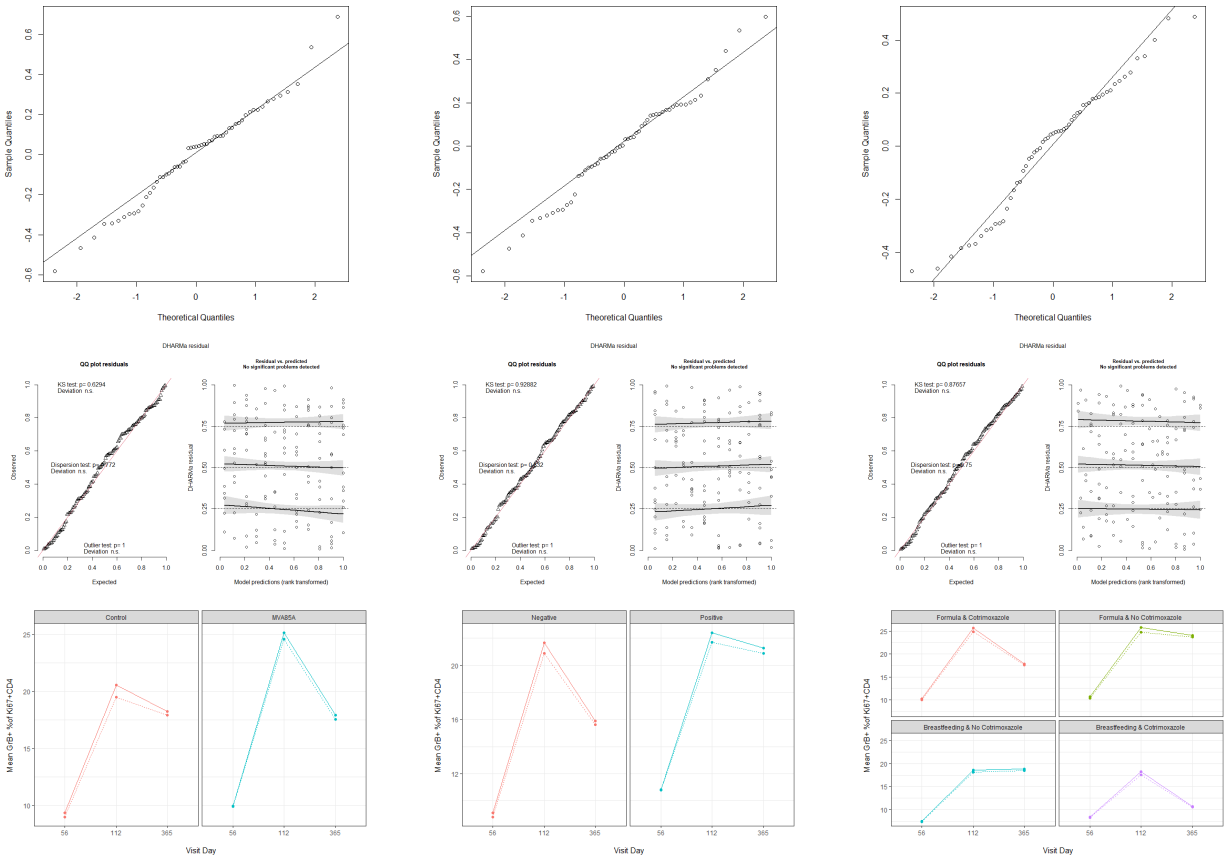
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.27: The percentage of proliferating Ki67$^+$ Natural Killer (NK) cells.



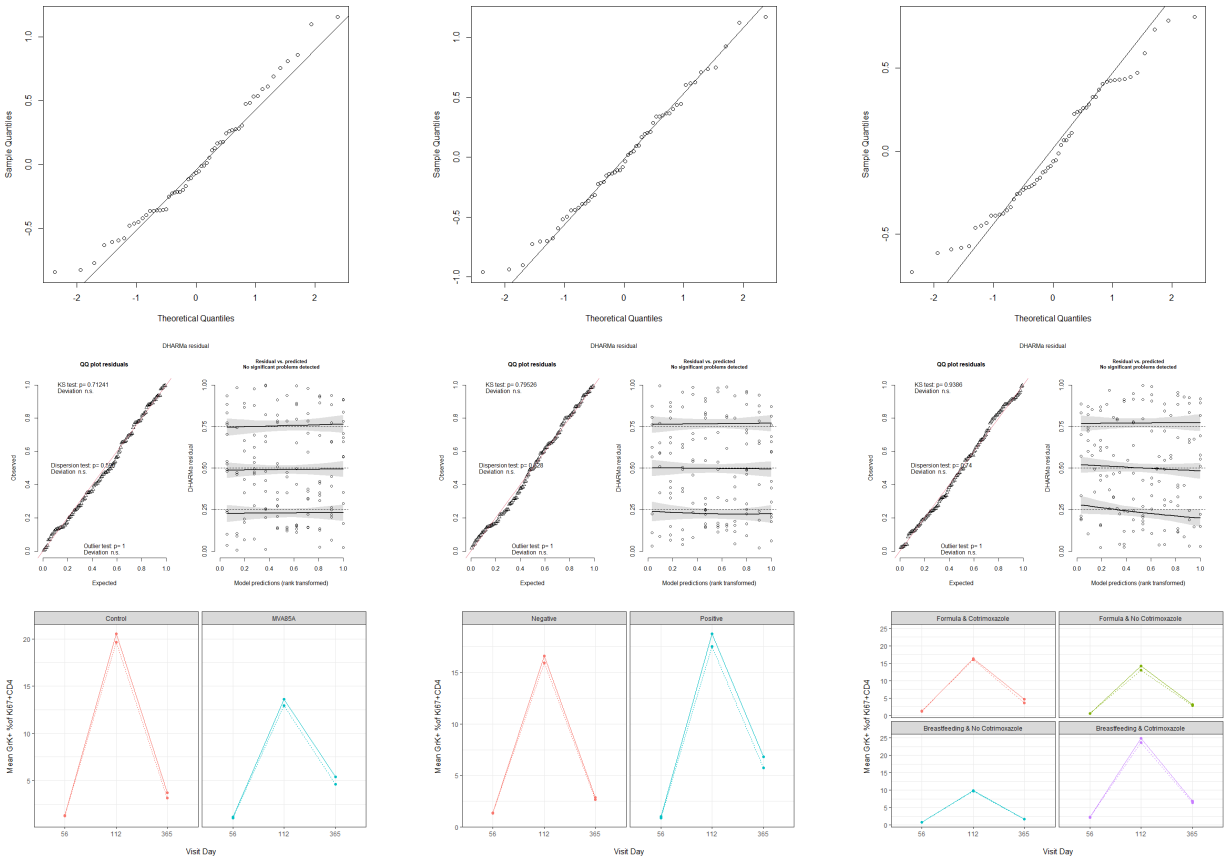
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.28: The percentage of proliferating Ki67$^+$ CD4$^+$ T cells expressing Granzyme A.



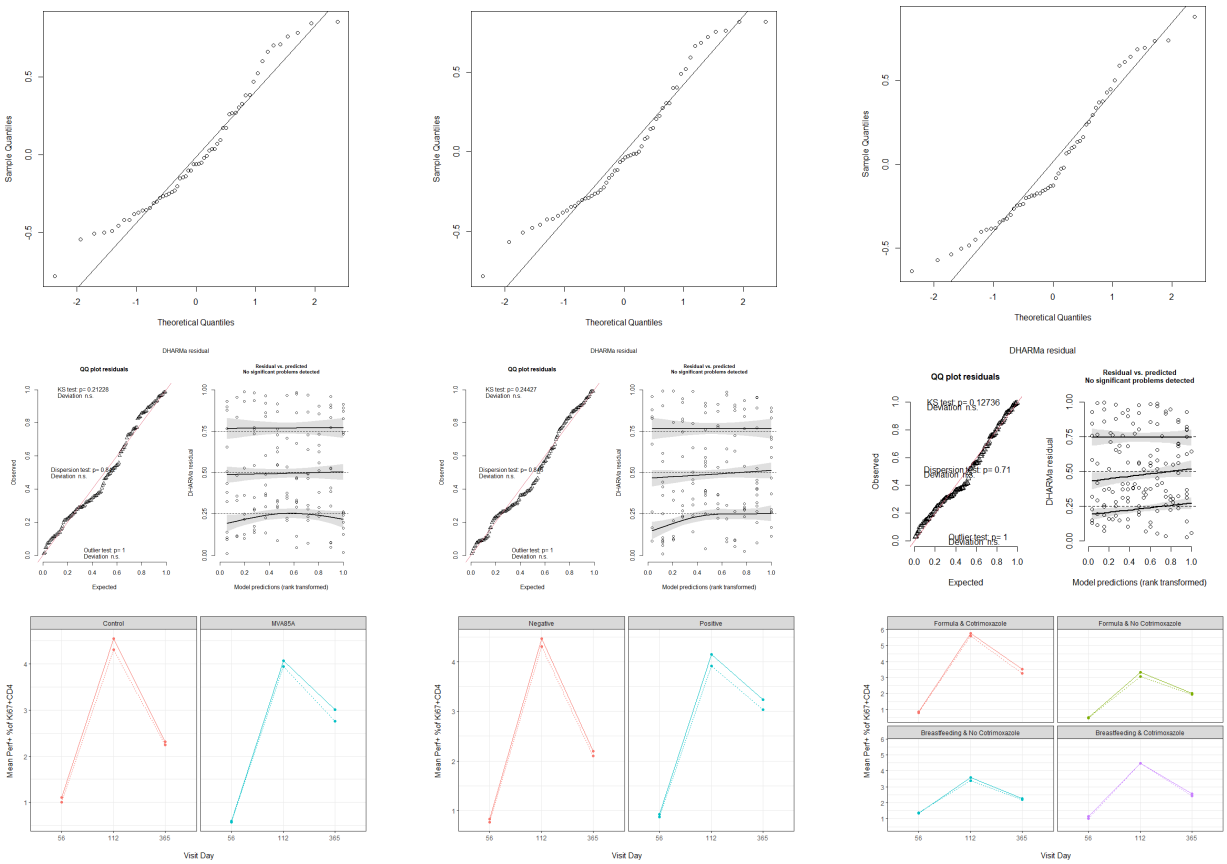
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.29: The percentage of proliferating Ki67$^+$ CD4$^+$ T cells expressing Granzyme B.



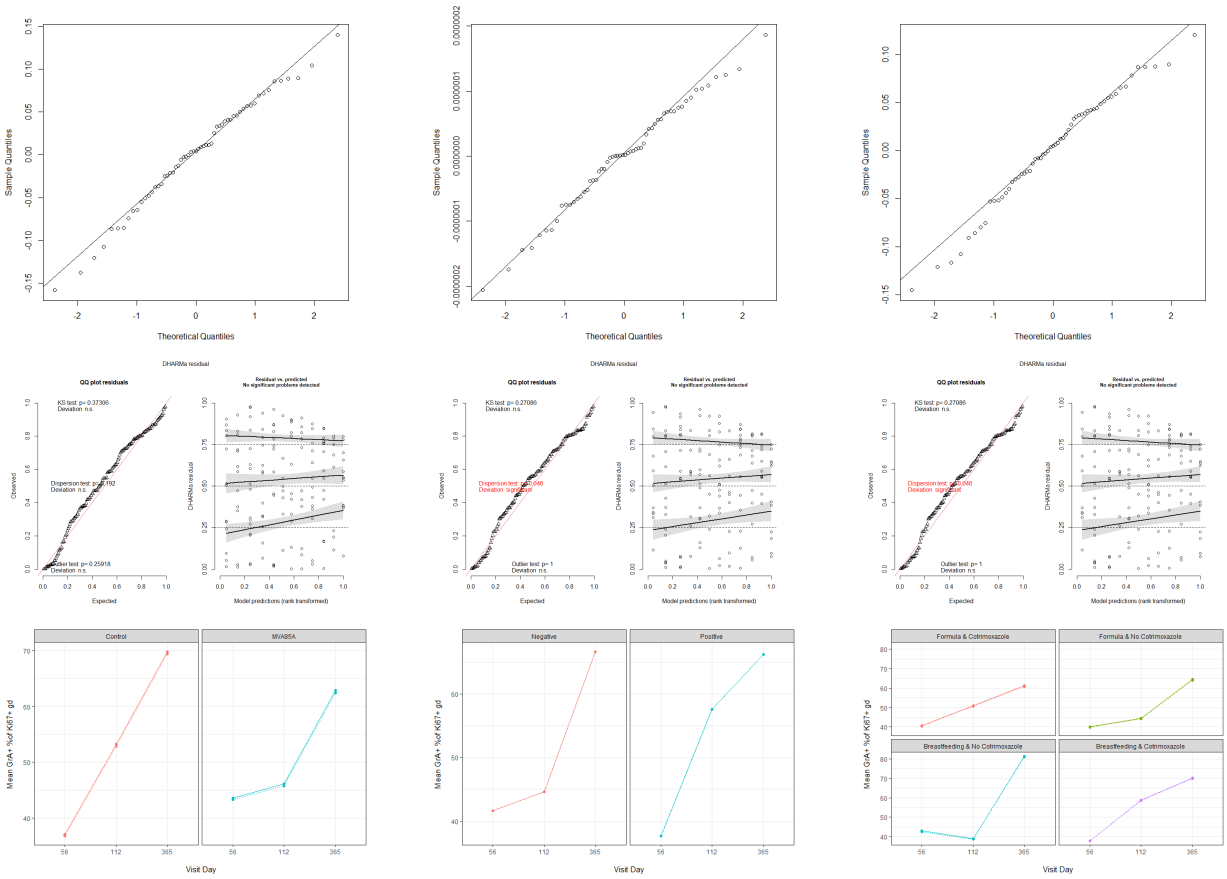
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.30: The percentage of proliferating Ki67$^+$ CD4$^+$ T cells expressing Granzyme K.



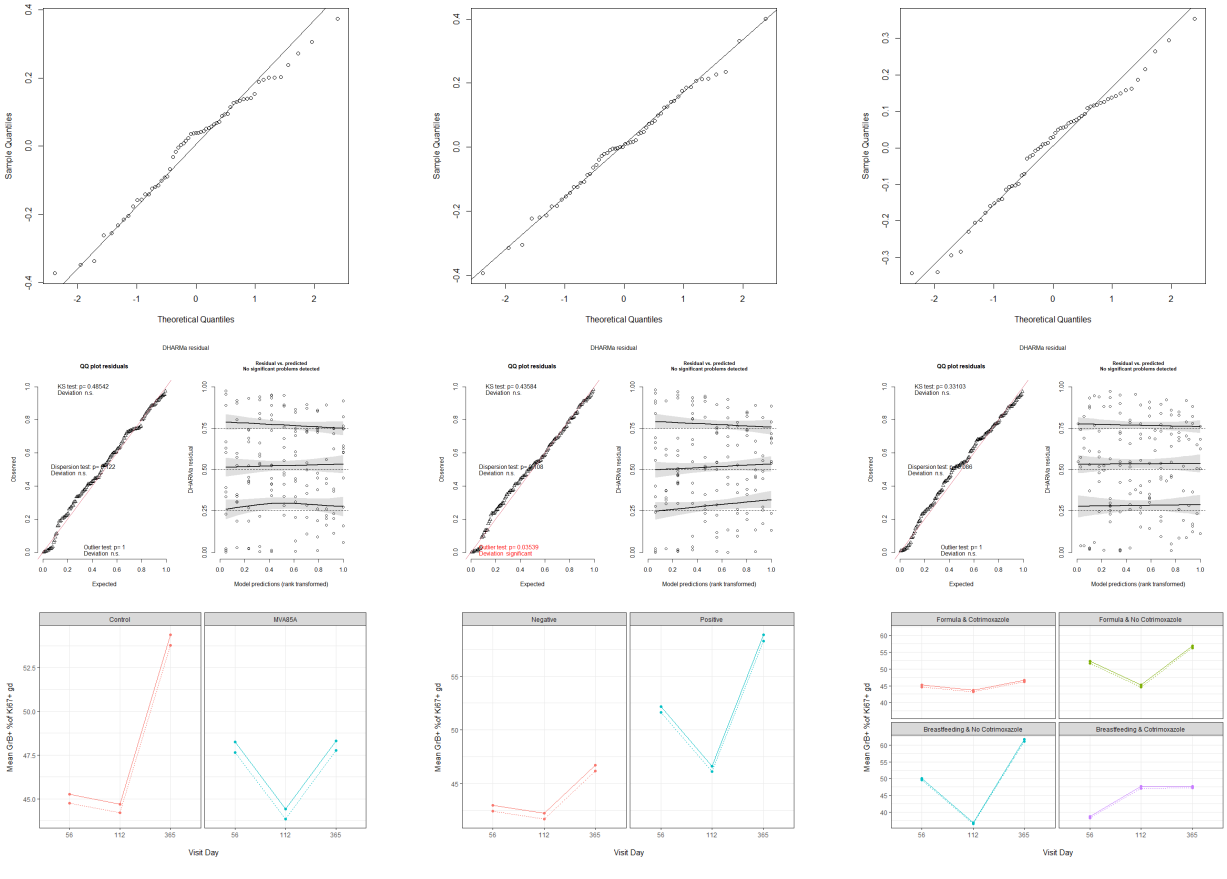
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.31: The percentage of proliferating Ki67$^+$ CD4$^+$ T cells expressing Perforin.



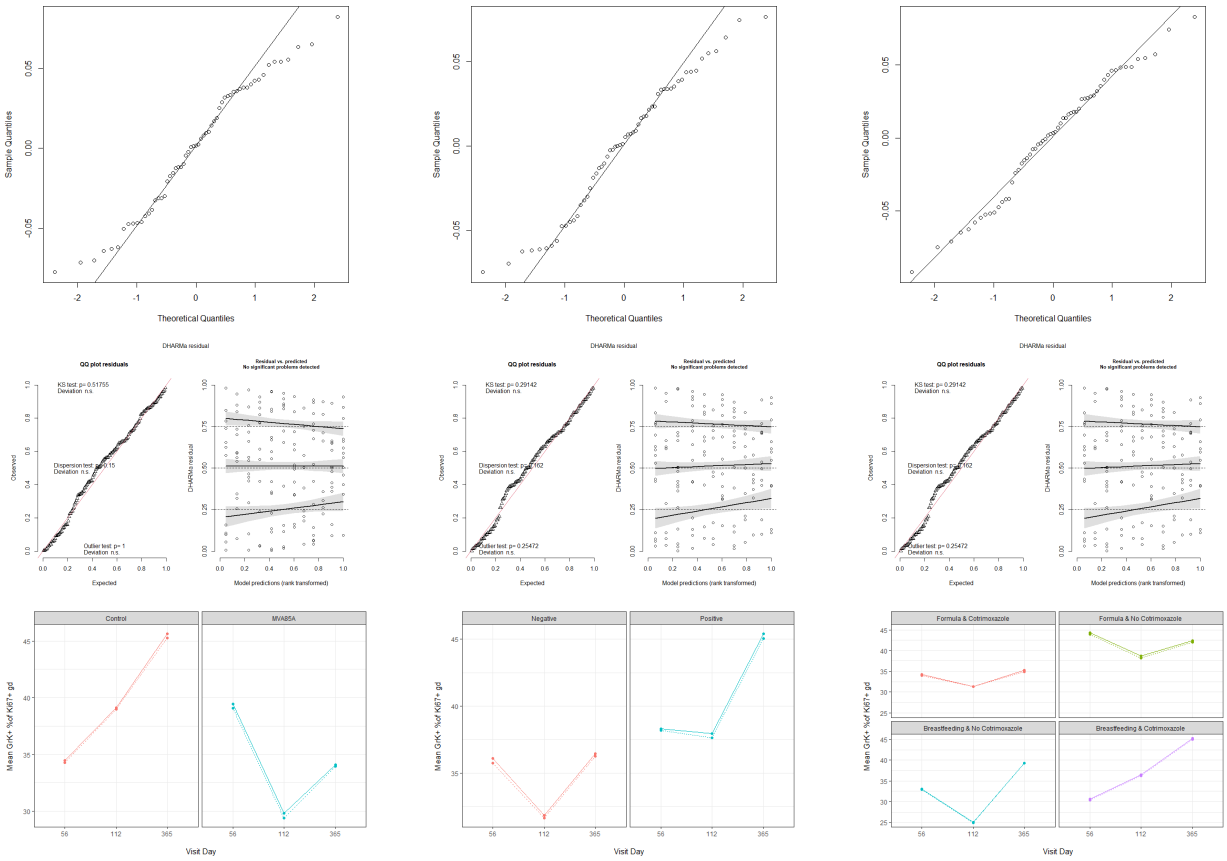
MVA85A priming Maternal QFT status Feeding & cotrimoxazole

Figure 9.32: The percentage of proliferating Ki67$^+$ $\gamma\delta$ (gd) T cells expressing Granzyme A.



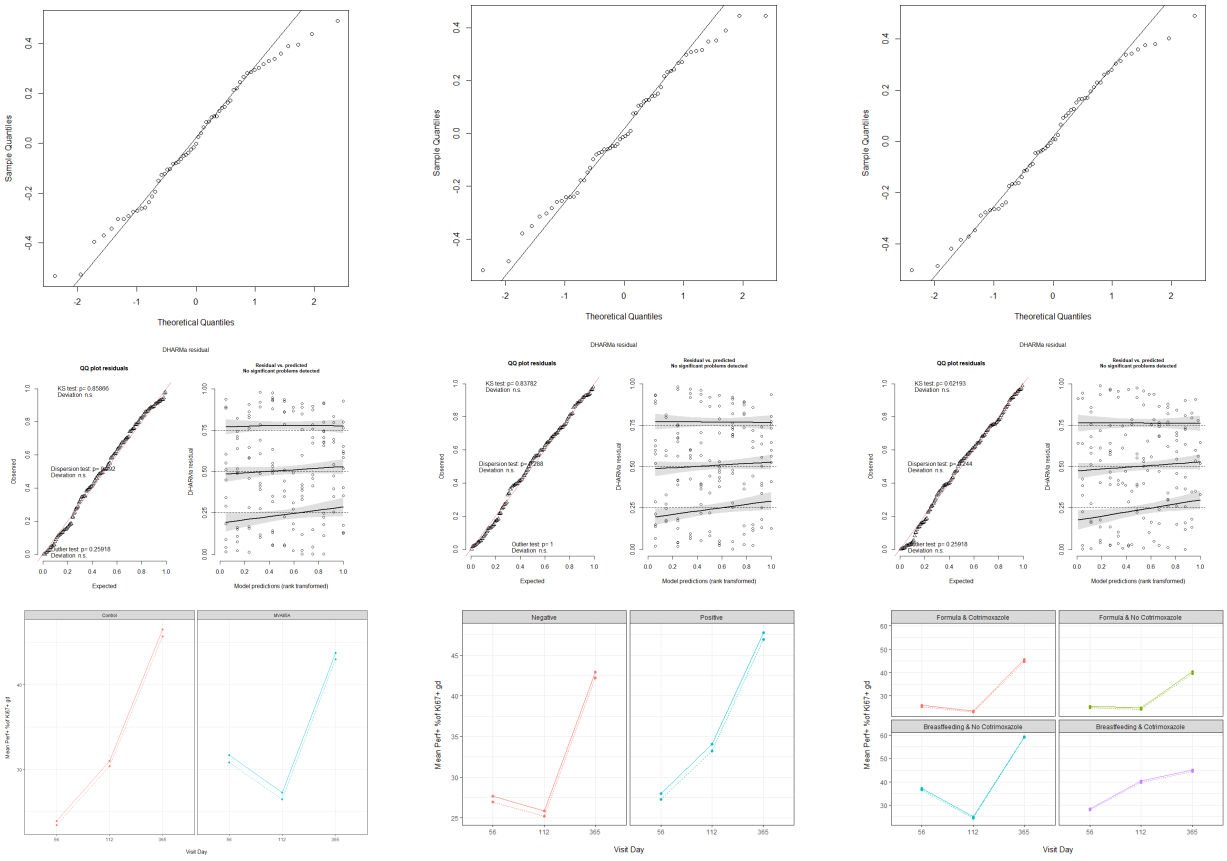
MVA85A priming Maternal QFT status Feeding & cotrimoxazole

Figure 9.33: The percentage of proliferating Ki67$^+$ $\gamma\delta$ (gd) T cells expressing Granzyme B.



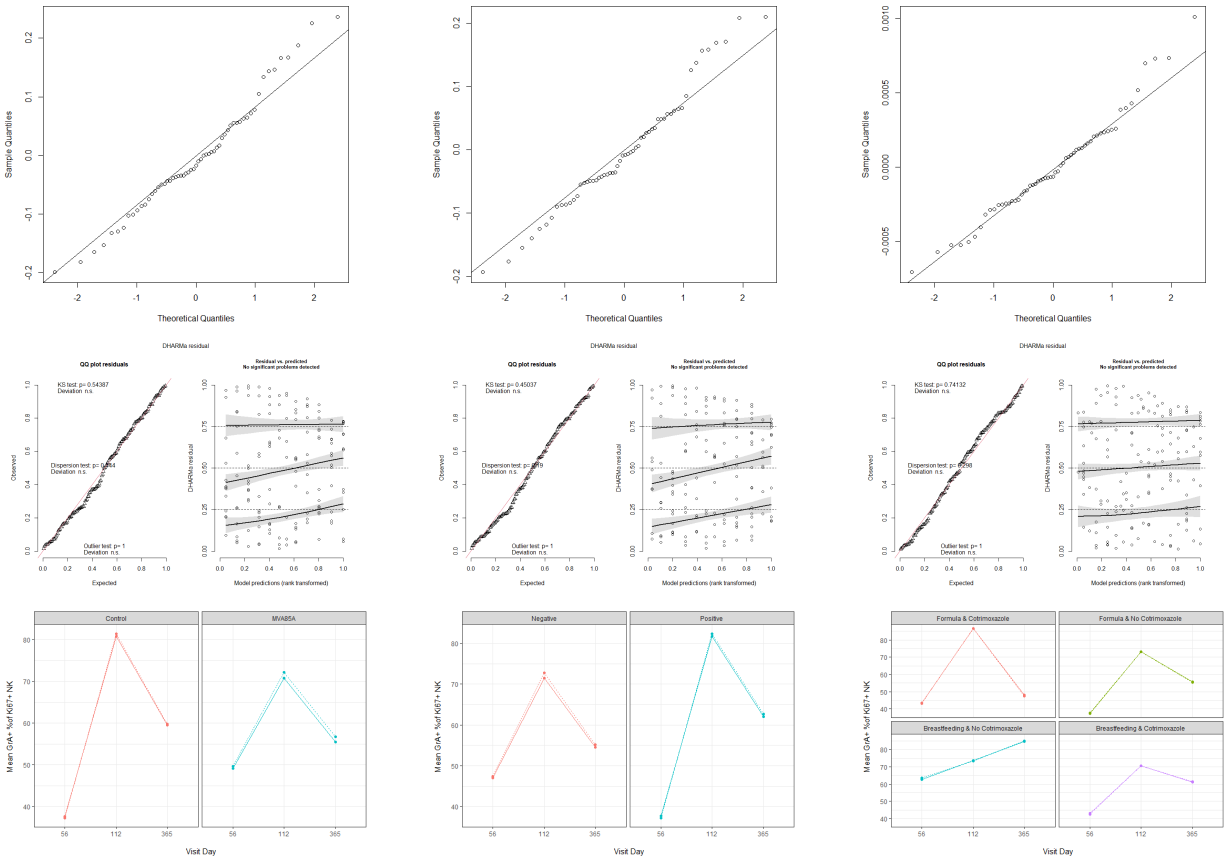
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.34: The percentage of proliferating Ki67$^+$ $\gamma\delta$ (gd) T cells expressing Granzyme K.



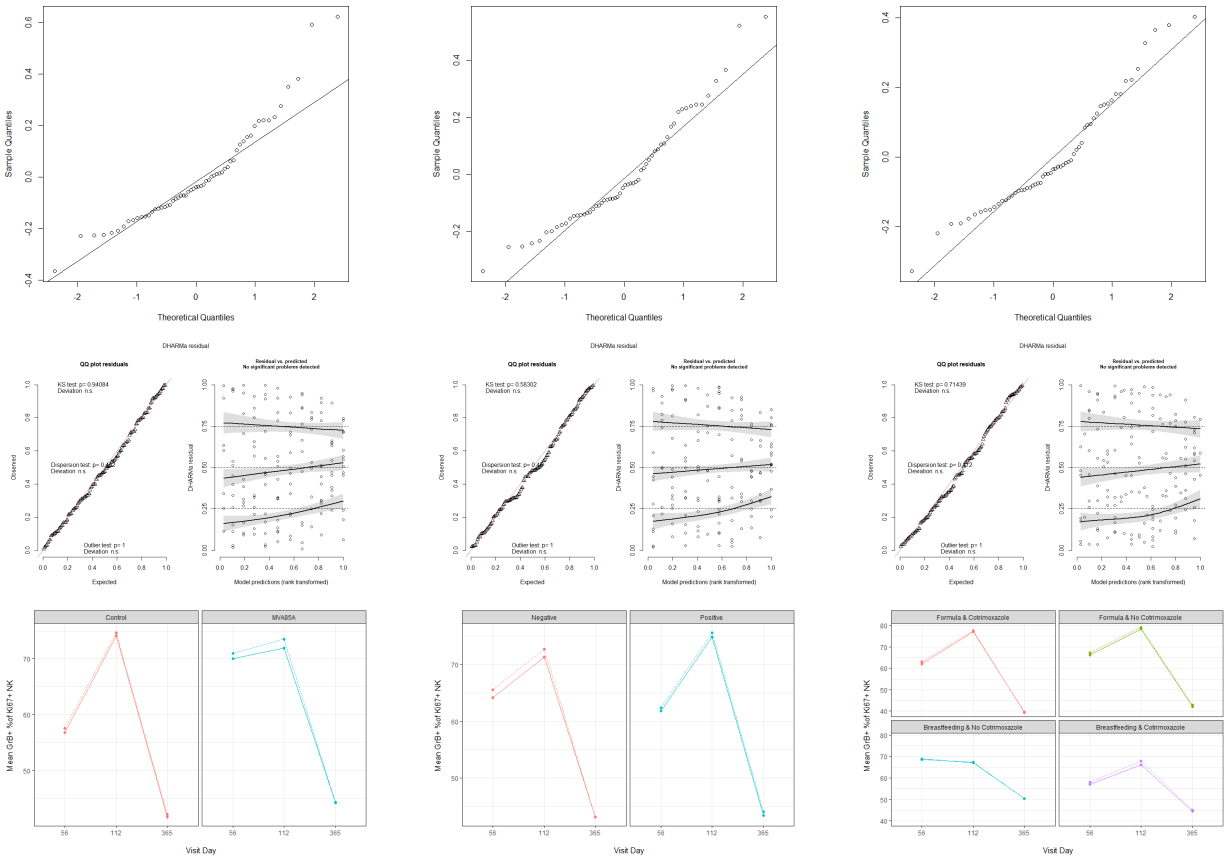
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.35: The percentage of proliferating Ki67$^+$ $\gamma\delta$ (gd) T cells expressing Perforin.



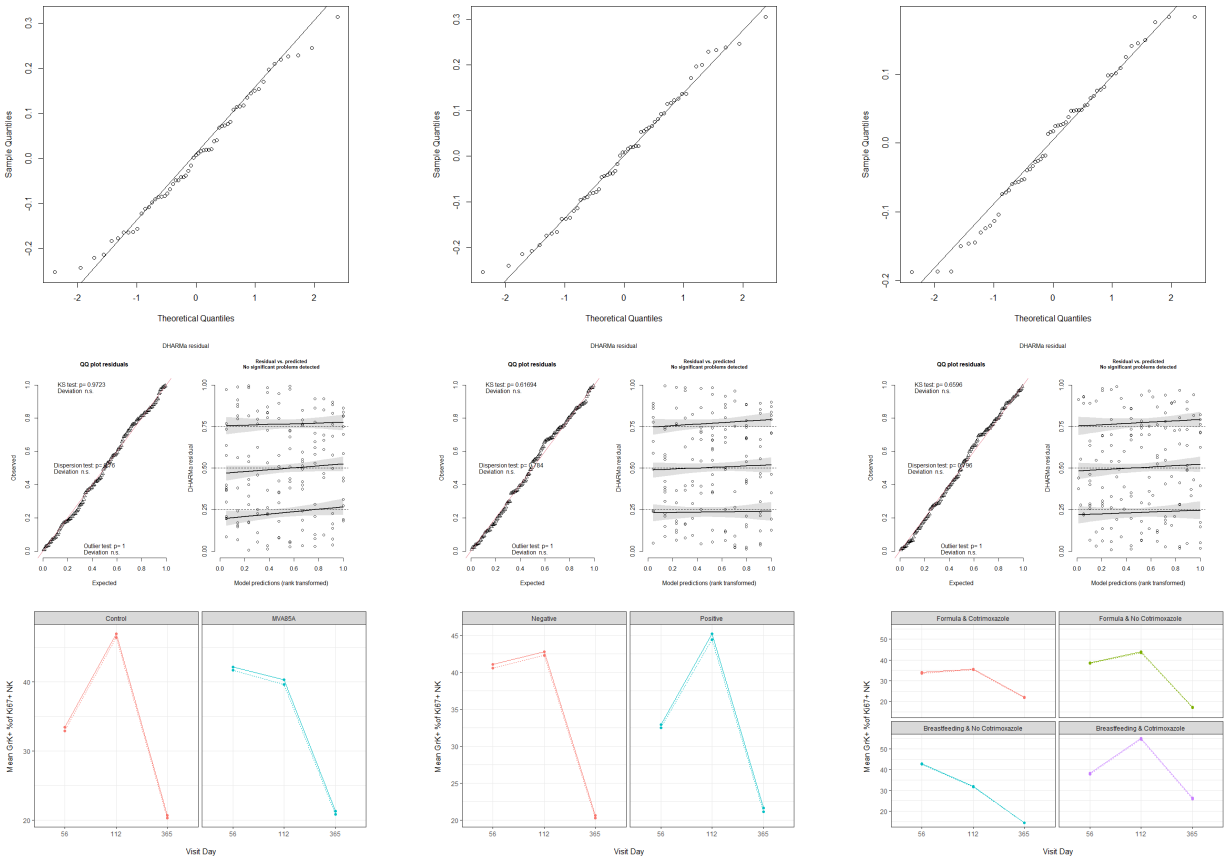
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.36: The percentage of proliferating Ki67$^+$ Natural Killer (NK) cells expressing Granzyme A.



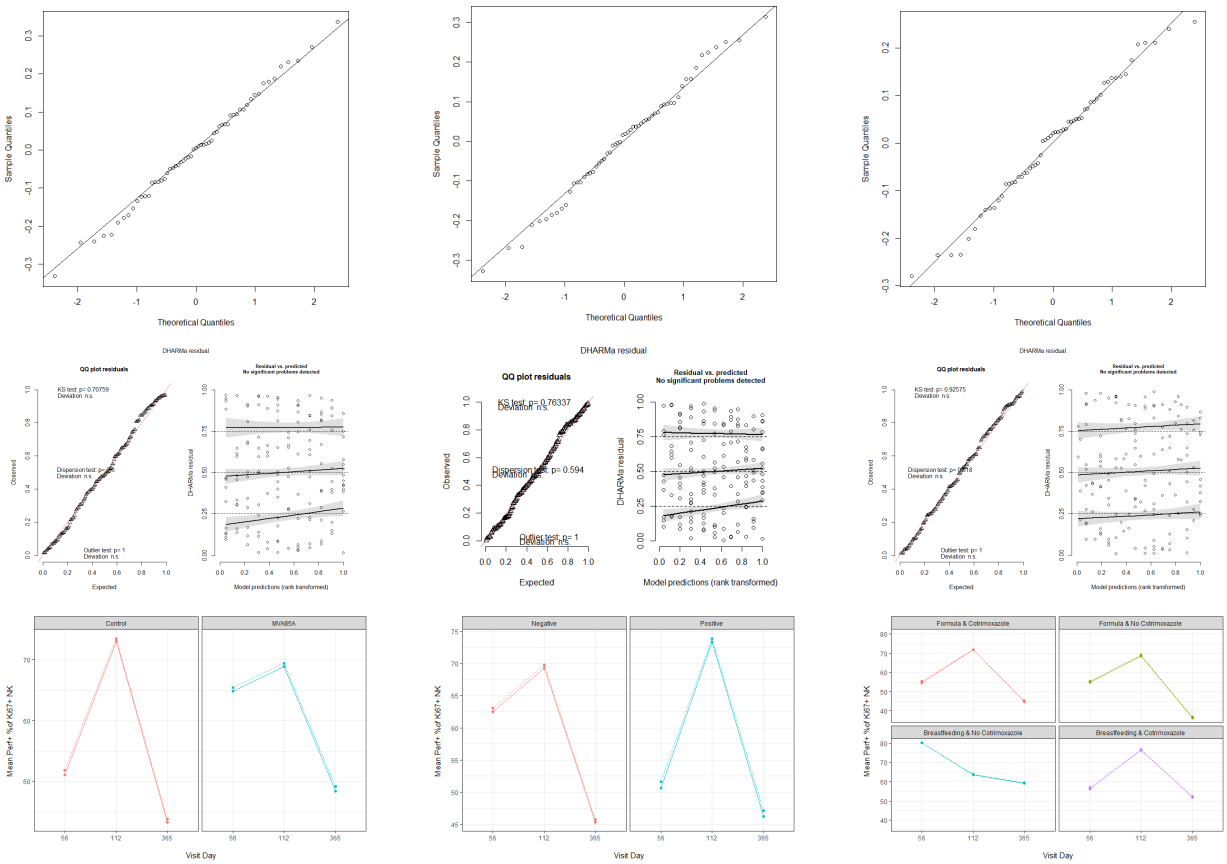
MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.37: The percentage of proliferating Ki67$^+$ Natural Killer (NK) cells expressing Granzyme B.



MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.38: The percentage of proliferating Ki67$^+$ Natural Killer (NK) cells expressing Granzyme K.



MVA85A priming | Maternal QFT status | Feeding & cotrimoxazole

Figure 9.39: The percentage of proliferating Ki67$^+$ Natural Killer (NK) cells expressing Perforin.

Appendix H

Application of Dimension Reduction Techniques

PCA Only

Dimension reduction by PCA only is presented below for the exposures of interest *MVA85A* and *QFT* in Group A. Contributions to first $q^* < q$ total PCs explaining at least 60% of the variance in the input data were investigated to identify a subset of the complete set of immune outcomes for further analysis of the comparisons of interest.

MVA85A Priming

A PCA was performed on the correlation matrix of the following standardised model coefficients for the Group A outcomes: $MVA85A^{(1)}$, $MVA85A : Time1^{(2)}$ and $MVA85A : Time3^{(3)}$.

As 33 immune outcomes are measured in Group A, PCA is performed on a 33×3 input data set. This produces three principal components (PCs). The scree plot of variance in the standardised model coefficients explained by each component is included below.

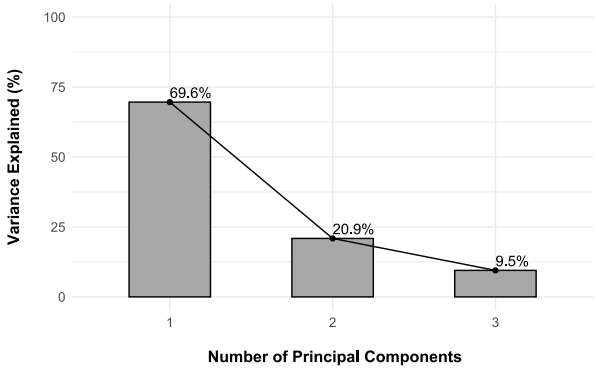


Figure 9.40: Scree plot of the percentage variance explained in the standardised model coefficients for MVA85A priming by the principal components.

The first PC explains 69.6% of the variance in the standardised model coefficients. In other words, for the 33 immune outcomes in Group A, PC1 captures most of the variance in the longitudinal effects of MVA85A priming. PC1 can act as a lower-dimensional representation of these effects in Group A.

Contributions of standardised model coefficients and immune outcomes were then investigated for PC1. The scree plot below shows which standardised model coefficients make large contributions to the variance explained by PC1.

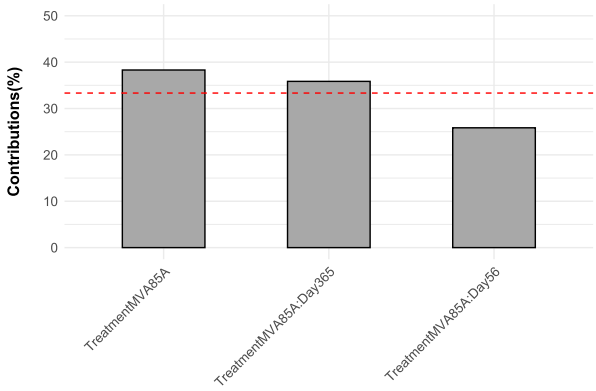


Figure 9.41: Scree plot of the percentage variance in PC1 contributed by the standardised model coefficients for *MVA85A*.

PC1 primarily describes variance in the standardised coefficients estimated for 8 weeks after BCG vaccination ($MVA85A^{(1)}$) and $\approx$ 44 weeks after BCG vaccination ($MVA85A : Time3^{(3)}$).

The scree plot below identifies nine immune outcomes that make a larger contribution to PC1 than would be expected under the assumption of uniform contributions.

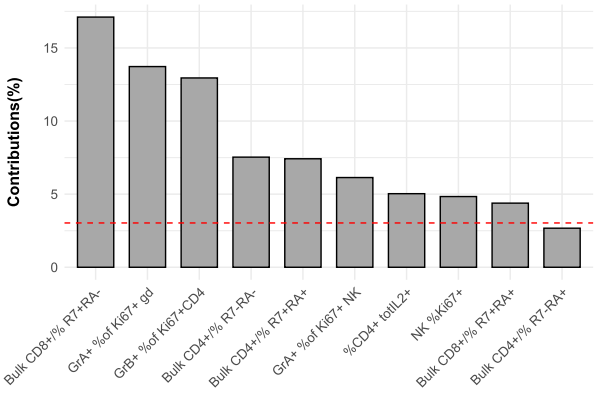


Figure 9.42: Scree plot of the percentage variance in *MVA85A* contributed by the Group A immune outcomes to PC1.

Hence, the nine immune outcomes making large contributions to PC1 are selected as an immune outcome subset for further analysis:

- $Bulk\ \%CD8^{+}\ R7^{+}\ RA^{-(1)}$
- $\%GrA^{+}\ Ki67^{+}\ gd^{(2)}$
- $\%GrB^{+}\ Ki67^{+}\ CD4^{+(3)}$
- $Bulk\ \%CD4^{+}\ R7^{-}\ RA^{-(4)}$
- $Bulk\ \%CD4^{+}\ R7^{+}\ RA^{+(5)}$
- $\%GrA^{+}\ Ki67^{+}\ NK^{(6)}$
- $\%CD4\ totIL2^{+(7)}$
- $\%NK\ Ki67^{+(8)}$
- $Bulk\ \%CD8^{+}\ R7^{+}\ RA^{+(9)}$

Instead of analysing the longitudinal effects of MVA85A priming for all 33 outcomes, the analysis can proceed with a subset of nine immune outcomes describing the majority of the variance in the effects.

Maternal QFT

A PCA was performed on the correlation matrix of the following standardised model coefficients for the Group A outcomes: $QFT^{(1)}$, $QFT : Time1^{(2)}$ and $QFT : Time3^{(3)}$.

As 33 immune outcomes are measured in Group A, PCA is performed on a 33×3 input data set. This produces three principal components (PCs). The scree plot of variance in the standardised model coefficients explained by each component is included below.

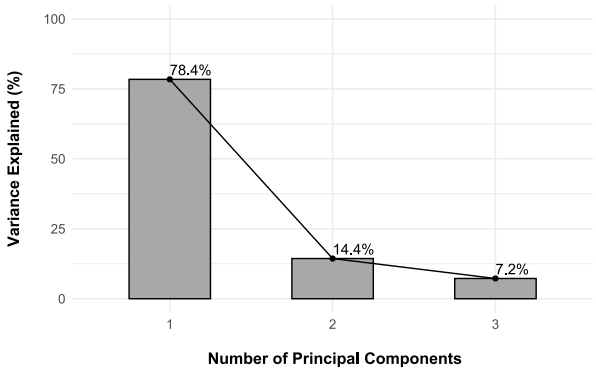


Figure 9.43: Scree plot of the percentage variance explained in the standardised model coefficients for QFT by the principal components.

The first PC explains 78.4% of the variance in the standardised model coefficients. In other words, for the 33 immune outcomes in Group A, PC1 captures most of the variance in the longitudinal effects of a positive maternal QFT. PC1 can act as a lower-dimensional representation of these effects in Group A.

Contributions of standardised model coefficients and immune outcomes were then investigated for PC1. The scree plot below shows which standardised model coefficients make large contributions to the variance explained by PC1.

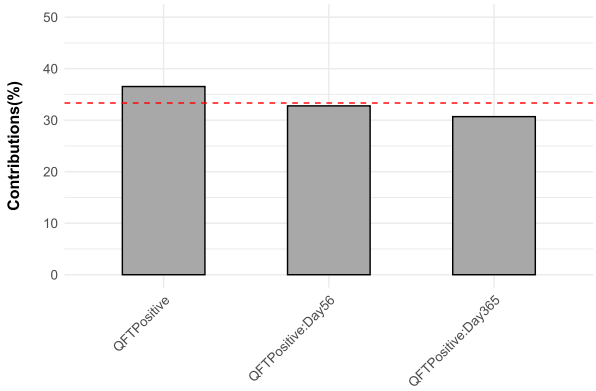


Figure 9.44: Scree plot of the percentage variance in PC1 contributed by the standardised model coefficients for QFT.

PC1 primarily describes variance in the standardised coefficients estimated for 8 weeks after BCG vaccination ($QFT^{(1)}$).

The scree plot below identifies eleven immune outcomes that make a larger contribution to PC1 than would be expected under the assumption of uniform contributions.

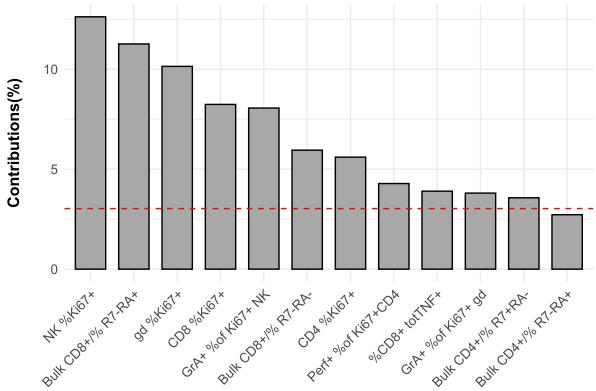


Figure 9.45: Scree plot of the percentage variance in *QFT* contributed by the Group A immune outcomes to PC1.

Hence, the eleven immune outcomes making large contributions to PC1 are selected as an immune outcome subset for further analysis:

1. $\%NK\ Ki67^{+(1)}$
2. $Bulk\ \%CD8^{+}\ R7^{-}\ RA^{+(2)}$
3. $\%gd\ Ki67^{+(3)}$
4. $\%CD8\ Ki67^{+(4)}$
5. $\%GrA^{+}\ Ki67^{+}\ NK^{(5)}$
6. $Bulk\ \%CD8^{+}\ R7^{-}\ RA^{-(6)}$
1. $\%CD4\ Ki67^{+(7)}$
2. $\%Perf^{+}\ Ki67^{+}\ CD4^{+(8)}$
3. $\%CD8\ totTNF^{+(9)}$
4. $\%GrA^{+}\ Ki67^{+}\ gd^{(10)}$
5. $Bulk\ \%CD4^{+}\ R7^{+}\ RA^{-(11)}$

Instead of analysing the longitudinal effects of a positive maternal QFT for all 33 outcomes, the analysis can proceed with a subset of eleven immune outcomes describing the majority of the variance in the effects.

HCA Then PCA

Dimension reduction by HCA followed by PCA is presented below for the exposures of interest *MVA85A* and *QFT* in Group A. Contributions to first $q^* < q$ total PCs explaining at least 60% of the variance in the input data were investigated to identify a subset of the complete set of immune outcomes for further analysis of the comparisons of interest.

MVA85A Priming

A hierarchical cluster analysis with average linkage was performed on the Pearson correlation distance matrix of the following standardised model coefficients for the Group A outcomes, adjusted for the average sex effect: $MVA85A^{(1)}$, $MVA85A : Time1^{(2)}$ and $MVA85A : Time3^{(3)}$.

As 33 immune outcomes are measured in Group A, HCA is performed on a 33×3 input data set. As the input data set contains 3 standardised regression coefficients, PCA can be performed on clusters containing three or more immune outcomes.

A cluster structure ($k = 6$) is extracted at a link height of $d_{ij} = 0.15$, corresponding to an average pairwise Pearson correlation of at least 0.85 between all immune

outcomes contained in any two clusters i and j. The dendograms below describe the intrinsic cluster structures of the standardised regression coefficients estimating these comparisons of interest related to MVA85A priming in Group A.

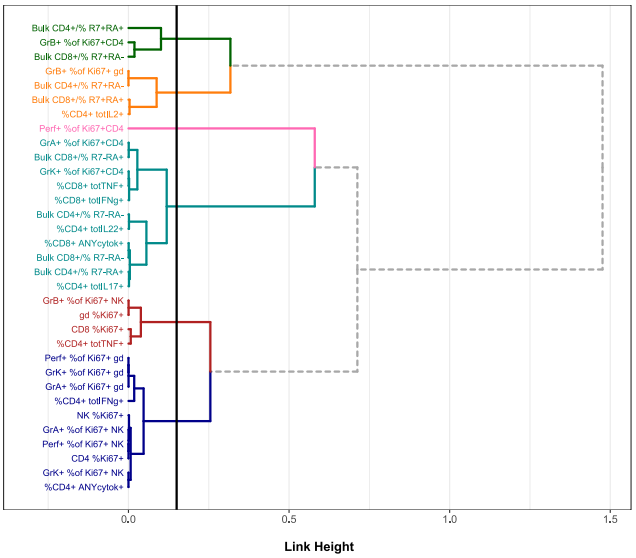


Figure 9.46: Dendogram of average-linkage hierarchical clustering of standardised regression coefficients related to MVA85A priming in Group A immune outcomes

At this link height, only one cluster (a singleton) contains fewer immune outcomes than the number of standardised model coefficients ($n_3^* = 1 < q$ (*pink*)). The immune outcome contained in this cluster ($\%Perf^+$ $Ki67^+$ $CD4^+$) is then selected for further analysis into subgroup differences described by levels of *MVA85A*. PCA can be performed to obtain lower-dimensional representations of the remaining clusters.

A PCA is performed on the standardised model coefficients for the members of Cluster 1 (*green*). As three immune outcomes are contained in this cluster, PCA is performed on a 3×3 input data set. The scree plot of the variance in standardised model coefficients explained by each of the 3 PCs is included below.

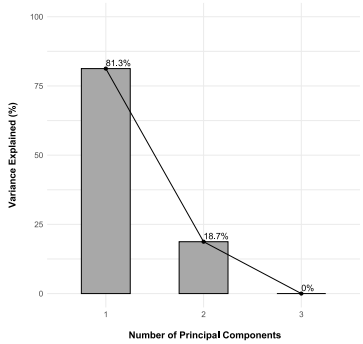


Figure 9.47: Scree plot of the percentage variance in standardised model coefficients for *MVA85A* explained by each principal component across Cluster 1 immune outcomes in Group A.

The first PC explains 81.3% of the variance in the standardised model coefficients for the cluster outcomes. In other words, for the 3 immune outcomes in this cluster, PC1 captures most of the cluster variance in the longitudinal effects of MVA85A priming. PC1 can act as a lower-dimensional representation of these effects in this cluster.

Contributions to the first PC were investigated to identify a subset of the cluster outcomes for further analysis. The scree plot below shows which standardised model coefficients (*left*) and immune outcomes (*right*) make large contributions to the variance explained by PC1.

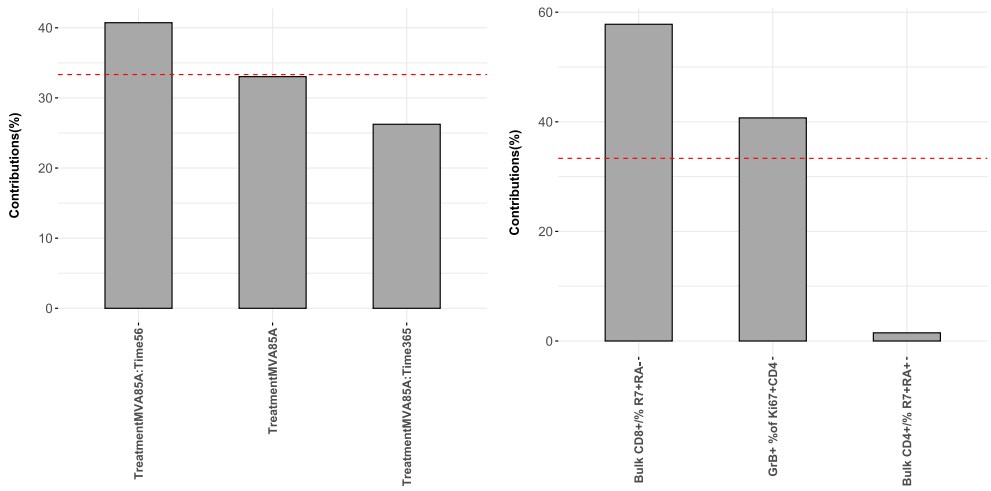


Figure 9.48: Scree plots of the contributions of standardised *MVA85A* model coefficients and immune outcomes to the Cluster 1 variance explained by PC1.

PC1 primarily describes variance in the standardised coefficient for the effect of MVA85A priming at $Time1$. The immune outcomes $Bulk\ \%CD8^{+}\ R7^{+}\ RA^{-}$ and $\%GrB^{+}\ Ki67^{+}\ CD4^{+}$ make the largest contributions to this variance and are selected for further analysis.

A PCA is performed on the standardised model coefficients for the members of Cluster 2 (*orange*). As four immune outcomes are contained in this cluster, PCA is performed on a 4×3 input data set. The scree plot of the variance in standardised model coefficients explained by each of the 3 PCs is included below.

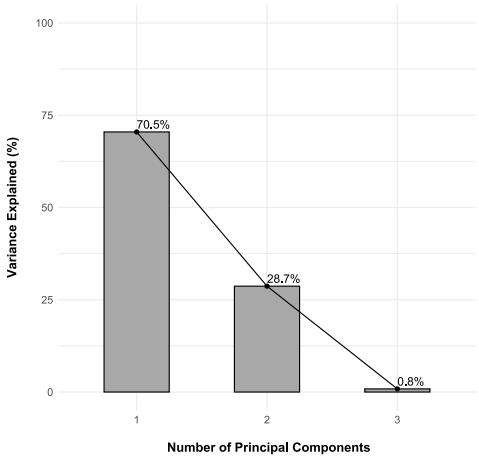


Figure 9.49: Scree plot of the percentage variance in standardised model coefficients for *MVA85A* explained by each principal component across Cluster 2 immune outcomes in Group A.

The first PC explains 70.5% of the variance in the standardised model coefficients for the cluster outcomes. In other words, for the 4 immune outcomes in this cluster, PC1

captures most of the cluster variance in the longitudinal effects of MVA85A priming. PC1 can act as a lower-dimensional representation of these effects in this cluster.

Contributions to the first PC were investigated to identify a subset of the cluster outcomes for further analysis. The scree plot below shows which standardised model coefficients (*left*) and immune outcomes (*right*) make large contributions to the variance explained by PC1.

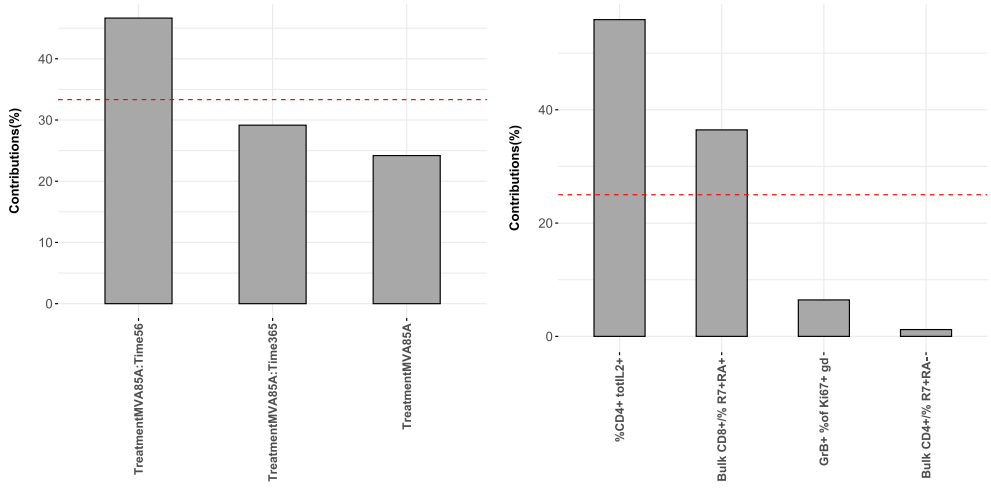


Figure 9.50: Scree plots of the contributions of standardised *MVA85A* model coefficients and immune outcomes to the Cluster 2 variance explained by PC1.

PC1 primarily describes variance in the standardised coefficient for the effect of MVA85A priming at $Time1$. The immune outcomes $\%CD4^{+}$ $totIL2^{+}$ and *Bulk* $\%CD8^{+}$ $R7^{+}$ RA^{+} make the largest contributions to this variance and are selected for further analysis.

A PCA is performed on the standardised model coefficients for the members of Cluster 4 (*orange*). As 11 immune outcomes are contained in this cluster, PCA is performed on a 11×3 input data set. The scree plot of the variance in standardised model coefficients explained by each of the 3 PCs is included below.

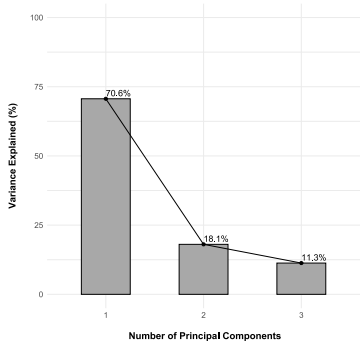


Figure 9.51: Scree plot of the percentage variance in standardised model coefficients for *MVA85A* explained by each principal component across Cluster 4 immune outcomes in Group A.

The first PC explains 70.6% of the variance in the standardised model coefficients for the cluster outcomes. In other words, for the 11 immune outcomes in this cluster, PC1 captures most of the cluster variance in the longitudinal effects of MVA85A

priming. PC1 can act as a lower-dimensional representation of these effects in this cluster.

Contributions to the first PC were investigated to identify a subset of the cluster outcomes for further analysis. The scree plot below shows which standardised model coefficients (*left*) and immune outcomes (*right*) make large contributions to the variance explained by PC1.

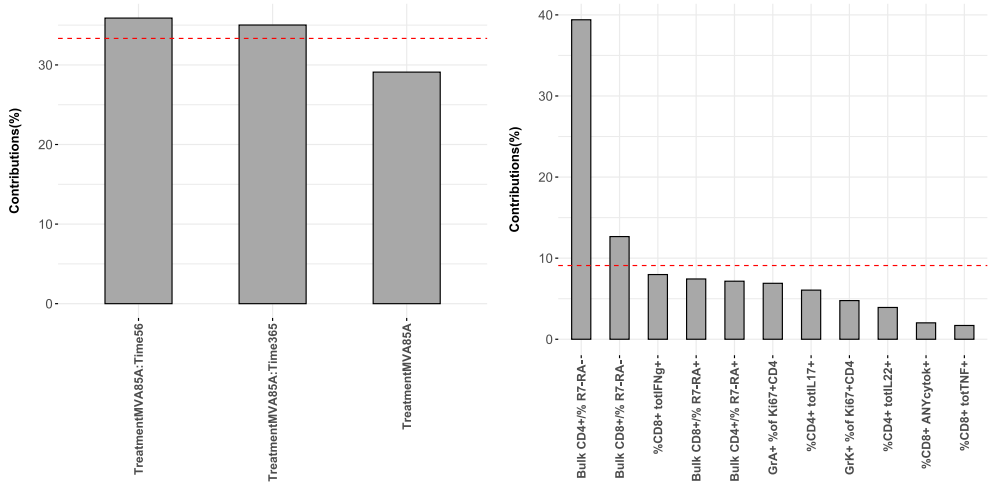


Figure 9.52: Scree plots of the contributions of standardised *MVA85A* model coefficients and immune outcomes to the Cluster 4 variance explained by PC1.

PC1 primarily describes variance in the standardised coefficient for the effect of MVA85A priming at $Time1$ and $Time3$. The immune outcomes *Bulk* $\%CD4^+$ $R7^-$ RA^- and *Bulk* $\%CD8^+$ $R7^-$ RA^- make the largest contributions to this variance and are selected for further analysis.

A PCA is performed on the standardised model coefficients for the members of Cluster 5 (*orange*). As four immune outcomes are contained in this cluster, PCA is performed on a 4×3 input data set. The scree plot of the variance in standardised model coefficients explained by each of the 3 PCs is included below.

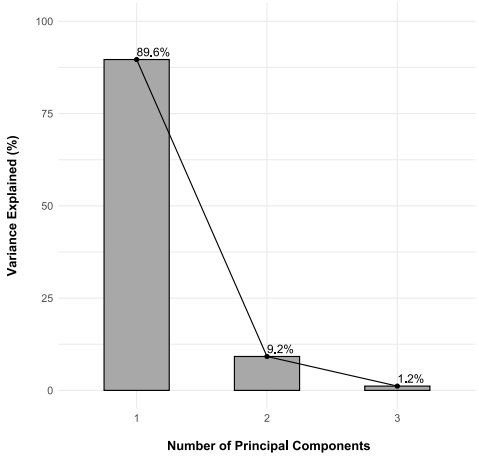


Figure 9.53: Scree plot of the percentage variance in standardised model coefficients for *MVA85A* explained by each principal component across Cluster 5 immune outcomes in Group A.

The first PC explains 89.6% of the variance in the standardised model coefficients for the cluster outcomes. In other words, for the four immune outcomes in this cluster, PC1 captures most of the cluster variance in the longitudinal effects of MVA85A priming. PC1 can act as a lower-dimensional representation of these effects in this cluster.

Contributions to the first PC were investigated to identify a subset of the cluster outcomes for further analysis. The scree plot below shows which standardised model coefficients (*left*) and immune outcomes (*right*) make large contributions to the variance explained by PC1.

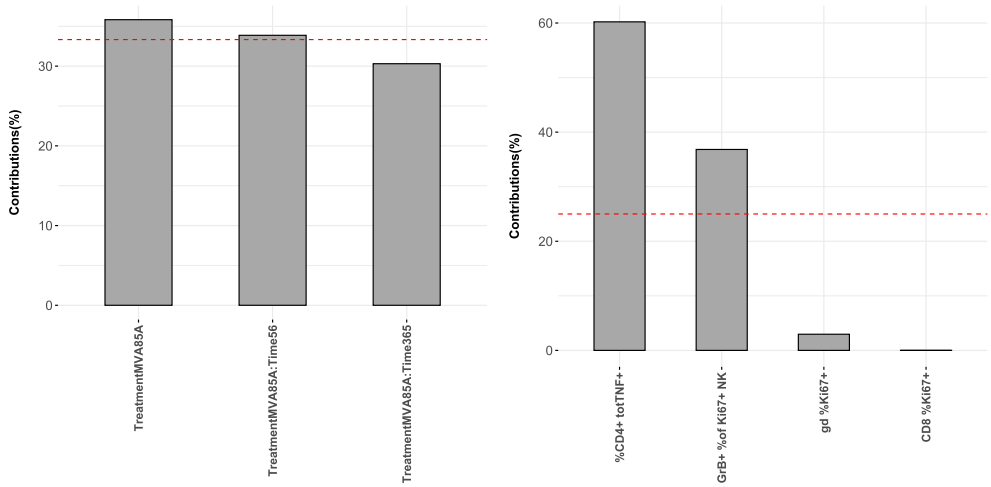


Figure 9.54: Scree plots of the contributions of standardised *MVA85A* model coefficients and immune outcomes to the Cluster 5 variance explained by PC1.

PC1 primarily describes variance in the standardised coefficient for the effect of MVA85A priming at $Time2$ and $Time1$. The immune outcomes $\%CD4^+$ $totTNF^+$ and $\%GrB^+$ $Ki67^+$ NK make the largest contributions to this variance and are selected for further analysis.

Instead of analysing the subgroup differences described by levels of *MVA85A* for all 33 outcomes simultaneously, the analysis is performed in five subsets of immune outcomes identified from clusters describing immune outcomes with similar behaviour over time. A total of 12 unique immune outcomes are analysed further.

Maternal QFT

A hierarchical cluster analysis with average linkage was performed on the Pearson correlation distance matrix of the following standardised model coefficients for the Group A outcomes, adjusted for the average sex effect: $QFT^{(1)}$, $QFT : Time1^{(2)}$ and $QFT : Time3^{(3)}$.

As 33 immune outcomes are measured in Group A, HCA is performed on a 33×3 input data set. As the input data set contains three standardised regression coefficients, PCA can be performed on clusters containing three or more immune outcomes.

A cluster structure ($k = 7$) is extracted at a link height of $d_{ij} = 0.15$, corresponding to an average pairwise Pearson correlation of at least 0.85 between all immune outcomes contained in any two clusters i and j. The dendograms below describe the

intrinsic cluster structures of the standardised regression coefficients estimating these comparisons of interest related to MVA85A priming in Group A.

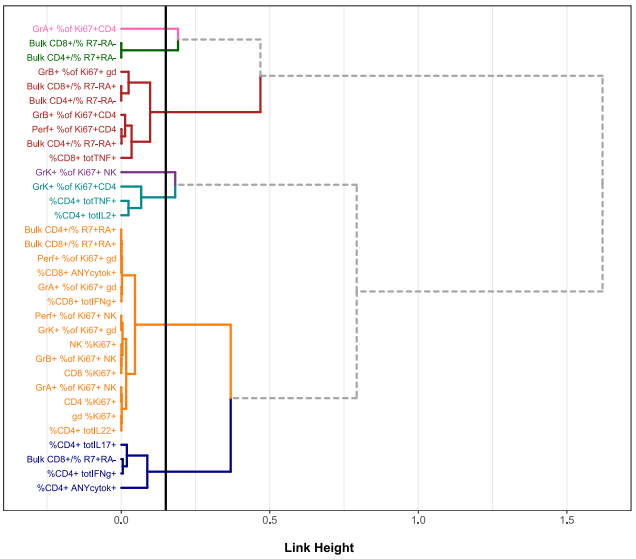


Figure 9.55: Dendogram of average-linkage hierarchical clustering of standardised regression coefficients related to maternal QFT in Group A immune outcomes

At this link height, three clusters (two singletons; one cluster containing 2 immune outcomes) contains fewer immune outcomes than the number of standardised model coefficients ($m_1^* = 1 < q$ (*pink*); $m_2^* = 2 < q$ (*green*; $m_4^* = 1 < q$ (*purple*))). The immune outcomes contained in these clusters are selected for further analysis into subgroup differences described by levels of *QFT*. PCA can be performed to obtain lower-dimensional representations of the remaining clusters.

A PCA is performed on the standardised model coefficients for the members of Cluster 3 (*red*). As seven immune outcomes are contained in this cluster, PCA is performed on a 7×3 input data set. The scree plot of the variance in standardised model coefficients explained by each of the 3 PCs is included below.

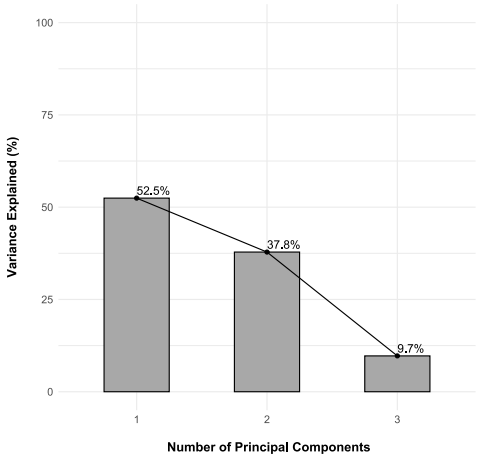


Figure 9.56: Scree plot of the percentage variance in standardised model coefficients for *QFT* explained by each principal component across Cluster 3 immune outcomes in Group A.

The first two PCs explain 90.3% of the variance in the standardised model coefficients for the cluster outcomes. In other words, for the seven immune outcomes in this cluster, PC1 and PC2 capture most of the cluster variance in the longitudinal effects of a positive maternal QFT. PC1 and PC2 can act as a lower-dimensional representation of these effects in this cluster.

Contributions to the first two PC were investigated to identify a subset of the cluster outcomes for further analysis. The scree plot below shows which standardised model coefficients make large contributions to the variance explained by PC1 ($left$) and PC2 ($right$)

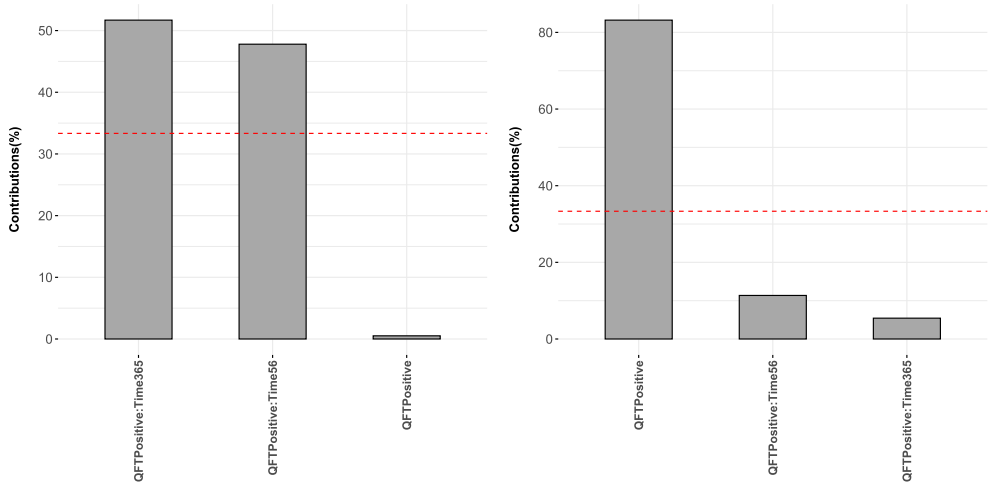


Figure 9.57: Scree plots of the contributions of standardised QFT model coefficients to the Cluster 3 variance explained by PC1 and PC2.

PC1 primarily describes variance in the standardised coefficient for the effect of a positive maternal QFT at $Time3$ and $Time1$. PC2 primarily describes variance in the standardised coefficient for the effect of a positive maternal QFT at $Time2$.

The scree plot below shows which immune outcomes make large contributions to the variance explained by PC1 ($left$) and PC2 ($right$)

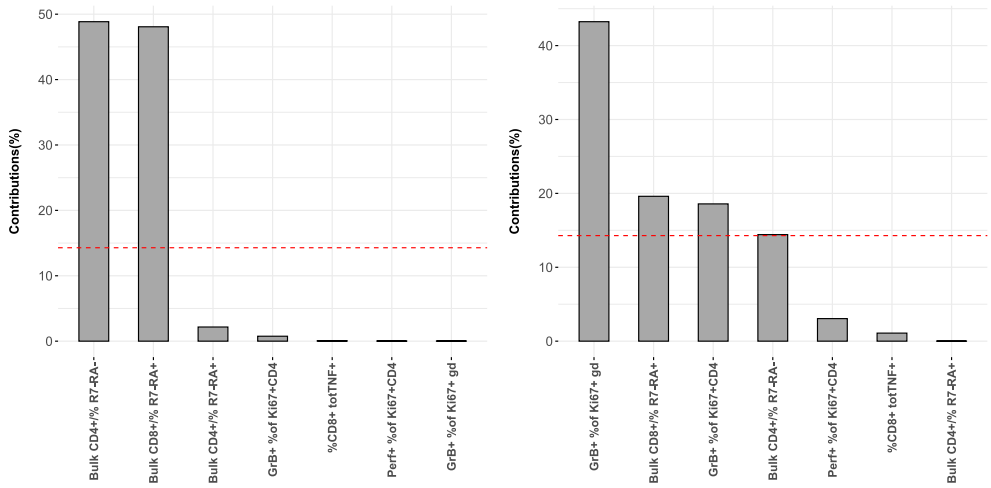


Figure 9.58: Scree plots of the contributions of immune outcomes to the Cluster 3 variance explained by PC1 and PC2.

Five immune outcomes making large contributions to this variance are selected for further analysis:

- *Bulk* $\%CD4^{+}$ $R7^{-}$ $RA^{-(1)}$
- *Bulk* $\%CD8^{+}$ $R7^{-}$ $RA^{-(2)}$
- $\%GrB^{+}$ $Ki67^{+}$ $gd^{(3)}$
- *Bulk* $\%CD8^{+}$ $R7^{-}$ $RA^{+(4)}$
- $\%GrB^{+}$ $Ki67^{+}$ $CD4^{+(5)}$

A PCA is performed on the standardised model coefficients for the members of Cluster 5 (*turquoise*). As three immune outcomes are contained in this cluster, PCA is performed on a 3×3 input data set. The scree plot of the variance in standardised model coefficients explained by each of the 3 PCs is included below.

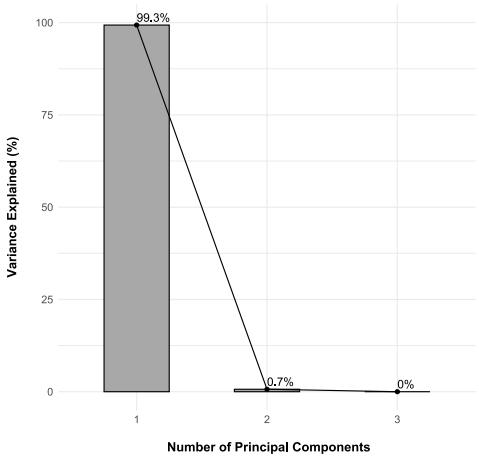


Figure 9.59: Scree plot of the percentage variance in standardised model coefficients for *QFT* explained by each principal component across Cluster 5 immune outcomes in Group A.

The first PC explains 99.3% of the variance in the standardised model coefficients for the cluster outcomes. In other words, for the three immune outcomes in this cluster, PC1 captures most of the cluster variance in the longitudinal effects of a positive maternal QFT. PC1 can act as a lower-dimensional representation of these effects in this cluster.

Contributions to the first PC were investigated to identify a subset of the cluster outcomes for further analysis. The scree plot below shows which standardised model coefficients (*left*) and immune outcomes (*right*) make large contributions to the variance explained by PC1.

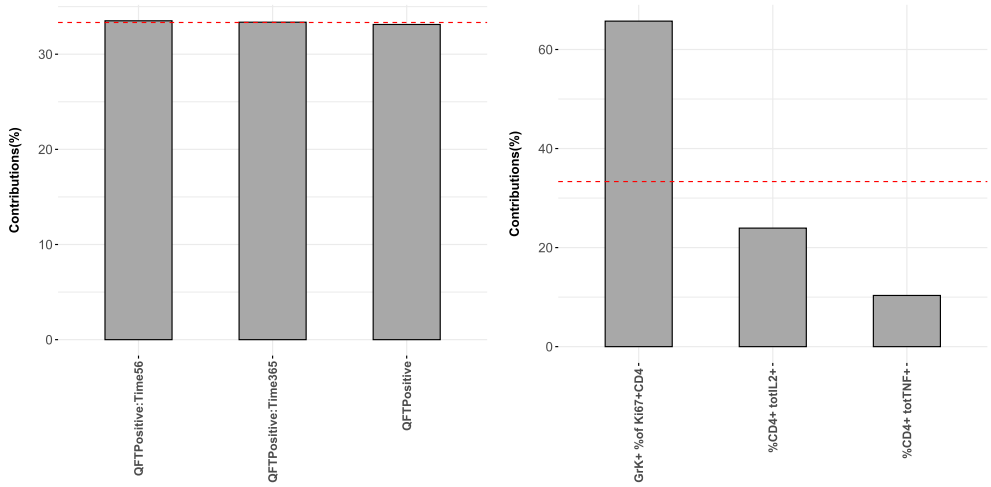


Figure 9.60: Scree plots of the contributions of standardised *QFT* model coefficients and immune outcomes to the Cluster 5 variance explained by PC1.

PC1 describes variance in all standardised coefficients for the effect of a positive maternal QFT. The immune outcome $\%GrK^{+}\ Ki67^{+}\ CD4^{+}$ makes the largest contributions to this variance and is selected for further analysis.

A PCA is performed on the standardised model coefficients for the members of Cluster 6 (*turquoise*). As 15 immune outcomes are contained in this cluster, PCA is performed on a 15×3 input data set. The scree plot of the variance in standardised model coefficients explained by each of the 3 PCs is included below.

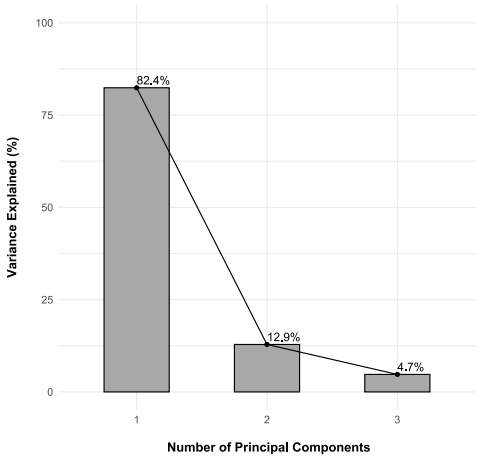


Figure 9.61: Scree plot of the percentage variance in standardised model coefficients for *QFT* explained by each principal component across Cluster 6 immune outcomes in Group A.

The first PC explains 82.4% of the variance in the standardised model coefficients for the cluster outcomes. In other words, for the 15 immune outcomes in this cluster, PC1 captures most of the cluster variance in the longitudinal effects of a positive maternal QFT. PC1 can act as a lower-dimensional representation of these effects in this cluster.

Contributions to the first PC were investigated to identify a subset of the cluster outcomes for further analysis. The scree plot below shows which standardised model

coefficients ($left$) and immune outcomes ($right$) make large contributions to the variance explained by PC1.

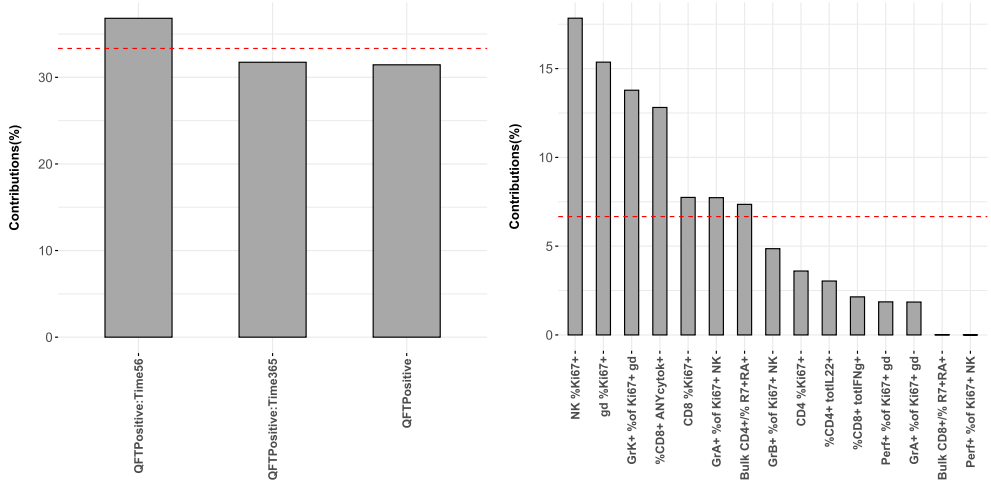


Figure 9.62: Scree plots of the contributions of standardised QFT model coefficients and immune outcomes to the Cluster 6 variance explained by PC1.

PC1 describes variance in the standardised coefficient for the effect of a positive maternal QFT at $Time1$. Seven immune outcomes making large contributions to this variance are selected for further analysis:

- $\%NK\ Ki67^{+(1)}$
- $\%gd\ Ki67^{+(2)}$
- $\%GrK^{+}\ Ki67^{+}\ gd^{(3)}$
- $\%CD8^{+}\ ANYcytok^{+(4)}$
- $\%GrA^{+}\ Ki67^{+}\ NK^{(5)}$
- $\%CD8^{+}\ Ki67^{+(6)}$
- $Bulk\ \%CD4^{+}\ R7^{+}\ RA^{+(7)}$

A PCA is performed on the standardised model coefficients for the members of Cluster 7 ($blue$). As four immune outcomes are contained in this cluster, PCA is performed on a 4×3 input data set. The scree plot of the variance in standardised model coefficients explained by each of the 3 PCs is included below.

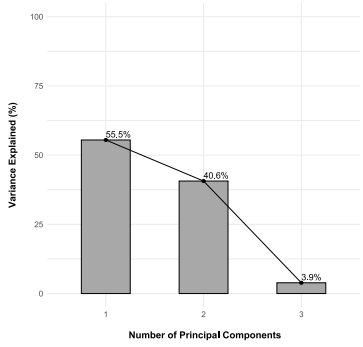


Figure 9.63: Scree plot of the percentage variance in standardised model coefficients for QFT explained by each principal component across Cluster 7 immune outcomes in Group A.

The first two PCs explain 96.1% of the variance in the standardised model coefficients for the cluster outcomes. In other words, for the four immune outcomes in this cluster,

PC1 and PC2 capture most of the cluster variance in the longitudinal effects of a positive maternal QFT. PC1 and PC2 can act as a lower-dimensional representation of these effects in this cluster.

Contributions to the first two PC were investigated to identify a subset of the cluster outcomes for further analysis. The scree plot below shows which standardised model coefficients make large contributions to the variance explained by PC1 ($left$) and PC2 ($right$)

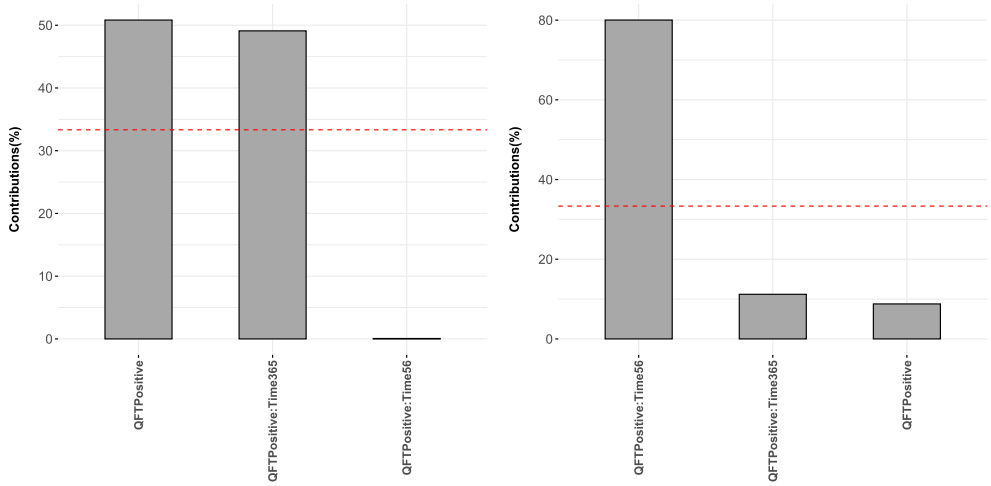


Figure 9.64: Scree plots of the contributions of standardised QFT model coefficients to the Cluster 7 variance explained by PC1 and PC2.

PC1 primarily describes variance in the standardised coefficient for the effect of a positive maternal QFT at $Time2$ and $Time3$. PC2 primarily describes variance in the standardised coefficient for the effect of a positive maternal QFT at $Time1$.

The scree plot below shows which immune outcomes make large contributions to the variance explained by PC1 ($left$) and PC2 ($right$)

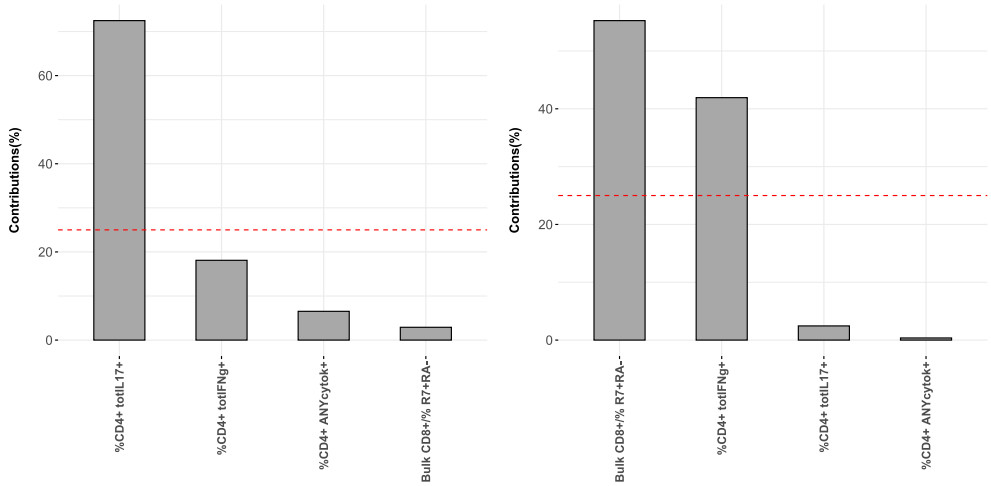


Figure 9.65: Scree plots of the contributions of immune outcomes to the Cluster 7 variance explained by PC1 and PC2.

The immune outcome $\%CD4^{+}\ totIL17^{-(1)}$ makes the largest contribution to the variance explained by PC1. The immune outcomes $Bulk\ \%CD8^{+}\ R7^{+}\ RA^{-(2)}$ and

$\%CD4^{+}$ $totIFNg^{+(3)}$ make the largest contributions to the variance explained by PC2. This subset of immune outcomes is selected for further analysis.

Appendix I

Feeding & Cotrimoxazole

Contrasts

The matrix of model coefficient estimates $\hat{\boldsymbol{\beta}}$ from the corresponding longitudinal models (Equation 4.3, p.36) is multiplied by the following contrast matrix (C) to estimate subgroup differences ($\hat{C}$) for this exposure of interest.

Contrast	FA0 T2	FA1 T2	FA2 T2	FA3 T2	FA0 T1	FA1 T1	FA2 T1	FA3 T1	FA0 T3	FA1 T3	FA2 T3	FA3 T3
FA0 T2 vs T1	−1	0	0	0	1	0	0	0	0	0	0	0
FA0 T2 vs T3	−1	0	0	0	0	0	0	0	1	0	0	0
FA0 T1 vs T3	0	0	0	0	−1	0	0	0	1	0	0	0
FA1 T2 vs T1	0	−1	0	0	0	1	0	0	0	0	0	0
FA1 T2 vs T3	0	−1	0	0	0	0	0	0	0	1	0	0
FA1 T1 vs T3	0	0	0	0	0	−1	0	0	0	1	0	0
FA2 T2 vs T1	0	0	−1	0	0	0	1	0	0	0	0	0
FA2 T2 vs T3	0	0	−1	0	0	0	0	0	0	0	1	0
FA2 T1 vs T3	0	0	0	0	0	0	−1	0	0	0	1	0
FA3 T2 vs T1	0	0	0	−1	0	0	0	1	0	0	0	0
FA3 T2 vs T3	0	0	0	−1	0	0	0	0	0	0	0	1
FA3 T1 vs T3	0	0	0	0	0	0	0	−1	0	0	0	1
T2 FA0 vs FA3	−1	0	0	1	0	0	0	0	0	0	0	0
T1 FA0 vs FA3	0	0	0	0	−1	0	0	1	0	0	0	0
T3 FA0 vs FA3	0	0	0	0	0	0	0	0	−1	0	0	1
T2 FA0 vs FA1	−1	1	0	0	0	0	0	0	0	0	0	0
T1 FA0 vs FA1	0	0	0	0	−1	1	0	0	0	0	0	0
T3 FA0 vs FA1	0	0	0	0	0	0	0	0	−1	1	0	0
T2 FA2 vs FA3	0	0	−1	1	0	0	0	0	0	0	0	0
T1 FA2 vs FA3	0	0	0	0	0	0	−1	1	0	0	0	0
T3 FA2 vs FA3	0	0	0	0	0	0	0	0	0	0	−1	1

PCA Only

Contrasts computing subgroup differences based on combinations of feeding practices and cotrimoxazole treatement were estimated from longitudinal models for a subset of Group A immune outcomes ($m_3^* = 15$, p. 64).

- There are 21 contrasts and 15 immune outcomes, meaning that the contrast p-values are adjusted for false discoveries in $21 \times 15 = 315$ comparisons.
- One-hundred-and-twenty-four contrasts are significantly different from zero for a Benjamini-Hochberg FDR of 15%.
- Three immune outcomes selected for this subset were reflected prior to longitudinal modelling (*Bulk %CD4$^+$ R7$^+$ RA$^+$*, *Bulk %CD8$^+$ R7$^+$ RA$^+$* and *%GrA$^+$ Ki67$^+$ NK*). Hence, their estimated subgroup differences were multiplied by -1 before exponentiating ($\exp(\hat{C}) = \exp(\text{-}\hat{C})$). These outcomes were marked with an asterisk when presenting the results.

The key results are summarised below after exponentiating contrast estimates for interpretation:

Rank	Outcome	Contrast	exp($\hat{C}$)	p-value	BH(q)	BY(q)
48/124	*Bulk %CD8$^+$ R7$^-$ RA$^+$*	*Time*2: *FA*0 vs *FA*3	4.209	0.007×10^{-2}	0.004×10^{-1}	0.003
52/124	*Bulk %CD8$^+$ R7$^-$ RA$^-$*	*Time*2: *FA*0 vs *FA*3	3.237	0.002×10^{-1}	0.001	0.009
59/124	*Bulk %CD8$^+$ R7$^+$ RA^{+*}*	*Time*2: *FA*0 vs *FA*3	0.487	0.006×10^{-1}	0.003	0.021
70/124	*Bulk %CD8$^+$ R7$^-$ RA^{+*}*	*Time*1: *FA*0 vs *FA*3	2.886	0.003	0.013	0.082
79/124	*%GrA$^+$ Ki67$^+$ NK**	*Time*3: *FA*2 vs *FA*3	0.398	0.009	0.037	0.234
80/124	*%GrA$^+$ Ki67$^+$ NK**	*Time*2: *FA*0 vs *FA*3	0.455	0.010	0.038	0.242
81/124	*Bulk %CD4$^+$ R7$^-$ RA$^-$*	*Time*2: *FA*0 vs *FA*3	1.705	0.010	0.038	0.242
84/124	*%GrA$^+$ Ki67$^+$ NK**	*Time*2: *FA*0 vs *FA*1	0.514	0.012	0.046	0.292
95/124	*Bulk %CD4$^+$ R7$^-$ RA$^-$*	*Time*1: *FA*0 vs *FA*3	1.636	0.017	0.056	0.357
98/124	*%GrB$^+$ Ki67$^+$ CD4$^+$*	*Time*3: *FA*0 vs *FA*3	0.510	0.020	0.066	0.417
100/124	*Bulk %CD4$^+$ R7$^+$ RA^{+*}*	*Time*2: *FA*0 vs *FA*3	0.650	0.025	0.079	0.498
104/124	*Bulk %CD4$^+$ R7$^+$ RA^{+*}*	*Time*1: *FA*0 vs *FA*3	0.667	0.033	0.099	0.627
105/124	*Bulk %CD8$^+$ R7$^-$ RA$^-$*	*Time*3: *FA*0 vs *FA*3	0.544	0.033	0.100	0.633
108/124	*Bulk %CD8$^+$ R7$^+$ RA^{+*}*	*Time*3: *FA*0 vs *FA*3	1.488	0.037	0.109	0.690
120/124	*%NK$^+$ Ki67$^+$*	*Time*2: *FA*0 vs *FA*3	0.524	0.048	0.127	0.804
123/124	*%GrK$^+$ Ki67$^+$ NK*	*Time*2: *FA*2 vs *FA*3	1.707	0.055	0.142	0.899

Table 9.21: Summary of differences in maternal QFT subgroups ($q < 0.15$) after correcting for multiple hypothesis testing with the Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) procedures.

There are 16 discoveries of differences in feeding-cotrimoxazole subgroups after following the BH procedure ($q < 0.15$). If the analysis followed scientific convention for statistical significance ($q < 0.05$), the four contrasts listed below would be identified as discoveries, assuming independence or positive dependence among the family of tests (BH).

- A difference in *Bulk %CD8$^+$ R7$^-$ RA$^+$* at *Time*2 between formula-fed and breastfed infants treated with cotrimoxazole (*FA*0 vs *FA*3)
- A difference in *Bulk %CD8$^+$ R7$^-$ RA$^-$* at *Time*2 between formula-fed and breastfed infants treated with cotrimoxazole (*FA*0 vs *FA*3)
- A difference in *Bulk %CD8$^+$ R7$^+$ RA$^+$* at *Time*2 between formula-fed and breastfed infants treated with cotrimoxazole (*FA*0 vs *FA*3)
- A difference in *Bulk %CD8$^+$ R7$^-$ RA$^+$* at *Time*1 between formula-fed and breastfed infants treated with cotrimoxazole (*FA*0 vs *FA*3)

The first three out of the discoveries listed above are also retained when the more conservative BY procedure is applied.

These estimates assume that *FA*, *MVA85A*, and *QFT* have independent effects on Group A immune outcomes. As demonstrated for *MVA85A* and *QFT*, the longitudinal models can be adjusted for the effects of multiple exposures.

Contrasts computing subgroup differences based on combinations of feeding practices and cotrimoxazole treatment were then estimated from these adjusted longitudinal models. The number of comparisons remains ($m_3^* \times 21 = 15 \times 21 = 315$ comparisons).

Rank	Outcome	Contrast	Adjustment	$\exp(\hat{C})$	p-value	BH(q)	BY(q)
48/124	*Bulk %CD8$^+$ R7$^-$ RA$^+$*	*Time*2: *FA*0 vs *FA*3	$+QFT$	4.322	0.007×10^{-2}	0.004×10^{-1}	0.003
54/124	*Bulk %CD8$^+$ R7$^+$ RA^{+*}*	*Time*2: *FA*0 vs *FA*3	$+MVA85A$	0.470	0.003×10^{-1}	0.002	0.011
58/124	*Bulk %CD8$^+$ R7$^-$ RA$^-$*	*Time*2: *FA*0 vs *FA*3	$+QFT$	3.130	0.004×10^{-1}	0.002	0.015
68/124	*Bulk %CD8$^+$ R7$^-$ RA$^+$*	*Time*1: *FA*0 vs *FA*3	$+QFT$	1.762	0.002	0.012	0.077
77/124	*Bulk %CD4$^+$ R7$^-$ RA$^-$*	*Time*2: *FA*0 vs *FA*3	$+MVA85A$	0.568	0.006	0.024	0.155
82/124	*%GrA$^+$ Ki67$^+$ NK**	*Time*2: *FA*0 vs *FA*3	$+MVA85A + QFT$	0.474	0.008	0.031	0.195
85/124	*Bulk %CD4$^+$ R7$^-$ RA$^-$*	*Time*1: *FA*0 vs *FA*3	$+MVA85A$	1.696	0.010	0.038	0.241
86/124	*%GrA$^+$ Ki67$^+$ NK**	*Time*2: *FA*0 vs *FA*1	$+MVA85A + QFT$	0.518	0.010	0.038	0.241
87/124	*%GrA$^+$ Ki67$^+$ NK**	*Time*3: *FA*2 vs *FA*3	$+MVA85A + QFT$	0.408	0.011	0.041	0.262
92/124	*%GrB$^+$ Ki67$^+$ CD4$^+$*	*Time*3: *FA*0 vs *FA*3	$+MVA85A$	0.488	0.013	0.045	0.286
98/124	*Bulk %CD4$^+$ R7$^+$ RA^{+*}*	*Time*2: *FA*0 vs *FA*3	$+MVA85A$	0.638	0.019	0.062	0.393
101/124	*Bulk %CD4$^+$ R7$^+$ RA^{+*}*	*Time*1: *FA*0 vs *FA*3	$+MVA85A$	0.652	0.025	0.078	0.493
106/124	*Bulk %CD8$^+$ R7$^-$ RA$^-$*	*Time*3: *FA*0 vs *FA*1	$+QFT$	0.538	0.030	0.090	0.569
115/124	*%NK$^+$ Ki67$^+$*	*Time*2: *FA*0 vs *FA*3	$+QFT$	0.516	0.045	0.123	0.776
120/124	*Bulk %CD8$^+$ R7$^+$ RA^{+*}*	*Time*3: *FA*0 vs *FA*3	$+MVA85A$	1.450	0.050	0.130	0.823
123/124	*%GrK$^+$ Ki67$^+$ NK*	*Time*2: *FA*2 vs *FA*3	None	1.707	0.055	0.142	0.899

Table 9.22: Summary of differences in feeding-cotrimoxazole subgroups ($q < 0.15$), adjusted for additional exposures, and corrected for multiple hypothesis testing with the Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) procedures.

No new subgroup differences are detected ($q < 0.15$) after adjusting the longitudinal models for additional exposures. If the analysis follows scientific convention for statistical significance ($q < 0.05$), ten of these contrasts would be significant discoveries, assuming independence or positive dependence among the family of tests (BH). This far exceeds the four discoveries identified for $q < 0.05$ under the same assumption without adjusting for additional exposures.

After adjusting for the average effects of *MVA85A* and *QFT*, a number of contrast q-values are smaller:

- *Bulk %CD8$^+$ R7$^+$ RA$^+$* at *Time*2 comparing *FA*3 to *FA*0 (BH: $q = 0.003$; $q = 0.002$; BY: $q = 0.021$; $q = 0.011$).
- *Bulk %CD8$^+$ R7$^-$ RA$^+$* at *Time*1 comparing *FA*3 to *FA*0 (BH: $q = 0.013$; $q = 0.012$; BY: $q = 0.082$; $q = 0.077$).
- *Bulk %CD4$^+$ R7$^-$ RA$^-$* at *Time*2, comparing *FA*3 to *FA*0 (BH: $q = 0.038$; $q = 0.024$; BY: $q = 0.242$; $q = 0.155$).
- *%GrA$^+$ Ki67$^+$ NK** at *Time*2, comparing *FA*3 to *FA*0 (BH: $q = 0.038$; $q = 0.031$; BY: $q = 0.242$; $q = 0.195$).
- *%GrB$^+$ Ki67$^+$ CD4$^+$* at *Time*3, comparing *FA*3 to *FA*0 (BH: $q = 0.066$; $q = 0.045$; BY: $q = 0.417$; $q = 0.286$).
- *Bulk %CD4$^+$ R7$^+$ RA$^+$* at *Time*1, comparing *FA*3 to *FA*0 (BH: $q = 0.079$; $q = 0.078$; BY: $q = 0.498$; $q = 0.493$).
- *%NK$^+$ Ki67$^+$* at *Time*2, comparing *FA*3 to *FA*0 (BH: $q = 0.127$; $q = 0.123$; BY: $q = 0.804$; $q = 0.776$).
- *Bulk %CD8$^+$ R7$^-$ RA$^-$* at *Time*3, comparing *FA*1 to *FA*0 (BH: $q = 0.100$; $q = 0.090$; BY: $q = 0.633$; $q = 0.569$).

However, after adjusting for additional exposures, a number of contrast q-values are larger:

- *Bulk %CD8$^+$ R7$^-$ RA$^-$* at *Time*2, comparing *FA*3 to *FA*0 (BH: $q = 0.001$; $q = 0.002$; BY: $q = 0.009$; $q = 0.015$).
- *%GrA$^+$ Ki67$^+$ NK* at *Time*3, comparing *FA*3 to *FA*2 (BH: $q = 0.037$; $q = 0.041$; BY: $q = 0.234$; $q = 0.262$).
- *Bulk %CD8$^+$ R7$^+$ RA$^+$* at *Time*3, comparing *FA*3 to *FA*0 (BH: $q = 0.109$; $q = 0.130$; BY: $q = 0.690$; $q = 0.823$).

HCA Then PCA

Contrasts computing subgroup differences based on feeding-cotrimoxazole combinations were estimated from longitudinal models for three subsets of Group A immune outcomes ($m_1^{**} = 4$, $m_2^{**} = 11$, $m_3^{**} = 5$). Once again, 21 contrasts are estimated for each subset. For each subset, contrast p-values are adjusted for false discoveries. These adjustments consider different numbers of comparisons:

- Cluster 1: $m_1^{**} \times 21 = 4 \times 21 = 84$
- Cluster 2: $m_2^{**} \times 21 = 11 \times 21 = 231$
- Cluster 3: $m_3^{**} \times 21 = 7 \times 21 = 147$

The estimated subgroup differences for outcomes reflected prior to longitudinal modelling were multiplied by -1 before exponentiating ($\exp(\hat{C}) = \exp(\text{-}\hat{C})$. As in the previous section, these outcomes are marked with an asterisk when presenting the results.

- Two immune outcomes selected for Cluster 1 were reflected prior to longitudinal modelling (*%GrB$^+$ Ki67$^+$ NK*; *Bulk %CD8$^+$ R7$^+$ RA$^+$*).
- Two immune outcomes selected for Cluster 2 were reflected prior to longitudinal modelling (*%GrA$^+$ Ki67$^+$ NK*; *Bulk %CD4$^+$ R7$^+$ RA$^+$*).
- One immune outcome selected for Cluster 3 was reflected prior to longitudinal modelling (*%Perf$^+$ Ki67$^+$ NK*).

Multiple testing corrections were applied to the contrast p-values for each of the seven clusters. The key results for a Benjamini-Hochberg FDR of 15% are summarised below:

Cluster	Rank	Outcome	Contrast	exp($\hat{C}$)	p-value	BH(q)	BY(q)
1	20/37	*Bulk %CD8$^+$ R7$^+$ RA$^+$**	*Time2*: *FA0* vs *FA3*	0.487	0.001	0.003	0.013
1	31/37	*%GrB$^+$ Ki67$^+$ CD4$^+$*	*Time3*: *FA0* vs *FA3*	0.510	0.020	0.056	0.278
1	33/37	*Bulk %CD8$^+$ R7$^+$ RA$^+$**	*Time3*: *FA0* vs *FA1*	1.488	0.037	0.095	0.477
1	37/37	*Bulk %CD8$^+$ R7$^+$ RA$^-$*	*Time2*: *FA0* vs *FA3*	0.689	0.065	0.147	0.739
2	41/59	*%GrA$^+$ Ki67$^+$ NK**	*Time3*: *FA2* vs *FA3*	0.398	0.009	0.051	0.308
2	43/59	*%GrA$^+$ Ki67$^+$ NK**	*Time3*: *FA0* vs *FA3*	0.455	0.010	0.052	0.314
2	46/59	*%GrA$^+$ Ki67$^+$ NK**	*Time2*: *FA0* vs *FA1*	0.515	0.012	0.061	0.368
2	58/59	*Bulk %CD4$^+$ R7$^+$ RA$^+$**	*Time1*: *FA0* vs *FA3*	0.650	0.033	0.109	0.655
3	21/69	*Bulk %CD8$^+$ R7$^-$ RA$^+$*	*Time2*: *FA0* vs *FA3*	4.209	6.81×10^{-5}	3×10^{-4}	0.002
3	37/69	*Bulk %CD8$^+$ R7$^-$ RA$^-$*	*Time2*: *FA0* vs *FA3*	3.237	2.31×10^{-4}	0.001	0.005
3	44/69	*Bulk %CD8$^+$ R7$^-$ RA$^+$*	*Time1*: *FA0* vs *FA3*	2.886	0.003	0.010	0.054
3	53/69	*%CD8$^+$ ANYcytok$^+$*	2.631	*Time2*: *FA2* vs *FA3*	0.011	0.031	0.173
3	56/69	*%Perf$^+$ Ki67$^+$ NK**	*Time1*: *FA2* vs *FA3*	0.469	0.019	0.050	0.278
3	60/69	*Bulk %CD8$^+$ R7$^-$ RA$^-$*	*Time3*: *FA0* vs *FA1*	0.544	0.033	0.082	0.455
3	65/69	*%GrK$^+$ Ki67$^+$ CD4$^+$*	*Time3*: *FA2* vs *FA3*	3.421	0.049	0.110	0.612
3	68/69	*Bulk %CD8$^+$ R7$^+$ RA$^-$*	*Time2*: *FA0* vs *FA3*	0.689	0.065	0.139	0.777

Table 9.23: Summary of differences in feeding-cotrimoxazole subgroups ($q < 0.15$) after correcting for multiple hypothesis testing by cluster with the Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) procedures.

There are 16 discoveries of differences in feeding-cotrimoxazole subgroups after following the BH procedure ($q < 0.15$). If the analysis followed scientific convention for statistical significance ($q < 0.05$), the six contrasts listed below would be identified as discoveries, assuming independence or positive dependence among the family of tests (BH).

- A difference in *Bulk %CD8$^+$ R7$^+$ RA$^+$* at *Time2* between formula-fed and breastfed infants treated with cotrimoxazole (*FA0* vs *FA3*)

- A difference in *Bulk %CD8$^+$ R7$^-$ RA$^+$* at *Time*2 between formula-fed and breastfed infants treated with cotrimoxazole ($FA0$ vs $FA3$)
- A difference in *Bulk %CD8$^+$ R7$^-$ RA$^-$* at *Time*2 between formula-fed and breastfed infants treated with cotrimoxazole ($FA0$ vs $FA3$)
- A difference in *Bulk %CD8$^+$ R7$^-$ RA$^+$* at *Time*1 between formula-fed and breastfed infants treated with cotrimoxazole ($FA0$ vs $FA3$)
- A difference in *%CD8$^+$ ANYcytok$^+$* at *Time*2 by cotrimoxazole treatment in breastfed infants ($FA2$ vs $FA3$)
- A difference in *%Perf$^+$ Ki67$^+$ NK* at *Time*1 by cotrimoxazole treatment in breastfed infants ($FA2$ vs $FA3$)

Two of the discoveries listed above are also retained when the more conservative BY procedure is applied.

- A difference in *Bulk %CD8$^+$ R7$^-$ RA$^+$* at *Time*2 between formula-fed and breastfed infants treated with cotrimoxazole ($FA0$ vs $FA3$)
- A difference in *Bulk %CD8$^+$ R7$^-$ RA$^-$* at *Time*2 between formula-fed and breastfed infants treated with cotrimoxazole ($FA0$ vs $FA3$)

However, these contrasts assume that *MVA85A*, *QFT*, and *FA* have independent effects on the immune outcomes. Referring back to the immune outcome subsets selected for the exposures of interest in Chapter 7 (p. 64), the longitudinal models can be adjusted for the effects of multiple exposures. The adjustment was applied to immune outcomes selected for *FA* and at least one of *MVA85A* and *QFT*.

Contrasts computing subgroup differences based on feeding-cotrimoxazole combinations were then estimated from these adjusted longitudinal models. The number of comparisons does not change, as the contrast is computed by averaging over the levels of the additional covariates.

After adjusting the longitudinal models for additional exposures (as required), multiple testing corrections were applied to the contrast p-values for each of the three clusters. The key results for a Benjamini-Hochberg FDR of 15% are summarised below:

Cluster	Rank	Outcome	Contrast	Adjustments	exp($\hat{C}$)	p-value	BH(q)	BY(q)
1	20/38	*Bulk %CD8$^+$ R7$^+$ RA^{+*}*	*Time*2: $FA0$ vs $FA3$	$+MVA85A$	0.485	6.7×10^{-3}	0.003	0.014
1	30/38	*%GrB$^+$ Ki67$^+$ CD4$^+$*	*Time*3: $FA0$ vs $FA3$	$+MVA85A$	0.505	0.018	0.051	0.251
1	33/38	*Bulk %CD8$^+$ R7$^+$ RA^{+*}*	*Time*3: $FA0$ vs $FA1$	$+MVA85A$	1.485	0.039	0.100	0.500
1	34/38	*Bulk %CD8$^+$ R7$^+$ RA$^-$*	*Time*2: $FA0$ vs $FA3$	$+MVA85A + QFT$	0.670	0.053	0.130	0.652
1	38/38	*%GrB$^+$ Ki67$^+$ NK*	*Time*2: $FA0$ vs $FA3$	$+MVA85A + QFT$	0.618	0.067	0.147	0.738
2	41/60	*%GrA$^+$ Ki67$^+$ NK*	*Time*3: $FA0$ vs $FA3$	$+MVA85A + QFT$	0.474	0.008	0.045	0.272
2	44/60	*%GrA$^+$ Ki67$^+$ NK*	*Time*2: $FA0$ vs $FA1$	$+MVA85A + QFT$	0.518	0.010	0.055	0.328
2	45/60	*%GrA$^+$ Ki67$^+$ NK*	*Time*3: $FA2$ vs $FA3$	$+MVA85A + QFT$	0.408	0.011	0.058	0.348
2	53/60	*%CD4$^+$ totIL2$^+$*	*Time*2: $FA0$ vs $FA1$	None	2.176	0.035	0.114	0.686
2	58/60	*%Perf$^+$ Ki67$^+$ CD4$^+$*	*Time*1: $FA0$ vs $FA1$	$+MVA85A$	0.502	0.027	0.139	0.836
3	28/70	*Bulk %CD8$^+$ R7$^-$ RA$^+$*	*Time*2: $FA0$ vs $FA3$	$+QFT$	4.566	3.19×10^{-5}	1.67×10^{-4}	0.001
3	37/70	*%Perf$^+$ Ki67$^+$ CD4$^+$*	*Time*2: $FA0$ vs $FA1$	$+MVA85A$	3.285	2.25×10^{-4}	0.001	0.005
3	41/70	*Bulk %CD8$^+$ R7$^-$ RA$^+$*	*Time*1: $FA0$ vs $FA1$	$+MVA85A$	3.141	0.001	0.005	0.029
3	52/70	*%CD8$^+$ ANYcytok$^+$*	*Time*2: $FA2$ vs $FA3$	$+QFT$	2.670	0.011	0.030	0.170
3	56/70	*%Perf$^+$ Ki67$^+$ NK*	*Time*1: $FA2$ vs $FA3$	$+QFT$	0.469	0.019	0.050	0.278
3	61/70	*%Perf$^+$ Ki67$^+$ CD4$^+$*	*Time*3: $FA0$ vs $FA1$	$+MVA85A$	0.553	0.037	0.088	0.492
3	62/70	*%GrK$^+$ Ki67$^+$ CD4$^+$*	*Time*3: $FA2$ vs $FA3$	$+QFT$	3.688	0.037	0.088	0.493
3	67/70	*Bulk %CD8$^+$ R7$^+$ RA$^-$*	*Time*2: $FA0$ vs $FA3$	$+QFT$	0.670	0.053	0.116	0.647
3	70/70	*%GrK$^+$ Ki67$^+$ CD4$^+$*	*Time*2: $FA2$ vs $FA3$	$+QFT$	2.235	0.067	0.141	0.787

Table 9.24: Summary of differences in feeding-cotrimoxazole subgroups ($q < 0.15$), adjusted for additional exposures, and corrected for multiple hypothesis testing by cluster with the Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) procedures.

After adjusting the longitudinal models for additional exposures, there are 19 discoveries of differences in feeding-cotrimoxazole subgroups after following the BH procedure ($q < 0.15$).

Estimated subgroup differences in *Bulk %CD8*$^+$ *R7*$^-$ *RA*$^+$ at $Time2$ are larger after adjusting the longitudinal models for the immune outcomes in Cluster 3.

- The estimated subgroup difference by breastfeeding in infants treated with cotrimoxazole for *Bulk %CD8*$^+$ *R7*$^-$ *RA*$^+$ at $Time2$ is considerably larger ($\exp(\hat{C}) = 4.566$; $\exp(\hat{C}) = 4.209$).
- In other words, mean *Bulk %CD8*$^+$ *R7*$^-$ *RA*$^+$ 8 weeks after BCG vaccination ($Time2$) is $\approx 357\%$ higher for breastfeeding in infants treated with cotrimoxazole, averaged over the distributions of *Sex* and *QFT*.
- When averaged over the distribution of *Sex* only, mean *Bulk %CD8*$^+$ *R7*$^-$ *RA*$^+$ at $Time2$ is $\approx 321\%$ higher for breastfeeding in infants treated with cotrimoxazole.
- Without adjusting the longitudinal models for additional exposures, the estimated subgroup difference is $\approx 36\%$ lower.
- After adjustments, the q-values for this contrast are much smaller (BH: $q = 1.67 \times 10^{-4}$; 3.19×10^{-5}; BY: $q = 0.002$; $q = 0.001$).

Similarly, estimated subgroup differences are larger for *%GrK*$^+$ *Ki67*$^+$ *CD4*$^+$ at $Time3$ after Cluster 3 adjustments.

- The estimated subgroup difference by cotrimoxazole treatment in breastfed infants for *%GrK*$^+$ *Ki67*$^+$ *CD4*$^+$ at $Time3$ is considerably larger ($\exp(\hat{C}) = 3.688$; $\exp(\hat{C}) = 3.421$).
- In other words, mean *%GrK*$^+$ *Ki67*$^+$ *CD4*$^+$ at ≈ 44 weeks after BCG vaccination ($Time3$) is $\approx 269\%$ higher for cotrimoxazole treatment in breastfed infants treated, averaged over the distributions of *Sex* and *QFT*.
- When averaged over the distribution of *Sex* only, mean *%GrK*$^+$ *Ki67*$^+$ *CD4*$^+$ at $Time3$ is $\approx 242\%$ higher for cotrimoxazole treatment in breastfed infants.
- Without adjusting the longitudinal models for additional exposures, the estimated subgroup difference is $\approx 27\%$ lower.
- After adjustments, the q-values for this contrast are smaller (BH: $q = 0.110$; $q = 0.088$; BY: $q = 0.621$; $q = 0.493$).

However, after adjusting for additional exposures, a number of contrast q-values are larger. Two discoveries are no longer significant ($q \geq 0.15$) after applying corrections to the adjusted longitudinal models.

- A difference in *Bulk %CD4*$^+$ *R7*$^+$ *RA*$^+$ at $Time1$ between formula-fed and breastfed infants treated with cotrimoxazole ($FA0$ vs $FA3$)
-
- A difference in *Bulk %CD8*$^+$ *R7*$^-$ *RA*$^-$ at $Time3$ by cotrimoxazole treatment in formula-fed infants ($FA0$ vs $FA3$)

If the analysis followed scientific convention for statistical significance ($q < 0.05$), the seven contrasts listed below would be identified as discoveries, assuming independence or positive dependence among the family of tests (BH).

- A difference in *Bulk %CD8⁺ R7⁺ RA⁺* at $Time2$ between formula-fed and breastfed infants treated with cotrimoxazole ($FA0$ vs $FA3$)
- A difference in *%GrA⁺ Ki67⁺ NK* at $Time2$ between formula-fed and breastfed infants treated with cotrimoxazole ($FA0$ vs $FA3$)
- A difference in *Bulk %CD8⁺ R7⁻ RA⁺* at $Time2$ between formula-fed and breastfed infants treated with cotrimoxazole ($FA0$ vs $FA3$)
- A difference in *%Perf⁺ Ki67⁺ CD4⁺* at $Time2$ by cotrimoxazole treatment in formula-fed infants ($FA0$ vs $FA1$)
- A difference in *Bulk %CD8⁺ R7⁺ RA⁺* at $Time1$ by cotrimoxazole treatment in formula-fed infants ($FA0$ vs $FA1$)
- A difference in *%CD8⁺ ANYcytok⁺* at $Time2$ by cotrimoxazole treatment in breastfed infants ($FA2$ vs $FA3$)
- A difference in *%Perf⁺ Ki67⁺ NK* at $Time1$ by cotrimoxazole treatment in breastfed infants ($FA2$ vs $FA3$)

This is one more discovery than was obtained by the PCA-only approach. Four of the discoveries listed above are also retained when the more conservative BY procedure is applied.

- A difference in *Bulk %CD8⁺ R7⁺ RA⁺* at $Time2$ between formula-fed and breastfed infants treated with cotrimoxazole ($FA0$ vs $FA3$)
- A difference in *Bulk %CD8⁺ R7⁻ RA⁺* at $Time2$ between formula-fed and breastfed infants treated with cotrimoxazole ($FA0$ vs $FA3$)
- A difference in *%Perf⁺ Ki67⁺ CD4⁺* at $Time2$ by cotrimoxazole treatment in formula-fed infants ($FA0$ vs $FA1$)
- A difference in *Bulk %CD8⁺ R7⁺ RA⁺* at $Time1$ by cotrimoxazole treatment in formula-fed infants ($FA0$ vs $FA1$)

This far exceeds the two discoveries identified for $q < 0.05$ under the same assumption without adjusting for additional exposures.

The subgroup differences in *%Perf⁺ Ki67⁺ CD4⁺* by cotrimoxazole treatment in formula-fed infants are only detected ($q < 0.15$) after adjusting the longitudinal models for additional exposures. The strongest result describes a significant difference (BH: $q = 0.001$; BY: $q = 0.005$) in *%Perf⁺ Ki67⁺ CD4⁺* 8 weeks after BCG ($Time2$) by cotrimoxazole treatment in formula-fed infants.

- Formula-fed infants treated with cotrimoxazole have 3.285-fold higher mean *%Perf⁺ Ki67⁺ CD4⁺* 8 weeks after BCG vaccination.
- There is also modest evidence (BH: $q = 0.088$) for lower mean *%Perf⁺ Ki67⁺ CD4⁺* in formula-fed infants treated with cotrimoxazole ≈ 44 weeks after BCG ($Time3$), assuming independence or positive dependence within the family of tests.
- At the time of BCG vaccination ($Time1$), there is weak evidence (BH: $q = 0.139$) for lower mean *%Perf⁺ Ki67⁺ CD4⁺* in formula-fed infants treated with cotrimoxazole, assuming independence or positive dependence within the family of tests.