Thesis

Quantifying biometrology operator data analysis subjectivity within Flow Cytometry using measurement uncertainty principles

Submitted by

Rebecca C. Grant

Wolfson School of Mechanical, Electrical and Manufacturing Engineering

In partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Loughborough University

Loughborough, Leicestershire, England,

Summer 2019

Doctoral Committee:

Supervisor: Jon Petzing

Supervisor: Karen Coopman

Internal Examiner: Rob Thomas

External Examiner: Joel Segal

The views and opinions presented within this thesis are the author's. This work has been reviewed internally and externally by supervisors and collaborators to ensure relevancy of the research.

The human biological samples were sourced ethically, and their research use was in accordance with the terms of the informed consents from the suppliers and Loughborough University human participant ethical sub-committee.

.

"*The subjectivity of consciousness is an irreducible feature of reality, and it must occupy as fundamental a place in any credible world view as matter, energy, space, time and numbers*".

Thomas Nagel, The View from Nowhere, 1986

"*The measure of a man's real character is what he would do if he knew he would never be found out.*"

1st Baron Thomas Macaulay, 1926

Abstract

A recent evaluation of medical error has shown it to be the third leading cause of death in the US, following heart disease and cancer. Better reporting and decision making could tackle this, but ultimately more accurate and precise measurement, with correct interpretation could make a significant difference to this unnecessary statistic. Clinical pathology measurement platforms are complex, requiring significant standardisation efforts to reduce false positives/negatives and the impacts these have on patient safety. Cell and Gene Therapy (CGT) manufacturing processes depend upon these platforms for measurement, with Flow Cytometry (FC) used for in-process and release metrics. However, the highly subjective nature of FC data analysis requires investigation to monitor impact on manufacturing and clinical decision making.

FC standardisation efforts have reduced variation from sample preparation and setup, however, no efforts have purely focused on the final post-analytical stage, to quantify the effect of subjective analysis of data files. This research has isolated this section of FC analysis, providing better measurement precision to build up a realistic uncertainty budget for FC measurements. Through a series of participant analysis studies that build in complexity, it has been shown that as FC data becomes more complex, the uncertainty contributions from inter-operator data analysis increase from 8 % to 34 %. This increase could mean the difference between a CGT treatment being provided at the right time, being discarded when it was suitable for administration, or an unsuitable treatment administered to the patient at an unsuitable time, having costly implications for all.

This variation does not correlate with operator experience or use frequency of the instrument, but is influenced by data visualisation effects, requiring further investigation at a later date to reduce this impact. Image parameters for other CGT measurement platforms are also impacted by subjective data analysis, requiring harmonisation to ensure the subjectivity is quantifiable, standardised and reduces manufacturing and hence medical error impacts to the patient and therapeutic product.

Acknowledgements

I would firstly like to acknowledge and thank all participants who took part in these Flow Cytometry analysis studies, without whom this research could not have been possible. A big thanks go to my supervisors: Jon Petzing, Karen Coopman and Nick Medcalf, for their continued support and guidance throughout this PhD, and for providing thorough feedback throughout the writing process.

I would like to thank all collaborators who have supported this work, from GSK: Bo Kara (formerly GSK), Aileen Kirkpatrick, Sebastian Mayer and Sandro Gomes. A big thanks to Sandro for helping to coordinate the participant studies at GSK and continued review of written work throughout the PhD, as well as the design and generation of the .fcs files for the complex model. From LGC thanks go to Julian Braybrook and Jonathan Campbell, for continued support and metrological insight into this work, as well as providing feedback on all outputs produced. I would also like to thank Tamara Lekishvili (formerly LGC) for Flow Cytometry training throughout this process, and to Richard Hughes and Jenny Hincks (formerly LGC) for helping to coordinate the participant studies at LGC.

I would like to thank all who have supported me within the EPSRC and MRC Centre for Doctoral Training in Regenerative Medicine (EP/L105072/1), especially Rob Thomas for providing access to Flow Cytometry equipment and David Williams for mentoring and development opportunities, as well as providing the cells for the basic model. Thanks to Jamie Holmes, for support running the colour vision work for the basic model. Thank you to all at the Center for Biomedical Innovation at MIT for providing an incredible secondment opportunity for the translational work, especially Jacqueline Wolfrum, Stacy Springs, and Anthony Sinksey.

I would also like to thank the IMechE, for their continued support throughout my Postgraduate Research Scholarship. Finally, I would like to thank my family and Loughborough friends who have provided continued support and enthusiasm throughout the PhD. A special thanks go to Anna, Emily, Laura, Karen, Jon and Jonathan, for looking out for me through the last stages of this journey.

Contents

List of Figures

List of Tables

Abbreviations

ANCOVA: Analysis of Covariance

ANOVA: Analysis of Variance

APC: Allophycocyanin

BSC: Biological Safety Cabinet

CAR-T: Chimeric Antigen Receptor T cell

CCR7: C-C Chemokine Receptor

CD: Cluster of Differentiation

CDC: Center for Disease Control

CFR: Code of Federal Regulations

CGT: Cell and Gene Therapy

CI: Continuous Improvement

CIP: Cancer ImmunoGuiding Program

CLL: Chronic Lymphocytic Leukaemia

CO_2: Carbon dioxide

COUHES: Committee on the Use of Humans as Experimental Subjects

CPA: Clinical Pathology Accreditation

CPD: Continued Professional Development

CQA: Critical Quality Attributes

CS&T: Cytometer Setup & Tracking

CV: Coefficient of Variation

DMAIC: Design-Measure-Analyse-Improve-Control

DMEM: Dulbecco's Modified Eagle Medium

DNA: Deoxyribose Nucleic Acid

DoE: Design of Experiments

EC 2102 Ep: Embryonal Carcinoma 2102 Ep Cell line

EDTA: Ethylenediamine Tetracetic Acid

EQAS: External Quality Assessment Schemes

EWMA: Exponentially Weighted Moving Average

FACS: Fluorescence Activated Cell Sorting

FBS: Fetal Bovine Serum

FC: Flow Cytometry

fcs: Flow Cytometry standard

FDA: Food & Drug Administration

FIRM: Future Investigators of Regenerative Medicine

FLOCK: Flow cytometry Clustering by K-means

FMH: Feto-Maternal Haemorrhage

FMO: Fluorescence Minus One

FSC: Forward Scatter

Gauge R&R: Gauge Repeatability & Reproducibility

GDPR: General Data Protection Regulation

GMP: Good Manufacturing Practice

GPG: Good Practice Guide

GSK: GlaxoSmithKline

GUM: Guide to the Expression of Uncertainty of Measurement

H_A: Alternative Hypothesis

H_0: Null Hypothesis

hESC: human Embryonic Stem Cell

HIV: Human Immunodeficiency Virus

HLA: Human Leukocyte Antigen

HPLC: High Performance Liquid Chromatography

HTA: Human Tissues Act

ICCS: International Clinical Cytometry Society

ICH: International Committee for Harmonization

ICS: Intracellular Cytokine Staining

ICSH: International Council for Standardization of Haematology

IQR: Interquartile Range

ISHAGE: International Society for Hematotherapy and Graft Engineering

IVD: In Vitro Diagnostics

ISO: International Organisation for Standardization

JCTLM: Joint Committee for the Traceability of Laboratory Medicine

k: coverage factor

KPI: Key Performance Indicator

K-S: Kolmogorov-Smirnov

LGC: Laboratory of the Government Chemist

MAD: Median Absolute Deviation

MANOVA: Multivariate Analysis of Variance

MedFI: Median Fluorescent Intensity

MHC: Major Histocompatibility Complex

MIFlowCyt: Minimum Information on a Flow Cytometry experiment

MRC: Medical Research Council

MRD: Minimal Residual Disease

MSA: Measurement Systems Analysis

MSC: Mesenchymal Stromal Cell

NHS: National Health Service

NIBSC: National Institute for Biological Standards and Control

NIST: National Institute for Standards & Technology

NK: Natural Killer

PBMC: Peripheral Blood Mononuclear Cells

PBS: Phosphate Buffered Saline

PDCA: Plan Do Check Act

PDS: Product Design Specification

pU: permissible Uncertainty

QbD: Quality-by-Design

QC: Quality Control

qPCR: quantitative Polymerase Chain Reaction

rCV: robust Coefficient of Variation

rSD: robust Standard Deviation

SCYM: Specialist in Cytometry

SD: Standard Deviation

Sig: Significance

SPC: Statistical Process Control

SSC: Side Scatter

SSEA: Stage Specific Embryonic Antigen

SOP: Standard Operating Procedure

S-W: Shapiro-Wilk

TCR: T-cell Receptor

t-SNE: t-Stochastic Neighbour Embedding

U: Expanded Uncertainty

U_c: combined Uncertainty

UKAS: United Kingdom Accreditation Service

UK NEQAS: United Kingdom National External Quality Assessment Scheme

V: Volts

WHO: World Health Organisation

7-AAD: 7-Aminoactinomycin D

List of Publications

Journal Manuscripts in Submission

Grant, R., Coopman, K., Medcalf, N., Silva-Gomes, S., Kara, B., Campbell, J. J., Braybrook, J., & Petzing, J. (2019). Understanding biometrology operator variability within Flow Cytometry data analysis for Quality Control of Cell and Gene Therapy Manufacturing. Measurement. Accepted for publication August 2019.

Grant, R., Coopman, K., Medcalf, N., Silva-Gomes, S., Kara, B., Campbell, J. J., Braybrook, J., & Petzing, J. (2019). Quantifying operator subjectivity within Flow Cytometry data analysis as a source of measurement uncertainty and the impact of experience on results. Cytotherapy. In Review.

External Conference Presentations

Grant, R., Coopman, K., Medcalf, N., Silva-Gomes, S., Kara, B., Campbell, J. J., Braybrook, J., & Petzing, J. (2019). Quantifying operator measurement uncertainty contributions within post-analytical Flow Cytometry data. 4th Parenteral Drug Association Europe Meeting, Amsterdam, the Netherlands, 26/06/19.

Grant, R., Coopman, K., Mayer, S., Kara, B., Campbell, J. J., Braybrook, J., & Petzing, J. (2018). Assessment of operator variation in flow cytometry measurements using gauge repeatability & reproducibility techniques. *Cytotherapy*,*20*(5), S77. ISCT 2018 Annual Meeting, Montreal, Canada, 2-5/05/18.

Grant, R., Coopman, K., Mayer, S., Kara, B., Campbell, J. J., Braybrook, J., & Petzing, J. (2017). Flow Cytometry Operator Variation Study. Future Investigators of Regenerative Medicine Symposium, Girona, Spain, 28/09/17.

Grant, R., Coopman, K., Mayer, S., Kara, B., Campbell, J. J., Braybrook, J., & Petzing, J. (2017). Flow Cytometry Operator Variation Study. Bioprocess UK, Cardiff, Wales, 28-30/11/17.

"The subjectivity of consciousness is an irreducible feature of reality, and it must occupy as fundamental a place in any credible world view as matter, energy, space, time and numbers".

Thomas Nagel, The View from Nowhere, 1986

"The measure of a man's real character is what he would do if he knew he would never be found out."

1st Baron Thomas Macaulay, 1926

Prelude: Thesis Context

Problem Statement

Subjective operator assessment is a common although often silent element of quantitative measurements. Flow Cytometry (FC) is a commonly used instrument within biological and particulate measurement, which requires manual subjective intervention for quantification and interpretation [1]. This platform is heavily used within clinical diagnosis and Cell and Gene Therapy (CGT) manufacturing and Quality Control (QC), with various standardisation and harmonisation efforts reporting large measurement variance [1,2]. This causes medical confusion, misdiagnosis and treatment error [3]. When used within CGT manufacturing, this poorly quantified measurement variation may be the difference between life and death. False positives, false negatives, opinions and inspector education cause medical error, with experts calling for international consensus for repeatability and reproducibility across sites [4]. As well as the impact on the patient, CGTs have time-, resource- and money-intensive manufacturing processes, so poor measures have significant impacts on product quality and therapy variability within international healthcare systems.

Context

The CGT market has accelerated over recent years, with the first licensed treatments becoming available across international healthcare systems. CGTs have been heralded as the fourth pillar of healthcare, supporting pharmaceuticals, biologics and medical devices. The opportunities for cure and treatment of many degenerative diseases such as cancer, diabetes and Parkinson's disease show promise for a new wave of healthcare innovation [5]. Cancer immunotherapy is leading the way for CGTs, with recent efforts demonstrating treatment of solid tumours and haematological malignancies [6]. Chimeric Antigen Receptor (CAR-T) therapies have provided the first treatments for B-cell malignancies, with Kymriah (tisagenlecleucel, Novartis) and Yescarta (axicabtagene ciloleucel, Kite Pharma) leading the way for many more regenerative medicine products [7]. However, these treatments are not without complications. These live products bring new challenges to reproducible manufacture and control, with a distinct lack of traceable reference material to

ensure accuracy and precision of measurement, impacting the clinical decision-making process [8]. Lack of traceability can make measurement more challenging across decentralised manufacturing systems, where the variability across sites also becomes an issues, in comparison to centralised manufacture where the CGT is manufactured on one site, removing inter-site reproducibility concerns.

The manufacture of CGTs is marking the turn of a new industrial revolution, however, to ensure future demand for these treatments needs to be met with 100 % efficacy and right-first-time delivery [9]. To ensure effective manufacture of the product, the measurement of the product must be to an appropriate level of accuracy and precision, to effectively monitor the product in all stages of in-process analysis and product release. This provides greater control of the product and a better understanding of the sources of variation. CGTs present a significant challenge to underpinning metrology, because there is a lack of traceable reference materials for live, biological products [10]. This is not only a difficulty when defining calibrations and higher reference, but also with core metrological definitions, such as accuracy and bias, because they require a 'true value' to be known. A true value is a traceable measurement up to primary SI units, that is not continually redefined due to finite reference materials. This is further complicated by measurement issues to detect rare cell populations or monitoring residual disease levels throughout the course of treatment, as they approach sensitivity limits of instruments. There have been unified efforts for volume traceability, with standards still being explored for individual cell counting.

FC is a core measurement platform used within the release and in-process measurement of CGTs, but it requires 'expert' interpretation of the data files on top of process optimisation [11,12]. Better quantification of this interpretation variability is required for greater manufacturing process control. This could be achieved using measurement uncertainty principles, required to report for ISO 17025 and ICH Q7 standards for laboratory accreditation and good competence monitoring and reporting [13,14].

Research Focus

This research attempts to quantify the variability contributed specifically from operator subjectivity within analysis of FC data. This complements other initiatives that focus on variation of other upstream factors such as instrumental setup or fluorescence calibration [15,16]. It provides more specificity to the variation contribution from FC gating, not just a measure of the whole process, which other gating and analysis studies have previously quantified [1,2]. These studies have focused on cytokine assays, whereas this novel research has a CGT focus, assessing inter-operator measurement variation across three models of complexity from three cell models that relate to CGT products.

Core Hypothesis and Objectives:

As the complexity of the cell models increase in the FC data for participants (as defined by increasing protocol sequences and FC dimensions), the between-participant ranges of Coefficient of Variation (CV) and measurement uncertainty will increase, indicating greater contributions of variation from the participant analysis as shown in Figure 1.

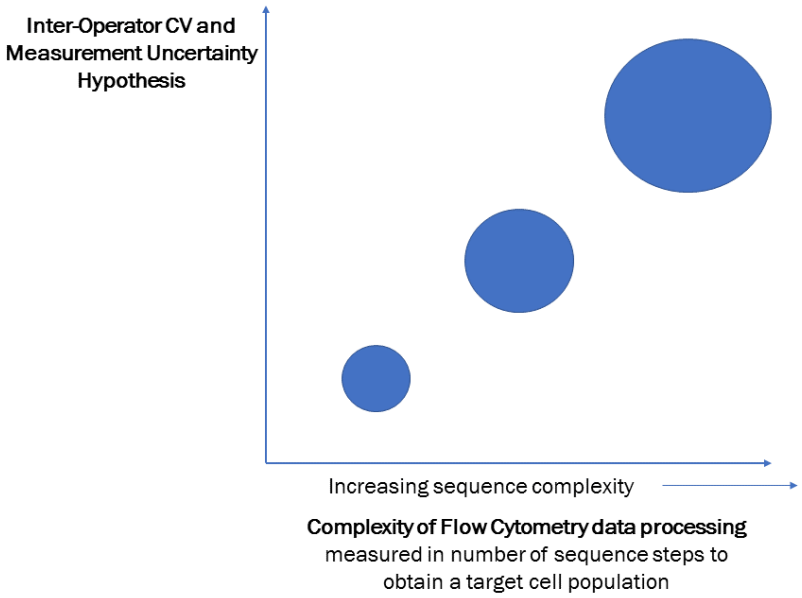


Figure 1 Core thesis hypothesis: as complexity of FC data increases, inter-participant analysis variation also increases.

The objectives of this research are as follows:

- Explore the integration of External Quality Assessment Schemes into CGT manufacture, to identify benefits already seen within clinical chemistry and clinical Flow Cytometry.
- Understand how measurement uncertainty works, and how it can be used to quantify variability components.
- Initially investigate how much variation is present between operators when they analyse the same univariate data.
- Identify differences in outlier specifications when different manufacturing performance criteria and log transformations are applied, to discuss whether harmonised boundaries can be achieved.
- Design human participant experimental studies to effectively capture measurement uncertainty across multiple complexity models, using a Gauge Repeatability & Reproducibility structure.
- Generate standard sets of FC data files to use for each representative complexity model, which increase in the number of gating steps to complete, due to increased data dimensionality.
- Quantify measurement uncertainty of participant FC gating across three models, with participants gating increasingly complex data as each model progresses.
- Investigate whether the use of operational procedures as protocols reduce inter-participant variability when gating.
- Compare the inter-participant ranges of variability between the three complexity models, defined as either inter- and intra-participant Coefficient of Variation (*CV*) or measurement uncertainty, and discuss the suitability of these measures.
- Compare human factor variables gathered within survey exercises to variability data from the complexity models to understand how experience affects variation of data.
- Identify other concerns of participants when gating FC data, and interpreting it from literature, to identify where future efforts for standardisation of measurement are needed.

- Identify other frequently used CGT measurement platform, to determine common operator variability and subjectivity within post-analytical phases, and demonstrate translational potential of this novel research

Thesis Structure

This thesis is broken into three core stages: a critical review of current work and theory, core experimentation to monitor changes in variance between FC participants, and a comparison of variances to human factor issues present in FC.

- **Chapter 1** presents a concise literature review of background knowledge and relevant research efforts around operator subjectivity and variation measures within FC.
- **Chapter 2** explains the relevant measurement statistics and measurement uncertainty principles used for data analysis in this research.
- **Chapter 3** presents an initial investigation of the magnitude of variation within operator subjectivity, by monitoring the inter-participant variance when gating univariate histograms. It also questions the use of different performance acceptance criteria in relation to outlier specification and transformation of data within CGT manufacturing scenarios.
- **Chapter 4** defines a basic data model to initially measure operator variance within FC analysis using a 3-step 2-colour FC exercise, gating a pluripotent Embryonal Carcinoma cell line population (2102 Ep).
- **Chapter 5** uses an intermediate model with a more complex 5-step, 3-colour FC panel, to stratify a naïve T-cell subset from a Peripheral Blood Mononuclear Cell (PBMC) population.
- **Chapter 6** presents the final complex model which uses an 8-step, 8-colour FC panel to identify the percentage of transduced engineered T-cells in the population.
- **Chapter 7** compares *CV* and measurement uncertainty results from Chapters 4 to 6.
- **Chapter 8** extends this comparison against participant questionnaire results, to identify any potential human factor correlations to variance measured.

- Chapter 9 presents a translational pilot study of the subjectivity issues identified within imaging platforms and qPCR, to show post-analytical subjectivity exists across CGT measurement platforms.
- Chapter 10 concludes the thesis, discussing novelty met, and future work identified to progress these research efforts.

Proposed Novelty of Research

- A critical review of current External Quality Assessment Schemes (EQAS), to identify opportunities for integration into CGT manufacturing.
- Relevance of application of manufacturing outlier definitions to define process control limits.
- Application of Gauge Repeatability & Reproducibility techniques to Flow Cytometry post-analytical variation.
- Use of measurement uncertainty for Flow Cytometry post-analytical variation.
- Use of measurement uncertainty for better measurement resolution and control through FC measurement which is representative of CGT analytical techniques.
- Quantification of participant subjectivity as a function of cell model complexity.
- Increased variability as a function of cell complexity.
- Development of a new performance monitoring diagram to aid continuous improvement of variation.
- Comparison of measurement variability metrics suitable for precision of FC measurements within CGT manufacturing contexts.

Chapter 1: Background Knowledge

1.0 Introduction to the Chapter

Chapter 1 presents a background and critical review of current FC standards and procedures in place to monitor participant variability in data analysis. It gives a short overview of the FC technique, and then reviews current efforts to standardise post-analytical data analysis and use of measurement uncertainty for the technique. Identification of sources of uncertainty within post-analytical data analysis are reviewed, to align with the core thesis hypothesis of measuring participant subjectivity contributions to the final measurement. Error components have already been attributed to sample preparation and instrumental setup, which have been investigated thoroughly by different working groups and therefore will not be covered within this specific review [16–20]. Current regulations for participant variation are also discussed, which currently exist for clinical applications of FC, and how this could integrate into a Cell and Gene Therapy (CGT) manufacturing context. Automation of FC data analysis is reviewed in comparison to manual gating efforts, because many recent efforts have attempted to remove the operator from the analysis, with different levels of success. Indeed, the use of automation is a recurring theme throughout this research, because of growing interest (but also scepticism) within the industry [20,21].

1.1 A short overview of the Flow Cytometry technique

Flow Cytometry is an analytical process used extensively within different biological fields, and within cell therapy it is mostly used for phenotyping [22]. Flow Cytometry is a laser-based technique that facilitates identification and analysis of individual cells and populations based upon size, granularity and expression of certain fluorochrome-conjugated markers, known to be indicative of specific cellular identification or function. Flow Cytometry can be used for a wide variety of applications such as assessing cell viability, analyzing DNA, cell death, pH and cellular membrane microparticles, cell surface or intracellular antigens and markers, chromosomes and specific

proteins [23,24]. However, a lack of traceability and reference materials within In Vitro Diagnostic (IVD) Medical Devices such as Flow Cytometers makes it much more difficult to provide confidence to results obtained [25]. If the measurement is ultimately incorrect (through a combination of added uncertainties and misinterpretations) then it becomes more challenging to define true representative values. Within a clinical context this can have severe knock-on effects, because if the metrology is wrong, the quantification and therefore diagnosis may be incorrect, which could be dangerous for a patient undergoing treatment [25,26]. This is also the case for CGT manufacturing, where Flow Cytometry is used as a core measurement technique at various manufacturing process steps, cell sorting and final therapy product release [27].

Analytical cytology dates back to the 1950s (and even earlier through the use of microspectrophotometers) [28] and as technology advanced through the decades, more advanced apparatus and data analysis software became available due to evolution of modern computer processing methods [28]. Modern Flow Cytometers are often combined with Fluorescence-Activated Cell Sorting (FACS) techniques, however, for the purpose of this review the main focus will be on Flow Cytometry without cell sorting, because this reduces the amount of variability in the result due to simplified processes.

1.1.1 Flow Cytometry Instrumentation

FC relies upon a laminar flow cell aligning cells through a laser pathway, depicted in Figure 2. Cell size is detected by forward scatter (FSC) of the beam, and granularity and fluorescence emission spectra of the cell-specific markers are detected by side scattered light (SSC), split through various optical filters [29]. This data can be used to generate a flow cytometry standard (fcs) file and plots of target populations, ready for identification and stratification by an operator [30]. Typically, a minimum of 30,000 cell events would be run through the Flow Cytometer, to ensure a valid data set was captured to conduct further analysis on.

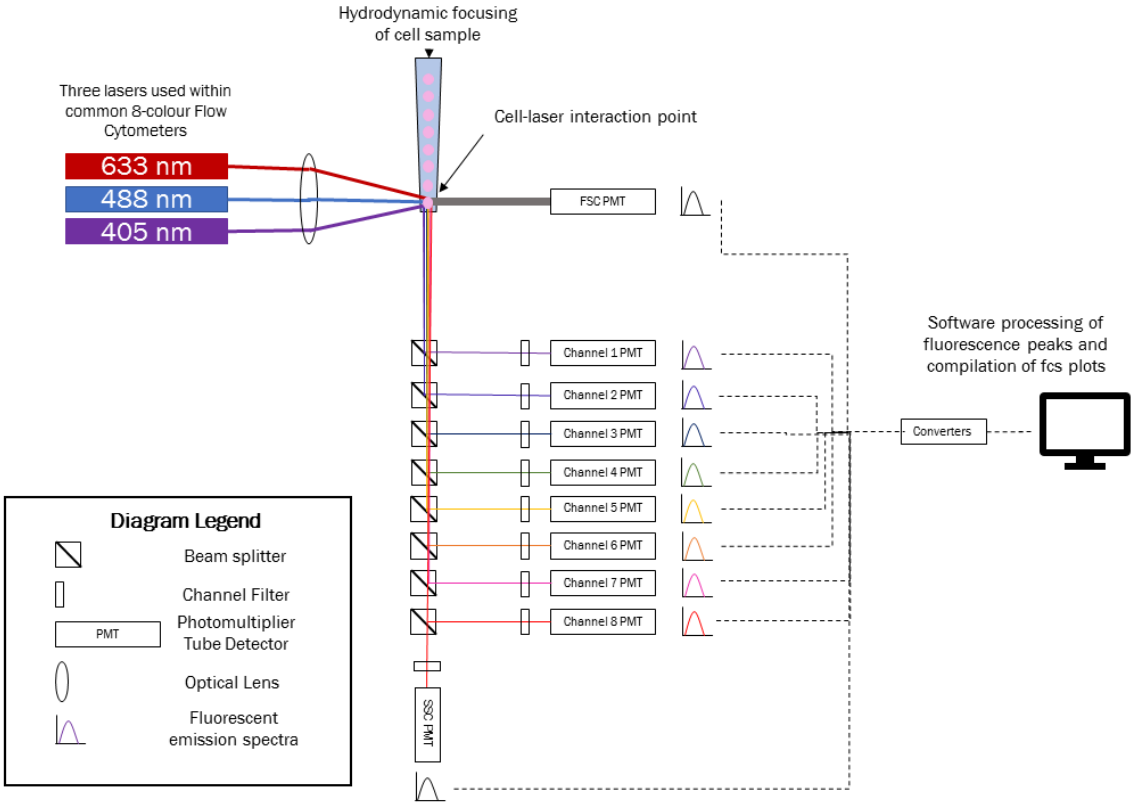


Figure 2 Flow Cytometer schematic

1.1.2 Flow Cytometry post-analytical data analysis

An example of fcs file data is shown in Figure 3, where each cell that is processed through the Flow Cytometer is shown as a single dot on each of the sequential plots, which build up to high (red) and low (blue) density regions indicative of cell populations. The location of the dot (or cell event) is determined by the respective fluorescence emission detected through each channel sensor [29]. The fluorescence intensity is plotted as a histogram (for univariate data), or a scatter plot of data can be created with two or more optical channels used for measurement, such as those seen in Figure 3. To distinguish a specific target population from the total sample, filtering or threshold gates are then applied manually by an operatpr (different shaped areas defined by boxes, ellipses, quadrants or polygons, working left to right in Figure) to further stratify the population. By stratifying the data, 'noise' created by debris and non-specific binding (for example) is removed, allowing relevant spectra to be analysed [31].

The emission spectra are excited by laser interaction with specific fluorophores conjugated to antibody markers which bind to extra-cellular or intracellular markers on the cell [29,32]. Other stains can be excited by laser interaction but are not bound to antigen markers on the cell surface, for example a variety of methods to stain cell debris or nuclei can be used to discriminate dead cells from the live population. Antibody markers are chosen based on the cell type in question and what specific sub-populations are being targeted. If these markers fluoresce, it indicates the marker has bound to the cell, so the target cell population can be identified because of higher or lower fluorescence intensity regions of the plots [29].

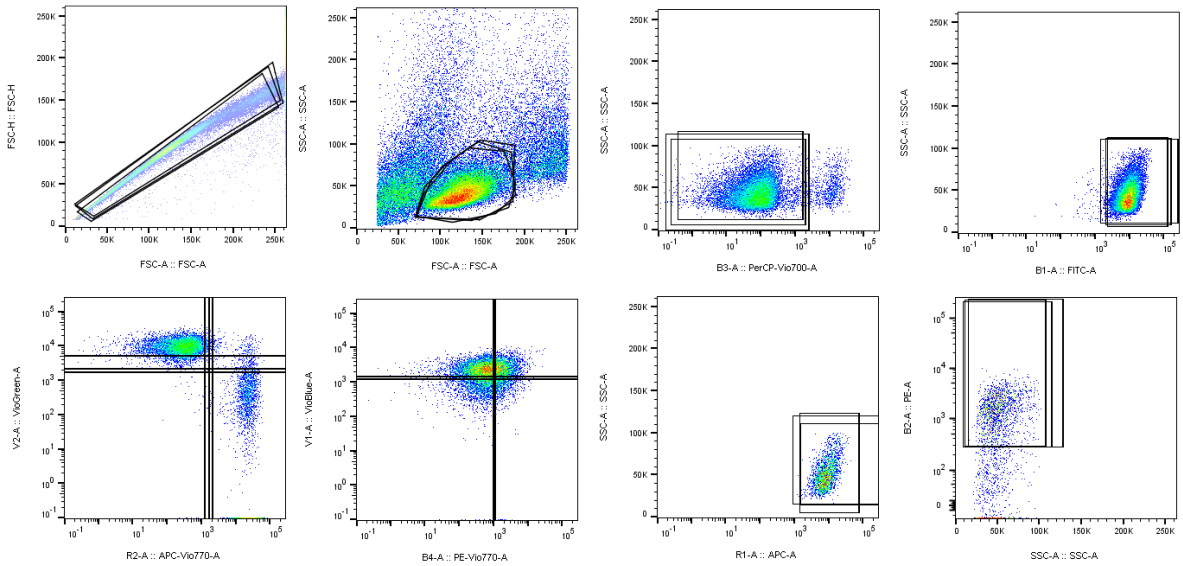


Figure 3 Example of Flow Cytometry gating within post-analytical Flow Cytometry

1.1.3 Flow Cytometry post-analytical controls

Control files can be created for each fluorescent marker, to help separate positive expressing populations from dim or negative expressing cells [33]. These controls are commonly referred to as Fluorescence Minus One (FMO) controls, because they contain a mix of all the markers being tested in a Flow Cytometry panel, except the respective marker for the optical channel in question. This gives a good indication of fluorescence spillover from markers in the panel when combined to the specific cell type in question, to set a threshold between positive and negative expression [34]. Examples of these controls are shown in Figure 4, with the fully stained sample shown with the final

gate applied in (a), and the FMO control files for each respective channel plotted on the axes shown in (b) and (c), demonstrating how the positive thresholds for the *x* and *y* axes are set. However, FMO controls cannot account for any background staining from the antibody used within that channel during analysis which could skew results, and is why isotype controls have been popular for many years [34]. Isotypes account for nonspecific staining of an antibody of a particular isotype conjugated to the required fluorochrome, because different isotypes can have different levels of background staining within the channel [32,33]. However, because these do not account for any other markers used within the Flow Cytometry panel, they are not preferred if they exhibit a lower amount of background staining than spillover from other markers within the panel [34].

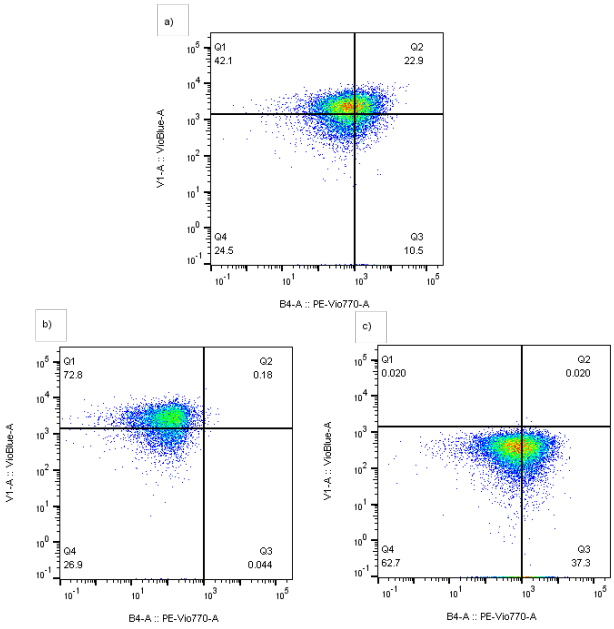


Figure 4 Example of how FMO gates are applied, a) Fully stained sample with gates applied from FMO controls, b) FMO control for the marker on the horizontal x axis, c) FMO control for the marker on the vertical y axis.

Placement of gates is aided by the various controls available to operators, but the decision of where the final gates are placed is mainly subjective, relying on the operator's perception of density, influenced by plot presentation, graphics, understanding of the cell type, use of controls and many other factors. This research aims to address and measure this subjectivity issue, to quantify the variability contributed from an operator's analysis of data. It is recognised that there are many other potential sources of variation in a FC measurement, but these are out of scope of this thesis and have been controlled within the experimentation in Chapters 3 to 6.

1.1.4 Flow Cytometry automated analysis

Many recent efforts to automate FC data analysis have shown the ability of machine learning algorithms to identify common cellular subsets, processing more data in less time than a manual operator [21]. This potential capability of processing large batches of highly-dimensional data still requires start and end manual screening processes, to ensure the algorithms can run effectively [35]. Downsampling (decimation) of data is required to reduce the computational time (by reducing the number of events processed) [36]. A percentage of the original data points are taken as a representative sample of the whole file population, if datasets exceed 20,000 events. t-Stochastic Neighbour Embedding (t-SNE) is a popular analysis algorithm, which 'plugs in' to various software platforms and will be used as an example here, plugged into Flowjo Version 10.0.8r1, because this platform has been used for subsequent FC studies [37,38].

There are different variables that still need to be defined by the user for clustering purposes; specifically, 'Perplexity', number of algorithm iterations and Eta. Perplexity essentially defines the distance between clusters using a sliding scale from 2 to 100. This requires optimisation to best represent the balance between local and global cluster populations using 'nearest neighbours' clustering. Algorithm iterations can be set between 300 to 3,000 iterations (the number of times the computation is applied), where higher iterations cause greater separations of final clusters. Eta (the learning rate of the algorithm) controls how much of the 'nearest neighbours' weightings are adjusted to obtain a minimum probability difference between data points [39]. Finally, when t-SNE has clustered the data, the user still has to manually gate to choose populations, which are unlabelled, increasing difficulty of validating subsets [39]. An example of different variable choices have been shown in Figure 5, to exemplify how different variable scales affect the shape of the t-SNE output, which still require gating. The data input was EC 2102 Ep cell line data, stained with 2 pluripotency markers, which are used in the basic uncertainty model in Chapter 4.

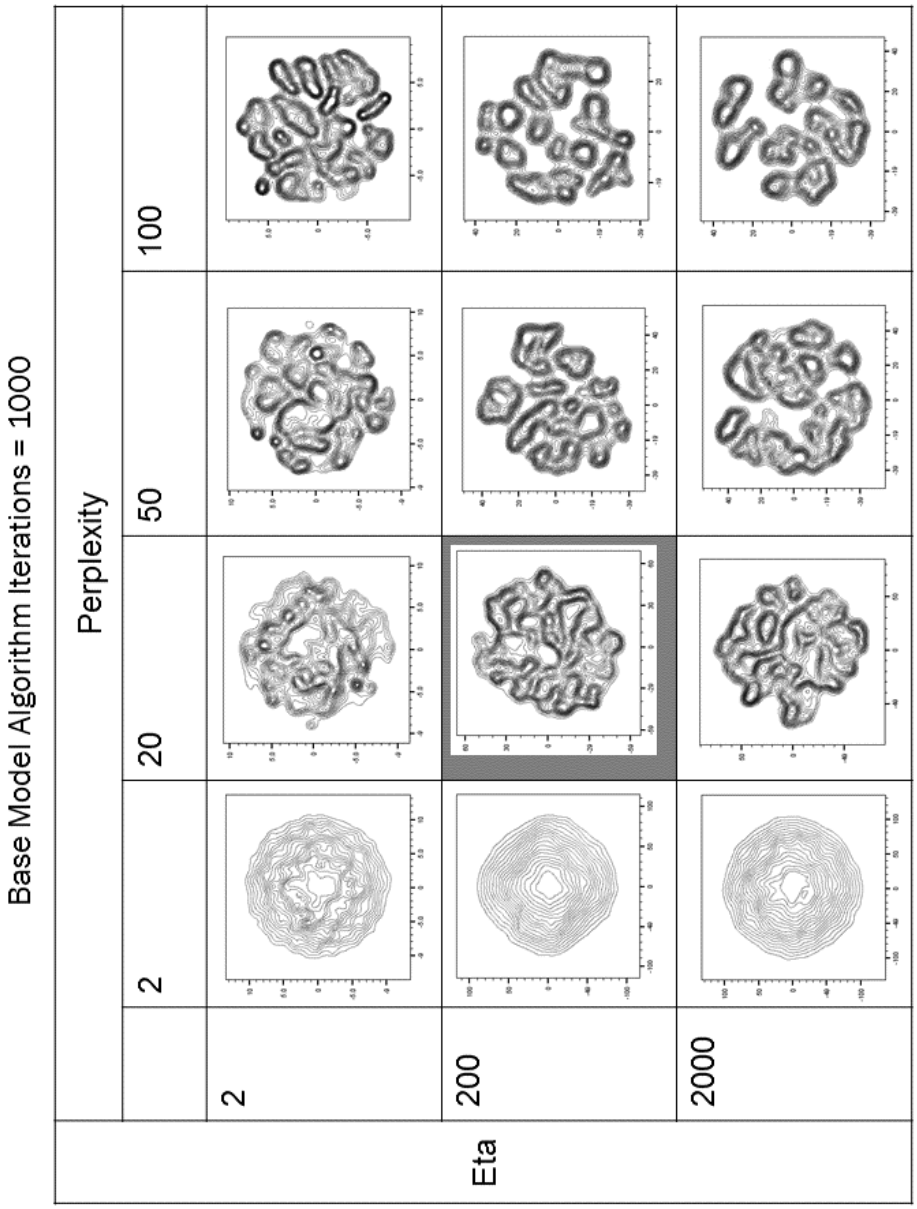


Figure 5 t-SNE iterations of a single FC data file, analysed with different variation limits, for 1,000 algorithm iterations. The grey cell indicates commonly used variables.

The algorithm repeatability has been explored to identify effects of t-SNE variables on simple clusters [39], concluding that multiple runs will produce the sample global shape, but certain data sets produced different cluster shapes on repeats, and there are no fixed number of repeats to validate settings. In addition, the t-SNE cluster size does not necessarily represent the actual size of the cell population, due to adaptation of the algorithm to create even cluster sizes to identify any smaller or rare subsets [39].

When considering the application of machine learning algorithms for cellular analysis within a Quality Control or Manufacturing scenario, the inverse relationship between quality/uncertainty and time needs to be addressed. Whilst t-SNE is a computationally powerful tool which enables users to analyse multiple samples made up of high dimensionality data, it still requires users to manually clean up the data and select the appropriate clusters within the analysis, exhibiting the same variation issues found within traditional manual gating, as well as inherent variation from the machine learning algorithm itself. Use of such a tool would require a high level of validation and optimisation, which could cost manufacturers in time and therefore money when quantifying uncertainty in the system, when pre-determined operator metrics may already be established over time.

1.1.5 Flow Cytometry uses within Cell and Gene Therapy manufacturing

Quality Control (QC) measurements are required for all stages of CGT manufacture, not just the final product release for patient infusion. The versatility of FC means it is heavily utilised for CGT product measurements at various points during the manufacture of a therapy for different factors. For example, a recent review highlights common considerations for CAR-T product release tests, due to their accelerating prominence within the CGT market [8]. The manufacturing process of an autologous CAR-T cell therapy follows the process in Figure 6, with sample measurements taken at defined QC points [40].

Phenotyping will be specific to the therapy being manufactured, however, it will always be conducted at the start of sample preparation to initially measure the donor material and achieve an idea of biological starting variation, which can range considerably [12]. Along with cell counts, this gives manufacturers an indication of expansion times for the product, or whether more starting material needs to be taken from the patient.

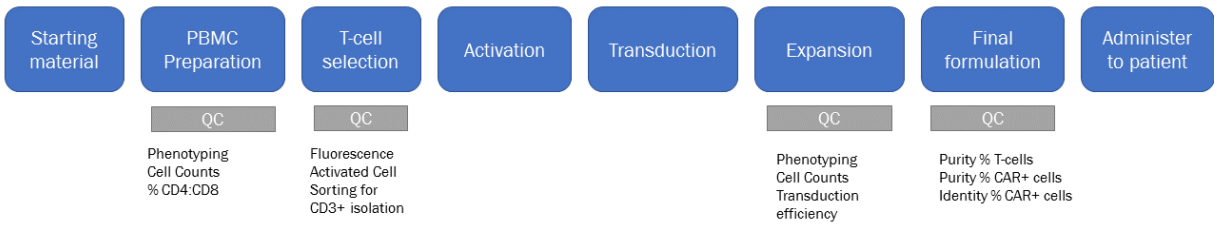


Figure 6 Flow Cytometry assays used within CAR-T manufacturing

When selecting T-cells, FACS is a Flow cytometry-based method used to sort a cell sample for CD3+ Lymphocytes if this configuration is available, although a common method for separation uses Dynabeads, because they are also used to activate the cells, to simplify processing stages. They have also been shown to have strongly correlated cell counts with Flow Cytometry, making them a comparable but cheaper alternative [41]. During expansion of the product, phenotyping will be conducted at regular intervals to monitor phenotypic changes to the CGT product. In addition to this, transduction efficiency is an important measure of viable cells expressing the CAR+ to measure transduction efficiency of the viral vector.

Once these checks are complete, the final therapy product will be formulated for administration to the patient. Final product release criteria has been reviewed in [8], with Flow Cytometry being used to identify the purity of the product, by counting the % T-cells and the % CAR+ cells, with this latter measure also being used to define the identity of the product. This composition needs to be quantified in order to deliver the correct dose to the patient, based upon their bodyweight. For example, to treat B-cell Acute Lymphoblastic Leukaemia (ALL) in paediatric patients < 50 kg, a dose of 1 to 3 bags containing 0.2 x 10^6 to 5 x 10^6 CAR+ viable T-cells/kg of body weight are required, and for those > 50 kg, this dosage increases to 1 to 3 bags of 0.1 x 10^8 to 2.5 x 10^8 cells/kg bodyweight [42]. This indicates the needs for repeatable, standardised measurements to ensure the therapy can be delivered effectively to the patient, having been optimised throughout manufacture.

1.2 Standardisation of Flow Cytometry Practice

1.2.1 Summary of efforts across FC measurement

Many efforts have been and are currently being made to standardise Flow Cytometry as a measurement technique, to make results more comparable between instruments and sites, with a variety of commercially available standards being sold to users [16,43–45]. Examples of these are SPHERO™ Rainbow beads used for cross-platform calibration of fluorescence [46], and NIBSC reference materials [47] for flow cytometry cell counting. A lot of emphasis has been placed on fluorescence standardisation [16] and instrument standardisation protocols [17], however, standardisation and traceability of these elements does not guard against downstream variation of manual data analysis and reproducibility concerns.

1.2.2 Acceptable variability of post-analytical FC data

Once the biological sample has been processed, the data is exported and stored in fcs file format, to allow processing across different software platforms with no data manipulation. The file standard provides uniform storage of data, for fair processing and reporting [30]. Software used for analysis need to comply with Part 11 of the Food and Drug Administration (FDA) Code of Federal Regulations (CFR), to ensure correct storage of electronic records for cell product traceability [48]. If publicly reporting Flow Cytometry experiments, the Minimum Information about a FC experiment (MIFlowCyt) guidelines should be followed for good reproducibility of data [49,50]. The minimum information project, also has guidelines to support further assay development and reporting, specific to T-cell and NK-cell experiments [51]. Within FC, there have been significant efforts to harmonise panels for experimental design, reporting and analysis to optimise experiments within the community [52], as well as harmonisation between instruments [53,54].

It has been recently reported that there is a large reproducibility crisis across science [55,56], and this exemplifies similar concerns for the development, manufacture and control of Cell and Gene

Therapy production [40,57–61], as well as standard Flow Cytometry experiments [20,62]. Reproducibility of measurements and products require traceable measurements and processes, and these must be validated and shown as part of regulatory Market Authorisation for CGT products [4,63,64]. Without reproducibility, there is less confidence in the measurement process, manufacturing decisions and more concern when releasing the product for the patient treatment [65]. Different metrics are commonly used to monitor variability, but may require more resolution to improve traceability, repeatability, reproducibility and therefore confidence and quality.

Repeatability of FC measurements are commonly reported as precision (standard deviation (*SD*) of repeats) or *CV* (%). *CV* is a combination of the mean and standard deviation of repeats and is often easier to monitor variability [29]), and these variation metrics are fully defined in Chapter 2, Section 2.2.2. Different studies have monitored *CV* contributed from different FC measurement components, to potentially provide guidance on experimental variability throughout the development of FC [66–68], and discussed the implications of validating FC in a regulated environment [62]. Different levels of acceptable variance have been defined across Flow Cytometry literature, with different metrics and resolution to address and reduce it over time. The International Council for Standardization of Haematology (ICSH) and the International Clinical Cytometry Society (ICCS) published a 5-part series summarising FC variability issues through the entire FC process [45,69–71]. Part V provides guidance on acceptable limits for *CV*, where < 10 % *CV* is suitable for measurements, or < 20 % *CV* where rare cell events are concerned [71].

Percent *CV* is preferred to *SD* as acceptance criteria by the ICCS, because it normalizes variations at lower levels of event detection. This imprecision metric should be taken from a minimum of five samples assayed in triplicate (at least) during the same analytical cycle. It has been noted that although the data is impacted by a subjective analysis, the technical assay performance is reproducible [71]. This gives better clinical assurity of assays requested, and it is easier for pathologists and clinicians to make decisions for the patient with more confidence in the measurement process. These references ranges would have been defined by clinical cases for

normal and specific disease states (typically 60 males and 60 females included in range analysis) to ensure clinical comparability for specific test results. Disease states are inherently more challenging to generate ranges for, however, patient cohorts in need of routine checks may be more readily available for tests, which could aid this data generation [71].

CV has been considered to quantify measurement uncertainty within the UK National Health Service (NHS), for example providing Key Performance Indicators (KPIs) of Feto-Maternal Haemorrhage (FMH) monitoring with Flow Cytometry. Measurement uncertainty is usually a combination of *SDs*, but in this instance the NHS have used *CV*, with KPIs of < 15 % for good uncertainty, 15 % to 20 % for satisfactory and > 20 % as unacceptable [72]. A recent international research review of Flow Cytometry CD4+ count measurements indicates that BD FACSCount beads used as reference standards show an inter- and intra-laboratory precision of <15 % [73]. A breakdown of uncertainty variation sources within phenotypic measurement gives further identification of contributing amounts of variation of the measurement [74].

Poor reporting and control of FC results leads to false positive and false negative results of samples, facilitating poor decision making of diagnoses and treatments, seen in many other laboratory measurements. A recent review has identified many pre- and post-analytical errors in laboratory haematology [3], specifically in automated cell counting. Haematology laboratory errors can contribute as much as 62 % and 23 % of measurement variability from these pre- and post-analytical errors respectively, which has not changed over 10 years [75]. If not caught, these issues cause incorrect or missed opportunity for diagnosis, but if they are caught before the result is issued, it causes delays in diagnosis which increase patient anxiety and it is a missed opportunity for diagnosis if a specimen cannot be retaken or accessed [76]. Diagnostic error is difficult to accurately estimate, but approximately 12 million USA citizens having suffered a diagnostic error, half of which were significant, and can impact patient safety [77].

Standardisation of measurement positivity thresholds is required across many platforms, exemplified by anaemia reporting across international bodies. The World Health Organisation (WHO) define lower limits of haemoglobin (Hb) concentration at 130 g/L in adult males, compared to the Centre for Disease Control (CDC) lower limit of 135 g/L [3]. Although the difference is small, it has a significant impact on referrals or missed treatments. Different External Quality Assessment Schemes (EQAS) have different performance criteria for CD4+ counts returned from participating laboratories, which can have a significant impact in patient safety, as discussed in the next section.

1.3 External Quality Assessment for reproducibility

1.3.1 Relevant standards for analytical competency

Both ISO 17025 and ICH Harmonised Tripartite Guideline Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients Q7 have grounds for personnel hired to work in the test environment [13,14], stating that personnel hired should have adequate qualifications for the role, and training should be provided regularly for operations performed by the employee. ISO 17025 provides more depth with regards to processes that require interpretation (such as Flow Cytometry), stating they should be performed by operators who are familiar with the technology, how it is used and possible defects that can occur, but there is no stated requirement to formally measure operator variability, which can impact results [13].

ISO 17025 is a standard which manufacturing and test laboratories ideally need to achieve accreditation with, providing patients, clinicians, healthcare trusts and companies with confidence that procedures are in place to deal with deviations and the site can monitor and control the product effectively. Uncertainty plays a part in this analysis, with ISO 17025 stating that all equipment used requires an uncertainty estimation (ISO 17025 Section 5.4) and any opinions or interpretations around measures need to be stated in test reports, under the basis in which they have been made (ISO 17025 Section 5.10.5). Flow Cytometry is a process that requires interpretation to generate a result/diagnosis, so this must be factored into release criteria, and is the subject of this research. Measurement uncertainty is explained in greater detail in Chapter 2.

Flow Cytometry operators can also obtain Cytometry Accreditation (Specialist in Cytometry, SCYM) [78], to ensure knowledge of data analysis such as understanding gating tools, statistical methods and assay validation methods, as well as standards and controls. Clinical Flow Cytometrists require experience and accreditation to correctly identify a variety of diseases and disease states from a sample. The manufacturing and QC scenario differ, because product conformance is required. Therefore, the FC will not necessarily be diagnostic, it will be used to assess a specification.

As the cell therapy manufacturing environment may not be clinical, FC operator certification requirements need to be scoped as well as potential participation in an External Quality Assessment Scheme (EQAS). This may be an established EQAS or an EQAS with different guidance. Comparability can therefore be assessed at different levels, but there may be inherent differences between sites and companies with regards to output results.

1.3.2 External Quality Assessment and accreditation

EQAS are used to verify correct implementation of protocols and measurements of particular targets across multiple centres, and are also known as External Quality Control, Interlaboratory Comparison Surveys or Proficiency Testing [79]. Proficiency testing exists to provide greater metrological traceability of measurements to primary SI units, to ensure that calibrators and instrumentation are not negatively affected by poor use [80].

EQAS addresses standardisation of practice in which measurement equipment is used (once a problem is identified), as well as commutability of the standardised measurand to enable comparison of these factors [25]. Clinically, centres that analyse patient material need to have validation through an EQAS, which are further traceable to ISO 17043 for conformity assessment [81]. This helps centres achieve ISO 15189 accreditation for quality and competence of medical laboratories [82], replacing Clinical Pathology Accreditation (CPA) in the NHS. In Vitro Diagnostic (IVD) medical devices, such as FC are also required to show metrological traceability of biological quantification to obtain CE marking [83]. This ISO 17511 compliance must conform to one of the defined traceability chains depending on the availability of different primary and secondary calibrators and procedures, illustrated in Figure 7.

CGT manufacturing centres would not necessarily need ISO 15189 accreditation, because they are not performing clinical pathology, although this is something that may translate from the clinical environment, due to different manufacturing scales. However the site, processes and QC would need to be Good Manufacturing Practice (GMP) compliant when manufacturing a licensed product,

and would need to follow guidelines set out in ICH Q7 [14] and gain ISO 17025 accreditation for competence of testing and calibration laboratories [13].

As previously stated, QC operators would need accreditation to ensure they remain GMP compliant, however, this calls into question how much further up the product development pipeline this level of competency and reporting is required. To design in quality to the product, all compliance should be adhered to from the start of product and process development stages, to ensure good translation further down the pipeline and to better maintain integrity and compliance. However, Process Development Scientists involved in later studies in this thesis (Chapter 8: Questionnaire results) have highlighted that this level of accreditation is not known or understood well. A lot of validation would be completed before handing over processes to further operations, however, this does not account for personnel accreditation.

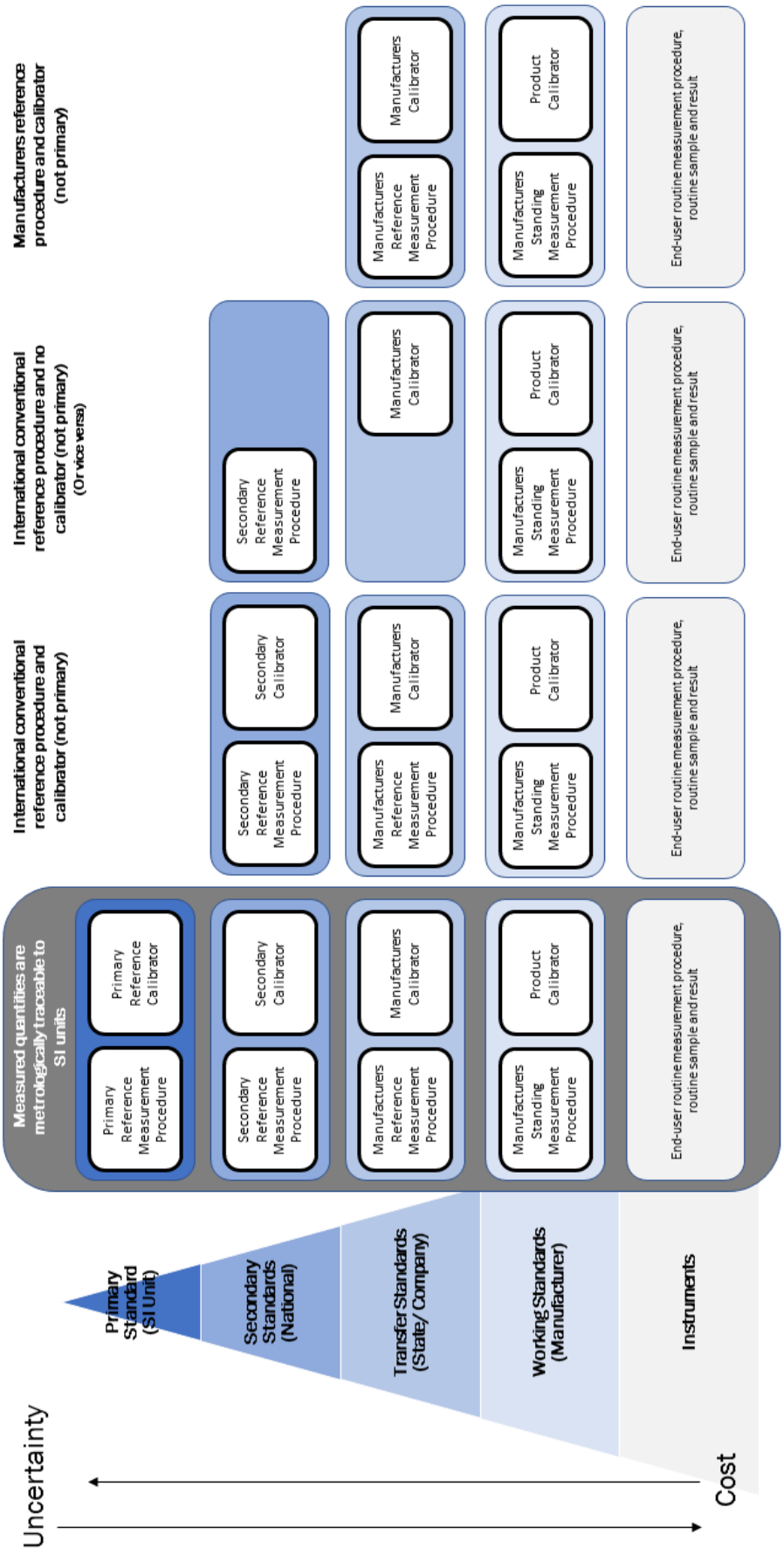


Figure 7 Traceability chains of ISO 17511 calibrators

Traceability of EQAS as part of the diagnostic measurement process has been more thoroughly investigated for clinical biochemistry. In recent years, this has qualified particular analytes and proteins [25,26,79,80], with the quantification of measurement uncertainty (required by ISO 15189 accreditation [82]) provided to indicate variability contributions to the measurements [84]. Many traceable calibrators have enabled these small uncertainties to be specifically defined, giving more measurement integrity and confidence to the analytical decisions [80]. These are underpinned by international metrology efforts to continually improve traceability of laboratory medicine, by the Joint Committee for Traceability in Laboratory Medicine (JCTLM) [85].

1.3.3 Flow Cytometry EQAS for CD4+ measurement

Currently there are many immune-related diseases that are looking to be addressed through a variety of different CGT treatments. From HIV treatment to various blood cancers, reliance on reproducible CD4+ T-Cell counts is necessary and will be used as an exemplar for EQAS here, because of CD4+ analysis undertaken within T-cell therapy measurement [86,87].

To establish comparability of T-cell measurements various External Quality Assessment Schemes (EQAS), have been in existence globally since the late 1980's [88]. These initially looked at cell quantification across dual platforms using various haematological analysers, but it was quickly concluded that a single platform measurement was desirable to reduce variability in CD4+ T-cell counting, with FC the preferred technique for quantification [89]. This correlates with the increase of HIV cases in 1980's, demanding higher throughput of blood samples and therefore treatment of patients. It has also been shown through various other studies that using a single platform of instrumentation reduces variation and error in measurement and diagnoses [54,90,91].

Various CD4+ FC EQAS are in clinical operation around the globe, with the most notable bodies and advancements in standards and teaching coming from the United Kingdom, central Europe and Canada, with output summarised in Table 1. These have developed over the last 20 to 30 years

with increasing laboratory participation and increased frequency correlating with a reduction in variation in participants [92]. Significant developments were made in EQAS in the 1990's and at the turn of the millennium. The most significant scheme developed is UK NEQAS (with international outreach) [93], which offers a variety of FC EQAS accredited by the United Kingdom Accreditation Service (UKAS) that relate to different Clinical assays required for haematology [92]. These include CD34+ Stem Cell Enumeration, Immune Monitoring, Leukaemia Immunophenotyping and Diagnostic Interpretation and Low Level Leucocyte Enumeration amongst other schemes for specific disease states [93].

Table 1 Summary of CD4+ EQAS

EQAS	Variability metric	Acceptance limit	Literature
UK NEQAS	*CV* Reduction in Absolute residuals	Trimmed mean ± 2*SD*	[91,92,94–98]
SIHON Score	Point system: majors and minors Discordant % positive	Within personalised boundaries for acceptable errors	[88,99]
Benelux EQA	Residuals	Robust multivariate regression	[100]
Central European Quality Control Program (CEQUAL)	*CV*	Mean ± 2*SD*	[101]
Gruppo Italiano di Citometria (GIC)	Resolution Index	95 % Confidence Intervals	[102]
Canadian QASI-QMS	Absolute deviation from Inter-laboratory mean	Linear regression	[103,104]
Thailand and South East Asia CD4+ EQA program	Mean, Standard Deviation Index (*SDI*), *CV*	*CV*< 15 %, -2 < *SDI* < 2, 95 % CI for PanLeucogating methods	[105,106]
EuroFlow Consortium	Personalised score using Medium Fluorescent Intensity and *CV*	95th Percentile	[15,17,18,18,107]
AFREQUAS Regional African EQAS	*CV*	Trimmed mean ± 2*SD*	[108]

Despite the excellent efforts made by various EQAS to harmonise measurement and procedures, the CD4+ monitoring schemes indicate how there can still be inconsistency, because although the different EQAS can monitor the same cell counts, their acceptance criteria differ. Some of these acceptance criteria differences were summarised by a review in which another way of defining acceptance (through linear regression) was suggested [109]. When this variance in processing is

extrapolated to cover the number of different types of EQAS scheme for just Flow Cytometry, the state of affairs becomes more confusing to navigate for EQAS participants. Furthermore, pathology centres often take part in more than one EQAS, where different scoring metrics and variation metrics are used, increasing confusion between reporting criteria. The statistical standard that governs EQAS procedures, ISO 13528, actually provides several different ways that variance can be quantified and monitored [110], as well as other monitoring factors, shown in Figure 8. It provides options for defining the reference value (a), performance criteria (b), calculation of performance criteria (c) and different graphics for reporting (d).

To initially consider how this could translate into CGT manufacturing, a change in performance statistics is required to align with ISO 15189 and ISO 17025 documentation. Currently most clinical CD4+ EQAS calculate performance as an estimation of deviation, albeit using a variety of metrics such as *CV*, residuals and self-defined statistics. To gain accreditation, laboratories also need to measure uncertainty, so inclusion of this in an EQAS not only provides better resolution of inter-laboratory processing to improve the community, it also allows participants to use those values as part of required uncertainty calculations. This provides more confidence in the measurement, and allows internal continuous improvement, to highlight the variation of operators in comparison to other instrumental and process components.

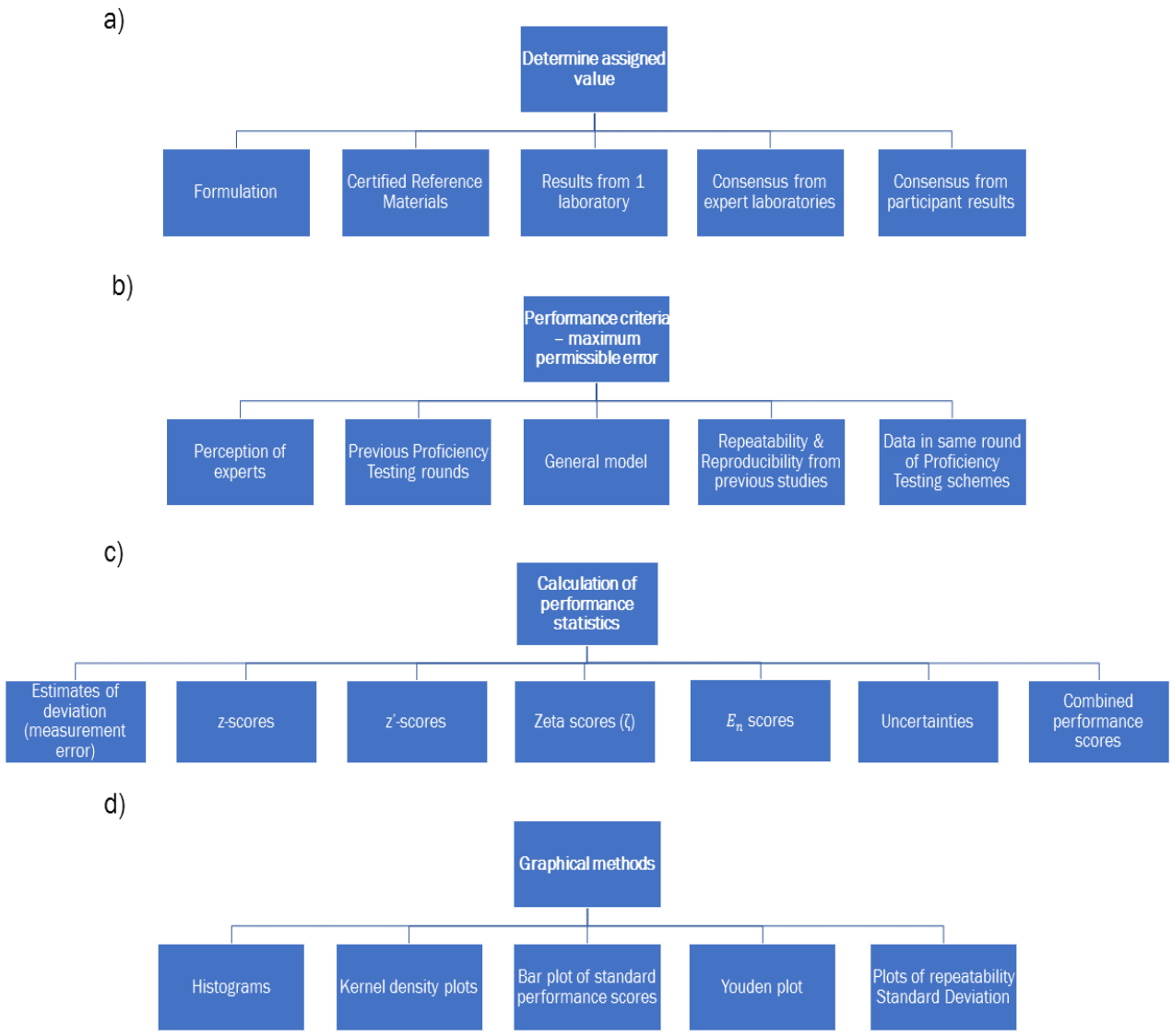


Figure 8 ISO 13528 options for EQAS providers, a) Determination of the initial reference value, b) Performance criteria, c) Calculation of performance statistics, d) Graphical methods to report results

Currently CD4+ EQAS allow participants to report back results for the respective schemes (absolute cell count, and the percentage of this population with respect to the original count in the sample). UK NEQAS have a patented blood stabilisation method used for reference wet sample send-out for analysis, qualified to ensure an assigned value and variance of this reference sample are known for comparison [111]. This provides a quantitative comparison of the reported laboratory results against the reference value, to quickly identify non-conformance. When non-conformance is identified, EQA providers can help respective centres find the root cause, however, because only

the final results are reported, this can be a long effort due to the many sources of contributing variation.

The EuroFlow Consortium took a different perspective for identification of haematological malignancies, by providing participants with detailed instructions to setup equipment and perform compensations [17], as well as providing lyophilised antibody tubes, to standardise the reagents used for defined panel markers [18]. EuroFlow is not an EQAS, however, has been used as a complementary tool to highlight variation in other areas of the FC measurement. By standardising upstream reagents and protocols, it aimed to identify the variation in the system if local blood donations were used (instead of standardised samples) for inter-laboratory comparison, to highlight that standardisation of upstream elements could be achieved.

It also noted how complex panels are very subjective, which can cause more variability in the downstream measurement [17,107], although new software efforts as part of the EuroFlow consortium aimed to standardise reporting for 'normal' cellular subsets in CD4+ and CD34+ FC panels. Both UK NEQAS and Euroflow schemes make assumptions about participants' cell identification abilities, where UK NEQAS assume if a participant can correctly classify the QA sample, they could also classify a leukaemia sample. Euroflow assumes if a participant can execute the Standard Operating Procedure (SOP) correctly and accurately assess lineage markers on normal cells, the laboratory could also assess these markers accurately on malignant cells [107].

Although UK NEQAS is based on final metrics rather than further controlling upstream variables, they have conducted studies to identify variation against standard analysis processes for FC data, in particular the ISHAGE protocol for CD34+ stem cell enumeration [112]. This protocol was established in 1996 by the International Society of Hematotherapy and Graft Engineering (ISHAGE) because lack of standardisation had led to divergent reporting of CD34+ stem cells in peripheral blood, which are considered to be extremely rare (0.01 % to 0.1 %). This standard recommended antibodies, gating strategies, cell separation and lysis techniques to use as well as reporting

mathematics [113]. The 2012 review sent out two stabilised samples to 255 clinical participants and asked them to analyse and report the required CD34+ data. 196 laboratories returned results, and 103 of these also returned dot plots for EQAS clarification. 83 of these participants claimed to use the ISHAGE protocol, but after further investigation only 57 % of these gating strategies were correctly setup. It was further shown that those using this protocol incorrectly on a single platform were twice as likely to fail EQA exercises, and those also using it in a dual platform scenario had a further two-fold increase in failure rate [112]. This indicates that even when standard samples, recommended panels and gating strategies are used, they can still be interpreted or used incorrectly, as well as being influenced by the measurement platform available. EQA always aims for 5 % of returns to be out of consensus, to ensure continuous improvement over time, however in this instance there were 13 % out of acceptance when following the protocol correctly, despite the further 41% of all used applying the strategy incorrectly [112].

1.4 Variability in Flow Cytometry Gating Studies

It has been noted through various EQAS publications documented that variation exists throughout the Flow Cytometry measurement process, with various efforts to standardise these elements for greater measurement precision. These efforts have isolated reagents, instrumental processes and between-instrument variation to understand contributions to measurement deviation. Little focus has been given on the post-analytical gating process within FC, despite being known to contribute a lot of subjectivity to final reported results [21].

The greatest efforts to measure subjectivity from gating have been through standardised cytokine FC assays, which monitored Intracellular Cytokine Staining (ICS) across cryopreserved Peripheral Blood Mononuclear Cells (PBMCs) and shipped whole blood material [114,115]. Inter-laboratory variability was reported as 28 % *CV* for cryopreserved PBMC material, which reduced to 23 % when data was analysed centrally rather than by the respective laboratories [115]. This reduction has also been reflected in standardisation efforts by the Human ImmunoPhenotyping Consortium [20].

Two gating strategies were tested within the Cancer Immunoguiding Program (CIP) for detecting and enumerating HLA-multimer binding cells, with group *CV*s reported as 53 % and 87 % for each strategy respectively, as well as noting that participant variability increased with low frequency T-cells (< 0.1 %) [2]. All of these studies do not completely isolate the gating process, because the process of staining and running the sample still had to be completed by participants, even though correct reagents and procedures were provided, similar to EuroFlow efforts, which returned an inter-laboratory *CV* below 30 % for Median Fluorescence Intensity (MedFI) of markers analysed [107].

In an attempt to standardise Minimal Residual Disease (MRD) Flow Cytometry assays for Chronic Lymphocytic Leukaemia (CLL), electronic data files were analysed by multiple experienced Flow Cytometry operators, to remove any upstream variation built into the fcs file and respective measurements [116]. 141 MRD files were generated with 40 files containing normal cells, 69 files containing 0.01 % to 0.1 % CLL cells and 32 files containing 0.0001 % to 0.01 % CLL cells, to represent MRD analysis scenarios and identify false-positive reporting scenarios. A pilot study with 26 files showed only 11 equivalent results returned, and then when given a more detailed operating procedure to re-analyse the files, equivalence increased to 23 cases. The use of a procedure to apply gates gave a 19 % improvement in accuracy, and a 44 % improvement in specificity, suggesting adoption of a physical protocol could reduce false positive rates in FC analysis [116].

This study catalysed efforts for standardised FC data analysis, through the use of automated machine learning platforms to remove the subjectivity and inter-laboratory variation element [117]. Numerous studies and consortia have evaluated different computational methods for FC data analysis, to identify lower variance with the same target-cell identification ability as a human operator [20,21,118–120]. However, whilst these methods are able to replicate operator gating for well-established cellular subsets, results are not ideal for subsets that are rare or difficult to separate, often requiring manual intervention for training or final analysis stratification [121], as

well as increasing chances of reported false positives for smaller populations, close to a cut-off limit.

Whilst automated platforms serve a clear purpose for FC data analysis as data moves to more complex, highly dimensional assays, there are still some usability issues that require refinement in order for these tools to become accessible and to be used correctly. Currently three core challenges need to be addressed for more successful adoption [122]. Firstly, few immunologists are aware of these tools because outreach of literature does not exceed bioinformatics or computational biology journals.

This is also the case for standardisation effort of FC experiments (MIFlowCyt) not leaving FC's inner circle and being presented in more cell-specific journals [50]. Secondly, although automation platforms are open source through a variety of platforms, these computation and software methods are not necessarily user friendly, and not easy to learn alongside full-time immunology roles. Third and finally, a lack of understanding of how the tools work can lead to scepticism or overconfidence in the methods, which can cause culture change issues, an oversight of quality control and issues with validation [122]. This issue will also be explored within the survey in Chapter 8.

Efforts to tackle the user interface challenge have been made by algorithms such as FLOCK, allowing participants to upload and test their own data on the platform, although this still requires an open mind set for trial and adoption [122]. Other software platforms such as Flowjo have algorithm plugins such as t-SNE, to allow participants to trial algorithms in a familiar environment [123]. Automated algorithms are still at the mercy of users, with improved analysis of data dependent on suitable data pre-processing, to clean the data and remove noise, reduce dimensionality to reduce computational time and sample back to the original cell count number [124].

A comparison study between individual participant gating, central gating and three automated platforms, listed different ways the data had to be 'pregated' or 'postgated' to ensure the data was appropriate for automated analysis or the correct cluster identification, therefore requiring manual intervention and subjectivity [1]. The data analysed here were Major Histocompatibility Complex (MHC) Dextramer ™ staining of T-cells recognising two different virus-derived epitopes (Epstein-Barr Virus and Influenza) in PBMCs from healthy donors. These are also markers analysed for engineered T-cell receptor (TCR) therapeutic products, providing significant affinity to this research within the CGT manufacturing industry. The highest *CV*s from gating individually and centrally were 122 % and 86 % respectively for one of the influenza epitopes, with the remaining *CV*s for other virus strains < 30 % *CV*. Again, analysis of the files was not isolated, with participants conducting staining and analysis of the files according to their own procedures and reagents available, contributing to downstream variation [1].

1.5 Variation impacts on CGT manufacturing and adoption of measurement uncertainty

Variation and advancement of measurement platforms has been identified as a core issue requiring attention for the development of CGTs [59,63], to ensure the successful validation of therapies requiring authorisation. This will only become more challenging when greater demand and throughput require more inline integration of measurement processes, to reduce the need for sampling. This creates a greater need for big data analytics, to ensure analysis can be completed quickly and reliably, as the panels for FC grow. Advances in measurement capability and resolution provide further understanding of the product, increasing measurement precision and control of the manufactured product [8,125]. Whilst FC analysis algorithms still require refinement for low level detection, they are a step in the right direction for integration into FC measurement against a pre-defined manufacturing specification. Comparison to the manual gating is still required for validation and more precise quantification of allowable variation, which this research addresses.

Whilst many publications address FC gating's significant contributions to overall measurement variance, this phase has not been fully isolated to purely quantify the user subjectivity within the post-analytical phase, which is investigated through experimentation in Chapters 3 to 6 of this thesis. A focus on purely subjective analysis of pre-prepared fcs files is being adapted by UK NEQAS for the Leucocyte Immunophenotying Scheme (Personal Communication with UK NEQAS) for clinical participants, as well as a monitoring scheme for image analysis of blood film morphology for participant Continued Professional Development (CPD) [126].

Future training of the required workforce has been identified as a big risk and challenge for the adoption of CGT manufacturing in the UK [127]. Operator application and position within CGT manufacture requires comparable consistency across multiple manufacturing sites, as well as within each centre, laboratory and team [9]. This presents an opportunity for proficiency testing schemes to enable continued training and development of staff across all CGT manufacturing sites, empowering operators to maintain good practice.

This is a timely issue for accreditation to ISO 15189 and ISO 17025, to ensure suitable levels of competency, as well as an opportunity to integrate education of measurement uncertainty. This not only satisfies the accreditation criteria for equipment, it provides greater resolution, precision and control of the measurement process, for training and continuous improvement. This in turn should improve quality and manufacturability of the CGT product, increasing the delivery of treatments to patients in a timely manner.

Whilst traceability of measurement and procedures through EQAS is important for analytical precision, it is also imperative for 'post-post-analytical' interpretation (clinician measurement interpretation), where results are used by clinicians for diagnosis and treatment decisions. Greater education and harmonisation of permissible limits for tests is needed for clinicians, due to poor assumptions that different numbers can be compared, leading to misinterpretation and impact on patient safety [128]. Harmonisation of clinical chemistry efforts have been pioneered by the

Australian Pathology units, to ensure general acceptance limits for analytes are the same across clinical centres to address this misinterpretation by clinicians [129].

Monitoring of correct units can be enforced through EQAS for harmony, and misinterpretations of limits are important to note for the sign off therapeutic products to the clinician to minimise issues due to misunderstanding. Greater focus on medical control limits for medical student training has been met with great interest, providing definition of the term “total error” of a measurement, which previously had negative connotations for clinicians, rather than being informative [130]. Measurement uncertainty has been recommended to be reported to clinicians as a set of acceptance boundaries for the test, to include reference limits, and remove any further mathematical interpretation error [130,131]. It should be seen as a quality indicator instead of a box-ticking exercise for pathologists and those completing in-process and release tests [132]. The use of measurement uncertainty will be tested within this research, to indicate its use as a quality and improvement tool within healthcare and Flow Cytometry post-analysis operations.

Chapter 2: Relevant Theory

2.0 Introduction to the Chapter

Chapter 2 provides an overview of relevant theory used to create the operator measurements for an Uncertainty estimate within Flow Cytometry analysis. This builds upon a 'Quality by Design' approach recommended by regulatory authorities, and it is one way of quantifying variation within a biomanufacturing Quality Control and measurement system. This continues to build upon the reviewed literature and FC background provided in Chapter 1, and describes the theory and methodology used for Gauge Repeatability & Reproducibility (Gauge R&R), and uncertainty methods, used within subsequent data Chapters.

2.1 Chapter Aims

This chapter provides an overview of the background theory for Uncertainty measurements, including process control structures for the measurements to be taken, underlying statistical reporting required and calculation of final Uncertainty values. This Chapter presents a structure that subsequent data chapters will follow for analysis.

2.1.1 Chapter Objectives

- Provide an overview of Gauge R&R from a Measurement Systems Analysis perspective.
- Investigate the use of Gauge R&R hybrid structure to enable measurement Uncertainty to be calculated.
- Present a statistical report format for use in subsequent chapters and definitions behind statistics used.
- Provide a detailed explanation of measurement Uncertainty and how it is calculated.

2.2 Measurement Systems Analysis

Measurement Systems Analysis (MSA) is the analysis of a "*collection of instruments or gages, standards, operation, methods, fixtures, software, personnel, environment and assumptions used to quantify a unit of measure or fix assessment to the feature characteristic being measured; the complete process used to obtain measurements.*" [133].

MSA falls under the 'Measure' category of the Design-Measure-Analyse-Improve-Control (DMAIC) Cycle, part of Six Sigma methodology frequently used to reduce variability and waste within a production system. An overview of the DMAIC cycle can be seen in Figure 9, along with different tools and methods that can be used within each stage of the Continuous Improvement (CI) cycle [134,135].

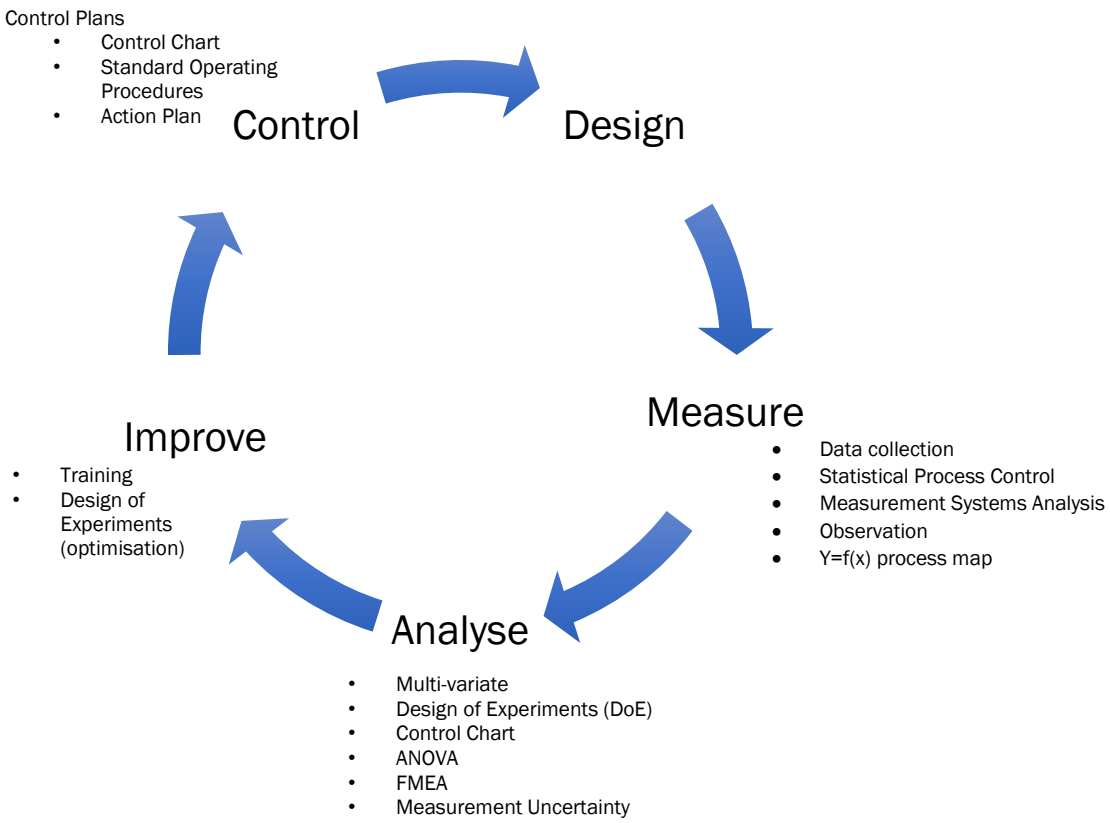


Figure 9 DMAIC Cycle

DMAIC is very similar to the Plan-Do-Check-Act (PDCA) Cycle developed by Walter Shewhart and championed by W.Edwards Deming, Quality Gurus and pioneers of Quality Control approaches [136–138]. This can also be used as a Lean manufacturing method to reduce variance in a system and improve output quality and is described in Figure 10.

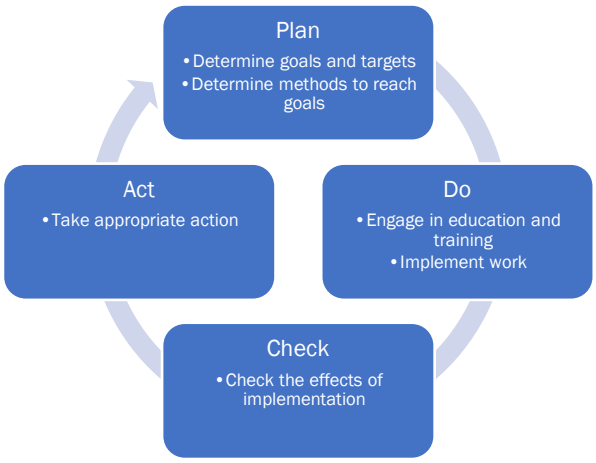


Figure 10 PDCA Cycle [136,138]

The DMAIC cycle tends to be prescriptive in its process, with clear tools that can be used at each point in the Continuous Improvement loop, so it is generally preferred within manufacturing control scenarios. Following this cycle helps CGT manufacturers to continually improve processes and quality, to comply with a 'Quality-by-Design' approach using risk analysis in line with Quality Management principles. This is a preferred strategy by major regulatory authorities such as the European Commission, placing more focus on front-end development to assure product quality [14,64,139].

Various tools can be used within MSA, most notably Statistical Process Control (SPC), Gauge Repeatability & Reproducibility (Gauge R&R), and Function (y=f(x)) Process Maps. SPC is a methodology used to monitor and continuously improve process performance and to reduce variability within important metrics. SPC is generally an on-line process that measures performance metrics in real time, so a better idea of measures and variability can be attained over time. Many

types of chart are used within SPC, but the most commonly used tools are process control charts, histograms or stem-and-leaf plots, check sheets, Pareto diagrams, cause-and-effect diagrams and scatter diagrams [134]. These allow trends or variance causes to be picked up over time, however it can be more difficult to measure the exact error or variance of a process item at one point in time. SPC is often used to measure multiple process parameters, so it can be costly in terms of time and resource to ensure SPC is setup correctly to identify the correct measurands, but also in terms of training and labour over time [140]. It relies on process operators correctly inputting data (and doing it honestly) if an SPC system is not integrated into the process itself.

Similar time and cost constraints are also found when using Function ($y = f(x)$) Process Maps. These are tools often used within Six Sigma methodology to identify a specific output measurand to quantify the process (y) and monitor this as a function of other process attributes (x) [135]. A good understanding of all attributes and measurands are required by operators, and honest input of results to ensure issues can be identified through trends in time. Operational definitions should be calculated when starting to use functional process maps, through analysis such as Gauge R&R measures to quantify isolated variances as a function of part or operator. This is a type of analysis session, run by an independent assessor that allows errors to be calculated at a point in time, for a process or application. This can then be repeated over time to check the alignment of tools such as SPC or Process Maps.

An adaptation of the Gauge R&R process will be used throughout this research to quantify variance and therefore uncertainty of operator gating subjectivity. Gauge R&R has been used because it generates measurements that show a combined estimate of a system's repeatability and reproducibility (sum of within-system and between-system variances). This can become part of an SPC log, because SPC often monitors a variety of control factors within a production environment such as mean and range of product measures and change over time to catch possible drift, whilst giving an accurate measure of a controlled variable without other system interactions to consider. However, Gauge R&R does not fully isolate variation components, often combining repeat

measures across a series of parts, which is not specific enough for combining in an uncertainty calculation.

2.2.1 Gauge Repeatability & Reproducibility

"*Gauge R&R is the variance equal to the sum of within-system and between-system variances*" [133]. Equation 1 depicts this, and in this instance, it is used to calculate intra- and inter-operator variation (Equation 2).

$$\sigma^2_{Gauge\ R\&R} = \sigma^2_{repeatability} + \sigma^2_{reproducibility}$$

Equation 1 Gauge R&R

$$\sigma^2_{Operator\ Gauge\ R\&R} = \sigma^2_{intra-operator} + \sigma^2_{inter-operator}$$

Equation 2 Operator Gauge R&R

The role of Gauge R&R contributes to a broader Measurement System context in Equation 3 and Equation 4, for measurement capability and performance respectively.

$$\sigma^2_{capability} = \sigma^2_{bias\ (linearity)} + \sigma^2_{Gauge\ R\&R}$$

Equation 3 System Capability

$$\sigma^2_{performance} = \sigma^2_{capability} + \sigma^2_{stability} + \sigma^2_{consistency}$$

Equation 4 System Performance

Gauge R&R is a form of MSA that gives a combined estimate of Measurement System Repeatability and Reproducibility. Repeatability and Reproducibility are respectively defined as:

"*Measurement precision under a set of repeatability conditions … that includes the same measurement procedure, same operators, same measurement system, same operating conditions*

and same location, and replicate measurements on the same or similar objects over a short period of time." [141]

"*Measurement precision under reproducibility conditions of measurement … that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects."* [141]

Repeatability measures precision within batches, whereas reproducibility measures precision between batches, operators, gauges etc. Figure 11 illustrates this difference within a manufacturing context. In a manufacturing facility, numerous replicate products are created every single day, by multiple operating staff on the shop floor. To calculate repeatability, *n* replicate measures would be taken at one time, by one operator on one product. According to the Guide to the expression of Uncertainty of Measurement (GUM), this should be between 3 and 10 replicates, with a higher number of repeats providing a better representation of the precision of the measurement [142]. However, within a CGT manufacturing scenario, this may not be economical or necessary. This measure of repeatability or variance is the *SD* of repeat measures taken at the same time.

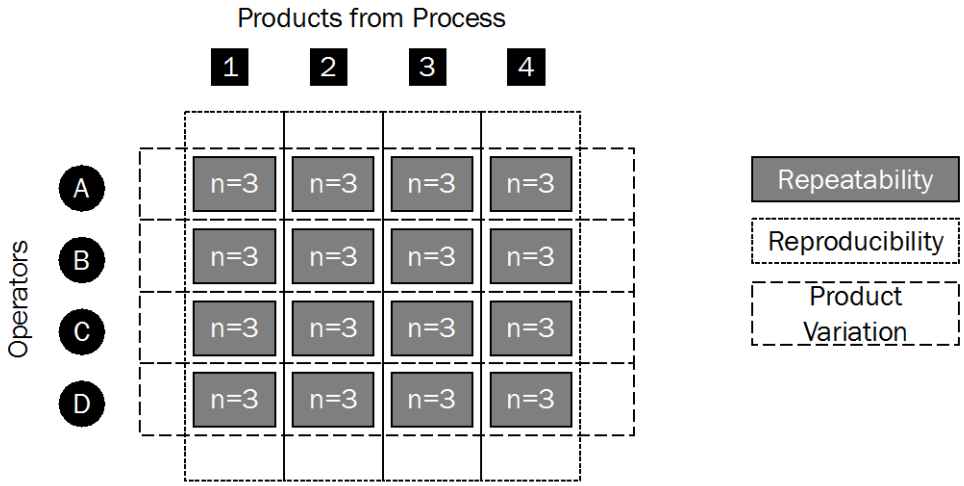


Figure 11 Illustration of repeatability, reproducibility and product variation within a manufacturing facility

Reproducibility compares these measures of precision across the different factors described earlier: - operators, environment, time etc. It can often show up the differences between operating shifts, operators or environments to understand the uncertainty or variance in a system when measures of a product fall out of specification and control. Comparing repeatability across different products, i.e. product variation, would require the same operator to perform the repeat measures on different products to ensure multiple sources of variance did not interact. Gauge R&R helps to generate an acceptable variation over the range of expected results. It can be used to support estimates of measurement system capability (random error over a short time period) by combining errors from linearity and uniformity of a system. It can also support estimates of measurement system performance, which is the effect of all variation sources over time [133].

Traditionally, there are three ways a Gauge R&R study can be conducted to populate a data collection template as seen in Table 2; the Range method, the Average and Range method, and the Analysis of Variance (ANOVA) method. The ANOVA method is usually preferred because it can measure operator-part interaction (in this instance 'parts' refers to .fcs files of the data for operators to analyse), whereas the other two options do not do this. However, for this thesis this interaction is not required because the part is kept uniform between participants, to create a Gauge

R&R-Uncertainty measurement hybrid. In addition, the ANOVA method assumes a normal distribution of measurements, so if the distribution of datasets used in this research are non-parametric this method is not ideal. The Range method is used to provide a quick estimate of measurement variation and is generally used to ensure the Gauge R&R has not changed over time. Traditionally, this method compares two appraisers and 5 components to be measured, where each appraiser measures each part once [133]. The range between appraiser A and B is calculated and a total Range average is taken from these part ranges ($\bar{R}$). Total variability is identified in Equation 5, where d_2^* is looked up using a d_2^* table [133].

$$Total\ variability = \bar{R} \times d_2^*$$

Equation 5 Total Variability using the Range method

Table 2 Example Gauge R&R Data Collection Sheet

Appraiser		PART										AVERAGE
		1	2	3	4	5	6	7	8	9	10	
A	1											
	2											
	3											
Average												$\bar{\bar{X}}_a =$
Range												$\bar{R}_a =$
B	1											
	2											
	3											
Average												$\bar{\bar{X}}_b =$
Range												$\bar{R}_b =$
C	1											
	2											
	3											
Average												$\bar{\bar{X}}_c =$
Range												$\bar{R}_c =$
Part Av												$\bar{\bar{X}} =$ $R_p =$
$\bar{\bar{R}} = \frac{[\bar{R}_a = \ ] + [\bar{R}_b = \ ] + [\bar{R}_c = \ ]}{[\#\ of\ appraisers]} =$												$\bar{\bar{R}} =$
$\bar{\bar{X}}_{DIFF} = [Max\ \bar{\bar{X}} = \ ] - [Min\bar{\bar{X}} = \ ]$												
Notes												

The Average and Range method ($\bar{X}\ and\ R$) provides estimates of both repeatability and reproducibility, splitting them into two measurements unlike the Range method's single measurement. Numerous appraisers can be used for this method and ideally more than 10 parts are preferred for measurement. Each appraiser takes turns to measure the components in a blind order, which can be pre-determined by a Design of Experiments (DoE) style randomisation [133].

All measurement equipment must be calibrated before the analysis begins and results are entered into the appropriate cells of the data collection sheet in Table 2. Subsequent appraisers measure the same components in different random orders and the results are further tabulated. From this Gauge R&R information both averages and ranges can be calculated for each part, for overall part averages and ranges to be determined. Various SPC charts can be plotted from this information, such as Average and Range Charts, to monitor the overall trend of the data and variance of the measurements, if required in a process.

This Gauge R&R process can be modified to enable uncertainty calculations to be taken for operators, which required multiple factors (including number of analysed parts) to be controlled. The framework of the Gauge R&R process will be used in subsequent participant studies, so participants are given multiple files to analyse, although only one of those will be used for uncertainty calculation.

Participants will be required to analyse the same data multiple times, to ensure a repeatability measure can be taken as an exemplar. It is possible to include more parts for operators to measure, however, due to potential time constraints with human participant research, Display Screen Equipment usage, attention span and cost of operator time this may not be feasible because repeat measures are more important to build confidence in the uncertainty metric.

2.3 Representative Metrics: Basic Statistics

The following section outlines basic statistics used to describe results, samples and populations. These form the basis of uncertainty measurements calculated for the Gauge R&R studies conducted within this thesis. IBM SPSS Statistical Software Version 24, supported by Microsoft Excel and Matlab R2019, will be used to calculate advanced statistical testing throughout this thesis, unless otherwise specified.

Many mathematical methods exist for the analysis and transformation of data sets and respective distributions. It is inefficient to provide full data sets to quantitatively describe a population, so measures of location, spread, skewness and normality have traditionally been used to represent distributions. These metrics enable processes and products to be measured over time to build confidence into Product Design Specifications (PDS), Process Performance Metrics or allowable tolerances for Critical Quality Attributes (CQAs). These basic statistics have been described in subsequent sections for their use in later statistical and uncertainty calculations.

2.3.1 Measures of Location

Measures of Location take one sample value to represent the outcome of an entire population or distribution. Measures of Location are absolute points and do not include judgement of spread or variance. Arithmetic mean and median are measures of central tendency, and all location values calculated are inclusive of all data points. Outliers can affect the final location metric but may not be excluded because they represent real analysis scenarios with patient data. They will, however, be investigated to understand the differences in technique.

Figure 12 shows how measures of Location can change depending on the spread of data within the population. When a normal distribution can represent the population, the mean, median and mode should all be equal, or very similar. The non-parametric distribution in Figure 12 is positively skewed and can be assessed quantitatively if the central tendency Location measures are unequal. Table 3 provides definitions and calculations for measures of Location that are used in subsequent chapters.

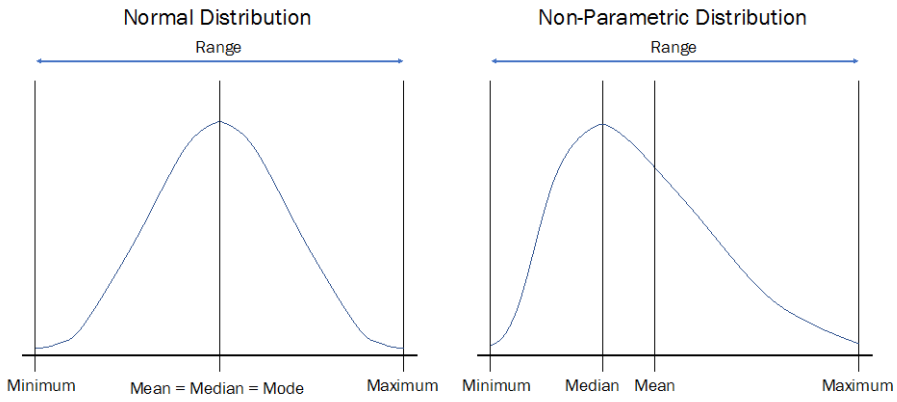


Figure 12 Measures of Location and Central tendency within normal and non-parametric distributions.

Table 3 Descriptions and equations to calculate measures of Location

Arithmetic Mean (denoted μ)	The total sum of results divided by the number of results present in the distribution. $\mu = \frac{1}{n}\sum_{i=1}^{n} a_i = \frac{a_1 + a_2 + \cdots + a_n}{n}$ *Equation 6 Arithmetic Mean*
Median	The 50th percentile of ordered results within the distribution.
Mode	The most popular or recurring value in a data set. This is less common in continuous data sets where lots of analysis results in different values.
Minimum	The lowest value generated within the data set.
Maximum	The highest value generated within the data set.

2.3.2 Measures of Spread

Measures of Spread or Dispersion monitor the amount of variance within a data set. They are used in conjunction with a Measure of Location to provide a statistical description of a population which can then be used to compare against other populations or data sets. Spread can also be used to show how well location measures represent the data [143]. For example, when plotting the arithmetic mean and *SD* on a non-parametric distribution, the mean would not be on the peak maxima, indicative of the central point of the population. Therefore, the median and absolute deviation could be used.

Table 4 provides definitions and calculations used to describe spread or variance of data throughout the remainder of the thesis. Figure 13 shows how these spread measures relate to normal or non-parametric distributions. Normal distributions are described by stating a location

measure of central tendency and a *SD* to represent the amount of variance. The normal distribution in Figure 13 details the percentage amount of data within the population that traditionally lie within standard deviation boundaries.

CV is a commonly used statistic within FC measurements, because it provide a single measurement used to represent variability, that considers both the *SD* and mean [29]. Due to its common use, this will also be calculated through this research in comparison with uncertainty, to identify potential differences between the two variance metrics.

Table 4 Descriptions and calculations for Measures of Spread

Range	The total width of the data set.	
	$$R = Maximum - Minimum$$ *Equation 7 Range*	
25th Percentile	The 25th percentile of ordered results within the distribution.	
75th Percentile	The 75th percentile of ordered results within the distribution.	
Interquartile range	The range of data across the central quartiles of data.	
	$$IQR = 75th\ percentile - 25th\ percentile$$ *Equation 8 Interquartile Range*	
Standard Deviation (denoted σ)	The unit of dispersion across a data set, which can be used to measure variation from the central point of a normal distribution. It can also be used to estimate confidence limits with 1 either side of the location metric representing 68%.	
	$$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{n}}$$ *Equation 9 Standard Deviation*	
Absolute Deviation	This is an alternative unit of dispersion that is used for data spread that include both positive and negative values and controls for the signs cancelling each other out.	
	$$Mean\ Absolute\ Deviation = \frac{\sum\lvert X-\mu\rvert}{n}$$ *Equation 10 Mean Absolute Deviation*	
Coefficient of Variation	This is a representative measure of variability, which considers both the mean and standard deviation of repeat measures, displayed as a percentage.	
	$$CV = \frac{\sigma}{\mu} \times 100$$ *Equation 11 Coefficient of Variation*	

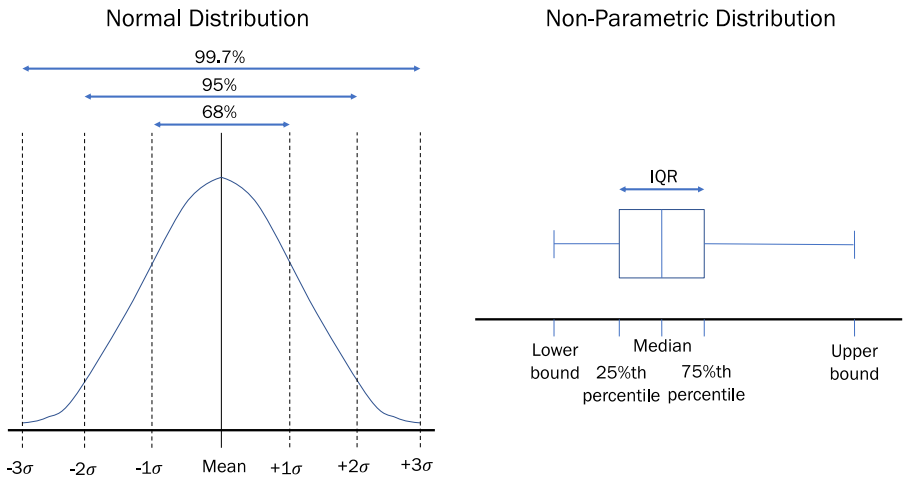


Figure 13 Measures of spread for normal and non-parametric distribution

2.3.3 Skewness

Skewness and kurtosis measurements are further characterisation metrics for a population distribution. They are often used as a quick judgement of normality of a distribution, because they are visual graphical representations of the data. However, whilst being good for a quick judgement, they are ill advised for small data sets as they may not represent a complete population. More detailed descriptions and calculations of skewness and kurtosis are listed in Table 5 and are used within statistical reporting throughout this thesis.

Table 5 Descriptions and calculations of Skewness and Kurtosis

Skewness	The measure of symmetry, or lack thereof, of a distribution. A positive skew will show the maxima of the distribution move towards the left of the arithmetic mean or central tendency. The mean will be on the right of the median for positive skewness. Examples of skewness can be seen in Figure 14. A negative skew will show the maxima of the distribution move towards the right of the arithmetic mean or central location point, with the median being to the right of the mean. SPSS provides a skewness metric and an associated standard error to calculate a skewness z-score. Using the following equation, a positive or negative result indicates a positive or negative skew respectively. If the *z*-score falls within ± 2.58 limits, as defined by SPSS, it is deemed as normally distributed. If greater than 2.58, the distribution has a strong positive skew and if less than -2.58 a strong negative skew [144].	
	$$skewness\ z-score = \frac{skewness}{standard\ error}$$ *Equation 12 Skewness z-score*	

Kurtosis	This is a measure of how heavy or light tailed the distributions are in comparison to a normal distribution which has a central tendency. It is also a measure of how the distribution is affected by outliers. Distributions with a heavy tail are more likely to be affected by outliers due to their frequency at one end of the data set. Low kurtosis would indicate small tails due to minimal effect from outliers. Examples of low and high kurtosis about a normal distribution are visualised in Figure 15. As with skewness in SPSS, a *z*-score can be calculated for kurtosis using the standard error of the data set. If the *z*-score falls within ± 2.58 limits it is deemed as normally distributed [144]. If outside of these limits, the distribution is strongly kurtosed.	
	$$kurtosis\ z-score = \frac{kurtosis}{standard\ error}$$ *Equation 13 Kurtosis z-score*	

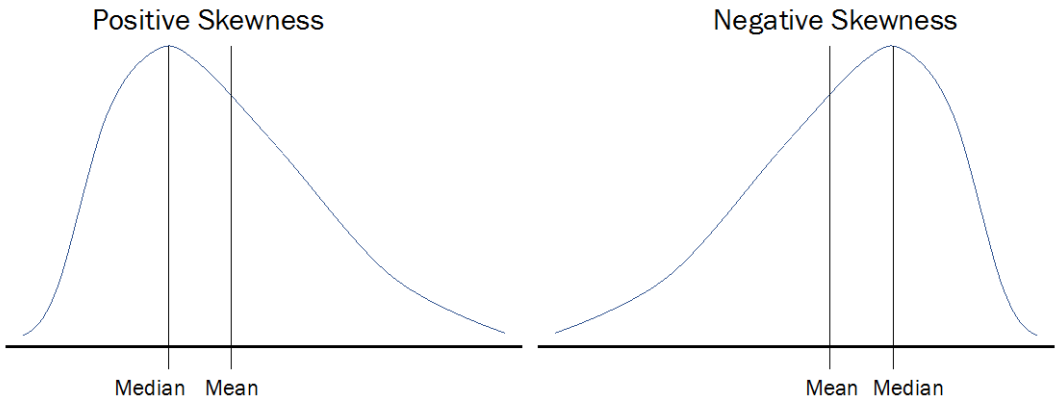


Figure 14 Examples of positive and negative skewness.

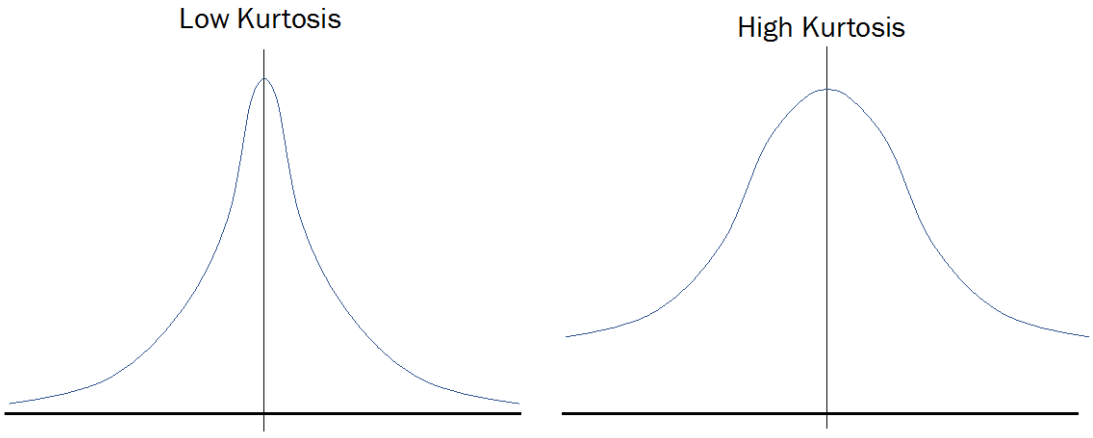


Figure 15 Examples of low and high kurtosis around a normal distribution.

2.3.4 Test for Normality

For a deeper statistical analysis of distribution shape, normality tests can be run within Statistical Software packages.

The Kolmogorov-Smirnov (K-S) test and Shapiro-Wilk (S-W) test both present null hypotheses that the distribution of a data set will be normally distributed, with an alternative hypothesis of a non-parametric distribution, requiring more investigation to understand its shape and parameters. A key difference between the K-S and S-W tests, is the K-W test has a much lower power, often leading to false results due to Type I error because it is more impacted by extreme values within datasets [145]. The S-W test has greater power when dealing with different distribution types, so may be the preferred normality test within this research, however this is dependent on the number of participants included in each study [146].

Type I errors (or α errors) are defined as:

"T*he probability of assuming that there is a difference or association between two or more variables when there is none. It is usually set at 0.05 or 5 % level*" [147].

In contrast, a Type II error (or β error) is defined as:

"*The probability of assuming that there is no difference or association between two or more variables when there is one. This probability is generally unknown*." [147]

2.3.5 Testing Statistical Power and required Sample Size

Power and sample size are defined for statistical tests where two or more conditions are tested to identify significant differences between the groups. Power is defined as:

"*The probability of a statistical test of finding a relationship between two variables when there is such a relationship. The maximum power a test can have is 1, and the minimum 0 with 0.80 indicating an acceptable level of power. The power of a test is generally greater for one- than two-tailed hypotheses, parametric than non-parametric tests, lower (e.g. 0.05) than higher (e.g. 0.001) significance levels and larger samples*" [147].

Sample size in the context of power analysis is defined as "*the minimum number of samples needed to run a study to find a desired effect*" [147].

Power is used when testing for a significant difference between two samples. Various parameters are considered when calculating power [148]:

- One- or two-sided test
- Level of significance, α
- Sample Size
- Effect Size relative to noise

2.3.5.1 One- or two-sided tests

One- or two-sided tests differ in their alternative hypotheses (H_A). Both tests have a null hypothesis (H_0) as follows:

H_0 = there will be no difference between the sample means.

$$H_0 = H_A$$

Two-sided tests simply have an alternative hypothesis that tests for difference, with no further specificity:

$H_{A\text{-}2}$ = There will be a difference between the sample means

$$H_0 \neq H_A$$

One-sided tests have an alternative hypothesis based upon size of the difference:

$H_{1\text{-}1}$ = One sample mean will be greater than the other.

$$H_0 < H_A$$

It is more common to use a two-sided test, because this simply states that means will be different, without giving an indication of differences in size or location. One-sided tests are used throughout this research, because it investigates a difference in variance, with respect to increasing complexity.

2.3.5.2 Errors and Significance Levels

Two types of error can be made when testing hypotheses, as defined previously and simplified below:

- Type I error (α): Null hypothesis is rejected when it is true
- Type II error (β): Null hypothesis is not rejected when it is false

α is also known as the level of significance chosen for a test, to specify a percentage confidence interval for the results. A 95 % (0.95) confidence interval is commonly chosen for tests and this carries a risk of 5% of cases where a Type I error may occur. This equates to $\alpha = 0.05$ because there is a 5 % risk in the null hypothesis being rejected when it should be accepted. For greater confidence, lower the α value towards 0.00, although this will impact other factors such as sample size in a power calculation [134].

2.3.5.3 Effect size

The effect size for statistical power is the between-group difference divided by the within-group standard deviation. For all statistical tests, because the effect size increases, the power will increase, if the other variables are kept constant. The greater the between-group difference, the less likely a Type II error becomes. The effect size is inversely related to sample size (if all other variables are constant), because small effects can only be detected with larger samples due to increased information, and large effects can often be detected with smaller samples as there may be less noise present.

Cohen's *d* statistic is a common calculation of effect size, computed using Equation 13 to show the difference in means (between two test groups, X_1 and X_2), divided by the *SD* (s, the error or 'noise') [149]. If two groups have different standard deviations, then the standard deviation used is often taken from the control group or it can be pooled.

$$\hat{d} = \frac{\bar{X}_1 - \bar{X}_2}{s}$$

Equation 14 Cohen's d statistic

General reference values for effect size can be used to compute other power statistics, based on whether a small, medium or large effect (respectively) is desired, but the context should always be considered when choosing an appropriate effect size [149].

2.3.5.4 Types of Power Analysis

Statistical power is calculated as follows, as the complement of Type II errors:

$$Power = 1 - \beta$$

Equation 15 Statistical Power

G*Power 3 is a statistical power software that will be used to estimate desired Power and sample size throughout this thesis. It has been routinely used throughout social, behavioural and biomedical sciences to provide an easy way to determine the correct parameters for a study, ensuring conclusions are as significant as possible [150].

The two common types of power analysis are A Priori and Post Hoc power analysis, relating to Power calculated before and after a data gathering exercise respectively. A Priori is completed before a study begins and it is the ideal choice because it provides users with options to control α and β [150,151]. It provides up-front calculation of how many participants to recruit to meet desired power and effect levels. Post Hoc analysis is performed after a study takes place, so the sample size can be used to calculate an observed power value. These are less desirable, because only α has been controlled, even though β has been assessed [150,151].

Throughout this research it is likely that A-Priori and Post Hoc Power Analysis will be computed for each study phase, where appropriate. To attain the appropriate power (0.80) [147], high sample sizes are often required (depending on effect size), which could be difficult to achieve considering constraints such as access to participants within the timeframe of the research. However, this research aims to quantify uncertainty, so Post Hoc Analysis can indicate the sample size needed for potential future studies to further confirm initial effects seen within this research, whilst giving a measure of current statistical power.

To calculate statistical power using A Priori or Post Hoc analysis, different tests can be used depending on the variable being compared. The most common tests are T-tests, F-tests and χ^2 tests and they often calculate power based on the following metrics [151]:

- T-tests: Linear bivariate regression; difference in means between test groups
- F-tests: Analysis of Covariance (ANCOVA); Analysis of Variance (ANOVA); Hotelling's T^2; Multivariate Analysis of Variance (MANOVA); Test of equality in variance
- χ^2 tests: Goodness-of-fit tests: Contingency tables

This research hypothesises an increase in variance is seen as increases in complexity (due to number of Flow Cytometry processing steps) are achieved through each study stage. It is most likely that F-tests will be used to assess statistical power in this case because this family of power tests are designed to measure variance rather than goodness-of-fit or a difference in mean location values. When looking at the cell population counts that operators achieve (absolute results rather than variance), it is most likely that a t-test will be used. The actual tests used are more likely to be applied Post Hoc, once an understanding of the population distribution has been attained, and a suitable test selected to represent this parameter. Where applicable, power analysis will be discussed after initial statistical reporting of results has been listed in each data Chapter.

2.3.6 Statistical Report Formats

The statistical calculations detailed above have been used to create Statistical Reports for data distributions occurring throughout this thesis, shown in Table 6. These reports provide a brief summary of the Absolute data results, showing the variance of actual final data values operators have defined to represent a sample. These reporting tools will also be used to define the measurement uncertainty of how these absolute values were attained. This gives a reporting viewpoint on the variation of the data sets and how they were calculated, rather than a single representative value.

Table 6 Statistical Report Format used throughout this Thesis

Arithmetic Mean	
Median	
Mode	
Minimum	
Maximum	

Range	
25th Percentile	
75th Percentile	
Interquartile Range	
Standard Deviation	
Median Absolute Deviation	

Skewness	
Skewness standard Error	
Skewness z-score	
Kurtosis	
Kurtosis Standard Error	
Kurtosis z-score	

Shapiro-Wilk statistic	
Significance	
Normal/Non-parametric	

2.4 Measurement Uncertainty

Measurement uncertainty is defined as:

"*A non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used*" [141]

This measure of spread provides users with a range containing the true value of the measurand, which has taken into account contributions or sources of variation that can cause the result to vary. This can provide users with greater certainty and confidence in their measurement result [152].

Within manufacturing, metrology and measurements are important to make key decisions, which are also true for many other scenarios. Within healthcare, clinicians make diagnosis or treatment decisions based upon the results they receive from pathology measurements. Therefore, an understanding of uncertainty of measurements is not only needed by metrologists and quality inspectors, it is also needed by those who can make influential decisions based upon the measurement reports they receive. This allows them to be more confident in their judgement of a situation, and an important trade-off with quality and cost of a product or service.

Sources of measurement uncertainty can arise from a number of factors, which all impact the variance directly or indirectly. These sources include, but are not limited to equipment, operators, time intervals, place, environment, chemicals and reagents and biology [153]. Figure 16 shows a fishbone diagram listing potential sources of uncertainty affecting FC results. This thesis focuses on factors affecting the operator and later chapters look at effects listed such as visual perception and gating strategies used.

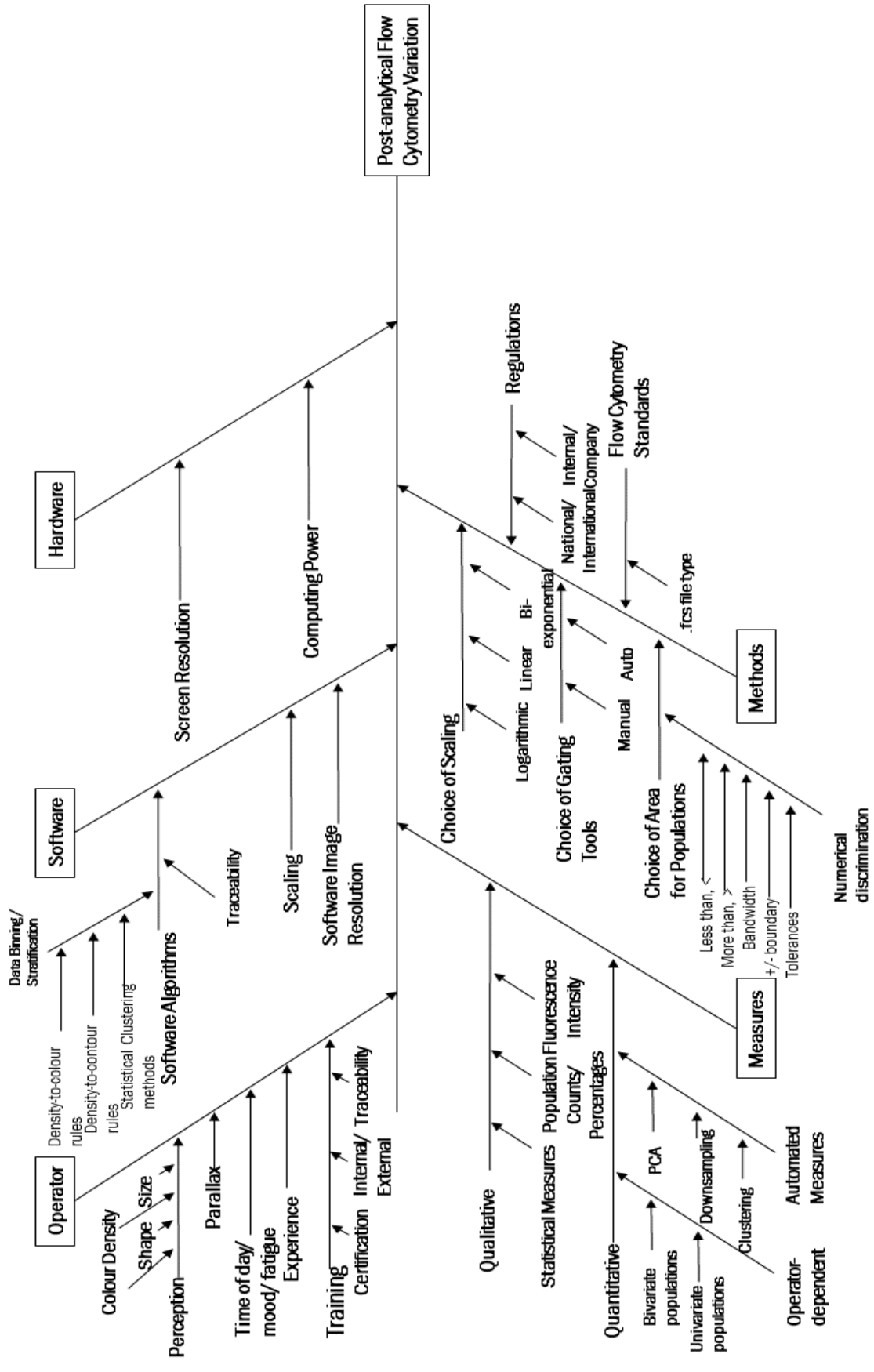


Figure 16 Fishbone diagram listing some sources of uncertainty within Flow Cytometry results [154]

2.4.1 Relationship between Error and Uncertainty

Calculating uncertainty provides a better understanding of the measurement methodology and the impact of each component, than measurement error. This provides opportunity for improvements in quality by tackling the largest identified sources of uncertainty. This also provides opportunity to increase product quality, save costs and there is an increased acceptability of results due to better understanding of the measurement and its attainment.

The true value of a product attribute cannot experimentally be known and be certain, therefore, there is always error associated with the observed value. Measurement error is defined as:

"*the result of a measurement minus the true value of a measurand*" [153]

A measurand is defined as:

"*A quantity intended to be measured*" [141]

Therefore, measurement error is a quantified difference, where a true value needs to be known. This differs from uncertainty because this is a defined range, calculated as a combination of variance from factors discussed above. Rather than being a stated number (like error), coverage factors are used to calculate the measurement tolerance, because these provide a range of confidence that the measured value will more certainly lie within. For example, a 95% confidence interval gives the user 95% confidence that the measurement lies within the ranges, defined by the combined and expanded measurement uncertainty calculation. That also means there is a 5% risk that the measurement does not fall within these stated boundaries. Due to this compilation of sources of variance and confidence, it provides a more realistic estimate of results than measurement error alone, because it is more representative of the true spread.

2.4.2 Quantifying Uncertainty

Quantifying uncertainty has been well documented in official literature from national and international standards coordinating bodies [141,152]. These methodologies will be followed to calculate the uncertainty of the operator variance within FC gating and is detailed in this section as methodology for subsequent chapters.

The process of quantifying uncertainty has been split into four sections to define the final uncertainty value:

- Identify the measurand and the process of obtaining it
- List sources of uncertainty for each stage of the measurement
- Quantify uncertainty components identified
- Combine values and calculate expanded uncertainty

These steps will be defined here for general process, theory and calculation and discussed with more specificity in the subsequent data chapters.

2.4.2.1 Measurand Identification and Methodology

To quantify variation within FC gating, each study phase will have a target cellular subpopulation that participants need to reach. This target will be accompanied by a process for participants to follow: a series of parent and child gates in a prescribed order to ensure measurement accuracy and comparability within- and between-participants. This will provide enough control for the measurement procedure to ensure variance comes from the participant application of gates, and not a difference in gating sequence.

To calculate variance, each absolute cell count representing each parent and child gate in the gating sequence will be used across a series of repeat measures. This structure has already been

described in Figure 11, depicting the repeat measures taken for each gate, and their combination to calculate a particular participant's reproducibility.

Within each gating stage, participants will repeat the gating process to define a measure of repeatability which can be used to compare against other operators. 3 repeats will be performed each time so that statistically a representative *SD* can be calculated, and session times do not exceed 1 hour with participants. This number of repeats were defined from a small pilot study with two operators (experienced and inexperienced) to monitor how long it took them to complete the gates. As only 3 repeats can be taken, *SD* will be used to compare variance, which assumes a normal distribution of data. Increased repeats may show a different distribution, but due to constraints of this experimental work a normal distribution is assumed, and *SD* is used to further calculate uncertainties.

This is a calculation of Type A uncertainty, by definition, as it is focused on the aim at hand [142,152]. Type A uncertainties are always calculated from a series of repeat measures. Type B uncertainty uses pre-determined uncertainty statements from documentation such as calibration certificates, historical records, manufacturer specification of guidelines from a data book [142,152].

A *SD* is calculated from repeat measures of the different hierarchical levels of the gating strategy participants followed during the different experimental studies. These representative variances for each level are then combined according to the root sum of squares rule for standard deviations, otherwise known as Pythagoras' theorem. Uncertainty sources cannot simply be added together, because this would be a misleading amount of variation, so it is combined in quadrature with the following equation:

$$u_c = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$$

Equation 16 combined Uncertainty

The combined uncertainty (u_c) provides an uncertainty range that reflects the variation from a specific measurement point. To give this measurement more confidence, it is then multiplied by a coverage factor (*k*) which represents confidence intervals of the normal distribution. Different coverage factors are noted in Table 7, but bespoke values can be calculated using the desired two-tailed confidence interval with known degrees of freedom to find the right coverage of value in a T-test table.

Table 7 Coverage factors (k) and respective Confidence Intervals

Coverage Factor (k)	Confidence Interval (%)
1	68
2	95
2.58	99
3	99.73

Expanded uncertainty (*U*) is calculated by multiplying the combined uncertainty with the chosen coverage factors according to Equation 16.

$$U = k \times u_c$$

Equation 17 Expanded Uncertainty

This value gives the user greater certainty that the measured value lies within the specified uncertainty limits calculated. When reporting this, the final expanded uncertainty would be written as the value following the calculated average or measurement (often denoted after a sign, instead if just the *SD* of measurement repeats).

2.5 Chapter Conclusions

This Chapter has provided a concise review of relevant measurement theory and how it will be applied to data analysis throughout this thesis. Gauge R&R is a form of MSA that can be used in a variety of methods to understand variance within a system. This becomes a core function of production SPC with various control charts that can be displayed to monitor quality over time. The Gauge R&R method will be abridged to ensure measurement uncertainty can be calculated for operators in multiple analysis scenarios, shown in subsequent chapters. This could then provide a basis to calculate a full Gauge R&R analysis in future.

To open the results analysis a basic statistical report format has been generated, to provide an initial basis for further investigation and root-cause analysis. This will be applied to absolute reported results and measurement uncertainty results in each Chapter, to better define the distributions being dealt with. An initial understanding of the data-set populations and respective distributions are becoming more important within CGT manufacturing as many biometric populations are not-normally distributed, a model which has been used extensively through traditional manufacturing paradigms. This non-parametric distribution issue is explored in greater detail within Chapter 3, to highlight the effect of control limit choice and potential transformation of data.

Finally, measurement uncertainty has been defined in detail here, but more specificity will be provided in subsequent chapters as the method is applied in each scenario. Measurement uncertainty is becoming more common to calculate, due to its inclusion within key regulatory documentation that manufacturers are required to follow. However, cell counts and reported data within the community still report measurement *CV*, which only focuses on the end of an analysis pipeline, whereas uncertainty can combine variances from the whole measurement process. Often this is calculated for measurement equipment, however, where subjective interpretation is concerned a combined uncertainty should be calculated to capture this variance.

Chapter 3: Pre-Study Variation Investigation

3.0 Introduction to the Chapter

Chapter 3 presents a preliminary study conducted before measurement uncertainty models were completed to investigate range in variance contributed by participants in a very uniform data analysis gating exercise. This was also an opportunity to define the structure of participant studies and ensure time parameters were correct, whilst confirming the procedures used were clear and repeatable to translate into future stages of work. The results of this Chapter were used to inform the structure of the uncertainty models used within Chapters 4 to 6. This Chapter compares the robust Coefficient of Variation (*rCV*) results of gates applied to a series of univariate histograms provided to participants for peak identification. This was conducted in a two-phase experimental study, to monitor the effect of participants applying gates based upon their own judgement, and when given a more prescriptive diagrammatical protocol to follow. This Chapter discusses the use of *CV* metrics which are heavily utilised within the FC industry.

Non-normality of data is present within CGT manufacturing data, so the effect of log transformations (often used to obtain a normal distribution) and different outlier boundaries will also be tested here to observe effects on the data and optimise for future study analyses presented later in this thesis.

A full data integrity check has been completed for all participant data used to create figures throughout this thesis. These have been independently verified by Loughborough University internal and collaborator external delegates and shows full traceability of data from the starting files, through participant analysis, to data extraction, interpretation and visualisation.

3.1 Chapter Aims

This Chapter aims to provide an initial understanding of whether there is variability contributed from participants when they analyse the same data. This will be completed with a series of peak separation diagrams, where participants will have to draw an area across where they believe one of the peaks resides. The variability in the participant results will be compared because they are all gating the same data, which will be used to inform future studies, with how much potential variation could be seen from the analyst alone. This also aims to monitor the effect of log transformation and different outlier classifications on a set of exemplar data.

3.1.1 Chapter Aims & Objectives

The Aims and Objectives of this Chapter can be defined as follows:

- Identify initial variation in data analysis contributed by operators when analysing the same data, on a simple univariate histogram example.
- Identify if diagrammatical protocols can reduce between-participant variability when gating univariate data.
- Identify if log transformations should be used in future analysis if results are non-normal.
- Validate the use of different outlier boundary estimators when dealing with non-normal distributions of data.

3.2 Methodology

The methodology for the development of this preliminary study and the respective developmental components are detailed in this section. An outline of the exemplar data set that participants analysed is provided, along with methodology on how this data was created, for reproducibility purposes. This expands upon how the studies were organised for participants and how the data was extracted and analysed.

3.2.1 fcs File Generation

A series of fcs 3.0 files were generated using a BD Bioscience FACSCanto II Flow Cytometer with 4-2-2 optical configuration, by running a suspension of BD Bioscience Cytometer Setup and Tracking (CS&T) beads (1 drop of beads within 500 µl of Phosphate Buffered Saline (PBS)), once a daily calibration was completed (Lot: 74538, Successful calibration) [155]. These files were generated because they provide a steady shift in fluorescence peak spectra required for this phase. The CS&T beads were a mix of dim, medium and brightly fluorescing polystyrene beads and are used to monitor baseline and daily performance of FC instruments. The beads were selected for use due to their lesser variation compared to biological material. The beads were run through the allophycocyanin (APC) channel, excited by the red laser (633 nm) at a medium flow rate (60 µL/min). The APC channel was used because it is a common channel across a wide variety of FC instruments.

Voltages, and therefore fluorescence peaks, were adjusted to determine the optimal analysis parameters in line with instrument sensitivity. The gain was changed using the '*voltage*' setting to alter the position of the fluorescence peaks seen through the APC channel on a univariate scale. Many methods have been developed to determine FC analysis parameters, but the most popular, referred to as the 'Peak 2 method' involves running a control sample at different voltage intervals [34], which has been used as the methodology here.

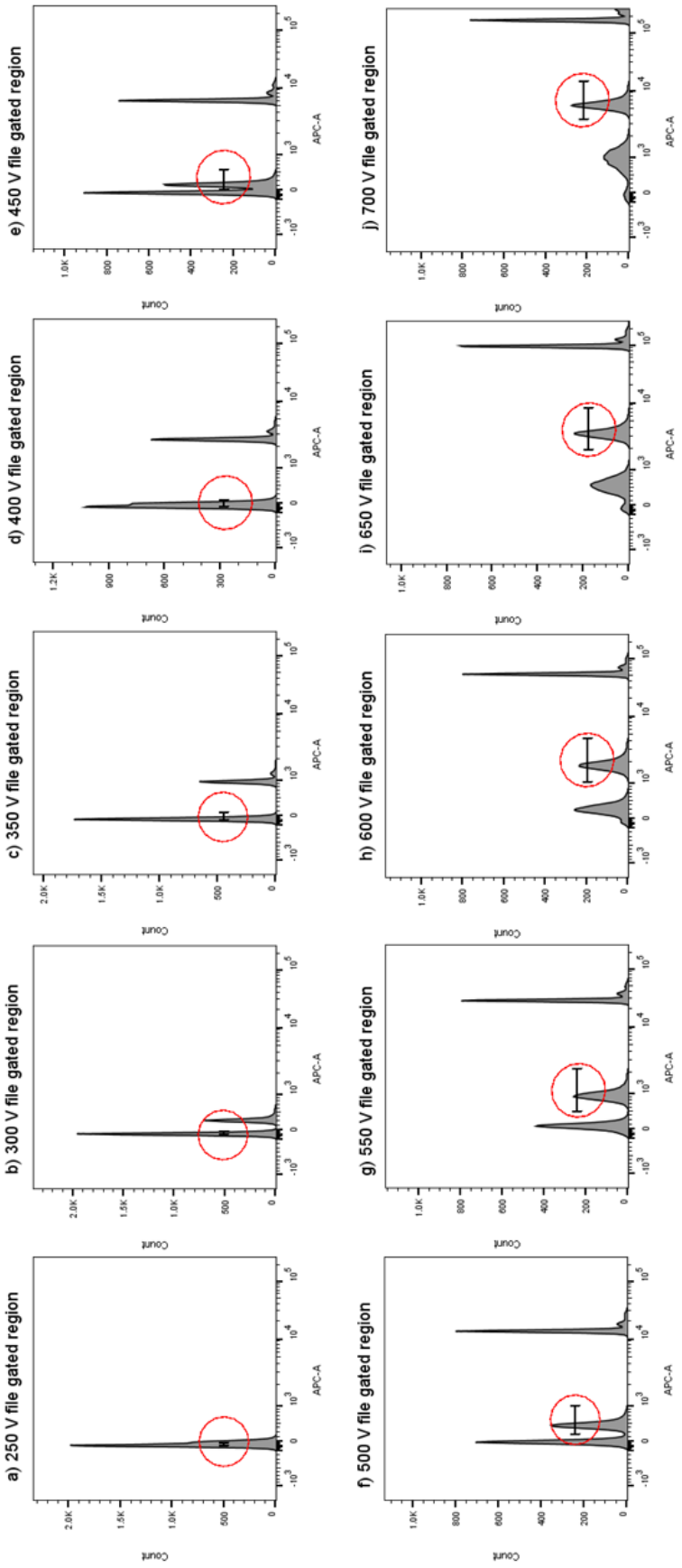


Figure 17 Images of histograms provided for participant gating studies [154]

The fcs files generated for this work (Figure 17) were gathered at ten different voltages from 250 V to 700 V, in 50 V intervals, allowing for resolution differences to be seen whilst not over-exerting the concentration of operators involved in the study. 10,000 beads were used to acquire each fcs file, which is the minimum number of events required for gathering bead-based data [156]. The fcs figure elements show the progression of the three data peaks across the univariate scale, in relation to the gain increase of the APC channel of the Flow Cytometer. Figure 17 also shows the range gates (red circle) applied to the histograms, which participants were asked to copy. The range gates were applied by the study coordinator (thesis author), to act as an independent reference point, so no participants were familiar with the data before taking part in the sessions. The univariate scale was kept to bi-exponential, because this is the default scaling used within Flowjo Version 10.0.8r1, the analysis software, and provided resolution of the three peaks across all voltage files created.

3.2.2 Flow Cytometry Study Organisation

36 participants across 3 centres were enrolled in this study to analyse the histogram files; 10 from an academic institution, 19 from an industrial cell therapy process development team and 7 as part of a data gathering exercise at the 2017 Future Investigators of Regenerative Medicine (FIRM) Symposium [157]. Ethical and GDPR approval was obtained for this study by the local University Human Participant Ethics Sub-Committee, which can be seen in Appendix A and covers human participant research conducted throughout this thesis. Participants completed the study individually within a 30-minute time-slot and were asked to complete two phases of analysis within this session. The two phases aimed to monitor the effects of personal data interpretation, in comparison to following a pre-defined protocol, or interpretation thereof, with the hypothesis of reduced inter-operator variation for FC analysis when following the latter. It was not expected to find much variation throughout this exercise, but to provide an ‘observed effect’ as a starting point.

During Phase 1, participants were asked to apply range gates to the area of the plots where they estimated the medium fluorescing peak lay. Five of the plots were considered to have 'poor' peak separation (Figure 17 a-e), where there was difficulty in discerning peaks. The remaining five files were considered to have 'good' peak separation (Figure 17 f-j) and to be more representative of good FC data. Participants used the third-party software Flowjo to examine the .fcs files and apply gates, using default bi-exponential scaling and maintaining the histogram view setting to ensure inter-participant consistency.

For Phase 2, participants were provided with printed images of range gates (highlighted by the red circles in Figure 17) that had been pre-applied to the series of files and asked to adjust their gates to try and match these 'diagrammatical protocols'. These gates were applied to aid precision of application, not accuracy, because accuracy can be translated into different subject contexts. The 'poor' gated files (Figure 17 a-e) were included to investigate the potential impact and worth of using diagrammatical protocols in difficult analysis conditions.

3.2.3 Variation Calculation

Range gates applied to histograms of one fluorescence channel show variation in spread and location parameters, chosen by perception of the operator. The robust Coefficient of Variation (*rCV*) (as defined in Equation 17) is used to compare variation amounts between operators within the two phases and is a common metric within FC data analysis [29]. It helps the operator to understand and monitor variation within a system, because it combines both location and spread parameters into one measure. *rCV* was used in this instance instead of *CV* (as previously defined in Chapter 2, Section 2.2.2), because fluorescence peak spectra often have a non-parametric shape, so using robust statistics was more representative of the FC data gathered.

$$rCV\ (\%) = \frac{rSD}{median} \times 100$$

Equation 18 robust Coefficient of Variation (rCV) (%)

The *rCV* measures were exported from Flowjo and are the combination of the *rSD* and the median of the data within the gate they have applied. Robust *SD* is the *SD* of the data points, based around the median rather than the mean. These *rCV* measures have subsequently been used to calculate a total range of variation between participants, per histogram file, per phase. Range was used to monitor reduction and therefore improvement in between-participant variation attributed to the results, because it is an easily understood measure of spread and monitors the total range rather than an arbitrary measure of spread (such as *SD*). The difference in range between operators can be used to observe the effectiveness of protocol instigation, with percentage of variation reduction monitored alongside this.

3.3 Results & Discussion

3.3.1 Flow Cytometry Pre-study statistical reporting Phase 1

The results reported here for each analysis file are taken from the inter-participant gates applied to each file using their own judgement (Phase 1). The *rCV* of the gate applied was extracted and tabulated per file (columns), per participant (rows). Table 8, Table 9, Table 10 and Table 11 provide a descriptive statistics overview of the inter-participant distributions for each file in the sequence and Figure 18 shows the inter-participant *rCV* distributions for each of these sets of file results. These descriptive statistics have been calculated following the statistical methods and definitions provided in Chapter 2, alongside the mean inter-participant *rCV* values for each file.

Table 8 Measures of Location for Phase 1 inter-participant rCV results (%)

	250 V	300 V	350 V	400 V	450 V	500 V	550 V	600 V	650 V	700 V
Average	207.9	134.2	68.9	34.5	20.2	15.6	15.1	15.2	15.2	15.0
Median	122.0	64.7	42.8	32.6	21.0	16.3	15.9	15.9	16.1	15.9
Min	-650.0	0.0	4.0	3.3	15.4	8.7	6.4	8.7	6.4	5.7
Max	1500.0	400.0	146.0	107.0	21.4	16.7	16.7	16.9	16.7	16.3

Table 9 Measures of spread for Phase 1 inter-participant rCV results (%)

	250 V	300 V	350 V	400 V	450 V	500 V	550 V	600 V	650 V	700 V
Range	2150.0	400.0	142.0	103.7	6.0	8.0	10.3	8.2	10.3	10.6
25%ile	50.1	29.1	35.1	28.5	19.8	15.5	15.2	14.9	15.4	15.0
75%ile	304.0	283.0	137.8	38.4	21.4	16.6	16.4	16.8	16.5	16.1
IQR	253.9	253.9	102.6	9.9	1.6	1.2	1.2	2.0	1.2	1.1
SD	349.6	127.0	48.7	15.6	1.6	1.7	2.2	2.3	2.3	2.2
***CV*(%)**	168.2	94.6	70.7	45.1	7.9	11.1	14.3	15.3	15.3	14.9

Table 10 Measures of skew for Phase 1 inter-participant rCV results (%) (3dp used for resolution from SPSS)

	250 V	300 V	350 V	400 V	450 V	500 V	550 V	600 V	650 V	700 V
Skewness	1.300	0.646	0.683	2.791	-1.894	-2.821	-2.865	-2.008	-2.86	-3.134
Skew std error	0.403	0.403	0.403	0.403	0.403	0.403	0.403	0.403	0.403	0.403
Skew z-score	3.226	1.603	1.695	6.926	-4.700	-7.000	-7.109	-4.983	-7.097	-7.777
Kurtosis	5.661	-1.271	-1.210	13.318	3.483	8.728	9.073	3.345	8.785	10.955
Kurt std error	0.788	0.788	0.788	0.788	0.788	0.788	0.788	0.788	0.788	0.788
Kurtosis z-score	7.184	-1.613	-1.536	16.901	4.420	11.076	11.514	4.245	11.148	13.902

Table 11 S-W test for normality for Phase 1 inter-participant rCV results (3dp used for resolution from SPSS)

	250 V	300 V	350 V	400 V	450 V	500 V	550 V	600 V	650 V	700 V
S-W statistic	0.833	0.788	0.805	0.731	0.747	0.624	0.635	0.705	0.613	0.584
Significance	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Normal/Non-Parametric	Non-parametric	Non-parametric	Non-parametric	Non-parametric	Non-parametric	Non-parametric	Non-parametric	Non-parametric	Non-parametric	Non-parametric

Figure 18 x-axis limits are much wider for 250 V to 400 V files because of the large inter-participant range returned for these files. If all file histograms were plotted with the same axis limitations, resolution of distribution shape would be lost for 450 V to 700 V files.

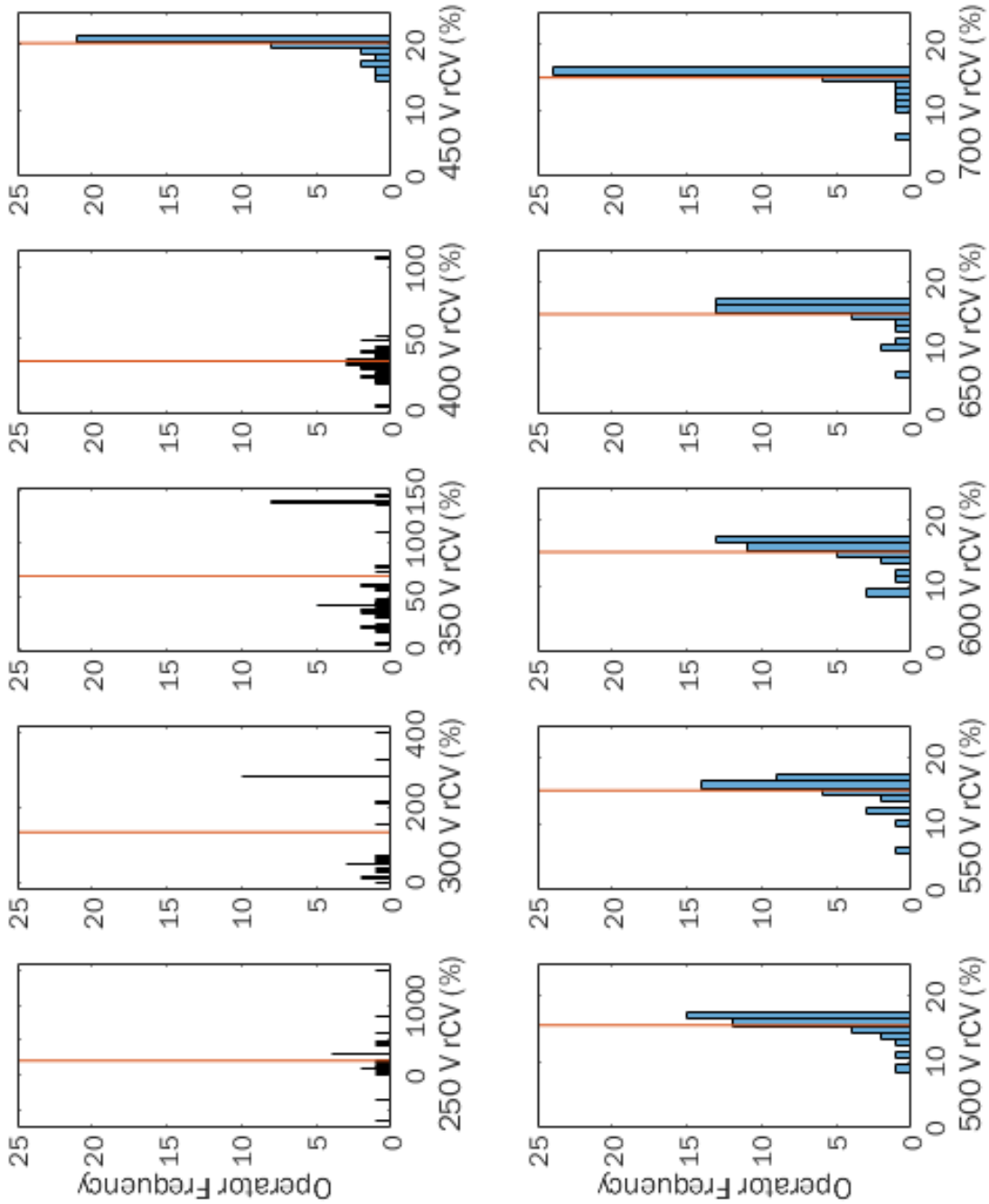


Figure 18 Inter-participant rCV distributions when gating fcs files using their own judgement (red lines represent mean rCV values for each file distribution).

Both average and median results for inter-participant gates appear to stabilise as the file voltage increases past 450 V and therefore peak separation increases. Inter-participant *rCV* averages are much higher for lower voltages due to little or no separation of the middle peak from other data. This difficulty in gating has created a wider range of inter-participant results when gating 'poor separation' files (Table 9), which in turn creates higher average and median values to represent the distribution.

The inter-participant ranges in Table 9 have a clear cluster of high and low ranges for low voltage and high voltage files respectively. This is also the case for interquartile ranges (IQR) for the set of files. Across all files, the IQR is small in comparison to the overall range, indicating a wide distribution of results, which suggests a location measure such as the mean or median could not confidently represent the datasets. The shape of the distributions further indicates this because the skewness and kurtosis *z*-scores fall out of the ± 2.58 limitations for normality for all file inter-participant distributions (Table 10), aside from the 300 V and 350 V files. Upon visual inspection of these distributions in Figure 18, the distributions appear to be bimodal, indicating a potential split in how participants have applied these gates, causing some to have a higher result *rCV* and some to have a lower result *rCV*. The Shapiro-Wilk normality test results in Table 11 further confirms this non-normality, with all file distributions having a significance value $p < 0.0005$. This rejects a null hypothesis of a normal distribution, accepting an alternative, non-parametric distribution (as described in Chapter 2, Section 2.2.4).

This data supports the use of correct upstream equipment and process validation because poor voltage choices can significantly affect downstream subjective judgement on where to place a histogram gate, with higher variability occurring when it is more difficult to separate required data from the noise or unnecessary populations at the limit of sensitivity.

Files with better peak separation (450 V to 700 V) also qualitatively appear non-normal due to a strong negative skew as also indicated in the skewness *z*-scores in Table 10. Qualitative

assessment of distributions adds significant value to this analysis because if basic statistics are used to represent data without considering distribution shape, mean values (shown in Figure 18 sub-plots, denoted as a red line for the mean *rCV* for each file) do not fall within frequency maxima bins on any occasion. This supports the use of robust statistical assessment when analysing FC data with medians better representing distributions overall. This will therefore be considered throughout each statistical data set in this research, if the distributions are non-normal.

3.3.2 Flow Cytometry Pre-study statistical reporting Phase 2

The results reported here for each analysis file are taken from the inter-participant gates applied to each file following a diagrammatical protocol, which participants had to follow and copy, shown in Figure 17. Table 12 to Table 15 provide a descriptive statistics overview of the inter-participant distributions for each file in the sequence and Figure 19 shows these distributions for each set of file results with the mean *rCV* values identified in each case.

Table 12 Measures of Location for Phase 2 inter-participant rCV results (%)

	250 V	300 V	350 V	400 V	450 V	500 V	550 V	600 V	650 V	700 V
Average	163.5	118.2	83.7	35.4	21.0	16.5	16.3	16.6	16.4	15.9
Median	207.0	96.2	76.8	31.9	21.1	16.6	16.4	16.7	16.5	16.1
Minimum	-1300.0	31.5	43.6	15.1	17.6	15.8	15.7	15.6	15.6	14.7
Maximum	800.0	283.0	139.0	75.7	22.9	16.6	16.5	16.8	16.6	16.2

Table 13 Measures of Spread for Phase 2 inter-participant rCV results (%)

	250 V	300 V	350 V	400 V	450 V	500 V	550 V	600 V	650 V	700 V
Range	2100.0	251.5	95.4	60.6	5.3	0.8	0.8	1.2	1.0	1.5
25%ile	-93.8	71.1	61.9	29.0	20.8	16.4	16.3	16.6	16.4	15.9
75%ile	488.0	150.0	109.0	37.8	21.4	16.6	16.5	16.7	16.5	16.1
IQR	581.8	78.9	47.1	8.8	0.6	0.2	0.2	0.1	0.1	0.2
SD	443.5	66.2	28.7	11.8	0.9	0.2	0.2	0.3	0.2	0.3
***CV*(%)**	271.2	56.0	34.3	33.2	4.2	1.2	1.2	1.8	1.3	2.0

Table 14 Measures of skew for Phase 2 inter-participant rCV results (%) (3dp used for resolution from SPSS)

	250 V	300 V	350 V	400 V	450 V	500 V	550 V	600 V	650 V	700 V
Skewness	-1.355	0.905	0.406	2.064	0.588	-1.559	-1.372	-2.104	-2.763	-2.762
Skew std error	0.414	0.414	0.414	0.414	0.414	0.414	0.414	0.414	0.414	0.414
Skew z-score	-3.273	2.186	0.981	4.986	1.420	-3.766	-3.314	-5.082	-6.674	-6.671
Kurtosis	2.630	-0.130	-1.074	6.077	1.618	3.151	1.108	3.815	8.895	8.076
Kurt std error	0.809	0.809	0.809	0.809	0.809	0.809	0.809	0.809	0.809	0.809
Kurtosis z-score	3.251	-0.161	-1.328	7.512	2.000	3.895	1.370	4.716	10.995	9.983

Table 15 S-W test for normality for Phase 2 inter-participant rCV results (%) (3dp used for resolution from SPSS)

	250 V	300 V	350 V	400 V	450 V	500 V	550 V	600 V	650 V	700 V
S-W statistic	0.901	0.909	0.933	0.821	0.925	0.700	0.792	0.673	0.594	0.623
Significance	0.007	0.011	0.047	0	0.029	0	0	0	0	0
Normal/Non-Parametric	Non-parametric	Non-parametric	Non-parametric	Non-parametric	Non-parametric	Non-parametric	Non-parametric	Non-parametric	Non-parametric	Non-parametric

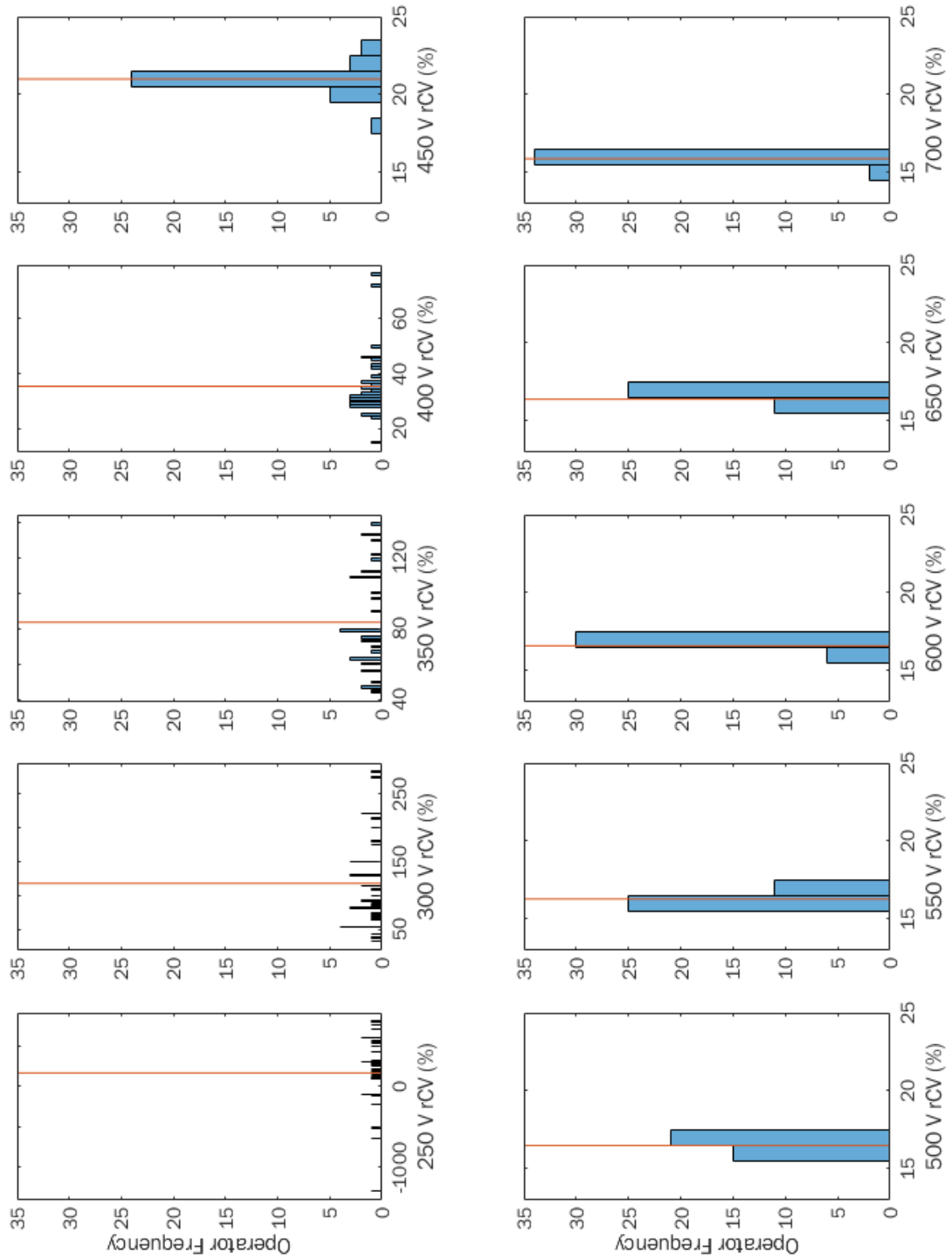


Figure 19 Inter-participant rCV distributions when gating fcs files following a protocol (red lines represent mean rCV values for each file distribution)

For files with notionally 'good separation' of peaks (Figure 17 f – j), following a protocol appears to converge the average and median *rCV* results for each file. Less difference between the average and median indicates central tendency and therefore a normal distribution. This can be seen by the red 'average' line markers plotted in Figure 19. Similar to the histograms in Figure 18, the red lines have been overlaid onto the inter-participant *rCV* distributions for each file. For each 'good separation' file, this average marker sits inside the peak maxima of the histogram in the centre of the distribution, because no bimodal distributions are present in this phase of analysis. It is difficult to qualitatively determine if the average better represents the distribution of the 'poor separation' files. However, the distributions for these files no longer appear to be bimodal (Figure 19), which could suggest using protocols to interpret data could aid conformance due to better inter-participant reproducibility, as defined in Chapter 2.

This would be true if the inter-participant *rCV* range reduced when following a protocol. This is the case for all voltage files, although, the 250 V file has only reduced slightly, but not significantly. However, this is negligible due to this voltage possibly being too low to run assays with confidence in this channel due to cellular autofluorescence [34,158], so participants would not face a triple-peak cluster in optimal, validated scenarios. The *rCV* inter-participant range for 'good separation' peaks is between 0.8 % - 1.5 %, which is considerably lower than when participants gated using their own judgement (8.0 % - 10.6 %). This shows that a diagrammatical protocol can possibly help in instances when gating is difficult.

The *CV* of measurements (as defined in Chapter 2, Section 2.3.2) is commonplace as a variance metric in FC and the *CV* of the inter-participant results ranges from 1.1 % to 2.0 % for 'good separation' files, far below the satisfactory criteria of < 10 % *CV* for a measurement, which is desirable [71]. Alongside the range of data, this supports using diagrammatical protocols to possibly reduce inter-participant variation within reported measurements.

The shift towards normality and a gaussian shape is not necessarily reflected in the skewness and kurtosis *z*-scores in Table 14. 300 V, 350 V and 450 V fall within the ± 2.58 boundaries for normality (defined in Chapter 2). However, all other files have a strong negative skew (aside from the 400 V file which has a strong positive skew). These files also have a high kurtosis, indicating a high peak within the data. Even though the 'good separation' files have a high kurtosis, in this instance that is desirable, because conformity of results that cause a centred spike shows better reproducibility between participants and less variation within the end result. The high skew and kurtosis results indicate a non-parametric distribution, confirmed by the Shapiro-Wilk tests results in Table 15. All distributions are significantly non-parametric.

A comparison of Phase 2 results to Phase 1 results will be completed in the following section. By individually analysing the descriptive statistics of each phase of results, the type of desired distribution can be questioned. Most analysis statistics are designed to work with a normal distribution, and this is often an 'ideal' data shape that manufacturing communities work towards. Not only does it allow use of normal statistical tools, it can be used to drive continuous improvement efforts to refine the process which causes data extremes. Strong skewness and kurtosis measures are therefore traditionally not ideal, but when looking at variability of data, they could be core indicators that show reproducibility of data, due to convergence of results.

3.3.3 Flow Cytometry Pre-study Outlier Discrimination Investigation

The ranges of inter-participant data have been used for comparison of Phase 1 and 2 data. This metric includes all participants and excludes no-one as an 'outlier', because they have produced a result under the same test conditions as other participants. Instead, any extremes of data will be investigated to understand the potential root cause of variation.

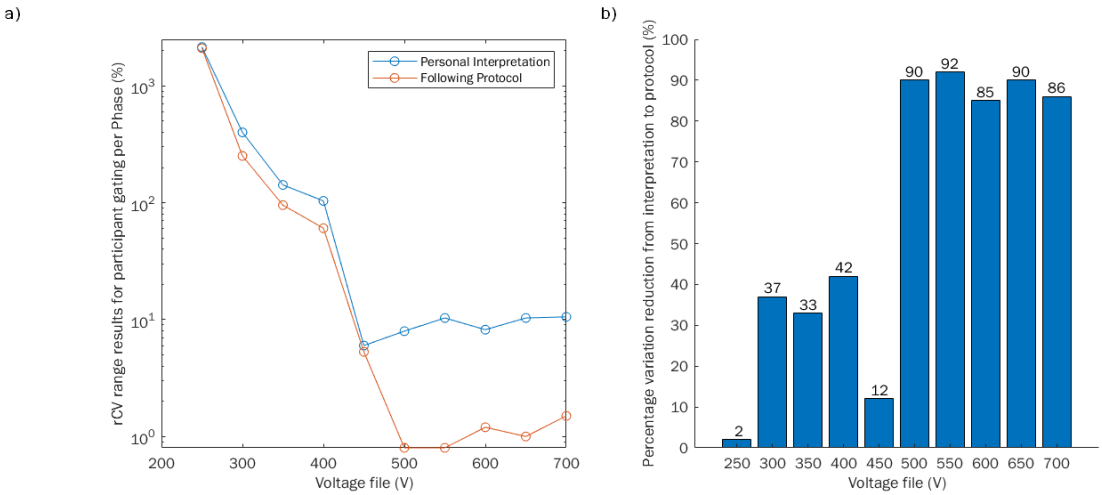


Figure 20 Comparison of inter-participant ranges when participants use their own judgement to gate and when they follow a protocol [154]

The range of inter-participant results when participants used a protocol is significantly smaller than when participants use their own judgement for all files analysed, as shown in Figure 20. The bar charts (b) show the percentage reduction of inter-participant range from Phase 1 to Phase 2. This was calculated using Equation 19 for each voltage file.

$$Percentage\ reduction\ (\%) = 100\ \% - (\left(\frac{Phase\ 1\ inter-participant\ range}{Phase\ 2\ inter-participant\ range}\right) \times 100)$$

Equation 19 Percentage reduction from Phase 1 to Phase 2 results

In addition, the *rCV* ranges for each voltage file have been plotted for Phase 1 (Figure 20a, Participant Interpretation of gate placement, blue line) and Phase 2 (Figure 20a, Gating Standard, orange line). These lines show the *rCV* ranges for each phase for each file, and are measured against a logarithmic y-axis, so all ranges and differences can be effectively seen.

The 'good separation' files from 500 V to 700 V show a large percentage reduction when participants use a gating protocol, despite the peaks in these files having a very clear separation. This variation reduction shows that using a protocol can help participants to gate histograms accordingly, even when noise affects the data at the other extreme, due to high amplification. This

noise can be seen in higher voltage files in Figure 17 (g – j) as small shoulders and longer tails on the edges of the main peaks. Within the 'good separation' peaks, the percentage reduction is high, although it appears to get slightly smaller as the voltage increases (from 90 % for the 500 V file to 86 % for the 700 V file). This could be because the higher the voltage, the more the noise is amplified, causing a bit more gating variability for the 'good separation' files as the voltage increases, due to increased visibility of peak shoulders and tails.

Despite the poor separation making gate application difficult within the 'poor separation' files (250 V to 450 V), this has shown to be aided by a diagrammatical image to copy and inform the correct scale points for gate placement. These files show poor separation and resolution of the middle peak, which should be identified in early upstream validation phases when using FC for analytical measurements. Therefore, these files are not representative of 'good data' that would be gated but do show difficulty when having to split clusters of beads or cells in analysis. Ultimately, thorough optimisation and validation of process steps and settings coupled with diagrammatical protocols can greatly influence and reduce the inter-participant range of reported results. This improves reproducibility of data between participants, converging to more confident results, analysis and interpretation.

Sign tests have also been executed using IBM SPSS Statistics Version 24 to test the use of gating protocols in comparison to participants gating using their own judgement. This tests for a significant difference in medians of the two test conditions and is used when population distributions are non-normal and subject to outliers. Sign tests have been used throughout this thesis when testing for differences between conditions. Table 16 summarises the results of Sign tests used for each voltage file, where the null hypothesis confirms no significant difference between the median inter-participant *rCV* ranges, and the alternative hypothesis accepts a significant difference between the two test conditions. Red highlighted cells indicate that the alternative hypothesis of a significant difference between the medians of the two groups has been rejected, so the null hypothesis of no

difference between the medians is retained. Green highlighted cells indicate that the alternative hypothesis has been accepted because of a significant difference between the group medians.

Table 16 Results of Sign test to compare medians of two-phase test conditions

Voltage File	Null Hypothesis	Test	Sig.	Decision
250 V	Median difference of P1 & P2 250 V files = 0	Related-samples Sign test	1.000	Retain null hypothesis
300 V	Median difference of P1 & P2 300 V files = 0	Related-samples Sign test	0.735	Retain null hypothesis
350 V	Median difference of P1 & P2 350 V files = 0	Related-samples Sign test	0.043	Reject null hypothesis
400 V	Median difference of P1 & P2 400 V files = 0	Related-samples Sign test	1.000	Retain null hypothesis
450 V	Median difference of P1 & P2 450 V files = 0	Related-samples Sign test	0.100	Retain null hypothesis
500 V	Median difference of P1 & P2 500 V files = 0	Related-samples Sign test	0.021	Reject null hypothesis
550 V	Median difference of P1 & P2 550 V files = 0	Related-samples Sign test	0.003	Reject null hypothesis
600 V	Median difference of P1 & P2 600 V files = 0	Related-samples Sign test	0.072	Retain null hypothesis
650 V	Median difference of P1 & P2 650 V files = 0	Related-samples Sign test	0.011	Reject null hypothesis
700 V	Median difference of P1 & P2 700 V files = 0	Related-samples Sign test	0.248	Retain null hypothesis

To accompany the Sign test to identify significant differences between the two testing conditions, the A Priori and Post Hoc power was calculated for each of the Voltage files. The variances for each file test condition calculated from SPSS were used, to test for equality of variance using an F-test within the G*Power software. Table 17 lists the variance for the inter-participant results for each test phase (Variance Phase 1 and Variance Phase 2), the ratio of which is used alongside effect size and significance level to define A Priori and Post Hoc power. The use of these Power analysis variables has been discussed within the methodology in Chapter 2, Section 2.2.5, and this structure is used in subsequent chapters. Even though this has been completed after data gathering, A Priori power calculated from the variances gives an indication of the sample size that would be required if access to more participants was available, to meet the desired power (0.80). The Post-Hoc power has also been calculated, which shows the 'actual power' calculated from the results. This is 1 – the probability of a Type II error being committed, which is when the null hypothesis fails to be rejected when it is false.

Table 17 A Priori and Post Hoc Power analysis (3 dp used for resolution specified in SPSS)

File	Variance Phase 1	Variance Phase 2	A-priori power	Sample size required	Actual power
250 V	122227.401	200720.528	0.802	103	0.430
300 V	15415.756	4699.454	0.813	20	0.969
350 V	2246.926	801.856	0.813	26	0.920
400 V	247.683	89.533	0.800	26	0.915
450 V	2.291	0.455	0.825	12	0.999
500 V	2.947	0.021	0.881	3	1.000
550 V	4.579	0.029	0.893	3	1.000
600 V	5.023	0.066	0.800	3	1.000
650 V	4.796	0.031	0.891	3	1.000
700 V	4.629	0.111	0.876	4	1.000

All tests conducted and differences seen between test conditions are significant in line with the power calculated for all files except for the 250 V files, which had a low 'actual power' of 0.430. To see a significant difference between test conditions with statistical confidence, 103 participants would be required. Due to the difficulty of this gate, it is likely that a small difference would be seen between the two test conditions, which requires a large pool of people for valid results. For all remaining files, less participants are needed to produce the variance ratio seen from the current data, which had 37 participants take part in Phase 1 and 34 of those participants take part in Phase 2. The gating completed using a protocol has a very noticeable difference for the well separated files, with an actual power of 1.00 for all files and only requiring 3 to 4 participants to capture this effect if repeated.

To actually use this data to choose an optimum voltage to run the APC channel, the 500 V file would be used, determined by the method described in [34]. This has been demonstrated in Figure 21, where the median inter-participant *rCV* values from each voltage file in Phase 1 were plotted to create a Stain Index *rCV* curve. A Stain Index curve is used to identify the optimum voltage for a specific channel, by plotting *rCV* values for the middle peak of each voltage file acquires with CS&T beads. Median inter-participant *rCV* values were plotted for each file due to non-parametric distributions of operator results for files across Phase 1 and 2 analysis. The inflection point that changes the curve to a plateau indicates the optimum voltage, minimising the effect of background

noise on the data, which can be seen as 500 V. This voltage file was selected to further investigate the distribution of an example data set, and the effects of different statistical methods of determining outliers and control limits.

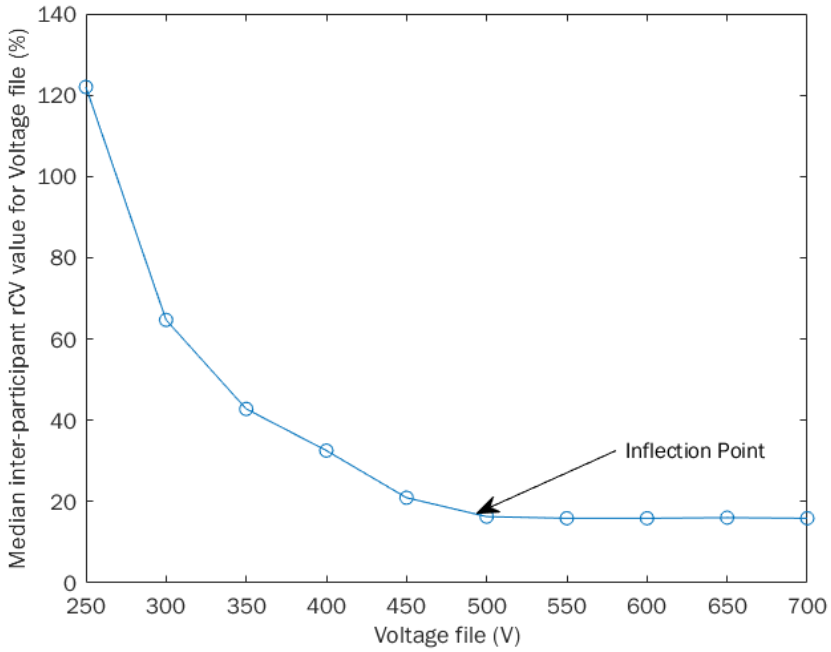


Figure 21 Stain Index rCV curve [154]

Having this 500 V variable selected as the optimum voltage reduces potential variability within the measurement due to amplification of noise or poor detection because of closeness to the limit of detection of the instrument [34,158]. Even though this is a low amount of variation carried forwards in the file, this variability can be further controlled when gating univariate histograms according to a prescribed protocol. A significant difference in range has been shown between participants when using this protocol, improving reproducibility to converge results, reducing variation input by the operator during data analysis.

3.3.4 Flow Cytometry Pre-study Outlier Discrimination Investigation

Normal distribution of data is commonplace to effectively use control limits within manufacturing and other product release criteria, however, in this instance all data is highly skewed, making normal control limit methods less robust to apply. These will be tested within this section to identify how well different methods identify potential outliers, and whether the log transformation process so commonly used shows equality to these.

Outliers are common within distributions and can be identified and removed using a series of control limit methods. These methods have been used within national and international proficiency testing schemes for FC, to maintain continuous improvement [88,95,98,109]. Generally speaking, a 95% Confidence interval is applied to the data set, so the 'out-of-specification' 5 % can be continually reviewed to refine the processes and identify variables that cause participants to fall outside of acceptance boundaries. However, more recently robust methods for outlier determination (based in median statistics) have been utilised within these schemes, generating a set of acceptance criteria, rather than position on a normal distribution [159]. The data presented throughout this Chapter is not normally distributed, despite what initial basic statistical measures may indicate. Often when data is non-normal, a logarithmic transformation can be applied to shift the distribution towards normality so normal statistical tools can be used. However, when transformations are applied, the distribution of the original data is lost, which can cause difficulty in judgement of outliers.

This can present a problem within CGT manufacturing, because if too much transformation of raw data occurs, it could lead to a loss of clarity of the data, potentially causing poor decisions and statistical tests to be applied to patient or product data. When undergoing a cell therapy transplant, there is limited amount of starting material that can be taken from the patient, so it is imperative that poor data analysis does not impact quality metrics of the final therapy product, because this could lead to dire consequences for the patient [160]. European and International regulations state that control limits for acceptance criteria should be set from validation runs of the product and

process [64,139]. However, there is limited validation that can be done on a patient-by-patient basis, due to the difficulties discussed.

Dealing with the raw data distribution rather than transforming to fit traditional norms gives a better understanding of the quality of the cell therapy product, even if the distribution is more unfamiliar. There are also a variety of control limits that can be set for in- and out-of-specification results which will be discussed here. Many traditional methods are used to fit normal distributions, but robust alternatives will also be considered due to the nature of the data distributions in this Chapter. There are no defined criteria for control limit selection, this would be the decision of the manufacturing and QC unit. This section applied the 500 V distribution seen earlier to different control limits to identify differences between in- and out-of-scope data, as well as transforming it to confer between control limits when data is approximated to a more 'normal' shape.

Outlier discrimination has been explored using the 500 V Phase 1 analysis data as an exemplar data set. Any extreme values can traditionally either be removed from analysis or the outliers are addressed to facilitate future improvement in process control.

Table 18 summarizes different boundary specifications used within a variety of industries including biomedical and cell therapy manufacture [96,100,164–170,101–103,109,159,161–163]. There is no one correct choice for outlier determination, with the choice being based on commercial guidelines, quality standards and/or method validation exercises. These boundary estimators indicate whether the statistics used are based upon normal or Gaussian distributions or are more robust to non-parametric shape. The 'Data Dimensionality' indicates whether the boundary estimators are used from one set of data (1 parameter) or if multiple sets are required, often seen in multiparametric analysis such as regression where commonality between two variables is considered. Finally, Table 18 indicates whether the error boundaries calculated by each method are used as strict cut-off limits for specifying outliers, so they could be removed from further

analysis, or whether the method applies a correction to 'out-of-specification' limits and the data can then be accommodated within further analysis steps.

Table 18 Error Boundary Estimators

Error Boundary Estimator	Normal/Robust Statistical Methods	Data Dimensionality	Outlier Limits/ Accommodates Data
Shewhart's Control Charts Mean ± 3*SD*	Normal	1 parameter	Outlier limits
Mean ± 2*SD* Calculation from Type I errors	Normal Normal	1 parameter 1 parameter	Outlier limits Outlier limits
Exponentially Weighted Moving Average (EWMA) charts	Robust	1 parameter	Outlier limits
10% trimmed Mean ± 2*SD*	Normal	1 parameter	Outlier limits
Median ± 2*MAD*	Robust	1 parameter	Outlier limits
Paxton's Criterion	Robust	1 parameter	Outlier limits
Bootstrapping	Normal, or log transformation of non-normal data	1 parameter	Outlier limits
Linear regression models	Normal	Minimum 2 parameters	Outlier limits
Robust Multivariate regression	Robust, but transforms data to normal	Minimum 2 parameters	Outlier limits
Longitudinal mixed effects models	Robust	Minimum 2 parameters	Outlier limits
M-Estimators	Robust	1 parameter	Accommodates data
Euroflow Personalised outlier estimator	Robust, but transforms data to normal	1 parameter	Outlier limits

Determining outliers is completed by calculation of location and spread values of a data set in question. Location parameters such as the mean, median or trimmed means are points used to represent a population and spread parameters such as *SD* or Absolute Deviation represent the amount of fluctuation around the defined location parameter. Historically, boundaries such as Mean ± 2*SD* have been used, because they approximately align to a 95% Confidence Interval, assuming normally distributed data.

Table 18 is not an exhaustive list but the majority use univariate data to understand distributions. However, regression analysis requires two input variables to determine a 95% Confidence Interval. Mean ± 3*SD* are traditional bounds, originally championed in the use of control charts to monitor

process quality over time [161]. These have been amended for use in analytical chemistry Quality Control [162] but were refined to Mean ± 2*SD* in the 1990's FC EQAS [101–103], alongside the use of 95% Confidence Intervals, defined by the Standard Error of the data set.

The Exponentially-Weighted Moving Average (EWMA) Chart is a robust alternative to the Mean ± 3*SD* method, that adjusts to the data set and is often preferred when detecting small shifts [134]. The use of EWMA charts to track data has been recommended more as an internal assessment for manufactured products [163], because it can be easily understood in a similar manner to a Levy-Jennings plot, a traditional control chart format used. Trimmed means are another alternative location parameter, involving recalculation of new spread metrics from a refined data set. They remove a certain percentile of data from the extremes, making the distribution more centralised [163]. This was popular in previous cycles of the UK NEQAS FC Schemes [96,164], allowing pathology laboratories identified as extremes to improve over time. However, if used incorrectly it can remove extremes that require investigation into causality, and still represent probable events occurring.

A robust alternative to Mean ± 2*SD*, is Median ± 2 Median Absolute Deviations (*MAD*) [165]. This uses the median value of residuals from the location median to estimate spread, which is more aligned to non-parametric distributions. Another traditional robust method was described for an early FC EQAS (Paxton's Criterion), that uses parameters calculated to also define Box and Whisker plots, from 25th and 75th percentiles, and subtraction and addition of the Interquartile Range respectively [166]. Bootstrapping is a resampling method used as an alternative technique for validation [167,168] that derives confidence intervals, making assumptions on the data probability distribution, and assumes a normal distribution or uses a log-transform of non-normal data. This is a suitable way of defining bounds if it fits the raw data or if the transform sufficiently equates.

Regression analysis is a common method for specifying data-fitness bounds. Linear regression analysis minimises the residual sum of squares and uses residual plots to detect outliers. Linear

regression has been used to compare FC EQAS approaches [109], with further developments building upon this by using robust multivariate regression [100], which is less sensitive to outliers due to transformation of non-normal data.

Longitudinal mixed effects models have been employed more recently by UK NEQAS [98,169]. This uses a mixture of pass criteria based upon different residual and deviate values, allowing for multiple stratification levels of the data, based upon robust statistics. An alternative method uses M-estimators, that accounts for all the data, but weights extreme values to shift them to an optimal boundary, also known as winsorizing means [170]. This does not technically exclude any data points as previously described methods do, but still uses normal statistics to define initial boundary points.

To demonstrate the effect that different estimators have on data, the Phase 1 500 V data set was used to calculate different outlier discrimination intervals, because this was defined as being the optimal voltage to run the APC channel from previous calculations. This data set was initially tested to describe the distribution, using IBM SPSS Statistics Version 24 software to calculate the descriptive statistics as discussed in previous sections.

To apply different error boundaries to the data set, the outlier discrimination methods were screened to include those that could be applied to a univariate data set for parameters that do not just related to manufactured products and those that act as a set boundary and do not transform the extreme values. Both normal distribution and robust statistics have been included. EWMA charts have been discounted, because there are multiple ways to calculated these control charts, with ambiguous variables which are hard to tailor to the topic in hand [134]. The effect of logarithmic transformation on the data has also been investigated, because this is a common method used to transform data to make it more normally distributed, applicable to a wider range of Gaussian-based statistical analyses and conforming to more traditional manufacturing statistical process control paradigms.

The calculations used for both the raw and transformed data can be seen in Table 19. Bootstrapping relies upon a 95% Confidence Interval calculation, so these have been accounted for together, and Mean ± 3*SD*, 2*SD* and 10 % trimmed mean ± 2*SD* all use the same *SD* calculation to generate control limits. These calculations were completed using Microsoft Excel (Windows 10), however, Matlab R2019a was used to calculate the 10 % trimmed mean and trimmed *SD*, and the 95 % Confidence Interval because it removed subjectivity from the population selection to define the extreme 10%.

Figure 22 and Figure 23 show four plots representing these non-parametric and transformed results. Figure 22 (a) shows a histogram of operator *rCV* values for the 500 V file where operators were asked to apply gates based upon personal interpretation (Phase 1). The grey stepped bands represent the calculated outlier discrimination boundaries derived using the information in Table 19. Any histogram column falling within these boundaries is determined as being in specification. Those falling outside of a boundary would then be classified as an outlier by that method.

The majority of participants fall within all bounds, but this shows the importance of correct selection of discrimination for the process. If chosen control limits are too fine, then allowable variation could be cut out, causing a high defect rate and false negatives. Too wide, and the process is allowed to head towards a more out-of-control state and false positives could be included in analysis, which is undesirable and makes location of the source of the variation more difficult.

Table 19 Outlier estimators, respective calculations and control limits defined [154]

Outlier estimator	Location Calculation	*Spread calculation*	Location value	Spread value	Lower Control Limit	Upper Control Limit	Control Limit Distance	Log transformed Lower Control Limit	Log transformed Upper Control Limit	Log transformed Control Limit Distance
Shewhart Control Charts Mean ± 3*SD*	For trimmed data, 10% of the ordered data are removed before calculation		15.62	5.19	10.43	20.81	10.37	1.02	1.36	0.35
Mean ± 2*SD*				3.46	12.16	19.08	6.92	1.08	1.31	0.23
10% trimmed mean ± 2*SD*			15.92	1.96	13.96	17.88	3.92	1.15	1.26	0.11
95% Confidence Intervals			15.62	0.56	15.06	16.18	1.11	1.17	1.21	0.04
Bootstrapping										
Med ± 2*MAD*			16.30	0.70	15.60	17.00	1.40	1.19	1.23	0.04
IQR Metrics				N/A	13.60	18.4	4.80	1.14	1.27	0.13

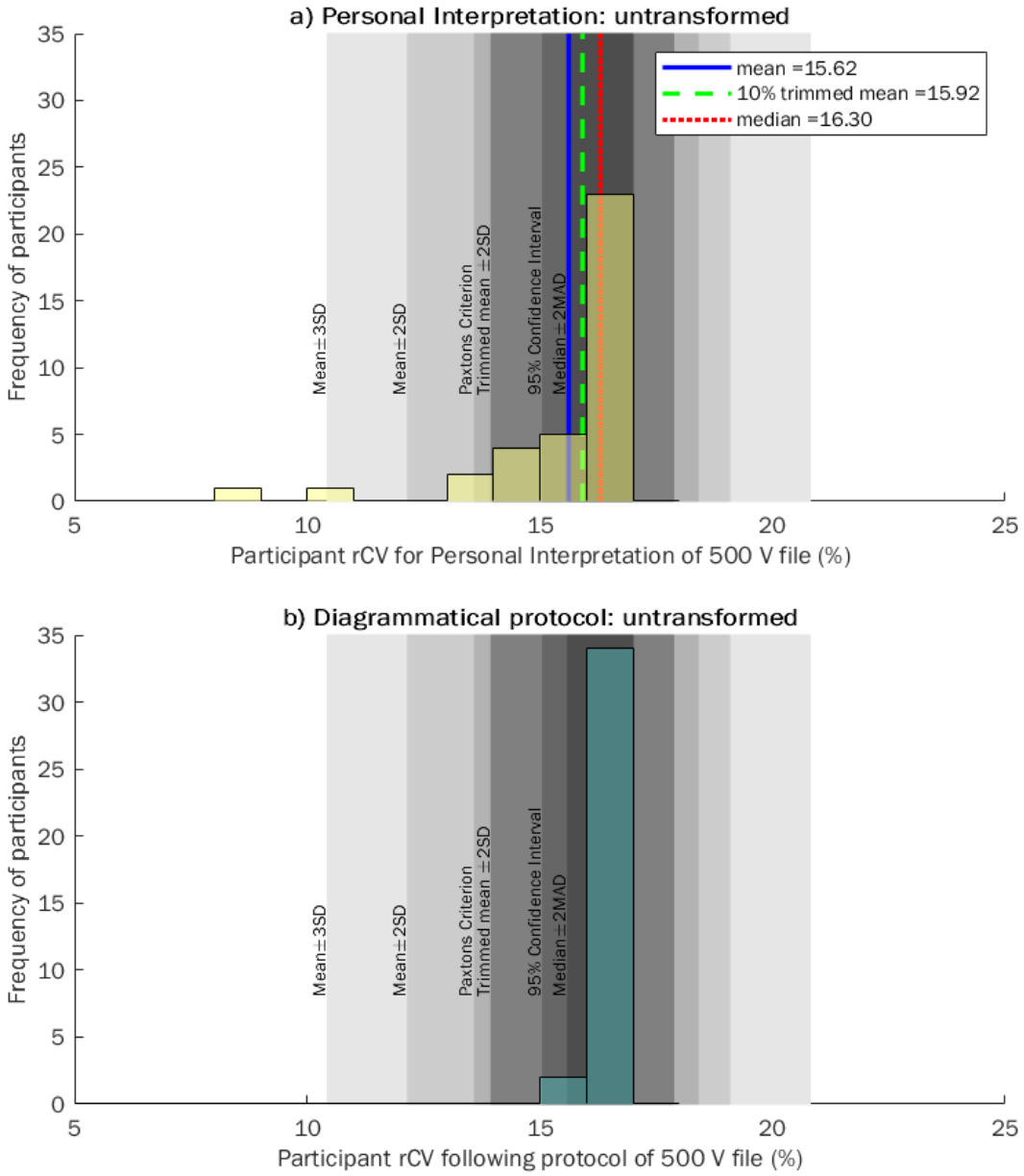


Figure 22 Application of difference control limits applied to data [154]

As seen in Figure 22, all the boundaries scale inwards and outwards, so overlaps can be seen on both sides. The mean, 10% trimmed mean and median have all been plotted on top, to show shift due to the distribution. The median value of the data fits best to represent the skewed distribution because it sits within the histogram maxima.

To show the impact of operators following use of a gating standard to obtain results, the Phase 2 data was overlaid onto discriminatory bounds calculated from Phase 1 data, as seen in Figure 22 (b). All participants fall within the tightest boundary, specified by the Median ± 2*MAD*. This confirms that causing operators to follow a visual protocol not only reduces inter-operator variation but also reduces the chance of 'outlier' results, regardless of how outliers are defined in this context. In the context of cell therapies and their manufacture, these outliers will represent key patient data metrics and cannot be excluded.

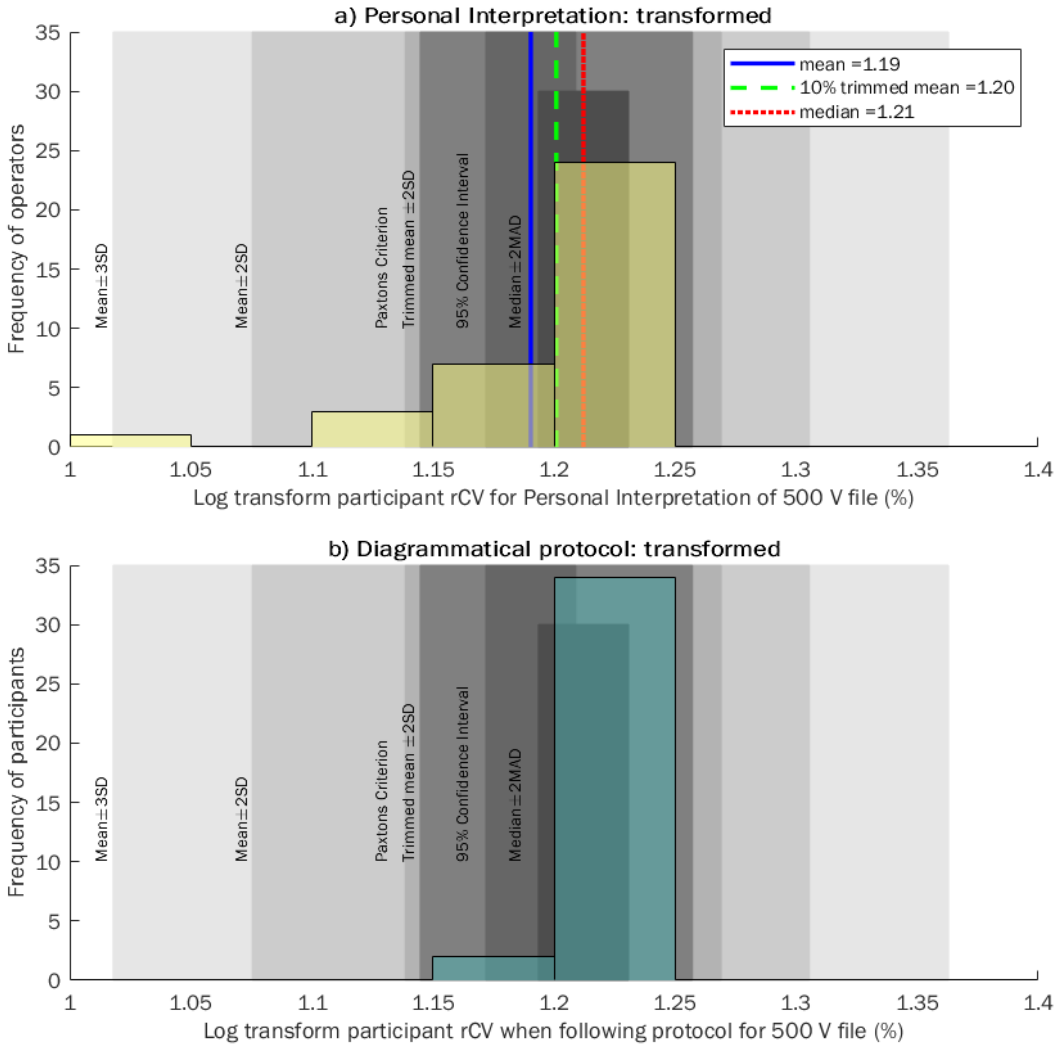


Figure 23 Application of difference control limits applied to log transformed data [154]

The boundaries in Figure 22 were recalculated when the skewed data was normalised with a log transform and this transformed data and boundaries are shown in Figure 23. Data was transformed according to Equation 19.

$$Transformed\ data = log_{10}(raw\ data)$$

Equation 20 Log transformation

The outer boundaries all fall within the same gradient pattern towards the centre, however the 95% confidence interval and the Median ± 2*MAD* bounds sit astride one another, discounting various operators as outliers. Application of Phase 2 (log transformed) data to these defined boundaries is shown again in Figure 23, with operators falling within the tightest boundaries when following a diagrammatical protocol to apply their gates. However, the shape of this distribution is not the same as the distribution when untransformed, potentially indicating a loss of participant traceability back to the raw data, which could cause data integrity difficulty if just using visual data.

To monitor equality of outlier discrimination between the bounds specified for raw and transformed data, Table 20 shows the different bounds used, and what outliers are specifically discriminated against using this method (outliers identified using anonymous coding). Log transformation is deemed appropriate in the context here where the outliers are identical for each context and therefore labelled as 'Equal'.

Table 20 Control limits used to specify outliers [154]

	Non-parametric Outlier frequency	Outlier code	Log transformed outlier frequency	Outlier Code	Equal Outliers	Reason for outlier
Mean ± 3*SD*	1	C04	1	C04	EQUAL	Range gate applied over smaller % of peak
Mean ± 2*SD*	2	C01, C04	2	C01, C04	EQUAL	Range gate applied over smaller % of peak
Paxton's Criterion	3	C01, C04, C07	4	A03, C01, C04, C07	UNEQUAL	Additional outlier in log transform data (A03)
10% trimmed mean ± 2*SD*	3	A03, C01, C04, C07	3	A03, B11, C01, C04, C07	UNEQUAL	Additional outlier in log transform data (B11)
95% Confidence Interval	29	A01-A10, B01-B05, B07, B08, B10-B14, B16, B17, C01, C03-C07	28	A01- A05, A07-A10, B01-B05, B07, B08, B10-B14, B16, B17, C01, C04,-C07	UNEQUAL	Additional outlier in raw data (A06, C03)
Median ± 2*MAD*	10	A03, A05, A06, B04, B09, B11, B15, C01, C04, C07	10	A03, A05, A06, B04, B11, B15, C01, C04, C07	UNEQUAL	Additional outlier in raw data (B09)

All outlier boundaries calculated up to Paxton's Criterion are equal in the number of outliers specified when using either the raw data or log transformed data. As these boundaries become more refined, outliers are included due to the area of the peak gated by the specific operators. Mean ± 3*SD* has one outlier quantified by the gate applied (participant C04, Figure 24) covering a small section of the peak. This does not include all events collected within the peak and therefore has a skewed *rCV* value as a result. This extreme can be compared to a median participant for this gate, because multiple people achieved the same *rCV* value (median = 16.3 %). This median is represented by participant B07, with their respective gate shown in Figure 25.

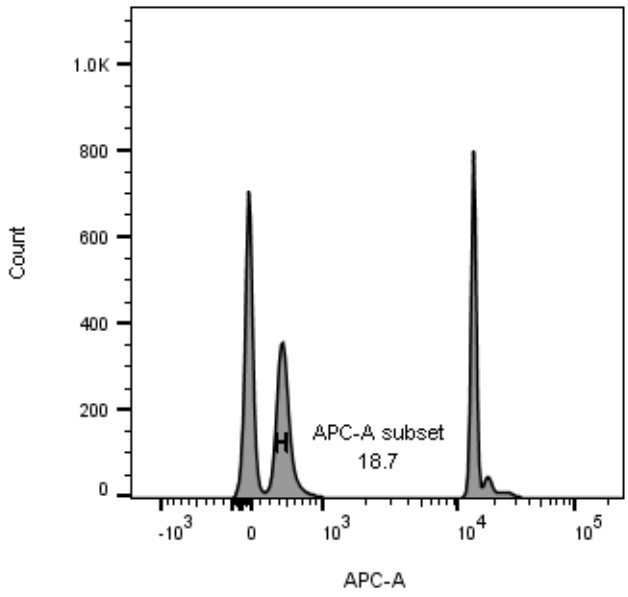


Figure 24 Participant C04 histogram gate applied to middle peak

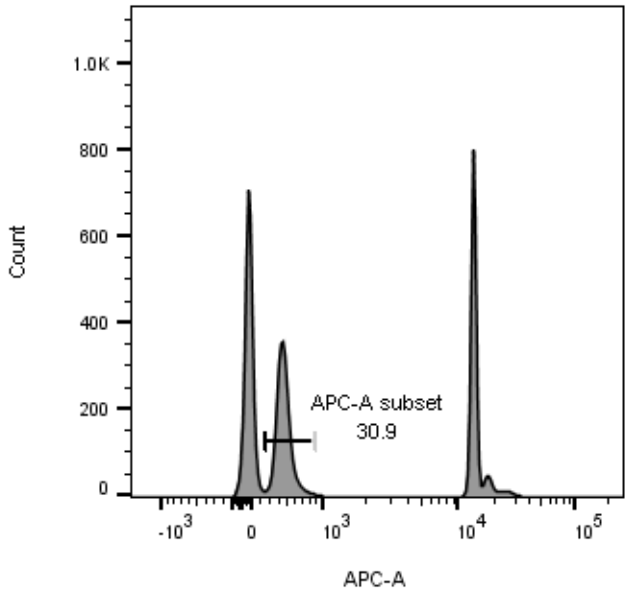


Figure 25 Participant B07 histogram gate applied to middle peak (representative median)

Mean ± 2*SD*, 10% Trimmed Mean ± 2*SD*, and Paxton's Criterion have all specified outliers due to gates not covering the entirety of the peak, and the gradual inclusion of these outliers through more refined control limits sees these outlier gates become wider to include more of the peak,

questioning the outliers have been defined due to differences in spread definition. Figure 26 shows the gate applied by participant C01, in addition to C04 they are out of specification for all boundary estimators except Mean ± 3*SD*.

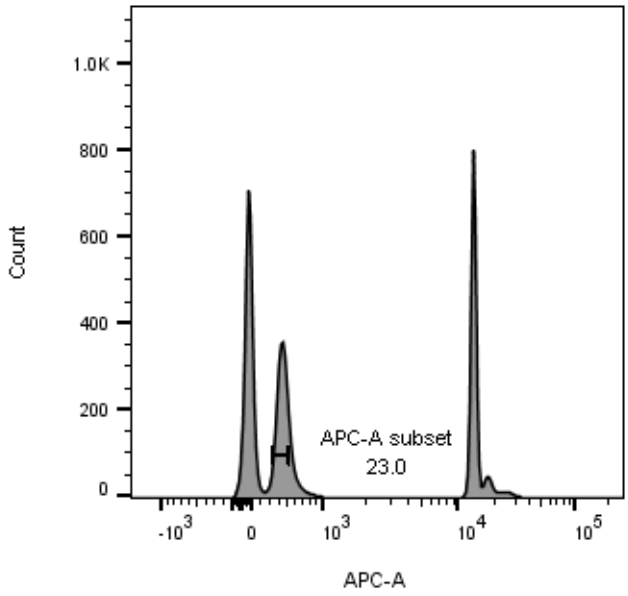


Figure 26 Participant C01 histogram gate applied to middle peak

The 10% trimmed Mean ± 2*SD* and Paxton's Criterion specify an additional outlier, participant C07, who falls outside of the Lower and Upper Control limits specified in Table 19. Figure 27 shows the gate applied by participant C07, which is slightly wider again than participant C04 and C01, but it still does not include the entirety of the peak.

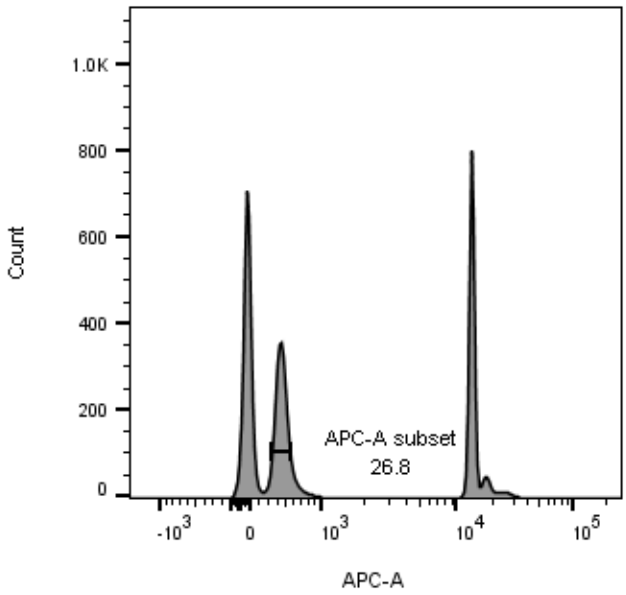


Figure 27 Participant C07 histogram gate applied to middle peak

Paxton's criterion specifies an additional participant A03 when the data has been log transformed, who falls outside of the Lower and Upper Control limits specified for the criteria in Table 19. Participant A03's gate can be seen in Figure 28 respectively, again with slightly wider boundaries applied to the peak, but not covering its entirety. The *rCV* value for this gate sits just outside of the *rCV* lower boundary, suggesting that there could be a difference to outlier definition when applying log transforms. Discrepancies can be seen between the raw and log transformed data for the different boundaries, due to data sitting very close to a raw or transformed boundary, causing it to become in- or out-of-specification. Some more examples have been visualised to identify this, because of the fact this data sits very close to the control limits.

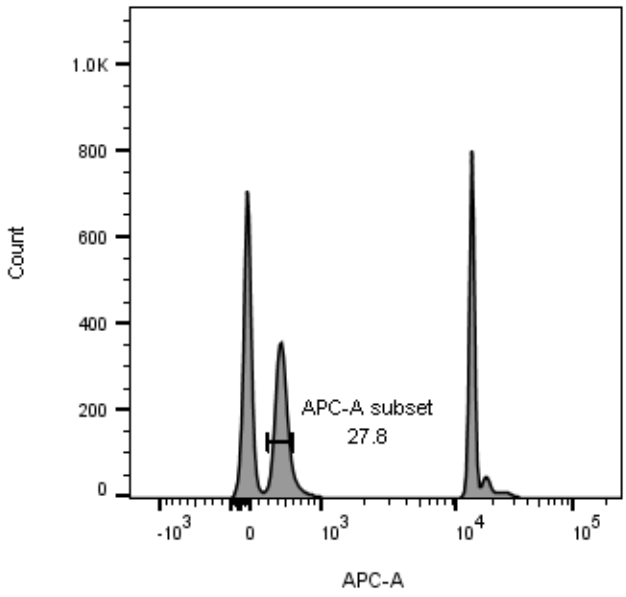


Figure 28 Participant A03 histogram gate applied to middle peak

Qualitatively, A03's gate and previous participant gates have slightly smaller boundaries than the median participant of the group, supporting the trend seen here that as the boundaries calculated become more refined, 'outliers' are those who have smaller *rCV* values because they have not captured as much of the peak as the rest of the participant population. The *rCV* value appears to correlate with the percentage cell count captured within each gate the participants have applied. This percentage cell count can be seen in the Participant gating figures in this section.

Participant B11 (Figure 29) has gated around a smaller section of the peak causing them to be a differential outlier in the log transformed control limits for the 10 % trimmed mean ± 2*SD*. B11 applied gating knowledge from other measurement techniques because they had previous experience with High Performance Liquid Chromatography (HPLC) measurement techniques, where the rule of thumb is to try and capture 80% of the peak to increase measurement confidence and minimise spectral overlap. Compared to more experienced Flow Cytometrists, who try and capture the entire peak, some differences could be seen in training carried over or assumed from other measurement techniques.

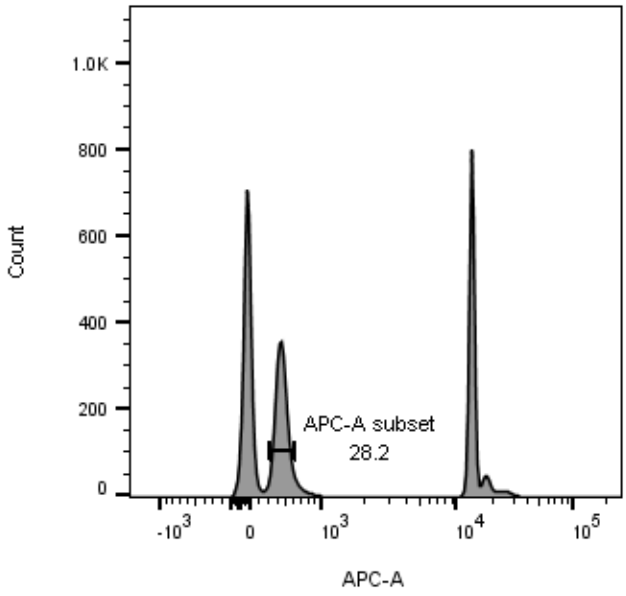


Figure 29 Participant B11 histogram gate applied to middle peak

Both highly refined control limits (95% Confidence Interval and Median ± 2*MAD*) have more outliers present when applying these boundaries to the data set. When calculating the 95% Confidence Interval the raw data contained one extra outlier, which may be expected due to this being a statistical test using normal metrics applied to a non-parametric data set. Both the raw and log-transformed control limits for the 95% Confidence Interval contain a high number of outliers (29 and 28 respectively), leaving only 7 and 8 respective 'inliers' to represent the distribution. Due to the distribution having a high kurtosis and skew, this could affect the calculation of this confidence interval, due to its dependence on the standard deviation that is not located at the peak maxima.

Participants A06 and C03 applied a gate which is unequal as an outlier when assessed as raw data applied to the 95 % Confidence Interval boundary. They are defined as an outlier within the raw data set, not the log transformed data set. As an example, Figure 30 shows A06's gating strategy, with the cell count and *rCV* following the same trend as the previous examples with wider acceptance boundaries. This cell count and *rCV* value are closer to the median value of the group, because this gate includes more of the peak than previous participants have included, but still not the full peak as the right tail of the middle peak has been excluded.

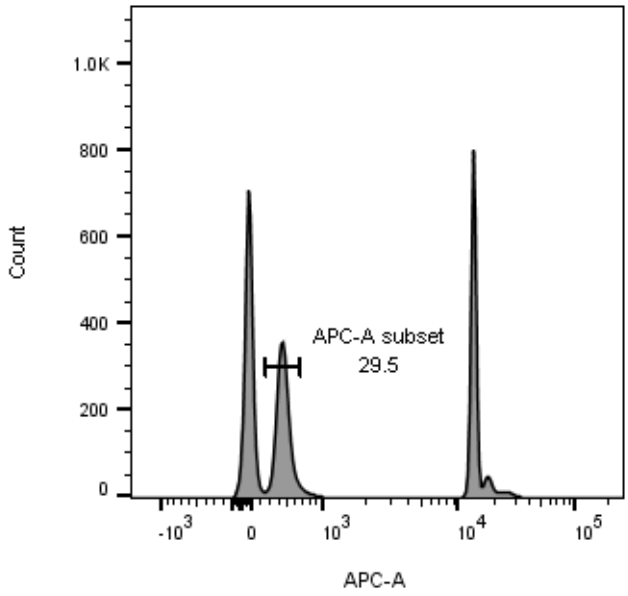


Figure 30 Participant A06 histogram gate applied to middle peak

The robust Median ± 2*MAD* control limits are unequal when assessed with both raw and log transformed data sets, and they also contain more outliers due to the tighter control limits. Participant B09 is the differential outliers specified by this raw data set, in comparison to the log transformed data limits. This is the tightest acceptance boundary, so any 'outliers' that have just fallen outside of boundaries are likely to still look very similar to those results that lie within the boundaries.

Qualitatively, the gate applied by B09 can be seen in Figure 31, and it looks as though the left side of the gate slightly crosses the trough between the dim peak (furthest left) and the middle peak. This could be listed as an outlier because of these other sections of the dim peak being picked up instead of the middle. It appears that this gate has shifted towards the left as the gate does not cover the right tail of the middle peak, like other participant gates have done.

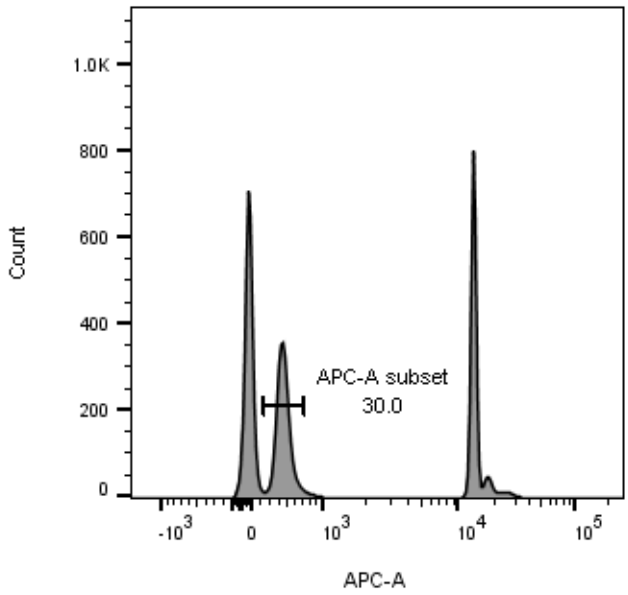


Figure 31 Participant B09 histogram gate applied to middle peak

To conclude, the raw untransformed data will be used for comparison in further studies, to enable a better understanding of the population which is more directly traceable to the study data gathered. Even though there is a small difference between outliers defined in raw and log transformed distributions, in some instances where there is inequality of outliers, it has shown that log transformation may not always be suitable for skewed data. It is important to confirm and understand the shape of the data before applying tranforms and control limits to identify outliers. These boundaries will be applied to further data sets to show the difference in control limit definitions,however, this will not be to specify a perferred option for defining outliers, because it requires more manufacturing validation and regulatory or industrial guidelines.

3.4 Chapter Conclusions

Chapter 3 shows the amount of variation seen between participants when applying histogram gates to the same data, indicating that this needs to be focused on and more specifically measured within future work. Across a range of different data histograms, the difference between participant analysis can vary greatly according to a number of factors which could depend upon the participant experience, how frequently they deal with histogram Flow Cytometry data and what they believe to be acceptable cut-off points for data inclusion. This can be used as evidence to formally measure baseline operator variability using measurement uncertainty methods when applying gates to data.

Chapter 3 has tested the use of gating protocols for univariate histogram data, providing a basis to practice running studies, validating this study design (3 repeats within a 30 minute - 1 hour time frame) for future uncertainty analysis trials, which can build in complexity from this univariate experiment.

Inter-participant variability was also present when participants had to apply gates following a diagrammatical 'gating protocol' instead. All the ranges of between-participant results dramatically reduced in cases where the peak separation was clear and an obvious split between the three peaks could be seen (percentage reductions ranging from 86 % to 92 % for 500 V to 700 V files). The gating images provided helped participants to align the gate to the selected upper and lower inclusion limits, although some participants expressed they would have applied the gate differently, using different tools or settings. However, this is all relative, because these factors were controlled to ensure comparability between participants.

Even when gating the poor separation peaks the gating protocol aided participants, because it reduced the range across all voltage files, however some were better than others (reduction when using the protocol between 2 % to 42% for 250 V to 450 V files). Those files with lower ranges were very difficult for participants to gate because there was no visual separation of the peaks to discern

the correct one. These lower voltages should not be used when setting up and running a flow cytometer, because they are too close to the limit of detection of the channel. This greatly affects the ability to discern populations within the data set and is an effective example of showing how upstream setup and variables need to be correctly optimised and validated to ensure confidence in final interpreted results.

This pre-study has also highlighted the importance of completing a thorough check on not only the basic statistics representing a population, but to check the distribution and more complex statistical tests because these give more information about departures from normality and potential clusters of data that can appear in bimodal instances. It has also called into question the need for a normal distribution, when trying to get reproducible data. If trying to get results within a certain boundary, measures such as skewness and kurtosis appear to be out-of-specification and extreme compared to SPSS boundaries (for normal distribution affinity), but they actually indicate a high convergence of results, because these specified limits indicate a normal distribution which may not be a necessary shape.

Due to the non-parametric distributions and skewness present, log transformations were investigated to see whether the transform could make the distributions appear more normal, to then apply various outlier boundary estimators to the data. This was applied to data distributions when participants applied gates using their own judgement and then following a protocol. However, when following a protocol the distribution was very convergent and therefore highly kurtosed, so not much difference was seen when transforming the data, other than flattening the two histogram bins present. More of a distribution shape was seen when participants used their own judgement and this is more indicative of real-world scenarios.

The power analysis completed on this data has shown that enough participants were recruited to identify the difference in variance for most of the files analysed. The one file which did not have enough participants, returning a power of 0.430, was the least optimal, so this generally would not

be used for FC analysis when setting up voltages for the APC channel. This power analysis structure will be used in further chapters where two different testing conditions are present, considering A Priori and Post Hoc power to ensure that the correct sample sizes and powers can be identified, even if they cannot necessarily be met.

Various error boundary estimators are used throughout manufacturing scenarios and were applied to a set of data here to highlight the difference in how many outliers were produced. Boundary estimators would usually be selected internally by manufacturers based upon historical validation data. However, these are traditionally based upon normal distributions being required, so these were tested on a non-normal set of data, alongside robust statistics which are designed to deal with more non-normal distributions. In some instances there is equality between the raw and log-transformed data when using error boundary estimators to identify outliers. However, instances where there was not equality between the two data representation methods has highlighted that differences could be due to using certain boundary methods such as 95 % Confidence interval. These differences are due to log transform boundaries being slightly different to those applied form the raw data, which could cause problems when trying to identify outliers in skewed distributions, depending on how the analyst interprets the results. Discrepancies between skewed raw and log transformed data has been discussed in other biomedical literature, showing that log transforms can cause discrepancies in data analysis and consideration from the raw data is perhaps more suitable [171,172]. This can give an appreciation of what the data looks like before trying to make it fit to a more traditional norm.

This will influence further work presented in this thesis, by ensuring that no transformation of data will be made without strong evidence of its use and without understanding the underlying statistical distribution first. It also calls into question error boundaries used to define outliers and enforce continuous improvement. As baseline uncertainty is measured, these limitations will need to be defined by a reputable error boundary discrimination method, which will depend on the shape of the distributions to ensure it is robust.

Ultimately, this pre-study has shown that there is significant variation present from analyst subjectivity, as well as potentially being able to control this with protocols, which will be investigated within subsequent Chapters. The investigation of log transformation of data has shown differences to occur in outliers, showing that this method needs to be fully understood before applying transformations to raw data. This can have a significant impact on CGT manufacturing, because a lot of cell and gene measurement distributions are not normal [173,174], and their measurement analysis as well as the data distributions themselves need much more attention and understanding to make informed decisions.

3.4.1 Consolidation of Objectives

- There is considerable variation in between-participant data analysis when they apply gates to the same data. This is seen through all data files given to participants, which show variability when peak separation is good (range of 8.0 % *rCV* for 500 V file) and when it is poor (range of 400 % *rCV* for 300 V). This has been shown on simple univariate data and can provide a session structure for subsequent chapters which detail more complex data and analysis sequences.

- In this instance, diagrammatical protocols provided to participants have reduced between-participant range of *rCV* results when copying the gates. This can be tested in future variation studies, but they will be more complex in visualisation and structure to monitor how the range in variation changes, so this must be taken into account. When peak separation is 'good', this reduction is up to 92 %, and when peak separation is 'poor', this reduction of variation is up to 42 %.

- Log transformation of the raw data has been tested here, with the conclusion that future data sets will not be transformed if they are non-normal, because there can be inequality of outliers and error boundaries between the two data presentation types. Therefore, raw data distributions shall be used, because these present relatable data distributions to work with, unless evidence suggests otherwise.

- Different outlier boundaries have been tested with the chosen data set here, with normal and non-normal parameters. It has highlighted for future work that the distribution of the data needs to be thoroughly tested before outliers are determined with a particular method, to ensure the data is distributed appropriately. It is likely that even if future distributions are normal or non-normal, robust methods may be used as these can be used as representative statistics in both cases.

Chapter 4: Basic Uncertainty Model

4.0 Introduction to the Chapter

Chapter 4 introduces the first uncertainty Gating Exercise, as part of the sequence of studies which monitor participant variance in comparison to complexity of data. As discussed in the Prelude, this first uncertainty exercise is basic, to set a baseline for understanding variance in non-complex analysis scenarios. This sets the basis for further uncertainty gating studies described in Chapters 5 and 6 with comparisons drawn in Chapter 7. The overall Gauge R&R study design as described in Chapter 2 has been followed to obtain statistics which can be taken forwards for uncertainty calculations. Further specificity for this basic study exercise is given in this Chapter. Previously in Chapter 3, only one variable had been considered, in histogram format, whereas this first uncertainty stage tackles FC data analysis in pseudocolour and bivariate formats using an immortalised pluripotent cell line (2102 EP Carcinoma line). This is much more prevalent within the community, therefore potentially providing a more relevant application of uncertainty to FC measurements.

4.1 Chapter Aims

This Chapter provides a basis for comparison of uncertainty in more complex FC analysis strategies found in future chapters. The fit of this Chapter to the thesis can be seen in Figure 32, specifically within the orange dashed box, providing a base for the core hypothesis: as complexity of FC data and processing increases, measurement uncertainty contributed from the participant will also increase. CGTs can have very complex analysis and quality control measures, so a better understanding of how the participant plays an influencing role in these results is essential for product release, continuous improvement and patient safety.

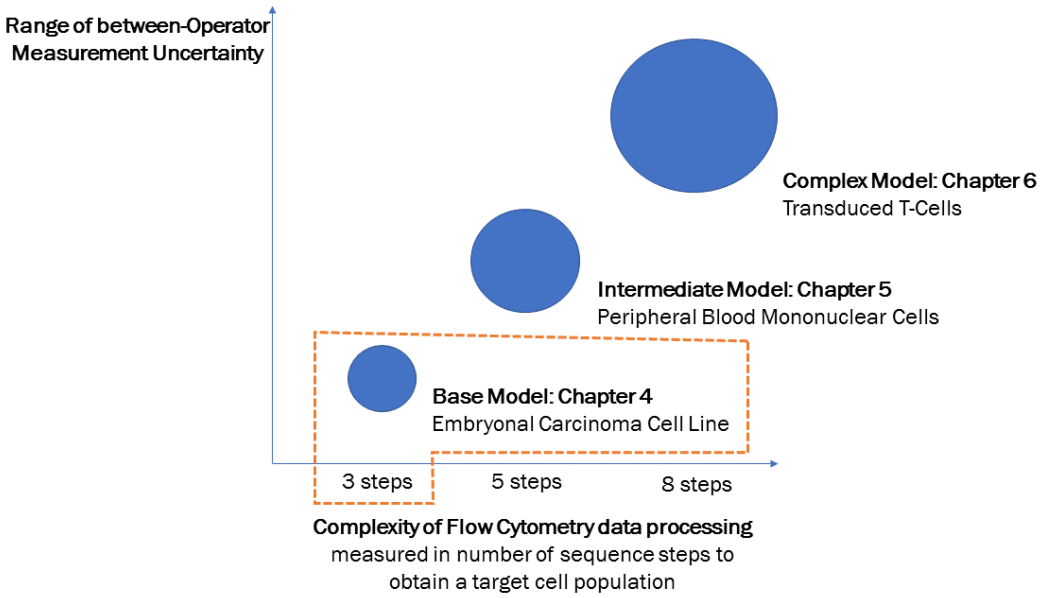


Figure 32 Diagram of Chapter position within the whole Thesis

4.1.1 Chapter Aims & Objectives

The Aims and Objectives of this Chapter can be defined as follows:

- Design a baseline study using a Gauge R&R structure to measure *CV* and uncertainty of participants when gating simple Flow Cytometry data.
- Identify what causes extreme values in absolute reported results and uncertainty measurements by reviewing extremes from this baseline study.
- Provide easy monitoring tools for understanding uncertainty within a larger participant study or facility.
- Investigate the impact on variance from data visualisation and different gating tools that can be used throughout Flow Cytometry Post-Analytical processing.

4.2 Methodology

The Embryonal Carcinoma (EC) 2102Ep immortalized cell line used has been suggested as a 'Ruler line' or 'Reference Standard' for culture and research, because it has an antigen expression profile similar to conventional human Embryonic Stem Cells (hESCs) [175]. This antigen expression profile was unchanging over 10 passages, and this biological stability make the cells a good example for quantifying downstream experimental uncertainty, due to minimal biological variation [175]. As such, they were chosen as the simple starting model for these studies.

4.2.1 Cell Selection and Culture

A vial of GlobalStem® EC 2102Ep cells (Passage 48) (5 x 10^6 cells) was removed from a liquid Nitrogen Cryobank (See Appendix B for Certificate of Analysis) and was thawed in a water bath for 3 minutes until a slither of frozen material remained. The material was topped up 1:1 with cell culture media, mixed slowly and seeded manually onto a T75 ThermoScientific™ Nunc™ Cell Culture Flask in 15 ml Gibco™ DMEM, high glucose, GlutaMAX™ supplement (Cat 61965, Lot 1813259), fortified with 10 % v/v Fetal Bovine Serum (FBS). Cells were maintained in a humidified incubator with 5 % CO_2 at 37˚C.

100 % medium exchange was carried out every 2 days after seeding or passage, and once confluent after 3 days, the seeded flask was manually passaged into 2 daughter T75 flasks at a seeding density of 6.7 x 10^4 cells/cm^2. This equated to a 1:3 split ratio, because cell counting was affected by the difficulty in dispersing cell clumps. To aid this dispersion, a 2-step disassociation process was used, cells were trypsinised with 1.5 mL Gibco™ Trypsin EDTA (0.25 %) (Cat 25200072, Lot 1814171) for 5 minutes in a humidified incubator at 37˚C with 5% CO_2, quenching with 3 mL of the fortified DMEM to stop the enzyme.

Cells were centrifuged at 300g for 5 minutes, waste supernatant aspirated, and the remaining cell pellet resuspended in 1.5 mL 0.25 % Trypsin EDTA and incubated for a further 5 minutes. After

further quenching and removal of the supernatant, cells were resuspended in the fortified DMEM, with cell counts, and viability assessed using a Nucleo-Counter® NC-3000™ and ChemoMetec Via1-Cassettes™, to stain and measure cells with Acridine Orange and DAPI dyes. Three repeat measures were taken at each count to obtain a mean cell count and viability, before reseeding into the next passage. Cells were passaged through 5 successions, with an average cell viability of 87 ± 3 % over the culture period.

4.2.2 . fcs File Generation

A series of fcs files were generated using the EC 2102 Ep cell line in culture through 5 passages and harvested as previously described. The cells were fixed, permeabilized and stained using the BD Stemflow™ Human/Mouse Pluripotent Stem Cell Analysis Kit (Cat 560477, Lot 7004890), according to the included method [156]. Enough cells were harvested to generate the respective isotype and Fluorescence Minus One (FMO) controls, alongside the stained cells. 1 x 10^7 cells were harvested and fixed in 1 ml 4 % BD Cytofix™ Fixation Buffer (Cat 51-9006276, Lot 7004890), incubated in the dark for 20 minutes and washed twice with 1 ml PBS (no Calcium, no Magnesium).

The cells were mixed slowly with 1 mL BD Perm/Wash Buffer (Cat 51-9006275, Lot 6232552) and incubated at room temperature in the dark for 10 minutes. After 2 washes, the cells were split into Eppendorf vials of 1 x 10^6 cells each, suspended in 100 µL Perm/Wash buffer and stained with the respective dyes: 15 µL BD Pharmingen™ PerCp-Cy5.5 Mouse Anti-Oct 3/4 (Cat 51-9006267, Lot 6232550, 10 µL BD Pharmingen™ Alexa Fluor 647 Mouse Anti-SSEA-4 (Cat 51-9006265, Lot 6316682) and 20 µL BD Pharmingen™ Mouse Anti-SSEA-1 (Cat 51-9006268, Lot 6316683).

Oct 3/4 is a member of the POU (Pit-Oct-Unc) family of transcription factors, which functions in the early stages of a pluripotent cell within hESC and EC Cell Lines. Oct 3/4 is expressed in undifferentiated cells, but loses expression as cells start to differentiate towards specific cellular subsets [176]. SSEA-4 (Stage Specific Embryonic Antigen–4) markers are also identifiers of undifferentiated ESCs and ECs [177], with the difference from Oct 3/4 being that SSEA-4 is a cell

surface marker. Oct 3/4 is intracellular, which requires the additional fixation and permeabilization step. SSEA-1 is a surface marker of differentiation for Stem cells, often used to compare with SSEA-4 and other pluripotency markers [178].

Once incubated for 30 minutes in the dark, cells were washed twice and transferred into BD Falcon™ Round Bottom 12 x 75 mm tubes (Cat 352063) and kept covered to minimise light exposure. Cells were run through a BD FACSCanto™ II Flow Cytometer, using the respective fluorescence channel and voltage: FSC 180 V, SSC 374 V, PerCp 420 V, APC (same detection range for Alexa Fluor 647) 450 V and PE 352 V, once a daily calibration was completed using Cytometer Setup & Tracking beads (Lot: 74538, Successful calibration). A viability stain was not included in the FC panel, due to viability being assessed with cell counts, and the need to keep the gating panel initially straightforward for participants, following the prescribed method in the Analysis Kit [156].

Each tube and respective fcs file were generated using a medium flow rate (60 µL/min) and by acquiring 30,000 cellular events. Multiple stained sample fcs files were generated to build a library of repeats to use within the variation studies. These are representative of the product samples described in the manufacturing scenario used to describe Gauge Repeatability & Reproducibility in Section 2.3.1. Files were exported as fcs 3.0 version types for use in Flowjo Version 10.0.8r1 third party analysis software [123] and saved as a workspace. This was repeated twice for each participant, so they could analyse data across a total of three workspaces.

4.2.3 Questionnaire Design & Ethical Approval for Study

Ethical approval was obtained from Loughborough University Human Participants Ethics sub-committee for the study and all participants were informed of the intentions of the study, as previously described in Section 3.2.2. Before the FC gating study commenced, participants completed an online questionnaire, to identify differences between participants and understand their experience background. This questionnaire was also given to participants to repeat at the end

of the FC complex gating exercises. Results comparing these two sections are discussed in Chapter 8, to identify factors that attribute to measured variance.

All participants and their respective data were anonymised at the point of data collection, and data stored in accordance to the ethical clearance obtained. All data analysis relating to questionnaire or gating results was completed anonymously, and participant coding was restructured from previous work to remove the possibility of analysis bias. Any questions requiring written text answers were analysed and qualitatively coded based upon prescribed manual coding methods [179]. These codes were counted to measure the frequency of issues reported, and more information on the qualitative coding can be found in Chapter 8.

4.2.4 Flow Cytometry Study Organisation

A total of 38 Participants from three separate centres (10 from an academic institution, 19 and 9 participants from separate industrial institutions) were invited to complete the study in a quiet analysis space, to avoid distraction and the possibility of others seeing the study content and analysis. Study sessions had a one-hour maximum duration, and participants were shown three Flowjo workspaces, which contained a series of fully stained EC 2102 Ep fcs files. One identical file was included in each workspace, and participants were instructed to gate through a three-plot sequence to identify target cells (using Forward Scatter (FSC) plot against Side Scatter (SSC)), then the option to gate single cells, and finally to apply a quadrant gate to the double positive stem cell marker population to identify the final percentage cell count of respective pluripotent stem cells. Flowjo was the choice of platform due to access of the software across all three collaborator and participant sites, meaning a higher number of participants were likely to be familiar with the platform.

Participants were also provided with isotype controls and FMO controls in each workspace to aid gate application and were allowed to use whatever manual gating tool on Flowjo they felt best to gate the population in hand. Due to the variety of ways in which people gate single cells, these axes

were left to the discretion of the participant and their preference. An overall schematic of the gating sequence they were asked to follow is shown in Figure 33, and participants gated each workspace of files separately to ensure a correct quantification of uncertainty through standard deviation calculation in accordance with Gauge R&R methodology principles described earlier.

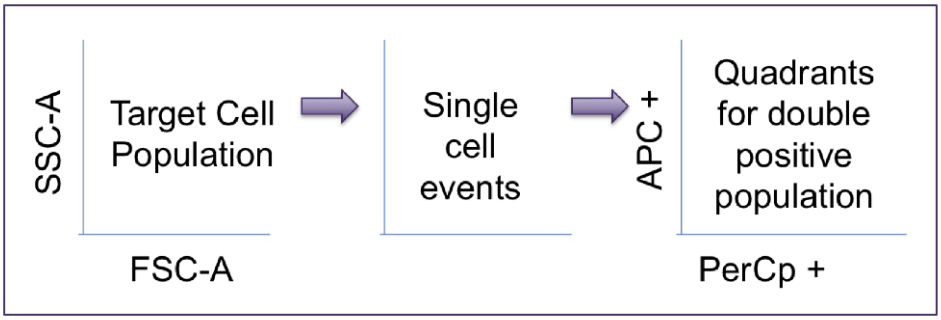


Figure 33 Gating sequence participants were asked to follow, to identify the target single cell population, with double positive expression for both Pluripotency markers

This data exercise aims to identify if variance does exist when reporting results from the same data and that uncertainty values can be calculated for participants to contribute towards an overall uncertainty estimation. This was designed to estimate a general absolute result median and uncertainty. Calculating significant differences between certain groups or testing conditions was irrelevant, meaning that standard power calculations to determine appropriate sample size could not be used. In addition, a one-sample T-test cannot be used because it requires a hypothesised value for the sample to differ from. Because this data set was supposed to be exploratory, the initial results from this uncertainty study can be used for future comparison.

When completing Gauge R&R studies it is recommended for repeats to exceed $n \geq 10$ [133], so that there is a good number of participants for inter-participant comparison of results. To calculate uncertainty, it is recommended that a minimum number of measurement repeats to take is between 4 to 10 to try and best eliminate anomalies whilst being time efficient [142]. Participants made 3 repeat measures, because when trialled, 4 repeats could not be successfully conducted within a 1-hour time slot, whereas 3 repeats could be comfortably achieved and this still allows for

SD to be calculated. A 1-hour slot was used because this was the maximum allowance given for each participant's time, agreed across the three institutions.

4.2.5 Uncertainty Calculation

Once studies had been completed, target cell, single cell and final percentage positive cell population metrics were extracted from the data, using the results from the identical repeated file situated in each Flowjo workspace. These were then used to calculate a mean cell count, *SD* and *CV* for each gating stage, per participant using Microsoft Excel software. Finally, a combined uncertainty (u_c) was calculated by combining these Type A uncertainties by summation in quadrature. The u_c value was expanded with a coverage factor of k = 2, representing a 95 % Confidence Interval for the uncertainty statement, which gave each participant a representative expanded uncertainty (*U*) figure, to show individual variance. The mathematical methodology used to calculate uncertainty metrics has been previously discussed in Section 2.4. An example of the data extraction through to calculation of metrics and uncertainty can be seen below.

		Total Cell Events within each gate in sequence			
Total Starting Cell Events		Repeated File	Target	Singlet	Double Positive
30000	Repeat 1	Wsp1, File 1	6027	5981	5972
	Repeat 2	Wsp2, File 2	6097	6041	6012
	Repeat 3	Wsp3, File 3	6608	6562	6515
		Respective Percentages of Cell Events in each gate as a function of total starting cell number			
			Target	Singlet	Double Positive
		Wsp1, File 1	20.09	19.94	19.91
		Wsp2, File 2	20.32	20.14	20.04
		Wsp3, File 3	22.03	21.87	21.72
			Target	Singlet	Double Positive
		Average	20.81	20.65	20.55
		STDEV	1.06	1.07	1.01
		CV	5.08	5.16	4.91
		k	2		
		uc	1.81		
		U	3.62		

Figure 34 Example of data extraction through to calculation of absolute results and uncertainty

4.3 Validation of Gating Tools and Plot Visualisations

A small pilot study with 8 participants from 1 site was conducted to see if the gating tools or plot visualisations used had an impact on the range of variation seen within the analysis. This fed into the main study to possibly restrain visualisation and gating tool variables that participants used in the software.

The repeated .fcs files used for repeats in Flowjo workspaces within the main study was analysed by the participants in a random order of repeats. Three visualisation methods were tested; a contour plot, a greyscale plot and a pseudocolour plot (Figure 35). The target cell population was gated using three different tool types: an ellipse, a manual polygon tool and an autogate tool (Figure 36). For each gating tool and visualisation used, three randomised repeats were taken.

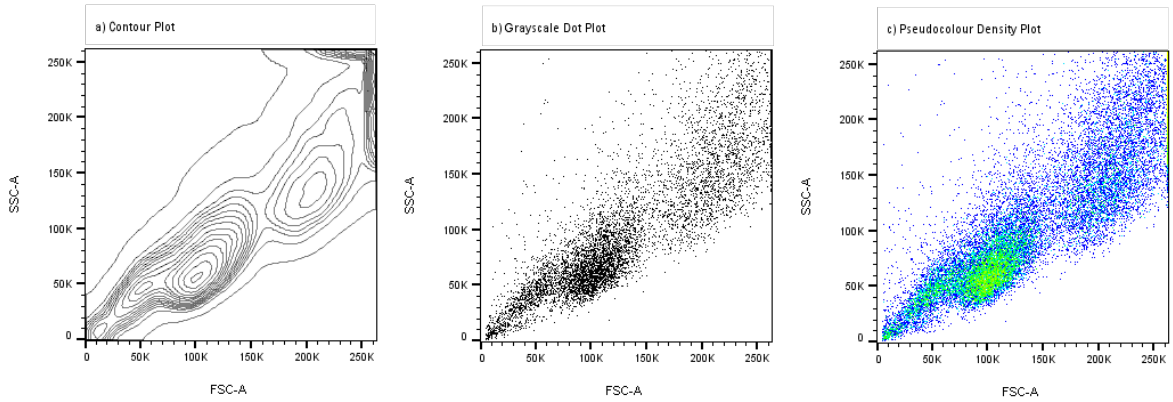


Figure 35 Visualisations tested within pilot to see if data visuals affect participant gating variance

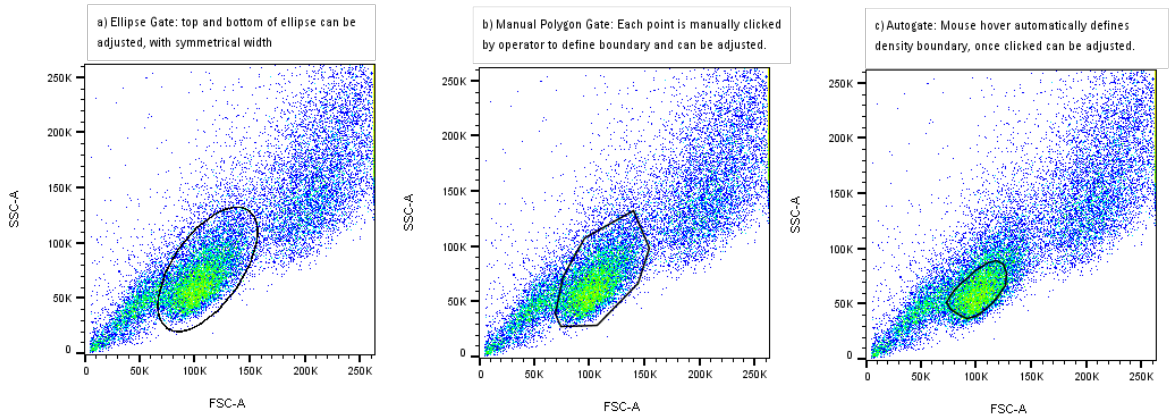


Figure 36 Gate tools tested to see if they have an impact on participant gating variance

This data was captured over 3 separate 30-minute sessions, as shown in Figure 37. No participants were deemed colour blind (when tested using a Farnsworth-Munsell D-15 test (Figure 38)), which is a potential source of variance for a participant and could be investigated as part of future work. A Farnsworth-Munsell D15 test uses 15 coloured discs that range from the blue end of the colour spectrum to the red end. These discs are randomly mixed up and participants had to arrange them from what they considered to be blue to red. These were then analysed by the study coordinator, because the reverse of the discs has a numerical order. If the order of the numbers is not correct, this can indicate colour blindness.

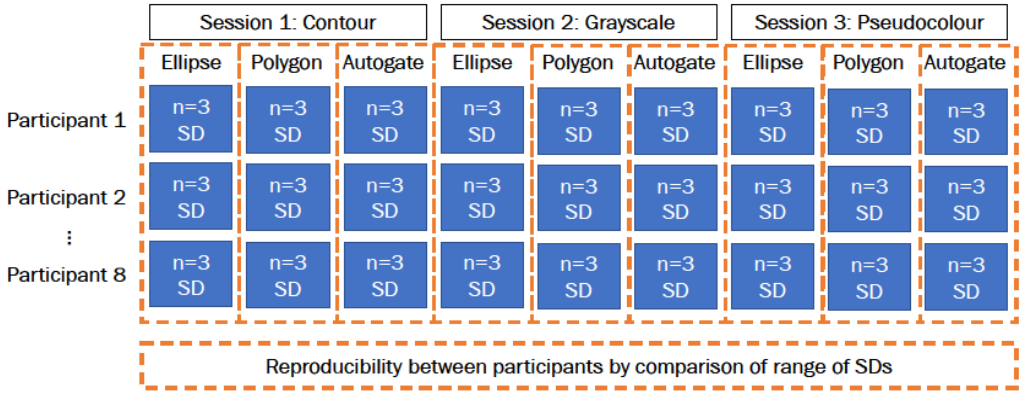


Figure 37 Visualisation pilot study structure diagram for comparison of gating tools and visualisation between participants

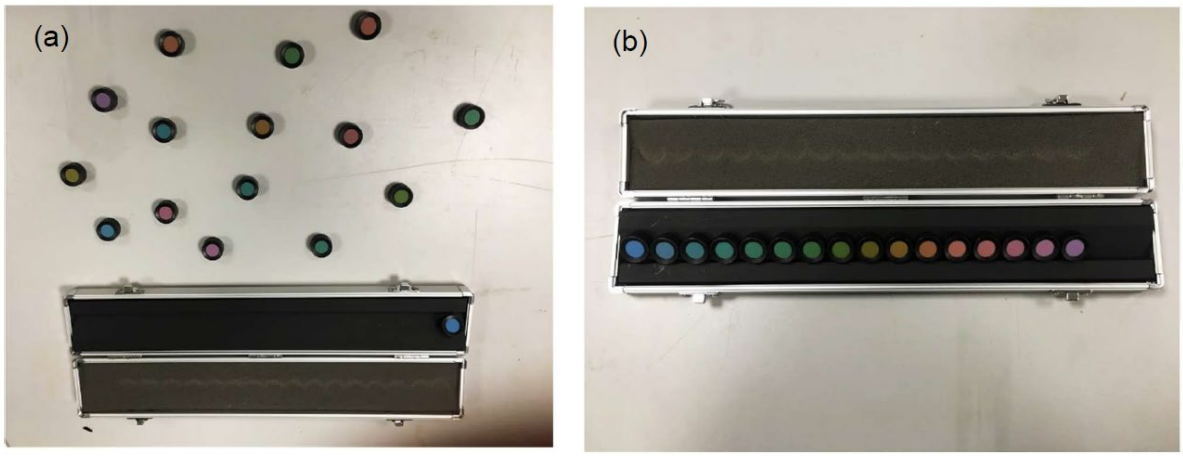


Figure 38 Farnsworth-Munsell D-15 Colour Vision Test; a) Random pieces for participants to order, b) Correct order for colour tags

The orange dashed boundaries in Figure 37 separate the reproducibility boundaries for each tool and plot, from which an understanding of gating range was looked into. The cell count percentage as a function of the original file cell event number (30,000 events) was used for comparison. *SD*s of the three repeats were compared for each tool and plot, with total inter-participant *SD* ranges for each configuration shown in Figure 37 for tool types and plots used [180], generated using a combination of Microsoft Excel and Matlab R2019a for processing and visualisation respectively.

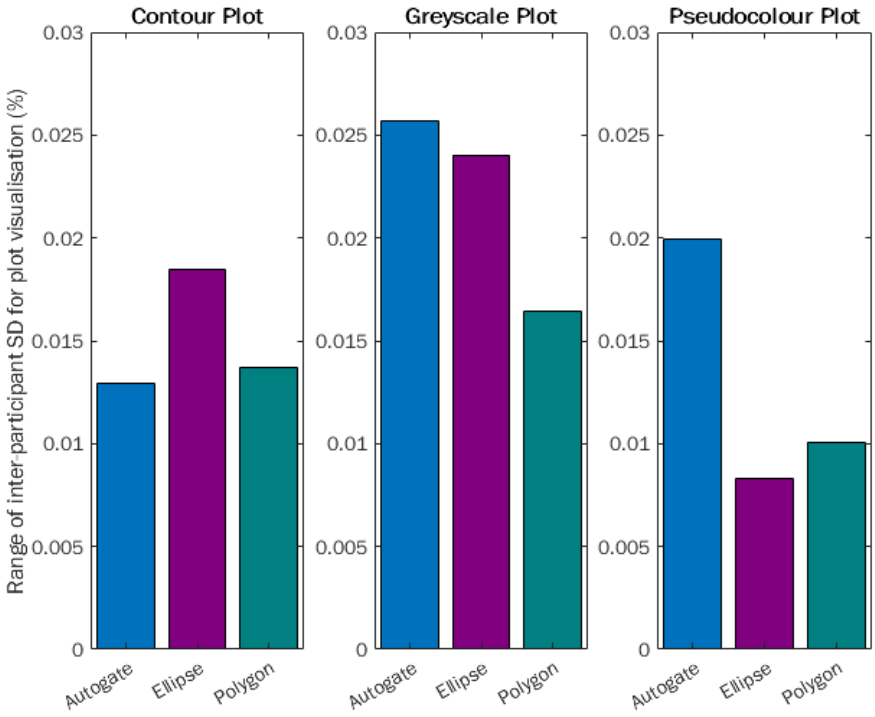


Figure 39 Ranges of inter-participant SD *for each gate tool on each visualisation plot*

Figure 39 shows that using a greyscale plot will potentially contribute greater variance to the overall measurement, especially if using an autogate tool. Unlike using this tool on a contour plot or a pseudo colour plot, greyscale provides no other perception of density of the cell events, leaving participants to make less informed decisions, increasing variability in their measurements. The autogate tool also had significantly higher between participant range when using pseudocolour density plots. The autogate tool could not effectively separate the boundary between the target cell population and the dying cells. This software issue led to participants making variable decisions on

where to place the gate. During the sessions, participants commented on the autogate tool, often feeling they either cut out too much of the population or included irrelevant cells from another.

Contour plots had the smallest difference between gate tool ranges, possibly because the contour lines provided guidance for participants to follow and place gates. Even though the ellipse gate fitted the target cell population very well, the manual polygon has consistently low variation range on each visual plot. This may be because the participant ultimately has more control over the shape of the gate and the placement of gate indices. Whilst the autogate tool allows adjustment once the gate has been placed, it produced so many indices that participants did not want to spend the time adjusting them all.

Ultimately, the tool used will be based on a variety of factors including but not limited to the general shape of the cell population, density, scaling, personal preference and knowledge of other tools (or lack of knowledge). Overall the *SD* ranges shown here are very small so the choice of visual plot or gating tool chosen may not have a significant effect on the overall measurement, however, the pseudocolour plot will be used for the main study, because it shows generally low variance with the gating tools tested and it is also the default layout in the Flowjo software used. One should be aware of other factors that can influence these measurements such as colour vision and time taken to complete these studies. In high-pressure situations where participants may not have full colour vision, results could be skewed or greater in uncertainty.

4.4 Results & Discussion

4.4.1 Flow Cytometry Basic Gating Exercise absolute results

The absolute results reported here are the targeted cell population that participants were asked to identify using the gating sequence defined in Figure 33. These are akin to what would be reported in literature for specific cell types, in this instance it is single, pluripotent stem cells. The uncertainty of the gating sequence will be discussed in the next section of this Chapter.

Table 21 Measures of Location for the absolute results of the Base Gating Study (%)

Arithmetic Mean	32.1
Median	32.5
Mode	N/A
Minimum	19.7
Maximum	51.3

Table 22 Measures of Spread for the absolute results of the Base Gating Study (%)

Range	31.6
25th Percentile	30.6
75th Percentile	33.9
Interquartile Range	3.3
Standard Deviation	5.7
CV	17.8
Median Absolute Deviation	1.9

Table 23 Measures of Skew for the absolute results of the Base Gating Study (%) (3dp for better resolution)

Skewness	0.492
Skewness standard Error	0.383
Skewness z-score	1.280
Kurtosis	3.271
Kurtosis Standard Error	0.750
Kurtosis z-score	4.560

Table 24 Measures of Normality for the absolute results of the Base Gating Study (%) (3dp for better resolution)

Shapiro-Wilk statistic	0.904
Significance	0.003
Normal/Non-parametric	Non-Parametric

Using descriptive statistics to give a general report on the size and shape of the data, the distribution approximates to a normal shape because the mean and median are very close together, as quoted in Table 21. However, the Shapiro-Wilk statistical test for normality (Table 24) significantly concludes the distribution is not normal. This non-parametric definition is most probably indicated by the spread of the distribution, rather than any specific location parameters. Measures of spread (Table 22) show that there is a wide range between the highest and lowest participant averages, but the IQR of the participant data fell within a boundary of 3.3%, which is 9.5 times smaller than the overall range.

The measures of skew tests further confirm the non-normality due to data extremes (Table 23). The actual skewness measure indicates the distribution is normal, but the kurtosis of the data set confirms that the distribution is heavily affected by 'outliers' in the data. These extreme values can be seen in Figure 40. Most participant averages lie close to the median, and the error bars show ± 1*SD* from each individual's repeated measures.

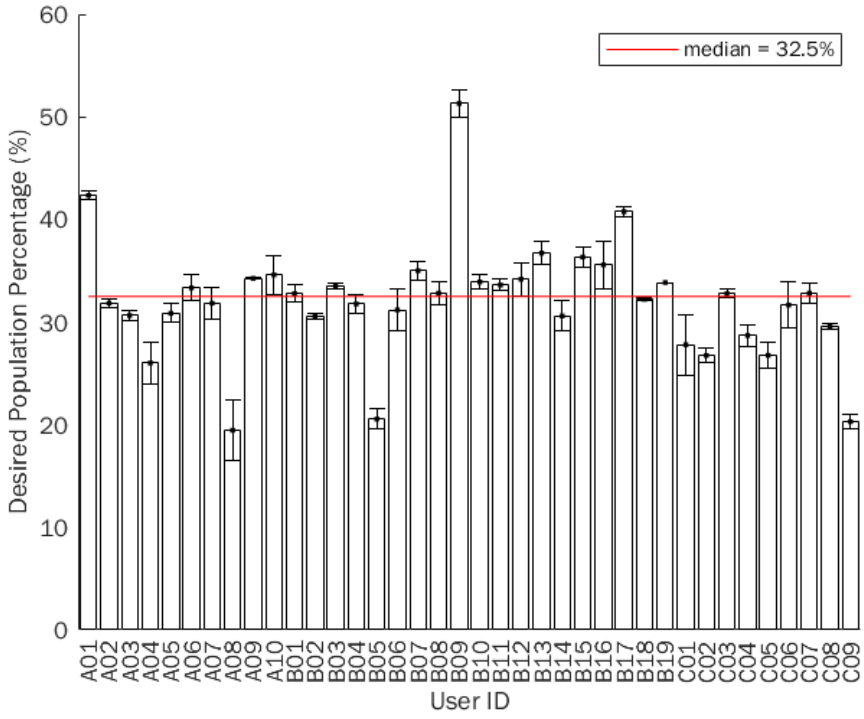


Figure 40 Absolute Results of Target Cell population, represented by each participant's average and ± SD.

Participant deviation from the median has been more clearly visualised in Figure 41, with bars depicting each participant's average from the median group value. The *SD* limits have also been plotted, because these are most commonly used within traditional manufacturing to define out of control/out of specification limits. 76 % of participants are within 1*SD* of the median, showing good corroboration of final results. Of those who fell out of bounds, two participants had results above + 1*SD*, one participant above + 2*SD*, one participant below - 1*SD* and three participants below - 2*SD*.

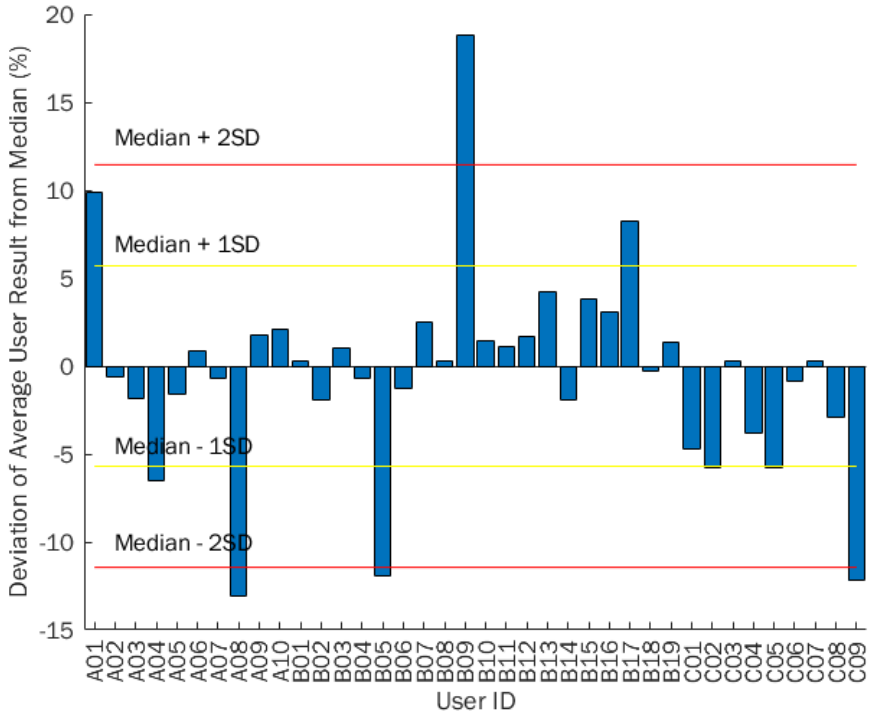


Figure 41 Participant average result deviations from overall group median.

The other control limits discussed in Chapter 3 have also been applied to this data, to further exemplify the difference between acceptance performance if different boundaries are specified by manufacturers. Figure 42 shows the control limits applied to this data set, cutting off different high and low extreme values, based on the mean and median of the data set. There is central tendency so there is no skewness of the data which could cause more outliers on one side of the distribution than the other.

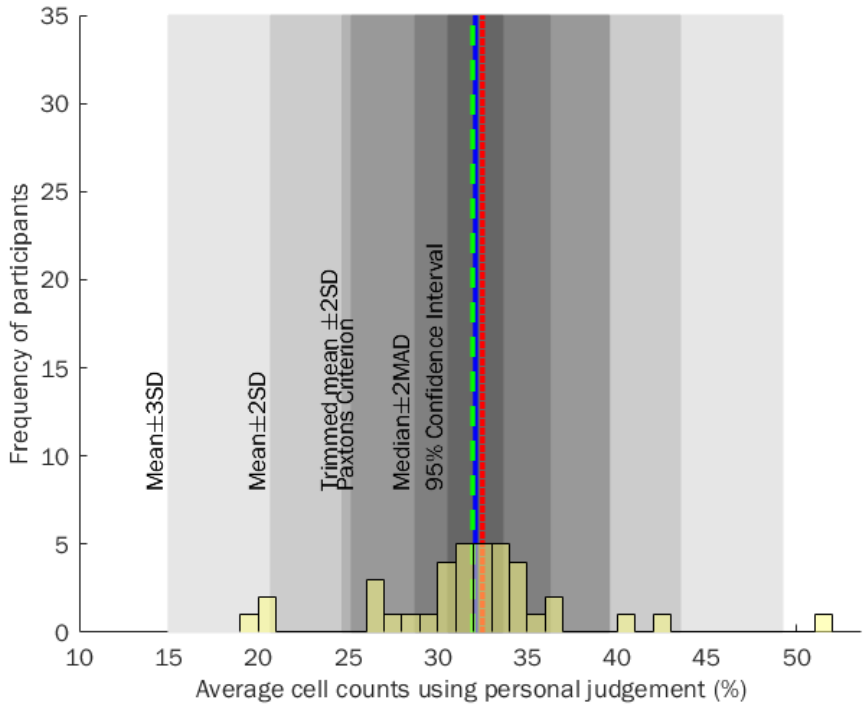


Figure 42 Acceptance boundaries applied to the Basic model averages to show difference in outlier classification.

To compare both positive and negative deviation data extremes, Figure 43 shows participant B08's gating strategy, one of the participants very close to the group median value. The three images show the gating sequence steps used to define the final population cell count, with B08's three repeats collated onto each sequence step image. This layout has been used for the remainder of qualitative participant analysis figures throughout this thesis. By comparing the extreme participants to a median participant, there are obvious differences between participants when identifying populations based on the visualised density. This is also coupled with personal preferences on inclusion or exclusion of data points to further refine the data set in search of a particular target.

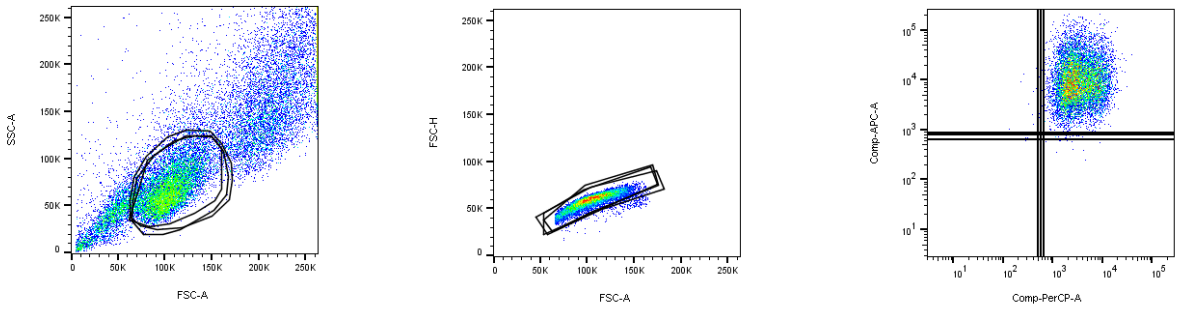


Figure 43 Participant B08 Gating Strategy interpretation, close to median result.

Focusing on those who fell outside of 2*SD*, participant B09 returned the highest population average (51.3%). Reviewing the participant's gating strategy has shown that most of bias is due to their initial conclusion of a larger cell population, consisting of doublets, which should have been excluded if following the defined sequence. This gating strategy can be seen in Figure 44, also showing a high variance in how this initial gate was applied.

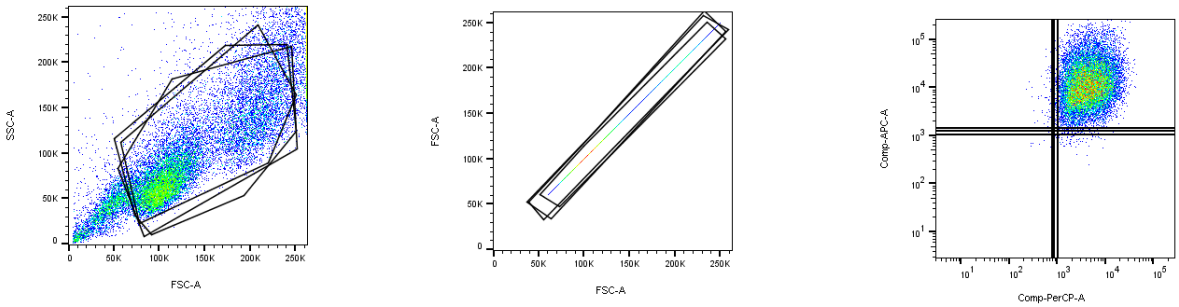


Figure 44 Participant B09 Gating Strategy interpretation

Participants A08, B05 and C09 also fell outside of this specification limit, with an average value significantly lower than the group median. Figure 45 to Figure 47 show their respective gating strategies, again noting that in each case, their first gate applied is situated within the target population, and therefore causing the low final result.

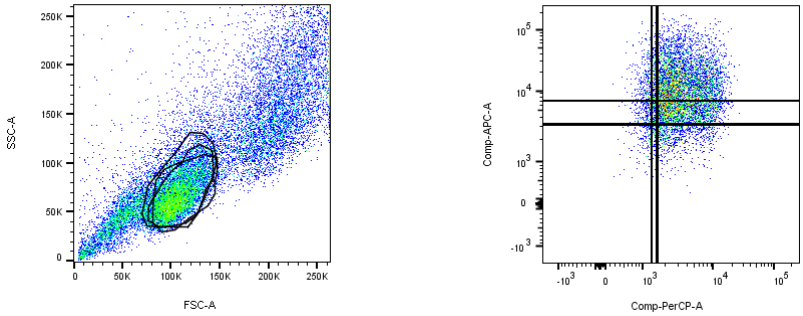


Figure 45 Participant A08 Gating Strategy interpretation

Participant A08 did not gate a single cell population because this was not something they have previously experienced in their own FC processes. Participant A08's final quadrant gates intersects the final population, because gates were applied against control files and not checked against the fully stained sample files. In this instance, poor gating 'clean-up' procedures and awareness of preliminary 'data-cleaning' processes have potentially caused this variance, alongside the participant perception of density.

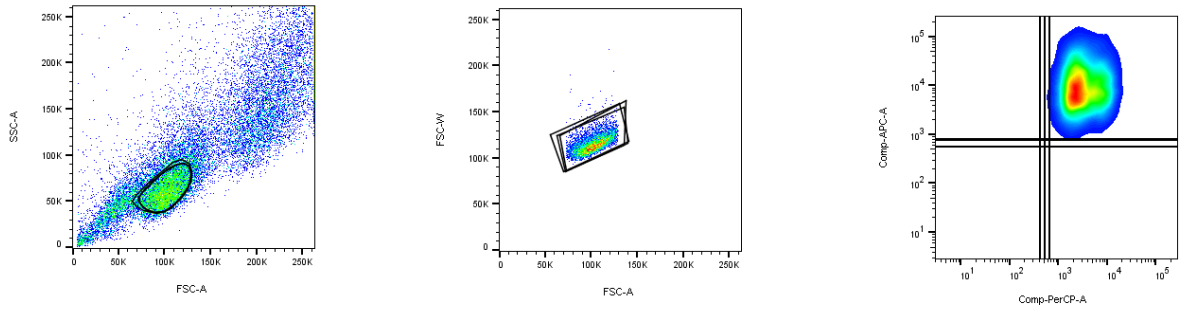


Figure 46 Participant B05 Gating Strategy interpretation

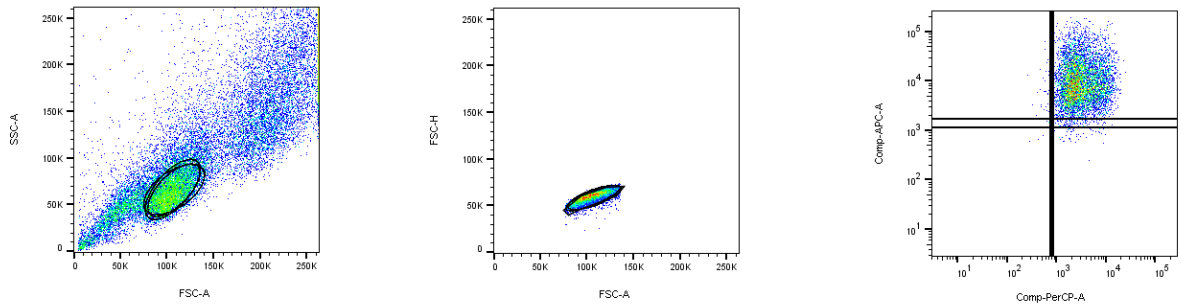


Figure 47 Participant C09 Gating Strategy interpretation

Participants B05 and C09 have applied their gates within the target cell population in the first instance. They all preferred to apply more refined gates to have better certainty of result to

removing any boundary outliers early on. These participants have all used different strategies to further clean the data to identify the single cell population. Participants were allowed to choose what axes they preferred to gate the single cell population, so even though each axis configuration does not have much noise, it will contribute a small amount of variation.

This analysis of the absolute results used to represent cell populations shows a 31.8 % cell count range between participants when determining final single pluripotent cell population percentages. The further qualitative analysis of the extremes identifies 4 participants who fall outside of initial control limits. These extreme participants account for over half of the cell population percentage range. If these 4 extreme participant values were removed, the range would fall to 16.3% (a difference in cell count of 4,890 cells) between participants (minimum value of 26.0% and maximum value of 42.3%).

Potential inclusion of more information of areas to gate would aid location of gate placement to improve general result conformity, although this may not necessarily improve intra-participant variability.

Variability of absolute results is commonly assessed using the *CV*, which combines the average and standard deviation of final cell count measurand. The distribution histogram of participant *CV* of reported results can be seen in Figure 48, plotted on top of 3 specification limits derived from the ICSH boundaries. The ICSH and ICCS have set imprecision criteria for Cell-based fluorescence assays as a desirable target of < 10 % *CV* or when dealing with rare-cell cases or minimal residual disease detection < 20 % *CV* is acceptable [71]. In this study 'Good Performance' is represented at 5 % *CV*, half of the ICSH guideline acceptance criteria. 'Satisfactory Performance' shows the amber region at 10 % *CV* and the 'Revision Required' limit is set to 20 % *CV*, because this basic model is not designed to detect rare cell events, it is designed to be easy for participants to complete with no questionable populations or events. Therefore, this amount of variance is not ideal in this

exemplar. These boundaries have been defined using the equation of a straight line in Equation 20, where *m* is the gradient of the line and *c* is the y-axis intercept.

$$y = mx + c$$

Equation 21 Equation of a straight line

The optimal scenario would have all 38 participants (*n*) with < 1 % *CV*, which sets the total height at *y*-intercept of the graph. Using the equation of a straight line, boundaries can be drawn from the optimal *CV* position, or *y*-intercept, to the uncertainty specification limits on the x-axis. A right-angle triangle shape shows progression towards the desired positive skew and optimal variation positions. Applying this to a traffic light style quality monitoring diagram would show performance levels within a facility and ideal variance which can be quickly and easily understood by all.

In a similar manner, this schematic has also been used for uncertainty of measurements throughout this thesis, because this is a measure of variability that combines *SD*s from the additional gating steps in the sequence, not just the final gate which is used to calculate *CV*.

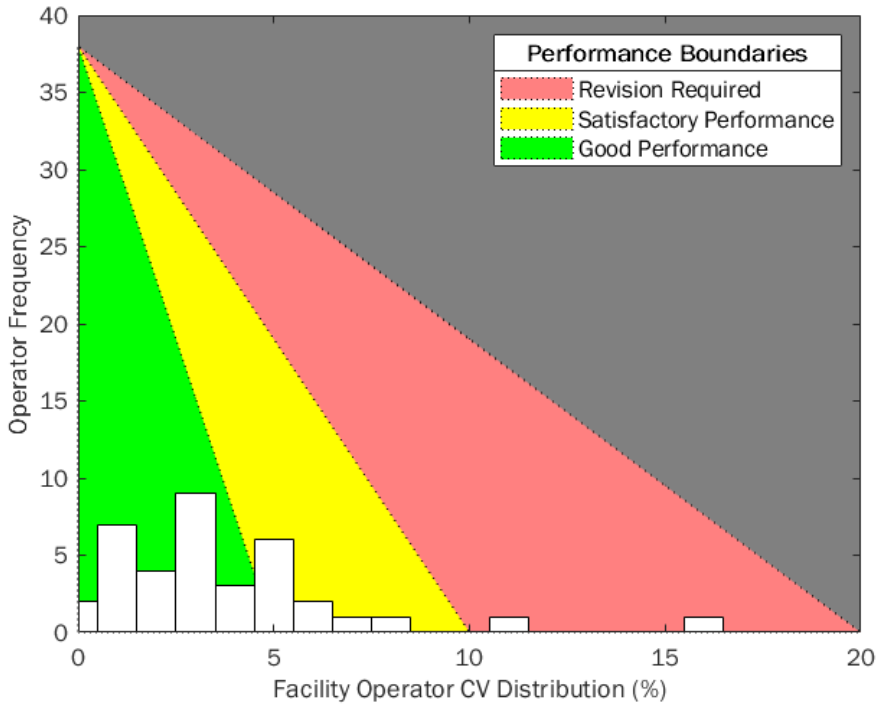


Figure 48 CV Performance of Participant Absolute results

Using these guidelines as boundaries, two participants fall outside of the 'good' and 'satisfactory' performance regions. These extreme outliers had more variation within the final quadrant gate they drew when identifying the final positive pluripotent cell population. Participant A08 had the highest *CV*, and their final gate can be seen already illustrated in Figure 45. A08 intersects the final population more than the rest of the participants and has one of the three repeated gates significantly higher than the rest, causing greater variation between the repeated measures.

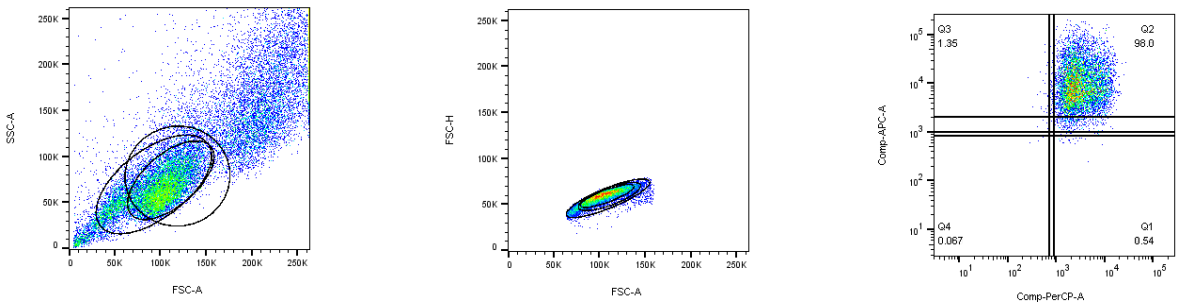


Figure 49 Participant C01 Base Model Study Gating interpretation

Participant C01 (Figure 49) also had a *CV* that fell within the 'Revision Required' region of the graph, and again the variation is caused in the final gate due to one of the repeats being placed significantly higher than the other two. It can be seen in the first two gates applied in participant C01's sequence that they are also very variable throughout the gating sequence, however, this is not captured within the representative population *CV* measurement, because it is only calculated from the final population cell counts derived using the whole sequence. Separate *CV* values could be calculated individually for each gate applied in the sequence. However, this can become cumbersome to use because the amount of data to analyse has tripled. In more complex sequences this would be even more time consuming.

Measurement uncertainty provides a way of combining variability measures (*SD*s) of each gate within the sequence, to provide a measurement that is more representative of the components of the gating sequence. When extremes in measurement uncertainty arise, uncertainty values can be easily deconstructed to identify which part of the gating sequence is responsible for causing variation within the measurement.

4.4.2 Flow Cytometry Basic Gating Exercise uncertainty results

The uncertainty results reported here are a combination of the three gating stages defined in Figure 33. The uncertainty values have been quantified following the prescribed methodology in Chapter 2 and Section 2.5. The uncertainty would better represent variance of measurements with greater confidence, because this combines variability from all gates applied in the sequence, not just the variance of the final gate applied.

Table 25 Measures of Location for Uncertainty of the Base Gating Study (%)

Arithmetic Mean	4.0
Median	3.6
Mode	N/A
Minimum	0.7
Maximum	13.1

Table 26 Measures of Spread for Uncertainty of the Base Gating Study (%)

Range	12.4
25th Percentile	2.0
75th Percentile	5.6
Interquartile Range	3.6
Standard Deviation	2.7
Median Absolute Deviation	2.0

Table 27 Measures of Skew for Uncertainty of the Base Gating Study (%) (3dp for resolution)

Skewness	1.288
Skewness standard Error	0.388
Skewness z-score	3.320
Kurtosis	2.311
Kurtosis Standard Error	0.759
Kurtosis z-score	3.045

Table 28 S-W test for normality for Uncertainty of the Base Gating Study (%) (3dp for resolution)

Shapiro-Wilk statistic	0.900
Significance	0.003
Normal/Non-parametric	Non-parametric

Similar to the descriptive statistics for absolute results for this study, the mean and the median are close together, indicating a normal distribution, as monitored in Table 25. The median is less than the mean, indicating a slight positive skew to the data. This is further supported by the Shapiro-Wilk test for normality, shown in Table 28, indicating that the distribution is non-parametric in shape, indicating skewness.

There is a wide range (12.4 %) between minimum and maximum participant uncertainties, which does not include Participant B19. Their Standard Deviation of zero (as seen in Figure 50), further showed that they had copied gates across the repetition workspaces, giving a comparable absolute result but no measure of precision, disqualifying them from this uncertainty analysis. Table 26 also shows the interquartile range as 3.6%, indicating a strong central tendency, because half of the data lies within 31% of the total distribution.

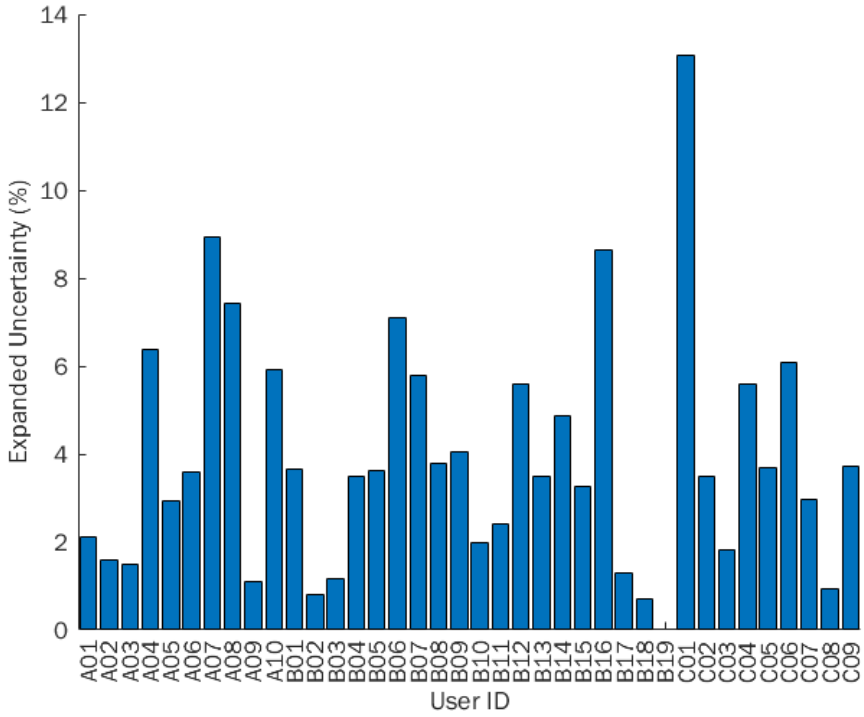


Figure 50 Expanded Uncertainty of all Participant Gating within the Base Model Study

The measures of skew tests (Table 27) further confirms this non-normality, because the skewness *z*-score also falls outside of the 2.58 bandwidth of normality (as described in Chapter 2, Section 2.2.2). The kurtosis *z*-score also falls positively outside of the 2.58 bandwidth (both skewness and kurtosis use the same scaling), although it is not as large as the skewness value, meaning the non-normality is more likely due to the shape of the data than extreme data points, although these can still have a significant impact on the final distribution statistics. This distribution shape can be observed within Figure 51, showing the positive skew with 2 or 3 larger uncertainty extremes.

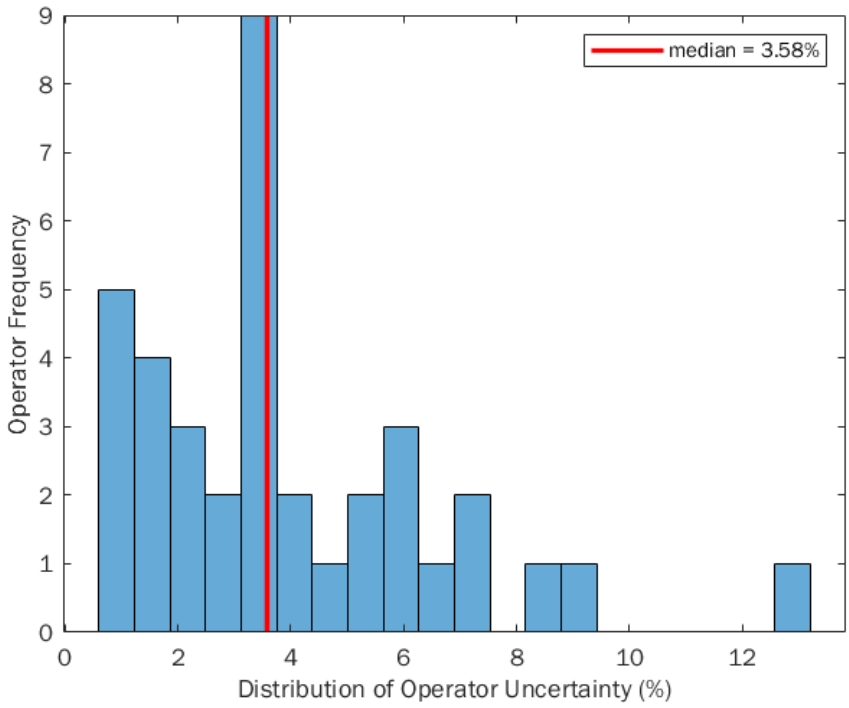


Figure 51 Histogram of Participant Uncertainty from repeats of Base Model Study

Whilst deviation from a median can help to explain the distribution parameters, when analysing variance, a positively skewed distribution towards zero is preferred. Comparing uncertainty in its size order allows boundaries to be set for permissible specification limits for product release/laboratory quality that increase in value.

The ICCH and ICS imprecision values described earlier for measurement *CV* have also been used here to define example specification limits if monitoring participant uncertainty. *CV* is a measure usually used to represent variability of a final metric because it considers both the mean and *SD* of a measurement. This only works for the final cell population count, whereas uncertainty combines in quadrature the *SD*s of all the gates applied in the sequence. In this instance the *CV* specification limits have been substituted for uncertainty, because there is a positive correlation (using a line of best fit) between the result *CV* per participant and their respective uncertainty for this cell model (Figure 52), and no other uncertainty specifications have been defined in the public body of knowledge from research or industry.

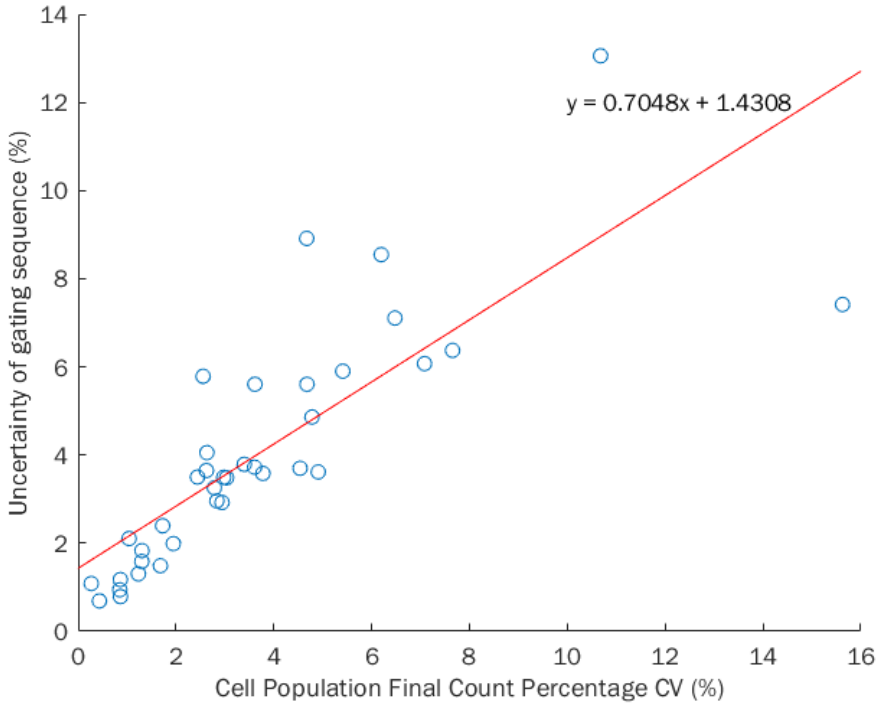


Figure 52 Final Cell Count Population Percentage versus Gating Uncertainty for participants

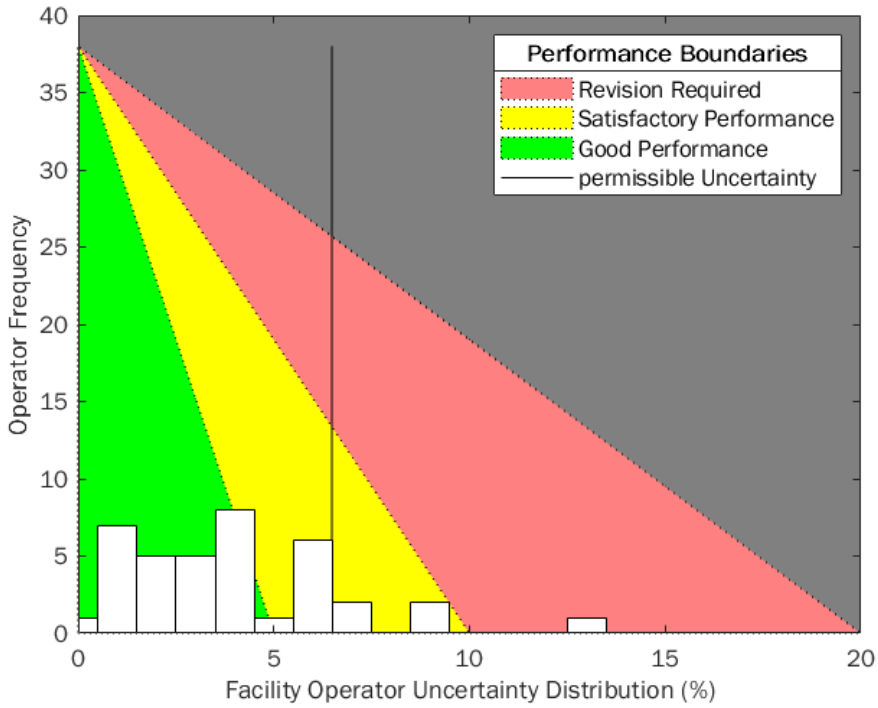


Figure 53 Participant performance monitoring diagram for uncertainty results of basic model

Figure 53 shows the base model uncertainty histogram from Figure 51, plotted on top of 3 specification limits derived from the ICSH boundaries. 'Good Performance' is represented at 5 %, half of the ICSH guideline acceptance criteria. 'Satisfactory Performance' shows the amber region at 10 % and the 'Revision Required' limit is set to 20 % once again. A permissible limit for uncertainty (*pU*) [181] has also been plotted as another variance discriminator, defined in Equation 2, assuming a 95 % Confidence Interval to the data.

$$pU = 2.39 \times SD$$

Equation 22 Permissible Uncertainty

Participant C01 (Figure 49) had an uncertainty that fell within the 'Revision Required' region, with participants A07 and B16 being at the higher variation end of 'Satisfactory Performance'. Figure 49, Figure 54 and Figure 55 show their respective gating strategies, compared with Participant B18 (Figure 56) who had the lowest uncertainty, exemplified as an ideal participant in this instance.

Participant C01's uncertainty can be attributed to the three gate phases drawn for target cells, single cells and the double positive pluripotent cells in a respective 40:33:27 % split. The *SD*s are the inter-participant group maximum for each gate applied in sequence. The target gate identifies the same population each time, but varies to capture smaller, possibly dying cells and varies again to make the ellipse more spherical. Consistency of gating shape could reduce this variability slightly, provided that the cell population boundary is well defined. Single cell gate variance is attributed to 2 of the 3 gates being consistently smaller in size, with the third gate capturing the majority of the population. C01 applied smaller gates on this window with each repeat, suggesting potential refinement (from memory) for each session repeat. The final gate had understandably smaller variance in comparison, being produced by one offset quadrant refining the density of the population.

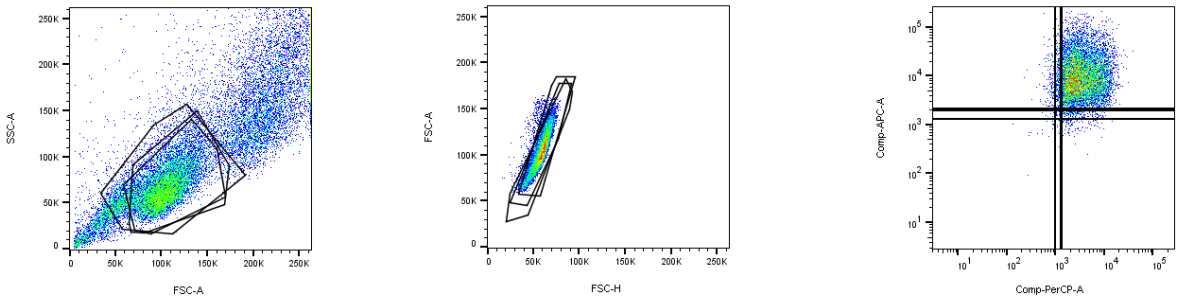


Figure 54 Participant A07 Base Model Study Gating interpretation

Participant A07's uncertainty is attributed to the target cells gate, single cells gate and double positive in the ratio of 44: 36: 20 % respectively. Most of the variance comes from the target cells, where differences can be seen in the gating at the boundary of dying cells, in a similar manner to Participant C01. The single cells gate is causing variance because of the gates being applied very closely to the corners of the dense cell population. A small fraction of the tip of the population has caused a significant difference in the number of cells captured in the gate each time. Finally, the quadrant gate has two repeatable gates applied, with one gate applied further out. Even though this captures more of the cell population this is still different from the other two gates drawn.

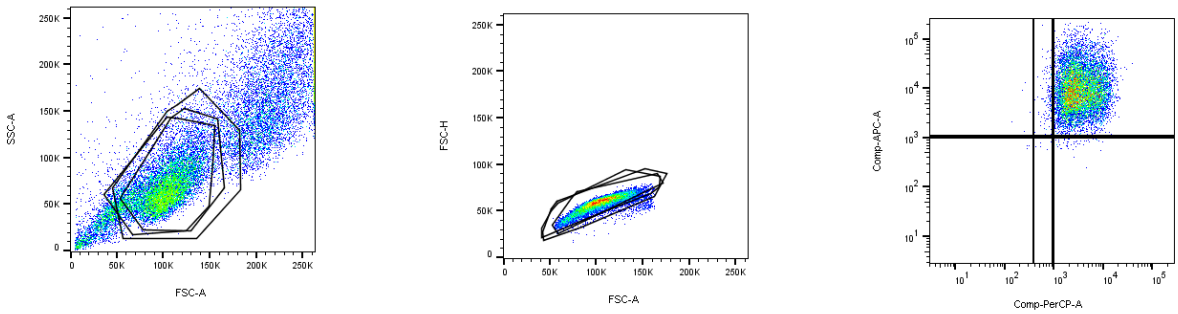


Figure 55 Participant B16 Base Model Study Gating interpretation

Both the target and single cell gates contribute 35% to B16's gating uncertainty. The target cell variance is likely due to varied inclusion of the smaller, possibly dying cells which caused variance for C01 and A07, because this is a region of gating difference overlaying more dense regions of data. The single cell variance is likely to be due to one gate applied which cuts off the top-right underside 'corner' of the data.

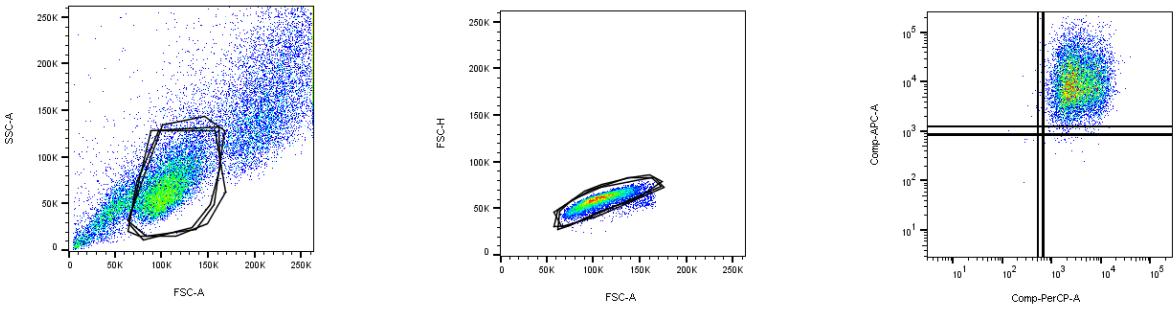


Figure 56 Participant B18 Base Model Study Gating interpretation

Participant B18 had the lowest uncertainty at 0.7 %, with their respective gating strategy shown in Figure 56. B18's initial target gate was repeatable in shape, as well as having minimal distance between gates at the cut-off for dead or dying cells, which has shown to cause variance for other participants. Again, their single cell gates were very similar in shape, and they captured the whole of the population without cutting any edges. Finally, the quadrant gates were very precise, having minimal effect because they do not intersect any high-density areas due to good use of control files provided. Control files provided were isotypes, Fluorescence Minus-One (FMO) controls and a negative sample, which were used to set the position of the gates using a positive/negative split on respective axes.

4.5 Chapter Conclusions

The EC 2102 Ep Cell line used was a good model for the baseline study because it has been shown to not differentiate over time, providing a good starting point for analysis due to simplicity of marker discrimination and very limited biological variation. Hence there are very limited/negligible upstream effects on downstream analysis. The studies run with the 3-workspace configuration for repeats worked well within the time available for participants so this structure should be used going forwards. The three-step process that each participant had to work through was also straight-forward to follow from the image, ensuring there was little deviance from the prescribed method.

Initial validation of plots and tools used has provided an interesting insight into how different layouts can affect subjectivity and therefore results. Greyscale plots are currently uncommon as

technology has progressed to allow more data dimensions to be shown at once, but the initial pilot has shown they should not be used over colour plots due to increased variance in analysis reproduction. A greater number of participants in this study would strengthen results and conclusions, but generally manual polygon tools provide more control over gate placement and exhibit a smaller difference in range across different visual scenarios.

The mean and median values (32.1 % and 32.5 % respectively) for the absolute cell count reported results are very similar, usually indicative of a normal distribution but upon further investigation the population is non-parametric, due to the distance of extreme values. These extremes appear to be attributed to variance in the first target cell gate applied. Due to the noise, debris and doublet populations present, more training should be given to Cytometrists when learning to gate, ensuring greater uniformity of results.

Knowledge of correct populations and decisions made on cell cluster boundaries could aid reproducibility between participants. Similarly, decisions on population density boundaries should be determined within a 'gating specification' possibly through the validation of population percentage and use of tools such as Gating-ML to determine gate shape and size parameters.

CV of measurements is a commonly used tool throughout Flow Cytometry to quickly monitor variation, however, this only accounts for repeat measures on the final gate of a sequence. International committees have defined acceptance criteria for Flow Cytometry *CV* measurements, which have been applied and adjusted to assess initial base model *CV* results here, although other variance criteria have been defined, as discussed in Chapter 1. Reviewing these extremes shows variability in the final gate applied, but it also highlights that other variability seen upstream of this gate is not taken into consideration, making measurement uncertainty a more suitable metric for accommodating variation throughout the whole gating sequence.

This base model with 3-step sequence has shown that calculating measurement uncertainty is possible for participants by using traditional measurement uncertainty methods. This was calculated successfully, by presenting participants with three repeated workspaces of data, informed by Gauge R&R testing layouts and extracting one repeated file located in each workspace randomisation. The Gauge R&R randomisation layout was useful to test but it caused some participants to exceed the 1-hour time limit due to the number of files used around the main analysis file. It also caused slightly more difficulty in traceability of the repeated file across each workspace. In subsequent Chapters this complexity of study design and randomisation will not be used, in order to have greater control over the data analysis variables which could affect the overall uncertainty contributed from participants.

The uncertainty was also skewed with an absolute median of 3.6 %, although it is preferable to have as low an uncertainty as possible to reduce measurement variation passed on to the final reported result. The overall range of participant uncertainty was 12.37 %, which is outside of the 'satisfactory performance' boundary assumed from the ICCH data, so revision of more extreme participants has been completed, showing that when data is more clustered and noisy at the start of the analysis sequence, this can have a knock on effect to their uncertainty contribution.

Acceptance limits determined from the ICSH have enabled this uncertainty data to be compared and put into manufacturing quality control context using the diagram generated in Figure 53. Simple traffic-light visuals have been proposed here and help show the quality status of a product/service/facility quickly and can be easily updated and adapted to suit the number of participants/participants and continuous improvement quality levels. A better knowledge of uncertainty visualised in this way potentially improves quality by addressing gating that falls outside of satisfactory boundaries and a good working culture that does not attribute blame can help a facility grow to tighten overall measurement uncertainties.

4.5.1 Consolidation of Objectives

- This study ran smoothly, acting as a good baseline for comparison of absolute reported results, *CV* and uncertainty measures. The session structures were suitable in time, and 3 repeats was suitable for participants to understand study context, but not become tired.

- A pilot study showed that different plot visualisations and gating tools contribute variance to a measurement, albeit a small amount. Grayscale plots are not recommended for use as they increase inter-participant variance, and they do not include as much information about density on the plot.

- Extreme values in absolute reported results were due to participants either over constraining or under constraining the initial target cell population within the first gate. In some cases, lack of knowledge of using controls to set gates led to variance in final pluripotent population metrics.
- Extreme values in uncertainty results were due to participant variability in applying a gate to separate the target cell population from the dead or dying cells.

- The performance monitoring diagrams visualised in Figure 48 and Figure 53 provides a straight forward way to monitor uncertainty performance with respect to the number of people in the study and defined quality satisfaction limits. These will be used in subsequent chapters to monitor uncertainty performance in more complex gating scenarios.

- This study defines participant uncertainty for a highly constrained, very stable 2 colour panel cell model, which can be used as a baseline for development into more complex cell models, to monitor potential growth of between participant uncertainty in more difficult analysis scenarios.

Chapter 5: Intermediate Uncertainty Model

5.0 Introduction to the Chapter

Chapter 5 introduces the second uncertainty gating exercise (intermediate model), as part of the sequence of studies which monitor participant variance in comparison to complexity of data. As discussed in the prelude, this second uncertainty exercise is intermediate, to further monitor uncertainty in a more complex analysis scenario. This is more representative of FC analysis and T-cell lineage markers used to monitor cell therapy products. This continues to build the pathway for further uncertainty gating studies described in Chapter 6 with comparisons drawn between all three models in Chapter 7. Previously in Chapter 4, only three gating steps had been considered with an immortalised cell line (Basic model), whereas this second uncertainty stage tackles more complex FC data analysis using primary Peripheral Blood Mononuclear Cells (PBMCs) which are used as a basis for many autologous engineering cell therapies. This becomes more relevant for the CGT manufacturing community, therefore potentially providing a more translational application of uncertainty to FC measurements. This Chapter uses a five-step analysis sequence, and similar to Chapter 3, sees participants analyse this data across two sessions using their own judgement and then using a diagrammatical protocol respectively.

5.1 Chapter Aims

This Chapter develops comparison of uncertainty in more complex FC analysis strategies. The fit of this Chapter to the thesis can be seen in Figure 57, specifically within the orange dashed box, providing development for the core hypothesis: as complexity of FC data and processing increases, measurement uncertainty contributed from the participant will also increase.

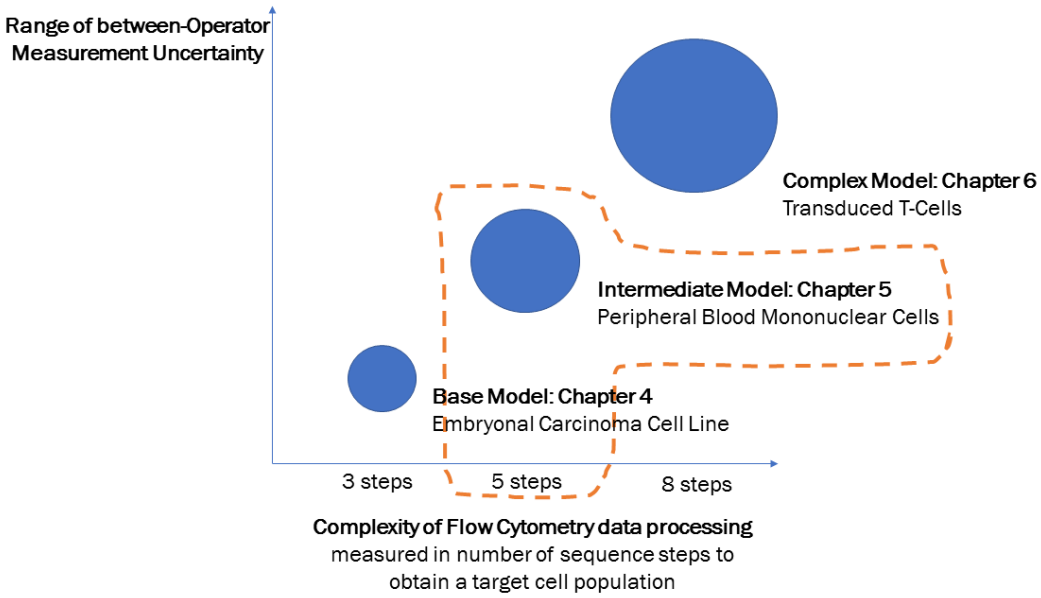


Figure 57 Diagram of Chapter position within the whole Thesis

5.1.1 Chapter Aims & Objectives

The Aims and Objectives of this Chapter can be defined as follows:

- Design an intermediate two-phase study structure to measure *CV* and uncertainty of participants when gating an intermediate complexity set of FC data.
- Identify whether using diagrammatical protocols to apply gates reduces between participant absolute reported results and uncertainties.
- Identify what causes extreme values in absolute reported results and uncertainty measurements by reviewing extremes from this intermediate study.
- Further test monitoring tools for understanding uncertainty within a more complex participant study or facility.

5.2 Methodology

Primary Peripheral Blood Mononuclear Cells (PBMCs) were acquired from LGC-ATCC cell banks for use within this phase of research (Cat Number: PCS-800-011, Lot number: 80628171).

5.2.1 Cell Selection and Culture

The vial of PBMC material acquired from LGC-ATCC cell banks was primary biological material, meaning this has been taken straight from the donor and separated to remove additional whole blood cell types. It has not been expanded in vitro before cryopreservation, so this material was handled in accordance to the Human Tissue Act (HTA) regulations [182]. Before acquiring this material, the Certificate of Analysis for the PBMC material was obtained, along with confirmation of donor consent from the supplier for this material to be used for research purposes. These documents can be found in Appendix C. Human Tissue Act training was also completed within the Centre for Biological Engineering, Loughborough University and through the Medical Research Council (MRC) online Research and Human Tissues legislation (Certificates in Appendix C), to ensure the correct handling, disposal and ethical treatment of the material was adhered to.

A vial of ATCC PBMCs (25 x 10^6 cells, suspended in 1 mL of cryoprotective fluid) was removed from a liquid nitrogen Cryobank and was thawed in a water bath for 3 minutes until a slither of frozen material remained. The material was topped up 1:1 with RPMI 1640 cell culture media (Cat Number 11875093, Lot Number: 1906058) (fortified with 10 % v/v Fetal Bovine Serum (FBS)), mixed slowly and transferred into a centrifuge tube. 1 mL of Flow Cytometry buffer fluid (Biolegend Cell Staining Buffer Cat Number 420201, Lot Number B228788) was used to wash the inside of the cryovial to remove any additional cells which remained after transfer. This 1 mL flow buffer suspension was also added to the centrifuge tube. The centrifuge tube was then topped up to 25 mL with an additional 22 mL of RPMI 1640 cell culture media, and a 250 µL sample was taken for an initial cell count. Three counts were completed using a Nucleo-Counter® NC-3000™ and ChemoMetec

Via1-Cassettes™, to stain and measure cells with Acridine Orange and DAPI dyes, calculating an average of 0.8 x 10^6 cells / mL, totalling approximately 20 x 10^6 cells within the resuspension.

The remaining cell suspension was centrifuged at 300 *g* for 5 minutes, transferred into a Biological Safety Cabinet (BSC), and the supernatant removed from the cell pellet. 29 mL of RPMI 1640 fortified media was added to a T75 flask, and the cell pellet was resuspended and slowly mixed with 1 mL of cell culture media, before being transferred into the T75 flask. This T75 flask was moved into a humidified incubator at 37 ˚C with 5% CO_2, for 24 hours to allow the cells to proliferate. To comply with HTA good practice, a sign was placed on the incubator to notify other users of the contents, and the material location was updated on the biological material database, Procuro.

After 24 hours had elapsed, cells were counted by taking a 250 µL sample from the mixed cell suspension. An average of 0.83 x 10^6 cells / mL was measured from the sample, totalling approximately 24.7 x 10^6 cells within the total resuspension, at an average viability of 94.8 %. The increase in cell count after 24 hours shows a successfully maintained cell population after being thawed, so the cells could be used for further analysis. In this instance, further cell culture or monitoring of certain cellular features was unnecessary because only 'snapshot' fcs files of the PBMC material were required. PBMCs are suspension cells, so no disassociation process was required to remove the cells from the surface of the tissue culture flask. The cell suspension was transferred into a 50 mL centrifuge tube and centrifuged for 5 minutes at 300 g to form a cell pellet.

5.2.2 .fcs File Generation

A series of fcs files were generated using the primary PBMCs kept in culture for the last 24 hours. Cells were resuspended in 2.5 mL of Cell Staining buffer and gently mixed, then recentrifuged to form a pellet and resuspended in 2.5 mL of Cell Staining buffer to wash the cells and remove any cell culture media remaining. 0.1 mL aliquots of the master cell suspension were placed into separate labelled microcentrifuge tubes so there were approximately 1 x 10^6 cells per tube (three

fully stained samples, one unstained sample, one live/dead stained sample, five single stain controls, five isotype controls, five FMO controls). The unstained sample was wrapped in foil and placed in a 4˚C fridge because this was not needed until the final analysis.

1 µL of Biolegend Zombie Aqua Viability dye (Cat Number: 423101, Lot Number: B243783) was added to all tubes and gently mixed, except for the unstained sample. The Eppendorf tubes were covered in foil to minimise light exposure and left to incubate for 20 minutes. Amine-reactive dyes or Live/Dead fixable dead cell stains cross the cell membrane of dead cells and react with free amines in the cytoplasm. Live cells exclude these dyes when they are intact so free stain can be washed away after staining, allowing for discrimination of live and dead cells in the population [183]. Amine-reactive dyes were used because they could be compensated for within the Flow Cytometry panel using amine-reactive compensation beads. Once incubated, all cells were washed twice with 1 mL Cell Staining Buffer, centrifuged (300 g for 5 minutes) and supernatant removed. Cells were then resuspended in 100 µL Cell Staining Buffer.

5 µL of Biolegend Human TruStain FcX™ F_c blocker (Cat Number: 422301, Lot Number: B235079) was added to all cell samples aside from the isotype controls and incubated covered in foil for a further 15 minutes. Human Fc receptors are expressed on a variety of cells and cells with Fc receptor expression can sometimes give false positives or false negative results within immunofluorescent staining because of Fc receptor non-specific binding. Human TruStain FcX™ is a blend of specialised IgG immunoglobulins that join to Fc receptors to stop non-specific binding occurring. Once incubated, all cells were washed twice with 1 mL Cell Staining Buffer, centrifuged (300 g for 5 minutes) and supernatant removed. Cells were then resuspended in 100 µL Cell Staining Buffer.

The live/dead sample then kept alongside the unstained sample in the fridge and was removed when Flow Cytometry analysis was undertaken. The remaining cells were then stained according to the following stain protocols in Table 29, Table 30, Table 31 and Table 32. Isotype controls for each

marker were stained with the respective antibody isotype control, single stain controls were stained with just the antibody marker for that specific stain and the FMO controls were stained with all stains aside from the stain aligned to that specific channel. This is to monitor any fluorescence spillover into the required channels from other markers being used.

The antigen markers used for the Single Stain controls (Table 29), FMO controls (Table 30) and Fully Stained Samples (Table 32) are Biolegend FITC anti-human CD3 antibody (Cat Number: 300306, Lot Number: B218086), Biolegend APC anti-human CD4 antibody (Cat Number: 357405, Lot Number: B223335), APC/Cy7 anti-human CD8 antibody (Cat Number: 300926, Lot Number: B231191), Biolegend PE anti-human CD45RA antibody (Cat Number: 362552, Lot Number: B210221) and Biolegend BV421 anti-human CD56 antibody (Cat Number: 423101, Lot Number: B246952).

Table 29 Staining volumes for Single Stain Controls

Channel	Antigen Marker	FITC Single Stain	APC Single Stain	APC/Cy7 Single Stain	PE Single Stain	BV421 Single Stain
633 nm laser, 660/20 filter	CD3 FITC	5 µL				
488 nm laser, 530/30 filter	CD4 APC		5 µL			
633 nm laser, 780/60 filter	CD8a APC/Cy7			5 µL		
488 nm laser, 585/42 filter	CD45RA PE				5 µL	
405 nm laser, 450/50 filter	CD56 BV421					5 µL

Table 30 Staining volumes for FMO Controls

Channel	Antigen Marker	FITC FMO	APC FMO	APC/Cy7 FMO	PE FMO	BV421 FMO
633 nm laser, 660/20 filter	CD3 FITC		5 µL	5 µL	5 µL	5 µL
488 nm laser, 530/30 filter	CD4 APC	5 µL		5 µL	5 µL	5 µL
633 nm laser, 780/60 filter	CD8a APC/Cy7	5 µL	5 µL		5 µL	5 µL
488 nm laser, 585/42 filter	CD45RA PE	5 µL	5 µL	5 µL		5 µL
405 nm laser, 450/50 filter	CD56 BV421	5 µL	5 µL	5 µL	5 µL	

The antigen markers used for the Isotype controls (Table 31) are Biolegend FITC Mouse IgG2a κ Isotype Control antibody (Cat Number: 400207, Lot Number: B235551), Biolegend APC Mouse IgG2b κ Isotype Control antibody (Cat Number: 400329), APC/Cy7 Mouse IgG1 κ Isotype Control antibody (Cat Number: 400127, Lot Number: B235070), PE Mouse IgG2b κ Isotype Control antibody (Cat Number: 400313, Lot Number: B246304) and Brilliant Violet 421 Mouse IgG1 κ Isotype Control antibody (Cat Number: 400157, Lot Number: B237449).

Table 31 Staining volumes for Isotype Controls

Channel	Antigen Marker	FITC Isotype	APC Isotype	APC Fire 750 Isotype	PE Isotype	BV421 Isotype
633 nm laser, 660/20 filter	CD3 FITC	5 µL				
488 nm laser, 530/30 filter	CD4 APC		5 µL			
633 nm laser, 780/60 filter	CD8a APC/Cy7			5 µL		
488 nm laser, 585/42 filter	CD45RA PE				5 µL	
405 nm laser, 450/50 filter	CD56 BV421					5 µL

Table 32 Staining volumes for Fully Stained Samples

Channel	Antigen Marker	Fully Stained Sample 1	Fully Stained Sample 2	Fully Stained Sample 3
633 nm laser, 660/20 filter	CD3 FITC	5 µL	5 µL	5 µL
488 nm laser, 530/30 filter	CD4 APC	5 µL	5 µL	5 µL
633 nm laser, 780/60 filter	CD8a APC/Cy7	5 µL	5 µL	5 µL
488 nm laser, 585/42 filter	CD45RA PE	5 µL	5 µL	5 µL
405 nm laser, 450/50 filter	CD56 BV421	5 µL	5 µL	5 µL

CD3 (Cluster of differentiation 3) is a T-cell co-receptor that helps to identify lymphocyte subsets and activate cytotoxic and helper T-cells. It transduces the activating signals to the cytoplasm of the T-cell [184]. CD4 (Cluster of differentiation 4) is an extracellular protein marker found on the surface of immune cells, specifically T helper cells [185,186]. This white blood cell subset signals to other immune cells to destroy foreign bodies found. If patients have low CD4+ counts, they are susceptible to lots of infections which can become difficult for the immune system to fight. CD8 (Cluster of differentiation 8) is a cell surface protein on cytotoxic T cells and also natural killer cells. These cells are able to kill virus-infected cells, cancer cells and can use cytokines to recruit other cells when mounting an immune response [187,188]. CD4:CD8 ratios are often used to measure the balance of the immune system [189].

CD45RA (Cluster of differentiation 45RA) is used to identify naïve T-cell subsets and is often compared with CD45RO which is used to monitor memory T-cells, because a cell cannot express both markers and this can be used to understand the population split of T-cells [190]. Finally, CD56 (Cluster of differentiation 56) is used to identify natural killer cells and is a marker for cytotoxicity [191].

These antibody quantities for the single stain controls, FMO controls and Isotype controls were added according to the quantities listed and incubated in the dark at 4˚C for 30 minutes. Once incubated, the cells were washed twice and transferred to BD Falcon™ Round Bottom 12 x 75 mm

tubes (Cat 352063) and kept covered to minimise light exposure. Cells were run through a BD FACSCanto™ II Flow Cytometer, using the respective fluorescence channel and voltage: FSC 310 V, SSC 400 V, FITC 389 V, APC 420 V, APC/Cy7 472 V, PE 350 V, BV421 300 V and BV510 451 V for the live-dead stain, once a daily calibration was completed using CS&T beads (Lot: 74538, Successful calibration).

Each tube and respective fcs file were generated using a medium flow rate (60 µL/min) and by acquiring 30,000 cellular events. 3 stained sample fcs files were generated to build a library of repeats to use within the variation studies, alongside the control files listed. Files were exported as fcs 3.0 version types for use in Flowjo Version 10.0.8r1 third party analysis software [123] and saved as a workspace.

5.2.3 Flow Cytometry Study Organisation

A total of 23 Participants from three separate centres (5 from an academic institution, 13 and 5 participants from separate industrial institutions) were invited to complete the study in a quiet analysis space, to avoid distraction and the possibility of others seeing the study content and analysis. As in the previous uncertainty exercise, participants made 3 repeat measures within a 1-hour slot, because this was the maximum allowance given for each participant's time, agreed across the three institutions.

Study sessions had a one-hour maximum duration, and participants were shown three Flowjo workspaces, which contained a series of fully stained PBMC .fcs files. Identical files were included in each workspace, and participants were instructed to gate through a five-plot sequence to identify target cells (using Forward Scatter (FSC) plot against Side Scatter (SSC)), single cells, live cells, CD3+ cells and finally to apply a quadrant gate to the double positive naïve T-cell CD4+ CD45RA+ population to identify final positive population cell counts. Flowjo was the choice of platform due to access of the software across all three collaborator and participant sites, meaning a higher number of participants were likely to be familiar with the platform.

Although an additional marker for CD56 was acquired, this was not used within the final analysis sessions for participants to identify natural killer subsets. To identify an additional population meant sessions would run over the allotted 1-hour time slot. The CD4+ CD45RA+ pipeline was kept, because of the suitability of this panel to current engineered T-cell product panels and to provide a good basis to increase complexity for the subsequent studies (Chapter 6, complex model) to more representative CGT T-cell product analysis.

Participants were also provided with isotype controls and FMO controls in each workspace to aid gate application and were allowed to use whatever manual gating tool on Flowjo they felt best to gate the population in hand. An overall schematic of the gating sequence they were asked to follow is shown in Figure 58, and participants gated each workspace of files separately to ensure a correct quantification of uncertainty through standard deviation calculation in accordance with principles described earlier in Chapter 2.

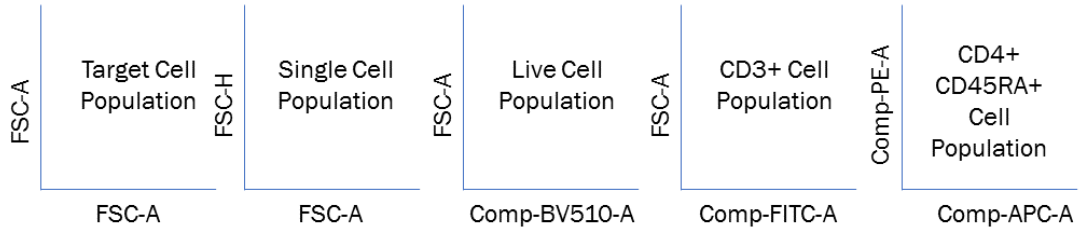


Figure 58 Gating sequence participants were asked to follow, to identify the target single live cell population, with CD3+ CD4+ CD45RA+ for naïve T-cells.

In a similar manner to the pre-study completed in Chapter 3, participants took part in a second phase, where they repeated the same gating process for the CD3+ CD4+ CD45RA+ cell population but were asked to copy a diagrammatical protocol to apply gates instead of using their own judgement, shown in Figure 59. To remove additional variability when placing these gates, participants were only given the three fully stained samples to apply the gates in each workspace,

so no control files could influence gate placement once the images had been copied, and participants followed the same gating sequence provided in Figure 58.

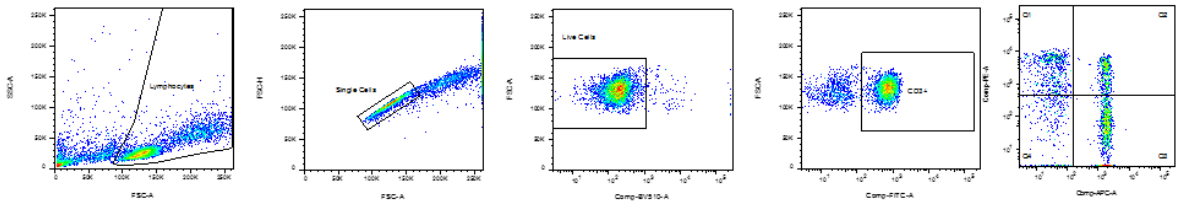


Figure 59 Diagrammatical protocol given to participants to copy gates instead of using personal judgement

All participants and their respective data were anonymised at the point of data collection, and data stored in accordance to the ethical clearance obtained. Participant coding was restructured from previous work to remove the possibility of analysis bias.

5.2.4 Uncertainty Calculation

Once studies had been completed, target cell, single cell, live cell, CD3+ and CD4+ CD45RA+ cell population metrics were extracted from the data, using the results from the identical repeated file situated in each Flowjo workspace. These were then used to calculate a mean cell count, *SD* and coefficient of variation (*CV*) for each gating stage, per participant using Microsoft Excel software (Office 16). Finally, a combined uncertainty (u_c) was calculated by combining these Type A uncertainties by summation in quadrature. The u_c value was expanded with a coverage factor of k = 2, representing a 95 % Confidence Interval for the uncertainty statement, which gave each participant a representative expanded uncertainty (*U*) figure, to show individual variance. The mathematical methodology used to calculate uncertainty metrics has been previously discussed in Section 2.5. An example of the data extraction through to calculation of metrics and uncertainty can be seen in Figure 60 for this intermediate model.

Total Cells
30000

	Workspace 1	Workspace 2	Workspace 3
Target Lymphocytes	16517	16754	16599
Single Cells	16384	16494	16390
Live Cells	14985	14779	14746
CD3+	10675	10545	10475
CD4+ CD45RA+	2072	2088	2074

Coverage factor	2

Percentages

	Workspace 1	Workspace 2	Workspace 3	AVERAGE	STDEV	CV
Target Lymphocytes	55.06	55.85	55.33	55.41	0.40	0.72
Single Cells	54.61	54.98	54.63	54.74	0.21	0.38
Live Cells	49.95	49.26	49.15	49.46	0.43	0.87
CD3+	35.58	35.15	34.92	35.22	0.34	0.96
CD4+ CD45RA+	6.91	6.96	6.91	6.93	0.03	0.42

CD4+CD45RA+		
uc		U
0.71		1.42

Stage 2

Absolute Values

Total Cells
30000

	Workspace 1	Workspace 2	Workspace 3
Target Lymphocytes	18936	18734	21639
Single Cells	14711	14685	14689
Live Cells	14255	14135	14270
CD3+	10375	10238	10366
CD4+ CD45RA+	1800	1832	1823

Coverage factor	2

Percentages

	Workspace 1	Workspace 2	Workspace 3	AVERAGE	STDEV	CV
Target Lymphocytes	63.12	62.45	72.13	65.90	5.41	8.20
Single Cells	49.04	48.95	48.96	48.98	0.05	0.10
Live Cells	47.52	47.12	47.57	47.40	0.25	0.52
CD3+	34.58	34.13	34.55	34.42	0.26	0.74
CD4+ CD45RA+	6.00	6.11	6.08	6.06	0.06	0.91

CD4+CD45RA+		
uc		U
5.42		10.84

Figure 60 Example of data extraction through to calculation of absolute results and uncertainty

5.3 Results & Discussion

5.3.1 Flow Cytometry Intermediate Gating Exercise absolute results – Phase 1 Personal judgement

The absolute results reported here are the targeted cell population that participants were asked to identify using the gating sequence defined in Figure 58, during the first gating session where they used their own judgement to apply gates. These are akin to what would be reported in literature for specific cell types, in this instance it is naïve T-cells. The uncertainty of the gating sequence will be discussed in subsequent sections of this Chapter.

Table 33 Measures of Location for the absolute results of the Intermediate Gating Study using personal judgement (%)

Arithmetic Mean	6.28
Median	6.01
Mode	N/A
Minimum	3.46
Maximum	7.99

Table 34 Measures of Spread for the absolute results of the Intermediate Gating Study using personal judgement (%)

Range	4.53
25th Percentile	5.65
75th Percentile	7.27
Interquartile Range	1.62
Standard Deviation	1.14
CV	18.15
Median Absolute Deviation	6.01

Table 35 Measures of Skew for the absolute results of the Intermediate Gating Study using personal judgement (%) (3dp for better resolution)

Skewness	-0.351
Skewness standard Error	0.481
Skewness z-score	-0.730
Kurtosis	0.042
Kurtosis Standard Error	0.935
Kurtosis z-score	0.045

Table 36 Measures of Normality for the absolute results of the Intermediate Gating Study using personal judgement (%) (3dp for better resolution)

Shapiro-Wilk statistic	0.945
Significance	0.231
Normal/Non-parametric	Normal

Using descriptive statistics to give a general report on the size and shape of the data, the distribution approximates to a normal shape because the mean and median are very close together, as quoted in Table 33. This is supported by the skewness and kurtosis *z*-scores (Table 35) and the Shapiro-Wilk statistical test for normality (Table 36) significantly concludes the distribution is normal. This normality definition is most probably indicated by the spread of the distribution, rather than any specific location parameters. Measures of spread (Table 34) show that there is a small range (in comparison to the Base Model Study) between the highest and lowest participant averages and the IQR of the participant data was just over half the size of the range, again indicating normality. However, because this model is based upon a more specific sub-population the ranges may not be as small as initially perceived, they may just be relative to the respective cell population.

These participant average cell count values can be seen in Figure 61. With most participant averages lying close to the median, and the error bars show ± 1*SD* from each participant's repeated measures.

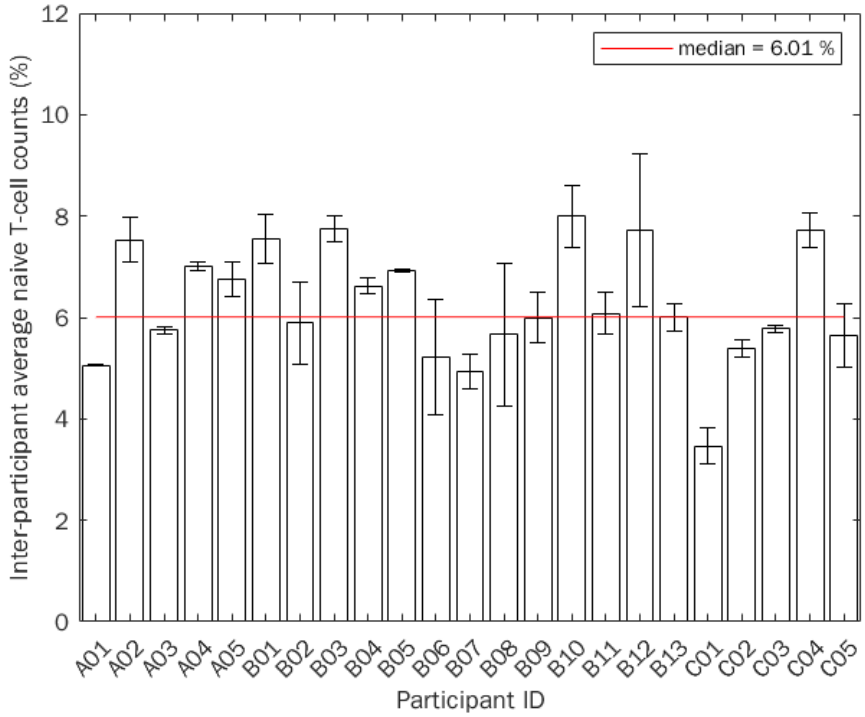


Figure 61 Absolute Results of Target Cell population, represented by each participant's average and ± SD.

Participant deviation from the median has been more clearly visualised in Figure 62, with bars depicting each participant's average from the median group value, or residual value (calculated by subtracting participant averages from the group median). The *SD* of the total group has also been plotted, because these are most commonly used within traditional manufacturing boundaries to define out of control/out of specification limits. 70 % of participants are within 1*SD* of the median, showing good corroboration of final results. Of those who fell out of bounds one participant had a result above +2*SD* and six participants below -1*SD*.

Figure 63 further shows this variability around the central location metrics to compare different outlier limits specified in Chapter 3, and further applied to the data in Chapter 4. The skew to the data shows how different boundaries would affect determination of outliers, however the data did fall within the most extreme boundary specified, which could indicate there is no real extreme in skewness or kurtosis of the data. This all depends on the acceptance criteria chosen for manufacturing distributions.

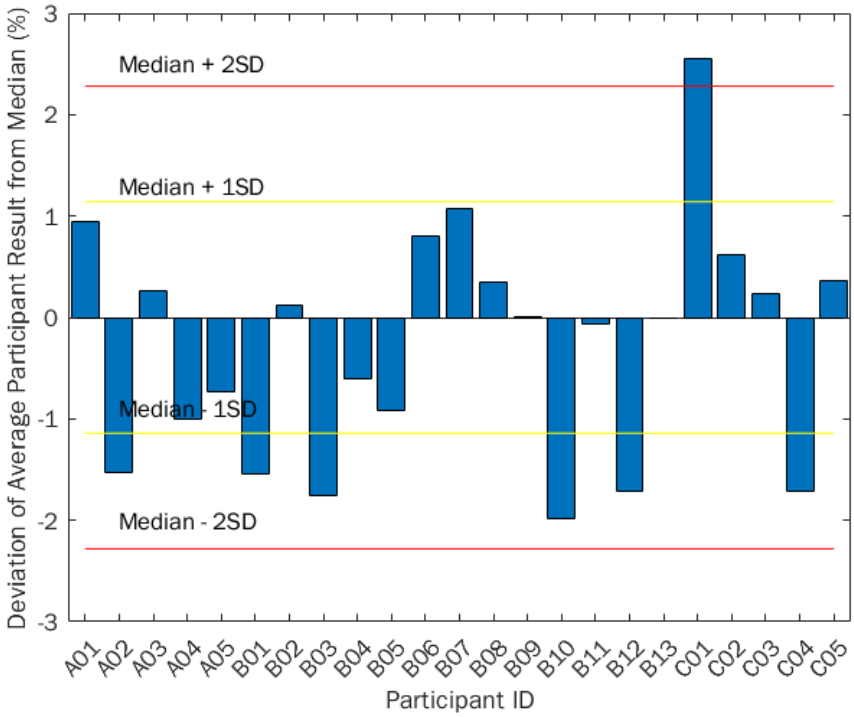


Figure 62 Participant average result deviations from overall group median.

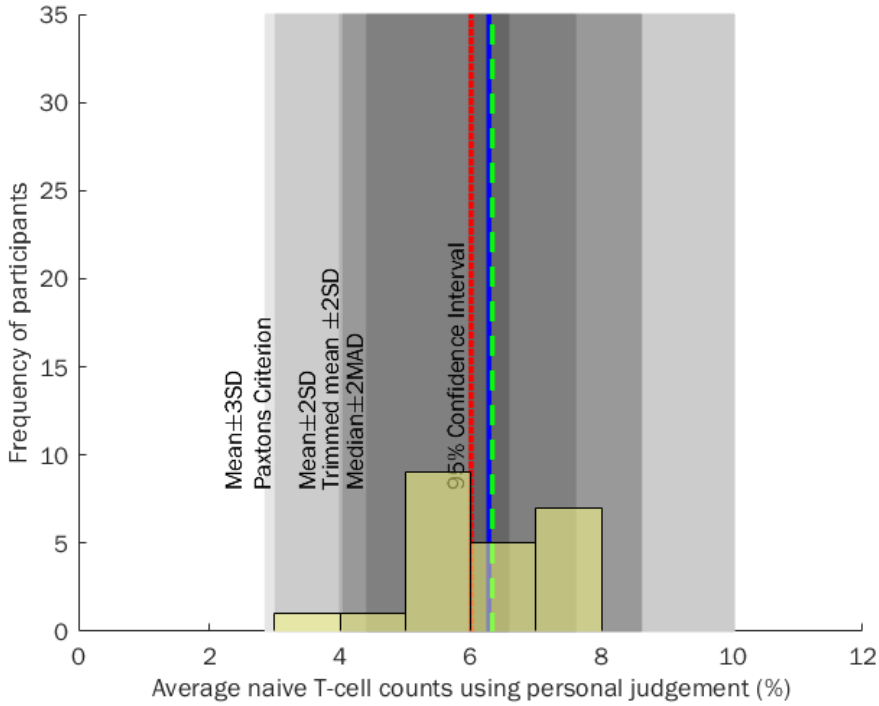


Figure 63 Acceptance boundaries for permissible limits of variability of Intermediate model personal judgement data

To compare both positive and negative deviation data extremes, Figure 64 shows participant B09's gating strategy, one of the participants very close to the group median value. By comparing the

extreme participants to a median participant, there are obvious differences between participants when identifying populations based on the visualised density. This is also coupled with personal preferences on inclusion or exclusion of data points to further refine the data set in search of a particular target. The count average comes from the final gate applied only, but further back in the gating sequence gates have been applied to capture most of the relevant populations. Variation in the vertical quadrant line does not impact the results because there is a very low cell count across this bandwidth. In contrast, the horizontal quadrant line shows less variance in the final cell count result because two of the repeats are identical in position. There appears to be a big difference in the CD3+ gate applied, because on one repeat B09 set the axes to bi-exponential instead of logarithmic, so when compiled, the gates appear different but actually capture similar data which has been transformed differently.

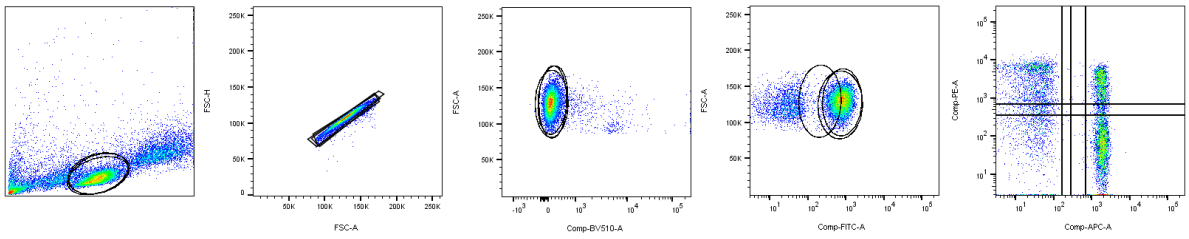


Figure 64 Participant B09 Gating Strategy interpretation, close to median result.

Focusing on those who fell outside of 2*SD*, only one participant (C01) had this much deviation from the median, with an overall percentage cell count of 3.5 %. Qualitatively reviewing the participant's gating strategy has shown that most of this bias is due to the gate applied to the live cells (BV510+, gate 3) which has been applied to a restrained proportion of the population (shown in Figure 65). C01 has applied this gate to follow the green density boundary, although there are more cells that could fall within this population that are less dense This gating strategy shown in Figure 65, identifies how under-constraining the cell populations leads to an overall lower cell count than the median.

Within a cell therapy manufacturing context, if a particular gate is under-constrained throughout the process, this will lead to an overall lower cell count population. Depending on the manufacturer-specific acceptance limit for cell counts, this could cause a therapy to have a false-negative measurement, potentially rejecting a treatment that is suitable to be delivered to a patient. Alternatively, if the therapy is given longer to culture due to a perceived low cell count, this becomes an inefficient and expensive use of resources.

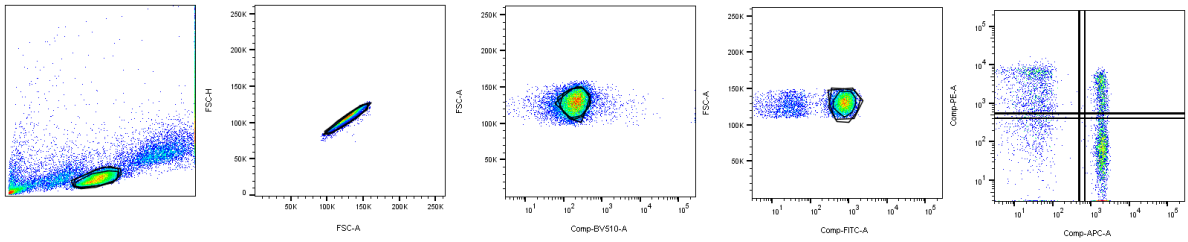


Figure 65 Participant C01 Gating Strategy interpretation

Participants A02, B01, B03, B10, B12 and C04 fall outside of the ± 1*SD* boundary, although all of these participants have average cell counts higher than the median value. Two examples have been selected to explore here; A02 has the smallest deviation (within this bandwidth) from the median and B10 has the greatest deviation (within this bandwidth). Figure 66 and Figure 67 show their respective gating strategies.

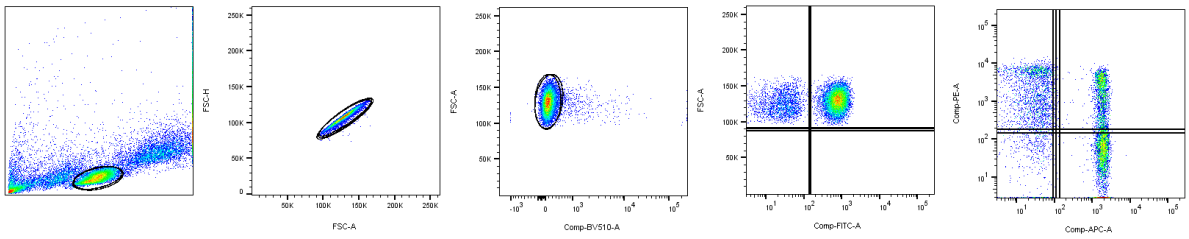


Figure 66 Participant A02 Gating Strategy interpretation

Participant A02 has a higher average result than the median inter-participant value because in the final step of their gating process (gate 5) they have included more cells than the median user. Whilst their final gate applied is very repeatable, the horizontal arm of the quadrant is lower than other median users, which has included a greater number of cells in the final count. In this instance,

A02 has applied gates very precisely, however they are potentially inaccurate, due to their use of control files to set the boundary limitations for the final fluorescence gates.

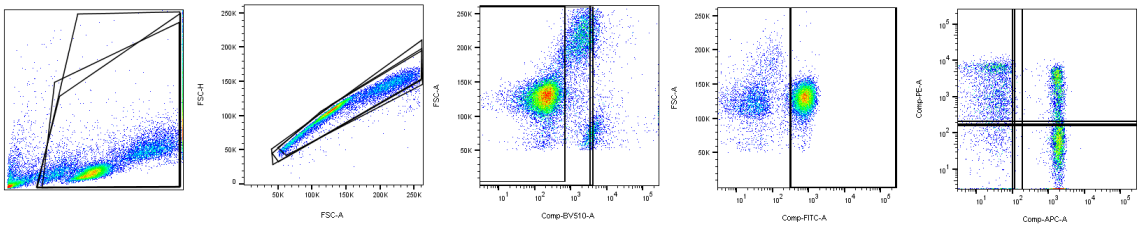


Figure 67 Participant B10 Gating Strategy interpretation

Participant B10 has again included additional cell populations which has caused the average cell count to be higher than the median. The doublets have not been excluded from the analysis in the first or second gates where they would usually be 'cleaned up' and they can be seen through the gating sequence as additional populations around the densest target population. They do appear to be gated out from the CD3+ gate (gate 4), however this is due to the compilation process of these images, with larger gates applied to the smaller one, so this population has been cut out.

This analysis of the absolute results used to represent cell populations shows a 4.5 % cell count range between participants when determining final naïve T-cell population percentages. The further quantitative analysis of the extremes identifies 1 participant (C01) who falls outside of initial control limits. This extreme participant accounts for one third of the cell population percentage range. If this extreme participant value was removed, the range would fall to 3.1 % between participants (minimum value of 4.9 % and maximum value of 8.0 %).

The distribution histogram of participant *CV* of reported results can be seen in Figure 68, plotted on top of 3 specification limits derived from the ICSH boundaries used within Chapter 4. The optimal scenario would have all 23 participants (*n*) with < 1 % *CV*, which sets the total height at the *y*-intercept of the graph.

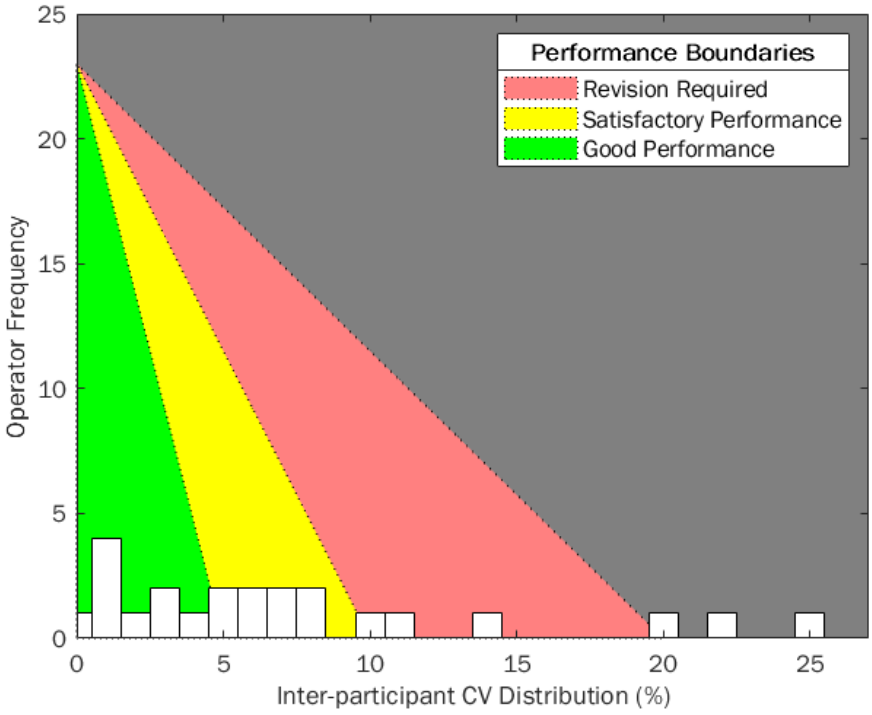


Figure 68 CV Performance of Participant Absolute results for personal judgement of intermediate model

Using these guidelines as boundaries, three participants were outside of the 'revision required' region due to very high *CV* and an additional three participants were outside of the 'good' and 'satisfactory' performance regions. These extreme outliers had more variation within the final quadrant gate when identifying the final naïve T-cell cell population. Participant B08 had the highest *CV*, and their final gate can be seen in Figure 69. B08 intersects the final population more than the rest of the participants with one gate significantly higher than the rest, causing this variation.

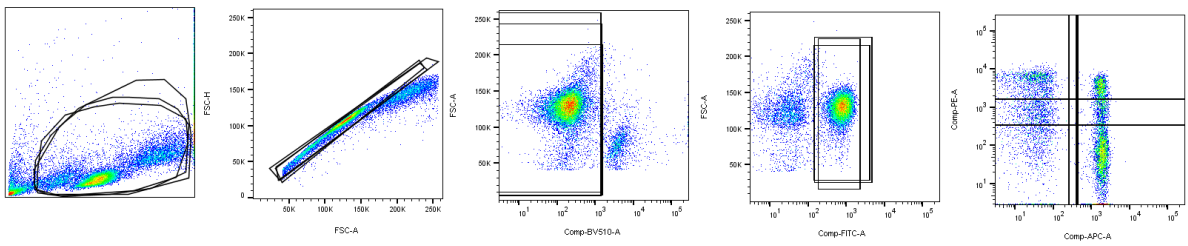


Figure 69 Participant B08 Intermediate Model Study Gating interpretation

Participants B06 and B12 (Figure 70 and Figure 71) also had *CV*s that fell outside the 'Revision Required' region of the graph, and again the variation is caused in the final gate due to one of the

quadrant repeats being placed significantly higher than the other two. Any gating variance observed throughout the gating sequence is not captured within the representative population *CV* measurement, because *CV* is only calculated from the final population cell counts derived using the whole sequence.

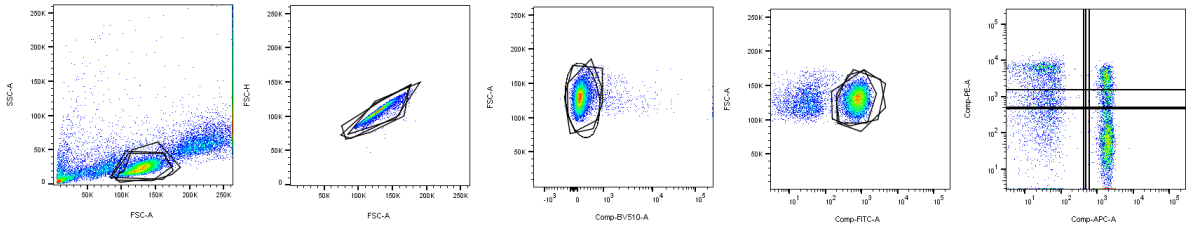


Figure 70 Participant B06 Intermediate Model Study Gating interpretation

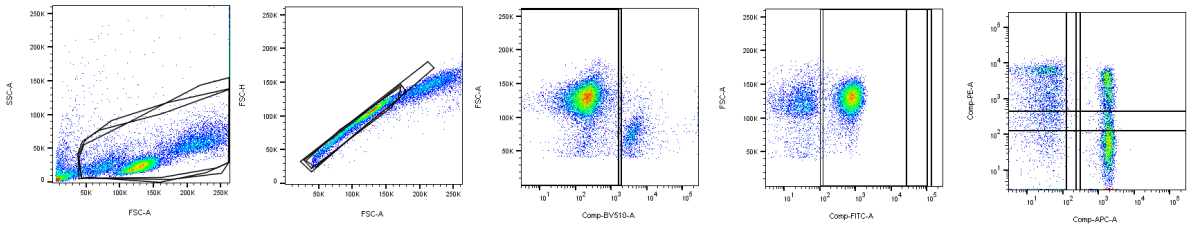


Figure 71 Participant B12 Intermediate Model Study Gating interpretation

Measurement uncertainty provides a way of combining variability measures (*SD*s) of each gate within the sequence, to provide a measure of variation that is more representative of the components of the gating sequence. When extremes in measurement uncertainty arise, uncertainty values can be easily deconstructed to identify which part of the gating sequence is responsible for causing variation within the measurement. Measurement uncertainty results for this phase are discussed in Section 5.3.4, once absolute results for phase 1 and 2 of this study have been reviewed and compared.

5.3.2 Flow Cytometry Intermediate Gating Exercise absolute results – Phase 2 Following Protocol

The absolute results reported here are the targeted cell population that participants were asked to identify using the gating sequence defined in Figure 58, during the second gating session where they used the diagrammatical protocol in Figure 59 to apply gates. A comparison of these results to Phase 1 and the uncertainty of the gating sequence is discussed in the subsequent section.

Table 37 Measures of Location for the absolute results of the Intermediate Gating Study using diagrammatical protocol (%)

Arithmetic Mean	6.17
Median	6.06
Mode	N/A
Minimum	5.70
Maximum	7.64

Table 38 Measures of Spread for the absolute results of the Intermediate Gating Study using diagrammatical protocol (%)

Range	1.94
25th Percentile	5.97
75th Percentile	6.25
Interquartile Range	0.28
Standard Deviation	0.39
CV	6.38
Median Absolute Deviation	6.06

Table 39 Measures of Skew for the absolute results of the Intermediate Gating Study using diagrammatical protocol (%) (3dp for better resolution)

Skewness	2.910
Skewness standard Error	0.512
Skewness z-score	5.684
Kurtosis	10.629
Kurtosis Standard Error	0.992
Kurtosis z-score	10.715

Table 40 Measures of Normality for the absolute results of the Intermediate Gating Study using diagrammatical protocol (%) (3dp for better resolution)

Shapiro-Wilk statistic	0.688
Significance	0.000
Normal/Non-parametric	Non-parametric

Using descriptive statistics to give a general report on the size and shape of the data, the distribution approximates to a normal shape because the mean and median are very close together, as quoted in Table 37. However, this is not supported by the skewness *z*-score (Table 39), and the high kurtosis *z*-score indicate a more non-parametric distribution due to outliers. The Shapiro-Wilk statistical test for normality (Table 40) also significantly indicates the distribution is non-parametric. This non-parametric definition is most probably indicated by a few outliers and a lot of uniform inliers, rather than any specific location parameters. Measures of spread (Table 38) show that there is a small range between the highest and lowest participant averages when participants followed a protocol and the IQR of the participant data was 86 % smaller than the total range, again indicating non-normality due to outliers.

These values can be seen in Figure 72. Most participant averages lie close to the median, and the error bars show ± 1*SD* from each participant's repeated measures.

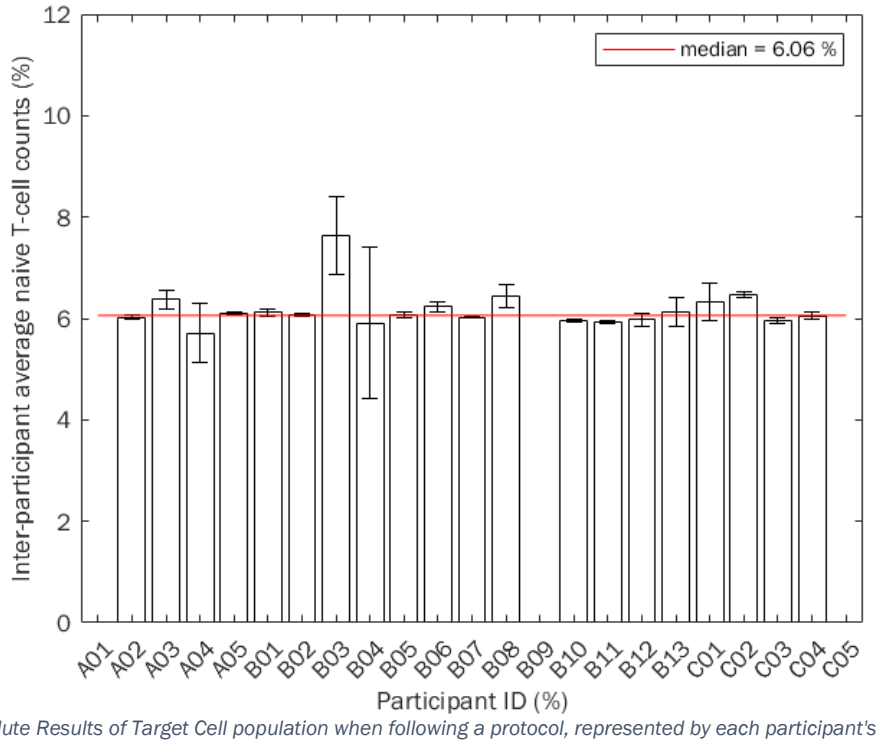


Figure 72 Absolute Results of Target Cell population when following a protocol, represented by each participant's average and ± SD.

Participant deviation from the median (residual values) has been more clearly visualised in Figure 73, with bars depicting each participant's average from the median group value (calculated by subtracting participant averages from the group median). The standard deviation limits have also been plotted, because these are most commonly used within traditional manufacturing to define out of control/out of specification limits. 95 % of participants are within 1*SD* of the median, showing good corroboration of final results. Of those who fell out of bounds one participant (B03) had a result outside of the - 2*SD* specification limit listed in Figure 73.

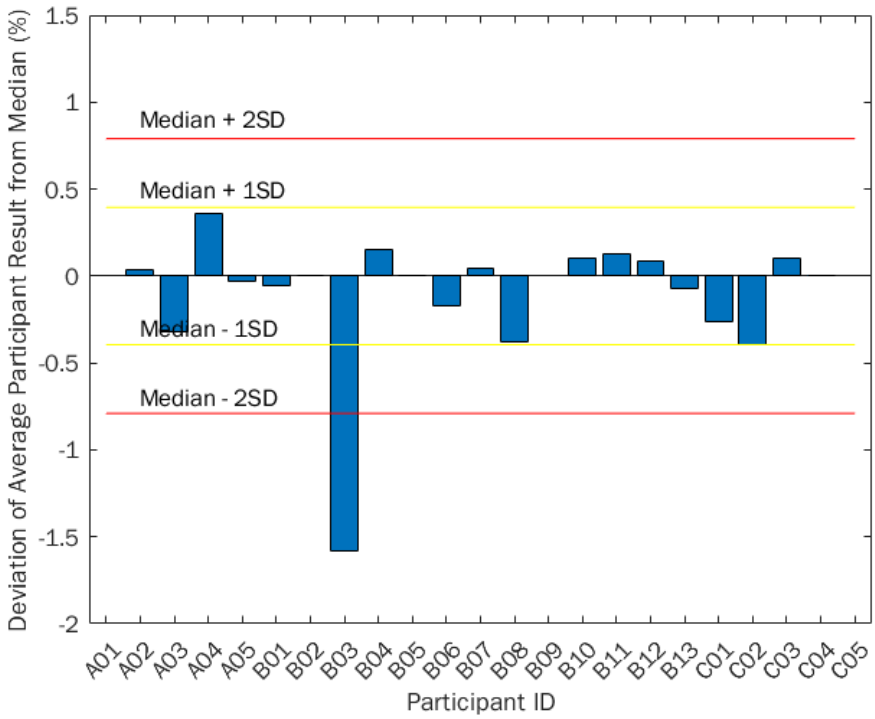


Figure 73 Participant average result deviations from overall group median.

The acceptance boundaries defined for the data collected when participants applied gates according to their personal judgement has been applied to the 'following protocol' data set in Figure 74. This data has become more refined when participants use a protocol, with the distribution falling within more of the acceptance limits. The 95 % Confidence Interval is very small, however the mean ± 2*SD*, trimmed mean ± 2*SD* and median ± 2*MAD* are all fairly similar in width, returning the same number of inliers and outliers. In this instance, further testing on the shape of the distribution should be complete before choosing the correct discrimination methods for extreme values.

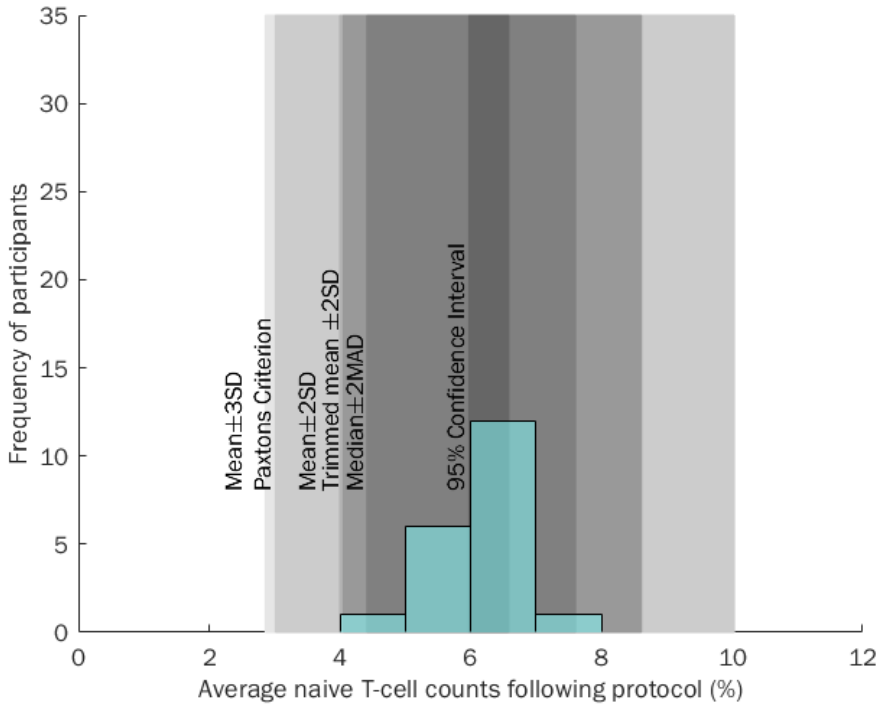


Figure 74 Acceptance boundaries of personal judgement data applied to data when participants followed a protocol

Qualitatively reviewing the extreme participant gating strategy has shown that most of this bias is due to the final gate applied to the CD4+ CD45RA+ cells (gate 5), because this gate includes more of the CD45RA- cell population that smears below the desired double positive population, with no clear density separation. This cuts across a dense region of this population, keeping a higher proportion of cells within the gate boundary. This gating strategy can be seen in Figure 75, which shows how over-constraining the final cell population leads to an overall higher cell count than the median.

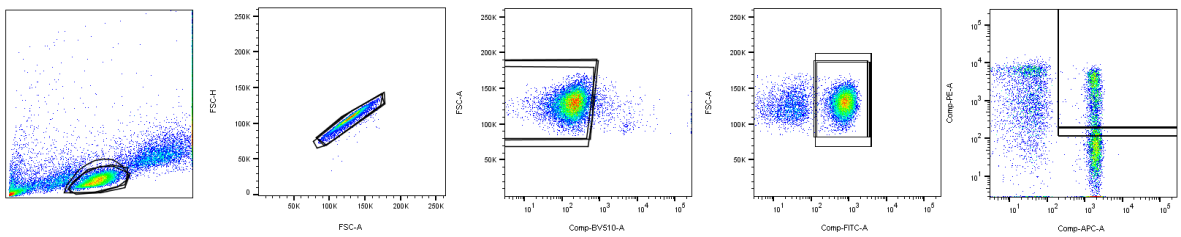


Figure 75 Participant B03 Gating Strategy interpretation

The distribution histogram of participant *CV* of reported results can be seen in Figure 76, plotted on top of 3 specification limits derived from the ICSH boundaries used throughout this thesis. Only 20 participants took part in this second gating session so the optimal scenario would have all 20 participants (*n*) with < 1 % *CV*, which sets the total height at the *y*-intercept of the graph.

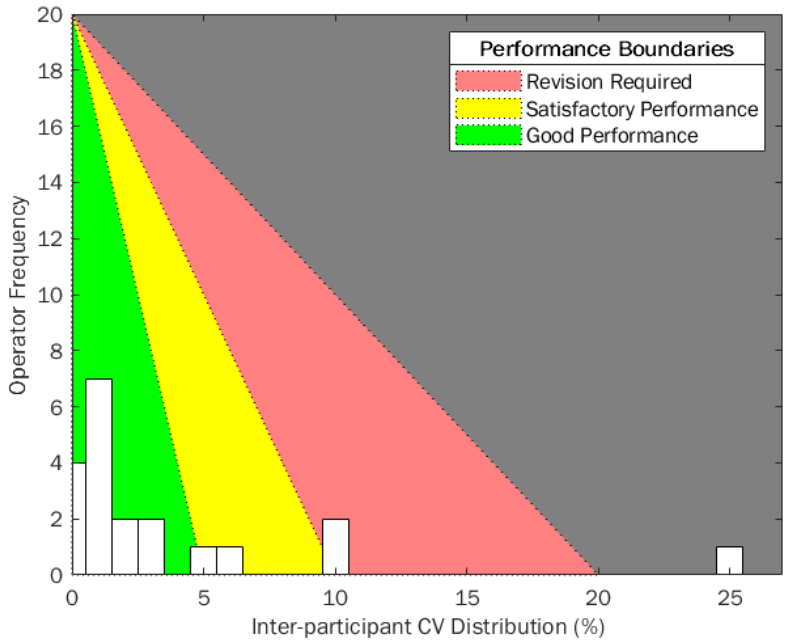


Figure 76 CV Performance of Participant Absolute results when following a protocol within intermediate model

Using these guidelines as boundaries, one participant fell outside of the 'revision required' region due to very high *CV* and two participants fell just outside of the 'good' and 'satisfactory' performance regions. These extreme outliers had more variation within the final quadrant gate (gate 5) they drew when identifying the final naïve T-cell cell population. Participant B04 had the highest *CV* when gating using the protocol, seen in Figure 77. B04 intersects the final population more and has one of the three repeated gates significantly higher than the rest, causing greater variation, similar to the variation seen in extremes in absolute cell count results.

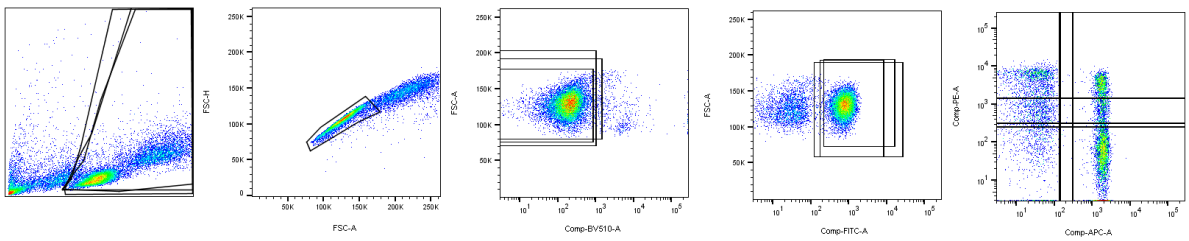


Figure 77 Participant B04 Intermediate Model Study Gating interpretation

5.3.3 Flow Cytometry Intermediate Gating Exercise absolute results – Comparison of Phase 1 and Phase 2

The average cell counts for each participant when they gated following their own judgement and then a protocol have been compiled into the histograms in Figure 78. Any dark orange areas are overlap of the two respective histograms. The range of cell counts has reduced by 2.59 % when participants followed the protocol, indicating that protocols could help participants conform to more reproducible cell counts. This is also reflected in higher skewness and kurtosis *z*-scores for the protocol data set. However, as discussed in Chapter 3, higher kurtosis could be a more suitable metric of conformity than trying to achieve a normal distribution when aiming for reproducible data.

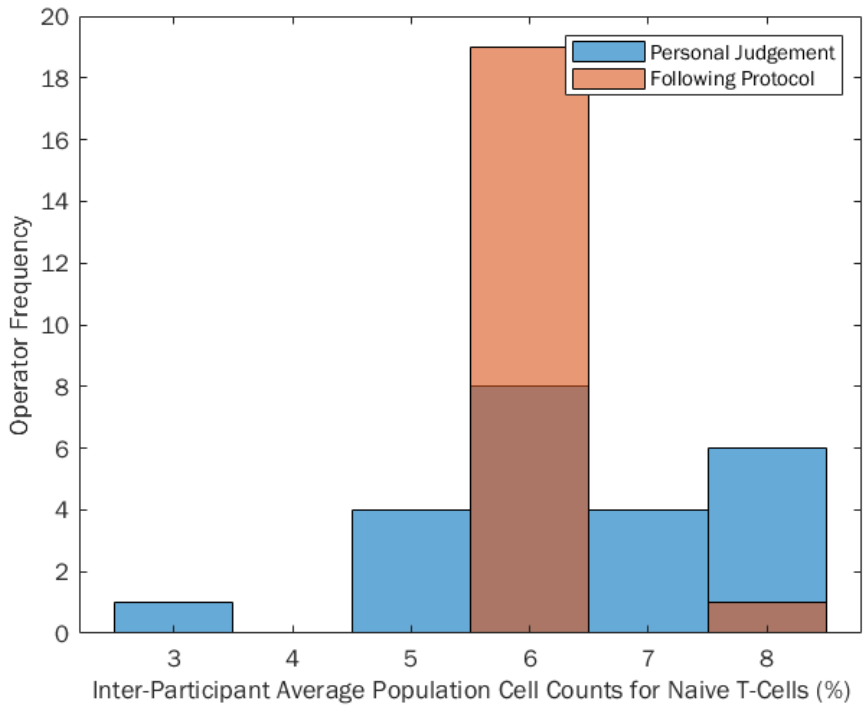


Figure 78 Comparison of inter-participant absolute cell counts when gating using their own judgement and when following a protocol (brown areas indicate overlap between the two distributions)

Further comparison of inter-participant data when gating using their own judgement and following a protocol has been completed with a Sign statistical test (Table 41), to compare equality of medians between each test condition. Although the Sign test results reject the alternative hypothesis of a difference in medians, this is potentially desirable when considering the cell count values themselves. In this instance there is no 'ideal correct' cell count answer for the cell population targeted for gating, so equality of the two testing conditions shows that protocols are comparable when identifying the target population and remove the variability from participant subjectivity around gating preferences to identify the population.

Table 41 Sign test results for comparison of Intermediate Gating Study Stages

Null Hypothesis	Test	Sig.	Decision
Median difference of P1 & P2 cell counts = 0	Related-samples Sign test	0.824	Retain null hypothesis

To further compare these two testing conditions the A Priori and Post Hoc power were calculated (Table 42) to identify whether a suitable number of participants had been gathered based upon the difference in variance of absolute cell counts in each test condition.

Table 42 A Priori and Post Hoc Power analysis for Intermediate Study absolute cell counts

Variance Phase 1	Variance Phase 2	A-priori power	Sample size required	Actual power
1.300	0.156	0.840	8	0.999

The A Priori and Post Hoc power analyses in Table 42 show that for the variances achieved between the two test condition average cell counts, only 8 participants would have been required to show this difference, to the required minimum power of 0.80. The actual power achieved through this study is 0.999, showing that the differences in variance of absolute cell counts are due to the two test conditions used and no other underlying factors.

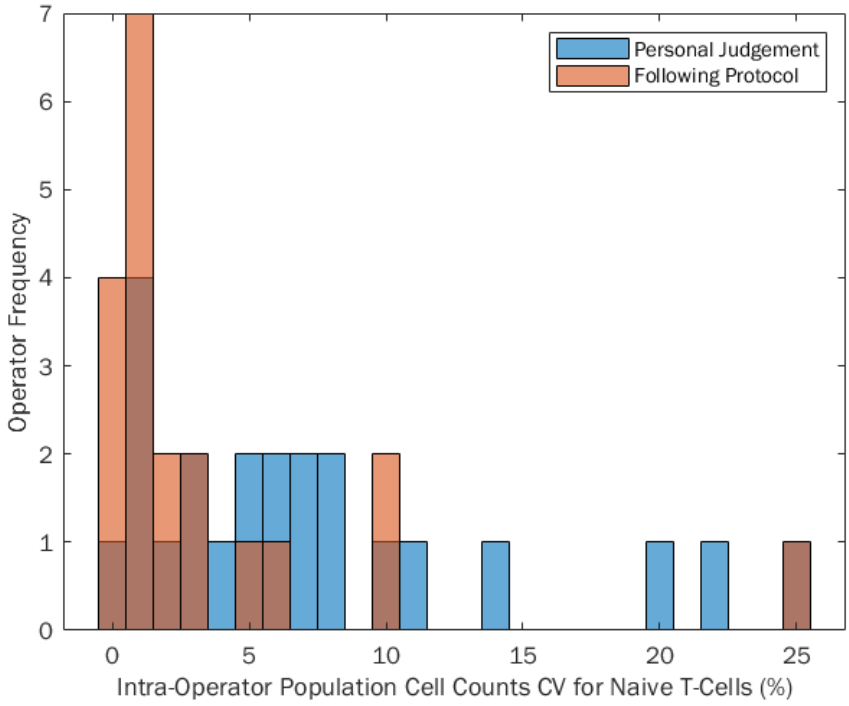


Figure 79 Comparison of participant absolute cell count CVs when gating using their own judgement and when following a protocol (brown areas indicate overlap between the two distributions)

Aside from one extreme value, the protocol appeared to reduce the range of results *CV* between participants when they followed this, but this only really considers the final gate applied within the repeats. Figure 79 shows that the *CV* is much more positively skewed towards 0 when participants followed a protocol, showing that following a protocol is more likely to reduce your final cell count gating variability and improve reproducibility of results overall.

5.3.4 Flow Cytometry Intermediate Gating Exercise uncertainty results – Phase 1: Personal Judgement

The uncertainty results reported here are a combination of the five gating stages defined in Figure 33 for gates applied when participants use their own judgement. The uncertainty values have been quantified following the prescribed methodology in Chapter 2, Section 2.5. The uncertainty would better represent variance of measurements with greater confidence, because this combines variability from all gates applied in the sequence, not just the variance of the final gate applied.

Table 43 Measures of Location for Uncertainty of the Intermediate Gating Study using personal judgement (%)

Arithmetic Mean	3.8
Median	2.1
Mode	N/A
Minimum	0.4
Maximum	16.1

Table 44 Measures of Spread for Uncertainty of the Intermediate Gating Study using personal judgement (%)

Range	15.7
25th Percentile	1.4
75th Percentile	3.8
Interquartile Range	2.3
Standard Deviation	4.3
Median Absolute Deviation	0.8

Table 45 Measures of Skew for Uncertainty of the Intermediate Gating Study using personal judgement (%) (3dp for resolution)

Skewness	1.942
Skewness standard Error	0.481
Skewness z-score	4.037
Kurtosis	2.899
Kurtosis Standard Error	0.935
Kurtosis z-score	3.101

Table 46 Shapiro-Wilk test for normality for Uncertainty of the Intermediate Gating Study using personal judgement (%) (3dp for resolution)

Shapiro-Wilk statistic	0.692
Significance	0.000
Normal/Non-parametric	Non-parametric

Unlike the descriptive statistics for absolute results for this study phase, the mean and the median are not close together, indicating a more skewed distribution, as monitored in Table 43. The median is less than the mean, indicating a slight positive skew to the data. This is further supported by the Shapiro-Wilk test for normality, shown in Table 46, indicating that the distribution is non-parametric in shape, indicating skewness.

There is a wide range (15.4 %) between minimum and maximum participant uncertainties. Table 44 also shows the interquartile range as 2.3 %, indicating a high kurtosis, because half of the data lies within 15 % of the total distribution range. This is supported by the skewness and kurtosis values in Table 45 with skewness and kurtosis *z*-scores falling outside of the ±2.58 boundaries specified for normality defined in Chapter 2. The raw data for each participant can be seen in Figure 80 with various extremes within the dataset. This distribution shape can be observed within Figure 81, showing the positive skew with a potential bimodal split and 4 larger uncertainty extremes.

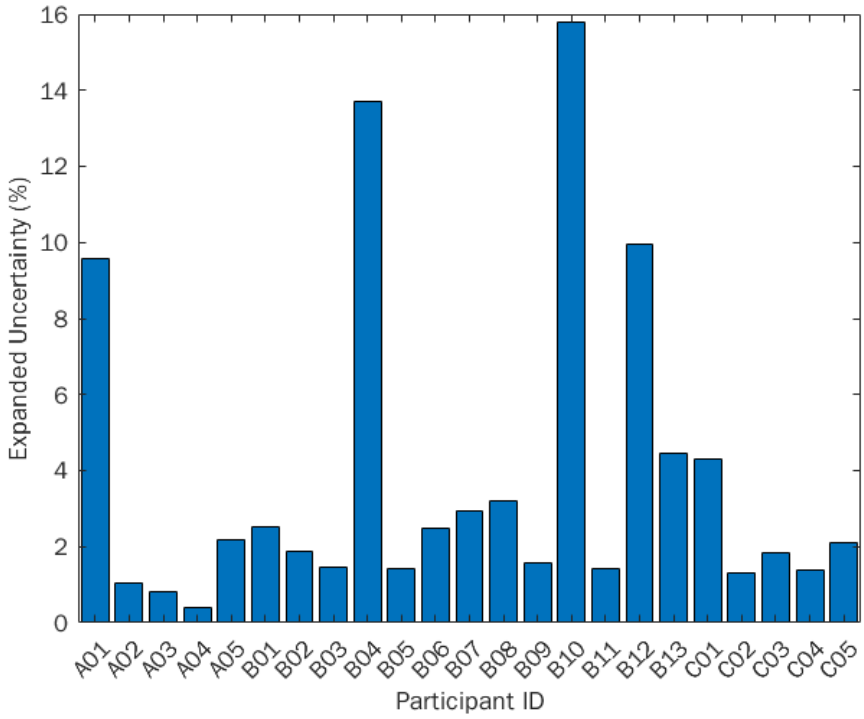


Figure 80 Expanded Uncertainty of all Participant Gating within the Intermediate Model Study

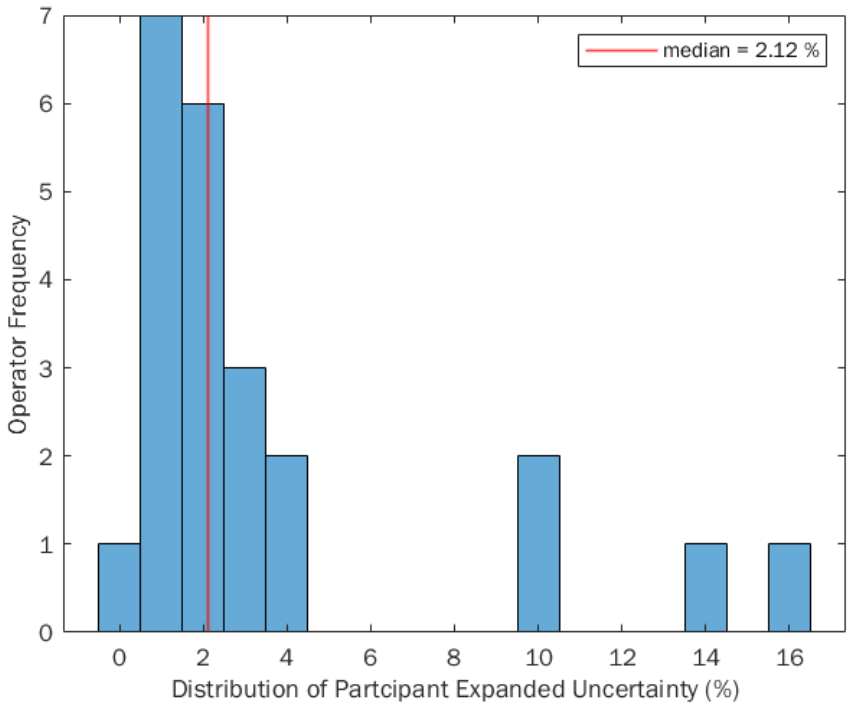


Figure 81 Histogram of Participant Uncertainty from repeats of Intermediate Model Study

Whilst deviation from a median can help to explain the distribution parameters, when analysing variance, a positively skewed distribution towards zero is preferred. Comparing uncertainty in its size order allows boundaries to be set for permissible specification limits for product release/laboratory quality that increase in value.

The ICCH and ICCS imprecision values have also been used here to define example specification limits if monitoring participant uncertainty. Again, in this instance the *CV* specification limits have been substituted for uncertainty (Figure 83), and no other uncertainty specifications have been defined in the public body of knowledge from research or industry. However, unlike the previous model, this correlation is not strong because of the four participants with extreme uncertainty values, which will be explored within the extremes identified within Figure 82. Even if these 4 'outliers' were removed, the slope (indicating correlation) only increases slightly to 0.0699x. Although this correlation was used to justify the use of *CV* boundaries for uncertainty measurements in Chapter 4, this correlation is not present here. However, for continuity of

reporting, the uncertainty distributions will still be binned according to these guidelines, to aid comparison between the uncertainty exercises.

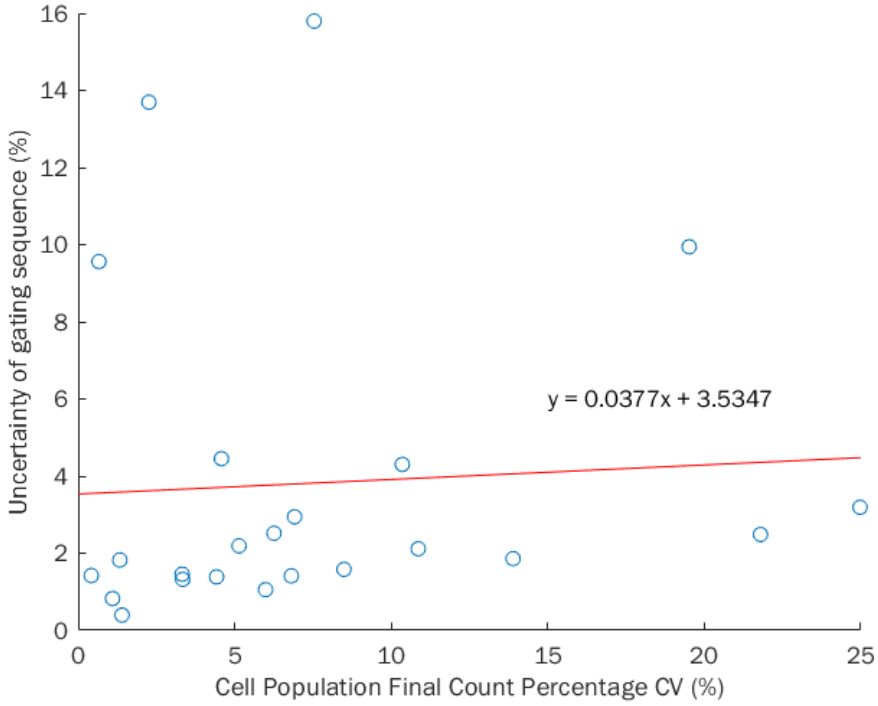


Figure 82 Final Cell Count Population Percentage versus Gating Uncertainty for participants

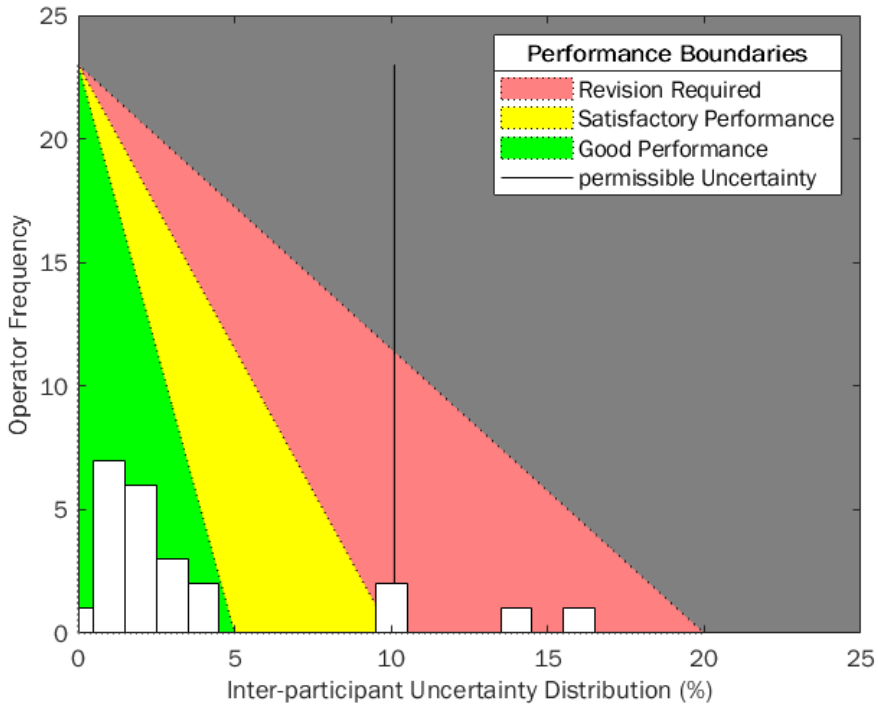


Figure 83 Participant Uncertainty performance monitoring diagram when using their own judgement during the intermediate model

Participant B10 had the highest overall uncertainty that fell within the 'revision required' region of the plot, followed by Participant B04, and participants C01 and C05 sitting on the boundary of 'satisfactory performance' and 'revision required'. A deeper review of B10's gating in Figure 84 shows their greatest variance came from the third gate applied to discriminate live/dead cells. The amine-reactive dye used stains dead or dying cells, so the live cells appear as the densest population on the left side of the plot. Reviewing B10's repeats has highlighted different ways this live/dead boundary was selected. B10 used the unstained control file to set areas for negative expression (live cells). The inclusion of the doublet cells and dying cells has caused confusion and therefore variability when choosing this boundary because they had not been previously 'cleaned up' from the data.

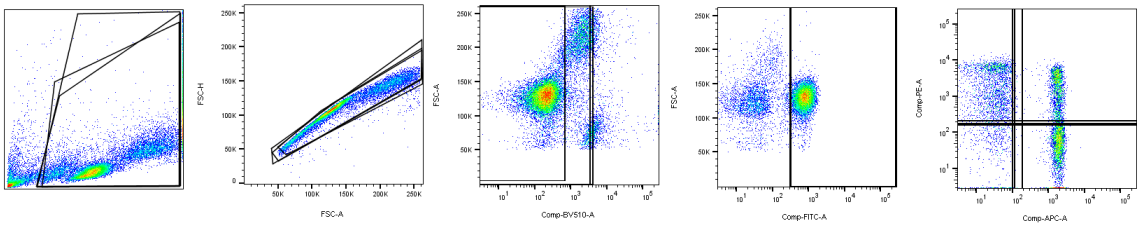


Figure 84 Participant B10 Intermediate Model Study Gating interpretation

This gate also caused the most variability for participant B04, who in one repeat also included this doublet population within their gate (shown in Figure 85, gate 3), causing variation and a higher population value as a result. B04 has also used a bi-exponential scale instead of a logarithmic scale, which could be an additional variable affecting their overall uncertainty.

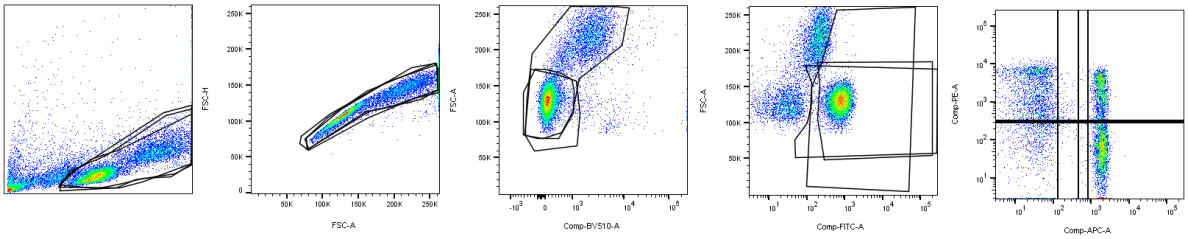


Figure 85 Participant B04 Intermediate Model Study Gating interpretation

The quadrant gates used to locate the final CD4+ CD45RA+ population were very precise, having minimal effect because they do not intersect any high-density areas through good use of control files provided. Control files provided were isotypes, FMO controls and a negative sample, which were used to set the position of the gates using a positive/negative split on respective axes. Control files were provided for other gates in the sequence but those who have extremes in their data appear to not have used these in the same manner to apply gates reproducibly.

5.3.5 Flow Cytometry Intermediate Gating Exercise uncertainty results – Phase 2: Following Protocol

The uncertainty results reported here (calculated as per Section 2.5 and Section 5.3.5) are a combination of the five gating stages defined in Figure 58 using the gating protocol provided, shown in Figure 59.

Table 47 Measures of Location for Uncertainty of the Intermediate Gating Study when following a protocol (%)

Arithmetic Mean	5.8
Median	2.2
Mode	N/A
Minimum	0.5
Maximum	12.6

Table 48 Measures of Spread for Uncertainty of the Intermediate Gating Study when following a protocol (%)

Range	12.1
25th Percentile	1.4
75th Percentile	10.9
Interquartile Range	9.5
Standard Deviation	5.1
Median Absolute Deviation	2.2

Table 49 Measures of Skew for Uncertainty of the Intermediate Gating Study when following a protocol l(%) (3dp for resolution)

Skewness	0.226
Skewness standard Error	0.512
Skewness z-score	0.441
Kurtosis	-2.065
Kurtosis Standard Error	0.992
Kurtosis z-score	-2.082

Table 50 Shapiro-Wilk test for normality for Uncertainty of the Intermediate Gating Study when following a protocol (%) (3dp for resolution)

Shapiro-Wilk statistic	0.757
Significance	0.000
Normal/Non-parametric	Non-parametric

Unlike the descriptive statistics for absolute results for this study, the mean and the median are not close together, indicating a more skewed distribution, as monitored in Table 47. The median is less than the mean, indicating a positive skew to the data. This is further supported by the Shapiro-Wilk test for normality, shown in Table 50, indicating that the distribution is non-parametric in shape, indicating skewness. There is a wide range (12.0 %) between minimum and maximum participant uncertainties. Table 48 also shows the interquartile range as 9.5 %, however, unlike previous distributions this supports a low kurtosis, because half of the data lies within 79 % of the total distribution. This is supported by the skewness and kurtosis values in Table 49 with skewness and kurtosis *z*-scores inside of the ±2.58 boundaries specified for normality defined in Chapter 2. The raw data for each participant can be seen in Figure 86 with various extremes within the dataset. This ordered distribution shape can be observed within Figure 87, showing the positive skew with a bimodal split and 9 larger uncertainty extremes. Qualitatively assessing the shape of the distribution shows a different shape to the data than the descriptive statistics, because the mean calculated does not reflect central tendency of the raw data as would be assumed. This bimodal split will be investigated further to understand the difference between the high variance and low variance clusters.

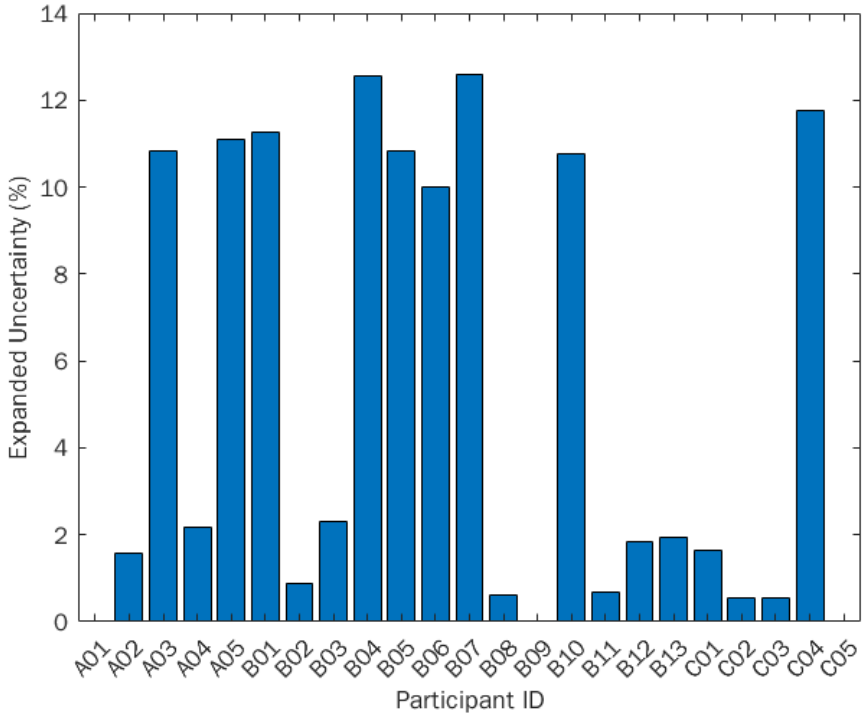


Figure 86 Expanded Uncertainty of all Participant Gating within the Intermediate Model Study when participants followed a protocol

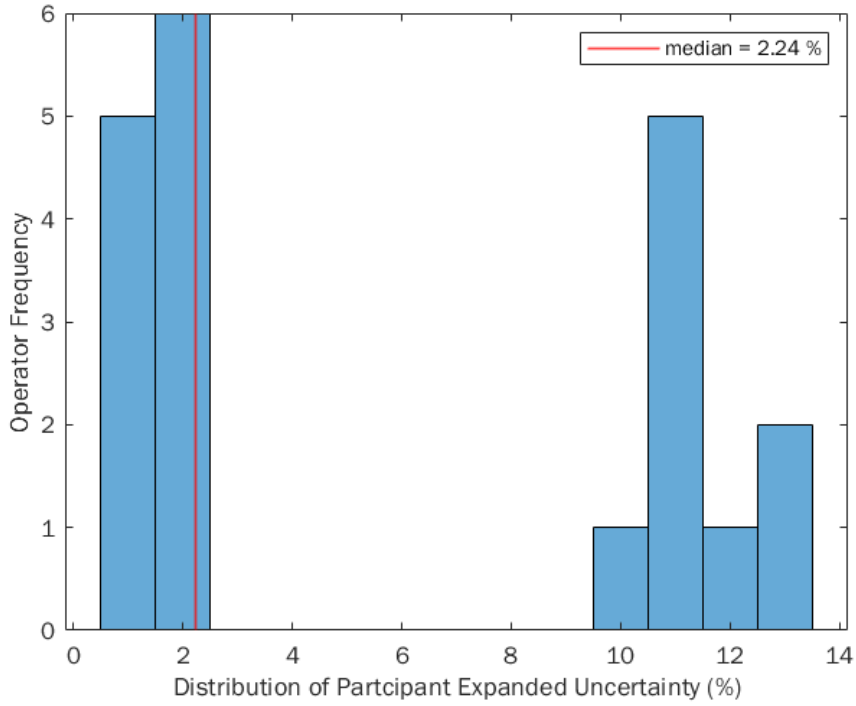


Figure 87 Histogram of Participant Uncertainty from repeats of Intermediate Model Study when participants followed a protocol

Whilst deviation from a median can help to explain the distribution parameters, when analysing variance, a positively skewed distribution towards zero is preferred. Comparing uncertainty in its size order allows boundaries to be set for permissible specification limits for product release/laboratory quality that increase in value.

The ICCH and ICS imprecision values described earlier for measurement *CV* and in previous Chapters have also been used here to define example specification limits if monitoring participant uncertainty. Again, in this instance the *CV* specification limits have been substituted for Uncertainty from this gating phase (Figure 88), because no other uncertainty specifications have been defined in the public body of knowledge from research or industry.

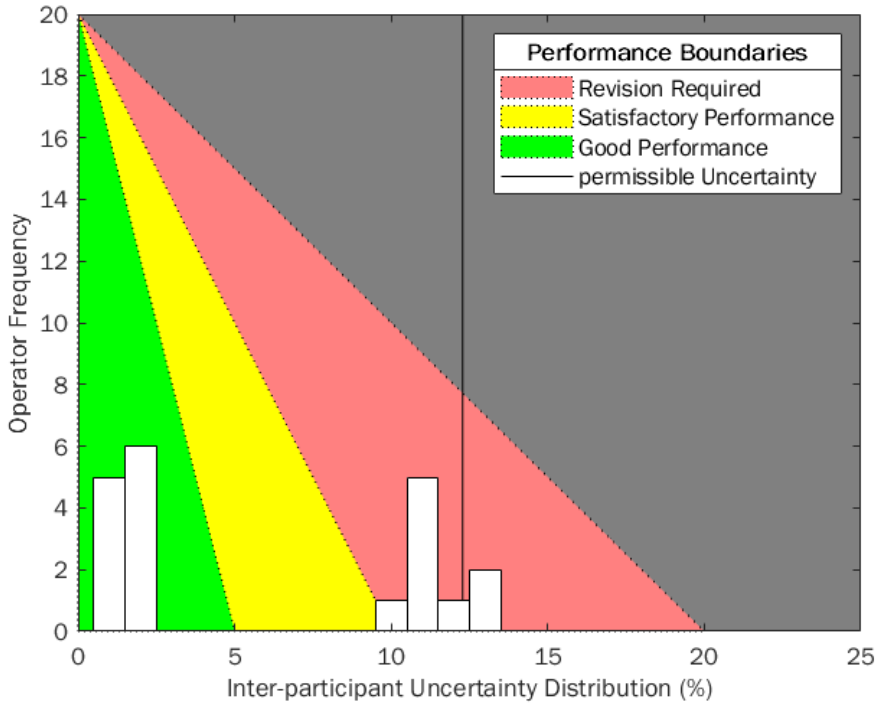


Figure 88 Participant Uncertainty performance monitoring diagram when following protocol during the intermediate model

Participants B04 and B07 had the highest overall uncertainty that fell within the 'revision required' region of the plot, followed by C04, who is just under the limit of permissible uncertainty on the

chart in Figure 88. However, there is a participant cluster of high variance which has its own maxima at 11 % Uncertainty, populated by participants A03, A05, B01, B05 and B10.

Participant B07's most variable gate was the on the first gate in the sequence. This has been enlarged in Figure 89 to better identify the source of this variation. The repeats are qualitatively very uniform in size and structure, with a repeatable cut-off separating the primary population from the dead and dying cells. The quantitative variability appears to come from differences in the right edge of the gate, where in one instance the gate applied includes a data spike that sits on the boundary of the plot. The axes of the plot have been scaled down slightly so this spike can be clearly seen. This data spike could be an amalgamation of all the data points that exceed the plot limits, so they have been complied and added to the boundary, however, there is no information from Flowjo on this visualisation effect.

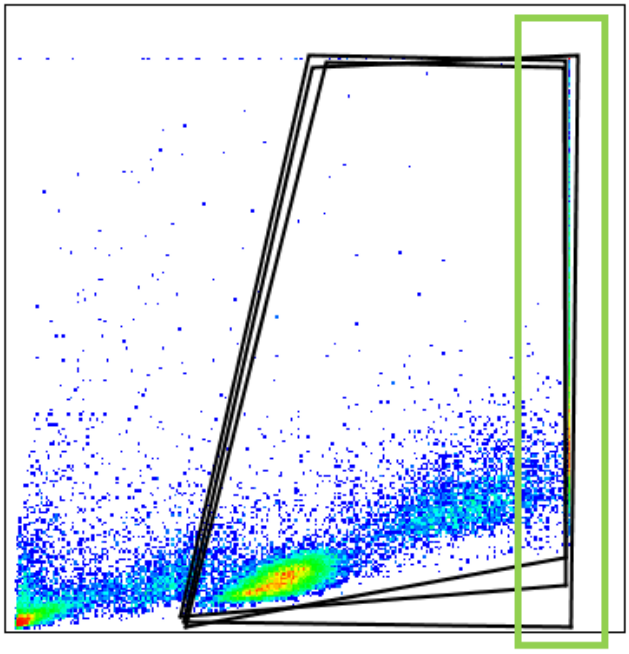


Figure 89 Enlargement of Participant B07's first plot and gates applied in Phase 2

Additionally, other Flow Cytometry data from other research teams using different cell models were checked to see if a boundary effect was present, and it appears that any cell events that exceed

the highest boundary scale point on the plots get concatenated in this way, making it seem to be a function of the software data binning and visualisation process. This intermediate data set was imported into two other Flow Cytometry Analysis Software packages, FlowLogic and FCS Express and qualitatively compared to the FlowJo output, seen in Figure 90. Not only do the other two software packages use different colour gradient scales to visualised cluster density, the boundary effect has been removed by FlowLogic, but is present and even more noticeable in FCS Express. This therefore falls to operator subjectivity, not only in which software they choose for analysis but whether they are aware of this boundary effect and if they consciously include or exclude it from their analysis.

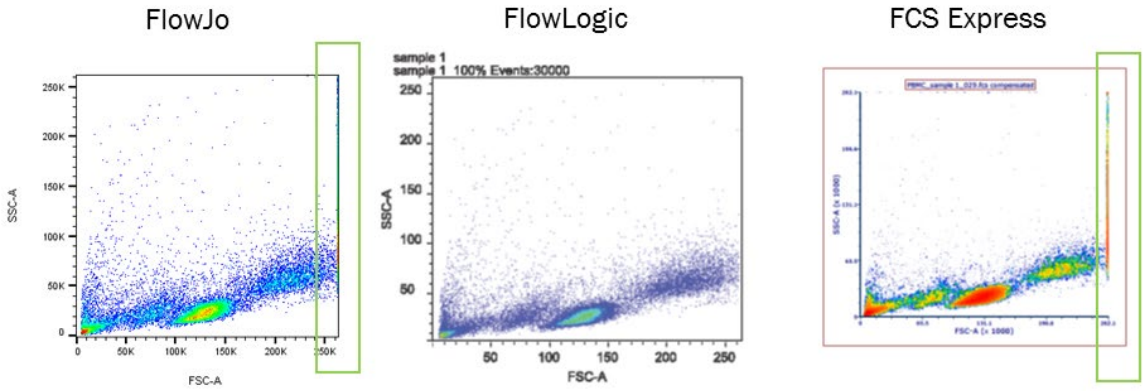


Figure 90 Visualisation of Intermediate Gating Study data in different Flow Cytometry Software

B04's Phase 2 gate has already been featured as an extreme in this Chapter (Section 5.3.2) due to high absolute cell count *CV*. Figure 77 depicts B04's gating strategy, with most of their gating variability introduced from the first gate applied to start 'cleaning' the data, to remove dead cells and debris. An enlargement of this plot has been provided in Figure 91. Although this plot has a lot of different overlapping populations, the repeated gates are very similar in size shape and where they intersect and cut off the dead and dying cells, because this area has a lot of clustering overlap, making it difficult to identify the main area of variation on this plot. Similar to B07's plots, the right edge sits along the boundary effect which could be causing some variability within the analysis. The majority of variation in the overall uncertainty of B07, C04, A03, A05, B01 and B05's gates were

also from this first plot, strengthening the need for further investigation into this variation and its cause.

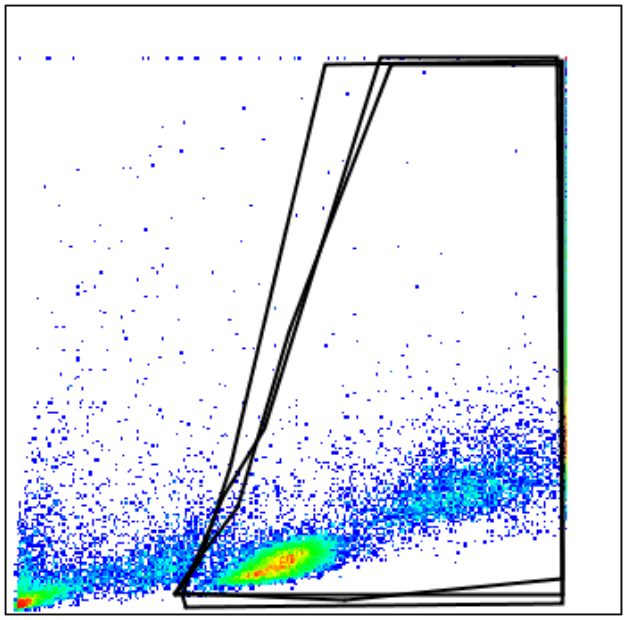


Figure 91 Enlargement of Participant B04's first plot and gates applied in Phase 2

To further investigate if the cause of this higher group cluster variability is a function of the first gate in the sequence, a selection of participants with low uncertainties were looked at to see if there was a difference or obvious understanding of this boundary effect, away from the data spike during their analysis. Participants B02 and B08 had very low uncertainties from their total gating process and both fall within the lower variance cluster. Their initial plots can be respectively seen in Figure 92 and Figure 93. Participant B02 has applied gates close to this boundary effect, however none touch this edge and they are very repeatable in size and shape. Upon closer inspection B08 has clearly not included any of this boundary effect and there is a clear distance from this edge and the right side of their gates. This initially shows that this difference in high and low variance groups could be down to participant knowledge and awareness of this boundary effect, however this was not something captured within the gating sessions themselves, nor could it be something that the participants were consciously aware of.

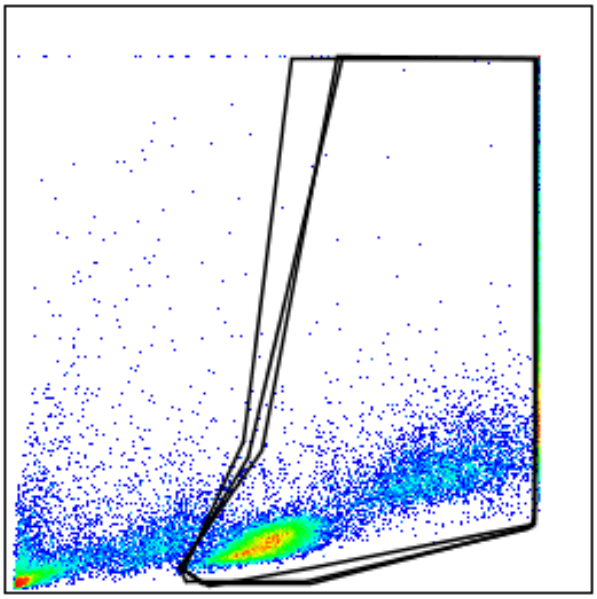


Figure 92 Enlargement of Participant B02's first plot and gates applied in Phase 2

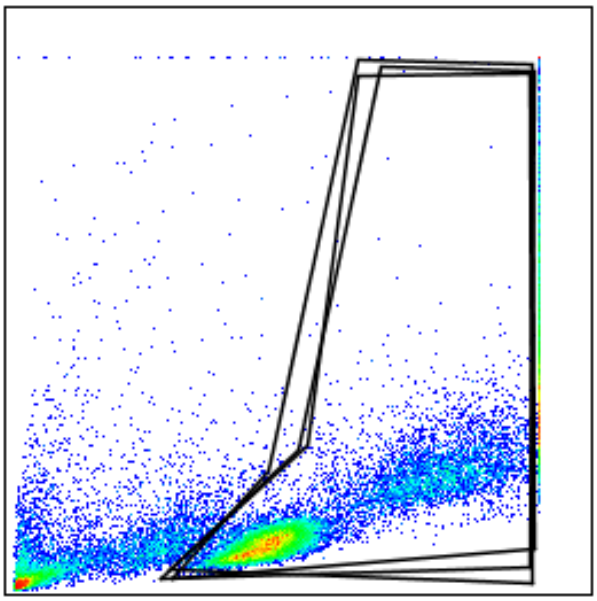


Figure 93 Enlargement of Participant B08's first plot and gates applied in Phase 2

A follow-up questionnaire was given to participants during the complex model (Chapter 6) to try and identify knowledge of this effect without asking leading questions and the results are discussed in more detail within Chapter 8.

5.3.6 Flow Cytometry Intermediate Gating Exercise uncertainty results – Comparison of Phase 1 and Phase 2

The uncertainties for each participant when they gated following their own judgement and then a protocol have been compiled into the histograms in Figure 94. Any dark orange areas are overlap of the two respective histograms. The range of cell counts has reduced by 3.4 %, indicating that protocols could help participants conform to reproducible cell count, however a bimodal distribution appeared so potentially more consideration needs to be applied to understanding how subjectivity and interpretation of a protocol and visual images impact the final uncertainty calculated. The skewness and kurtosis *z*-scores have reduced when participants use a protocol, however this is showing a tendency towards normality, which the distribution shape does not support in this instance.

When participants used their own judgement to apply gates, a split between higher and lower uncertainty groups appears, but the shape of the overall distribution is positively skewed, with only a few extremes exhibiting high variance, so it can be described as bimodal, similar to the protocol uncertainty distribution. The protocol participants were asked to copy included the boundary effect observed, which is potentially causing the difference in variance clusters, which is something that requires further investigation, but has been used to aid the instructions for the complex model in Chapter 6.

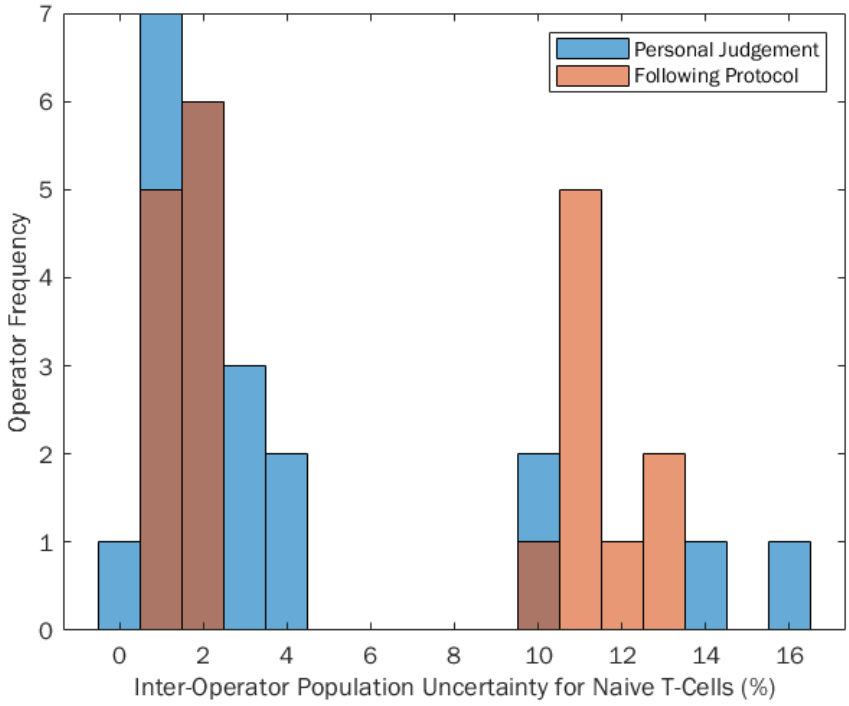


Figure 94 Overlaid histograms of participant uncertainties when gating the naïve T-cell population following their own judgment and then using a protocol (brown areas indicate overlap between the two distributions)

Further comparison of participant data when gating using their own judgement and following a protocol has been completed with a Sign statistical test, to compare equality of median uncertainties between each test condition, shown in Table 51. The Sign test results reject the alternative hypothesis, because the uncertainty medians of the two testing conditions are not statistically different. This does not statistically show that the uncertainties are smaller when participants use a protocol, however this only considered location values, so other measures of variability should also be considered before deciding whether a protocol effectively reduces inter-participant variability, or not.

However, the reduction in overall range shows that the distribution can potentially be controlled to minimise extreme values occurring in the data set, although use of this protocol is causing subjective behaviour to divide the population into high and low variance clusters.

Table 51 Sign test for median difference between Phase 1 and Phase 2 uncertainties

Null Hypothesis	Test	Sig.	Decision
Median difference of P1 & P2 uncertainties = 0	Related-samples Sign test	1.000	Retain null hypothesis

To further compare these two testing conditions the A Priori and Post Hoc power were calculated (Table 52) to identify whether a suitable number of participants had been gathered based upon the difference in variance of absolute cell counts in each test condition.

Table 52 A Priori and Post Hoc Power for Phase 1 and 2 uncertainties

Variance Phase 1	Variance Phase 2	A-priori power	Sample size required	Actual power
18.193	26.388	0.80	181	0.213

The A Priori and Post Hoc power analyses in Table 52 show that for the variances achieved between the two test conditions uncertainties, 181 participants would have been required to show this difference, to the required minimum power of 0.80. The actual power achieved through this study is 0.213. This low power indicates that any differences seen from the data have a low probability of being just due to the two test conditions used and no other underlying factors present. Again, in a manner similar to the Sign test, this needs to be considered carefully. A greater number of participants in the study could always benefit and provide more confidence in the results, however, the distributions are not normally distributed, so the variance calculated assumes a distribution with central tendency (whether this is normally distributed or non-parametric). In this instance, the range of data becomes more important to consider, due to the distribution shape and clusters appearing within the uncertainty data.

5.4 Chapter Conclusions

The primary PBMC material used was a good model for the intermediate study because it provided an appropriate step up in complexity of the analysis pipeline participants were required to complete, and it also provided affinity towards current cell therapy treatments which are T-cell based. The studies completed with the 3-workspace configuration in Flowjo for repeats continued to work well within the time available for participants so this structure has shown to be a good working model for these studies and also will be used for the final, most complex model. The five-step gating process that each participant had to work through was also straight-forward to follow from the gating sequence protocol and the diagrammatical protocol, to help control additional variance within the study.

When reporting the absolute cell count percentages for the results, the mean and median values for Phase 1 and Phase 2 were very similar (mean = 6.3 % and median = 6.0 % for Phase 1, mean = 6.2 % and median = 6.1 % for Phase 2). These results both indicate a normal distribution within both data sets, but it also shows that a protocol can achieve the same target cell counts compared to when participants used their own judgement. The difference when using a protocol is the reduction in between-participant cell count results. The protocol reduced the range of cell count results between-participants by 2.6 % of the overall cell count (777 cell events). This is a 57 % reduction in range from Phase 1 and Phase 2, indicating protocols can potentially aid reproducible cell counts between participants. The Sign test confirmed there was no significant difference between the medians of the two testing conditions, showing that protocols can produce similar results to the null testing condition, however, the power analysis completed indicated the variance of these two groups is different enough such that only 8 participants would have been required to attain the same distribution results.

CV is a common variation metric used within cellular measurements within Flow Cytometry, however, when considering the use of protocols, the range of *CV* in reported results remained the same as when participants used their own judgement (up to 25 %). However, the use of the protocol

made the distribution much more positively skewed towards 0 %, which is desirable for variation metrics, indicating that more participants were less variable (in their final cell count) when using a protocol to apply gates. Reviewing the extremes in *CV* using the adjusted traffic light diagram shows variability in the final gate applied, but it also highlights that other variability seen upstream of this final gate can have an impact on cell count. However, gate variability across the five gating stages itself is not taken into consideration in the *CV* calculation, making measurement uncertainty a more suitable metric for accommodating variation throughout the whole gating sequence and increasing confidence in the results.

This intermediate model with the 5-gate sequence has shown that calculating measurement uncertainty is possible for participants by using traditional measurement uncertainty methods. This was calculated successfully, by presenting participants with three repeated workspaces of data, and extracting one repeated file located in each workspace randomisation.

The uncertainties calculated to accompany the cell counts are more non-parametric than the absolute cell counts. The mean and median uncertainties went from 3.8 % and 2.1 % respectively in Phase 1 to 5.8 % and 2.2 % respectively in Phase 2. The medians are very close together between the two sessions and is a more suitable metric due to the skewness present. The means are unsuitable metrics because they do not represent the peak maxima, especially in Phase 2, where the bimodal distribution causes the mean to sit between two peaks. However, the range of inter-participant uncertainty reduced when using a protocol which shows this could possibly improve reproducibility of results between FC analysis. Whilst following a protocol reduced the range of uncertainty, the distribution shape separated into two peaks, indicating clusters of high and low variance participants.

Further investigation has shown a high probability of this variation being caused by a boundary effect within the data visualisation software. Cells in the file that have a fluorescence signal higher than the visualisation axes are concatenated on the boundaries. Inclusion of these (either in a

consistent or inconsistent manner) in the repeated analysis can skew the cell counts and uncertainty significantly. Those in the lower variance cluster have not included these cells in their analysis, or repeatedly have so the overall variation would be lower between repeats (although they would have a higher average cell count). Revision of extreme participants using the uncertainty traffic light diagrams has shown most of the variation is contributed within this first gate applied, where the boundary effect is initially seen, increasing the possibility that the boundary effect causes this higher variation.

Overall, the structure of these analysis sessions and data extraction processes works well, so this structure will be used to inform and run the subsequent complex model, which can increase in complexity due to the required cell population target and the number of gates required to obtain it. So far, this chapter has potentially shown affinity to the core hypothesis of this thesis. The overall range of uncertainty has increase from the basic model (Chapter 4, 12.37 %) to this intermediate model (Chapter 5, 15.7 %) showing that the range of uncertainty between-participants has increase as the data has become more complex. This needs to be further tested with a more complex model, presented in Chapter 6, and a diagrammatical protocol will also be tested in this instance, because it has shown to reduce the range of participant results within this intermediate model (12.1 %).

5.4.1 Consolidation of Objectives

- This study ran smoothly, acting as a good intermediate model for comparison of absolute reported results, *CV* and uncertainty measures. The session structures were suitable in time, and 3 repeats was suitable for participants to understand study context, but not become tired. The randomised Gauge R&R structure used for the previous model (Chapter 4) was not used here, because of the extra time incurred during analysis.

- Diagrammatical protocols used by participants during the second phase of this study have shown to reduce the range in absolute results reported and reduce the range of inter-participant uncertainties, showing promise for use in future analysis pipelines. This will be further tested in the subsequent complex model (Chapter 6).

- Extreme values in absolute reported results were due to participants either over constraining or under constraining the live cell population within the third gate, or being variable with the final gate applied, over- or under-constraining the quadrant around the desired double positive population. In some cases, lack of knowledge of using controls to set gates led to variance in population metrics.
- Extreme values in uncertainty results were due to participant variability in applying a gate to separate the target cell population from the dead or dying cells, alongside additional variation caused in this first gate by boundary effects on the edge of the visualisation plot, caused by concatenated data that would otherwise be outside the plot axes.

- The performance monitoring diagrams visualised continue to provide a straightforward way to monitor uncertainty performance with respect to the number of people in the study and defined quality satisfaction limits. These will be used in the subsequent chapter (Chapter 6) to monitor uncertainty performance in a more complex gating scenario.

- This study defines participant uncertainty for a highly constrained 5 colour panel cell model, which can be used as an intermediate model for development into more complex cell models, to monitor potential growth of participant uncertainty in more difficult analysis scenarios.

Chapter 6: Complex Uncertainty Model

6.0 Introduction to the Chapter

Chapter 6 introduces the third and final uncertainty gating exercise, as part of the sequence of studies which monitor participant variance in comparison to complexity of data. As discussed in the prelude, this final uncertainty exercise is more complex in panel design and therefore gating difficulty, to further monitor uncertainty in a more difficult analysis scenario. This is more representative of FC analysis pipelines and engineered T-cell markers used to monitor cell therapy products. This allows comparisons to be drawn between all three models in Chapter 7.

Previously in Chapter 5, a five-step pipeline was used to monitor naïve T-cells within primary PBMC populations. This Complex model expands upon the FC panel used in the previous chapter to monitor additional markers for engineered-product specificities, such as programmed cell death and transduction efficiency. This is relevant and translational for the community, therefore potentially providing a more representative application of uncertainty to FC measurements. This Chapter uses an eight-step analysis sequence, and similar to Chapter 5, sees participants analyse this data across two sessions using their own judgement and then using a diagrammatical protocol respectively.

6.1 Chapter Aims

This Chapter develops comparison of uncertainty in complex FC analysis strategies. The fit of this Chapter to the thesis can be seen in Figure 95, specifically within the orange dashed box, providing development for the core hypothesis: as complexity of FC data and processing increases, measurement uncertainty contributed from the participant will also increase. This Chapter further investigates the influence of using measurement uncertainty to quantify subjectivity, and if diagrammatical protocols can aid this situation.

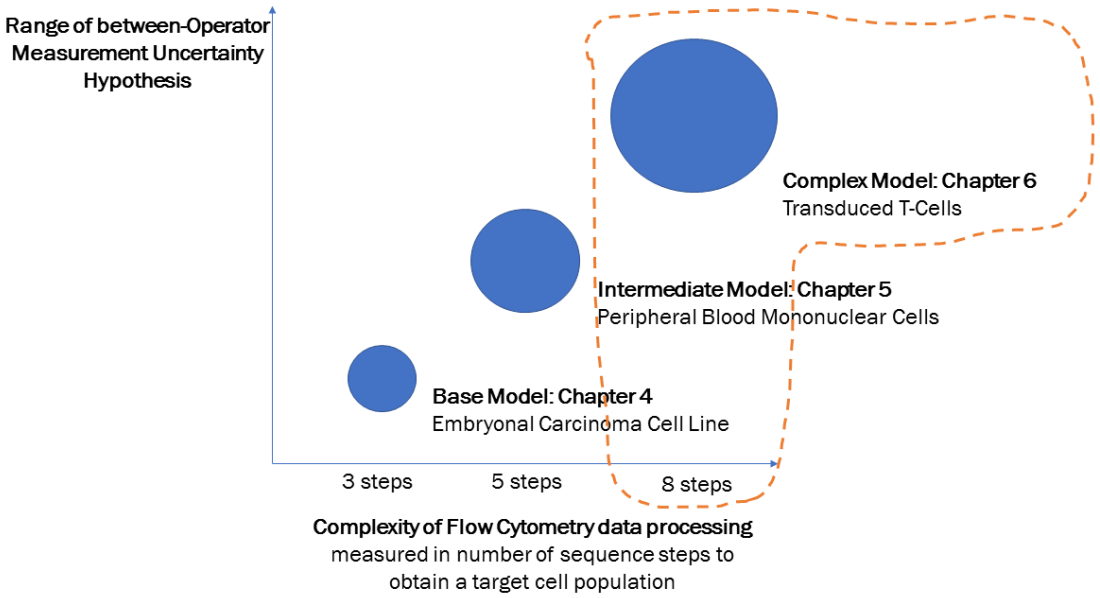


Figure 95 Diagram of Chapter position within the whole Thesis

6.1.1 Chapter Aims & Objectives

The Aims and Objectives of this Chapter can be defined as follows:

- Design a complex two-phase study to measure *CV* and uncertainty of participants when gating complex FC data.
- Identify whether using diagrammatical protocols to apply gates reduces inter-participant absolute reported results and uncertainties.
- Identify what causes extreme values in absolute reported results and uncertainty measurements by reviewing extremes from this intermediate study.
- Further test monitoring tools for understanding uncertainty within a more complex participant study or facility.

6.2 Methodology

Engineered T-cells, derived from Peripheral Blood Mononuclear Cells (PBMCs) were provided by a collaborator, and all staining, file generation and analysis templates were completed by the collaborator, in line with the collaboration agreement. Due to commercial sensitivity, further details on the engineered T-cell vector or cell cannot be provided. Likewise, specific volumes for staining and process steps cannot be provided.

Engineered T-cells have been used for this complex stage because they are representative of current and developing therapies on the market. Current treatments are based upon Chimeric Antigen Receptor (CAR-T) or engineered T-cell Receptor (TCR) treatment methods, using autologous T-cells, derived from patient PBMC material. By using a complex FC panel which develops upon the T-cell lineage panel in the previous model (Chapter 5), it has included more complexity (gating steps and difficulty) and more specificity to a Cell Therapy product. This shows an industrially relevant assay to define the population of transduced T-cells that exhibit appropriate markers for a TCR product.

6.2.2 fcs File Generation

A series of fcs files were generated using the engineered T-cell product created at the collaborator site. Genetically Modified Cells were washed with Cell Staining buffer to remove any cell culture media remaining. 0.1 mL aliquots of the master cell suspension were placed into separate labelled microcentrifuge tubes so there were approximately 1 x 10^6 cells per tube (3 fully stained samples, 1 unstained sample, 7 FMO controls). The unstained sample was wrapped in foil and placed in a 4˚C fridge because this was not needed until the final analysis.

To detect expression of engineered TCR, a PE-conjugated dextramer reagent specific for the engineered TCR was utilized, according to the manufacturer recommended protocol (Immudex, Denmark). The cells were then stained for additional antigen markers, according to the following

stain protocols in Table 53 and Table 54. Staining was performed according to manufacturer recommendations and standard flow cytometry staining protocol. Isotype controls for each marker were not produced because these are no longer recommended for use, as FMO controls are more effective and economical. The FMO controls were stained with all stains aside from the stain aligned to that specific channel. This is to monitor any fluorescence spillover into the required channels from other markers being used. Fully stained samples were stained with collaborator-optimised volumes of all the antigen markers in use.

The antigen markers used for the FMO controls (Table 1) and fully stained samples (Table 2) are:

- Miltenyi CD197 (CCR7) – VioBlue, human (Cat Number: 130-117-353)
- Miltenyi CD4 – VioGreen, Human (Cat Number: 130-113-221)
- Miltenyi CD3 – FITC, Human (Cat Number: 130-113-128)
- Miltenyi CD45RA – PE-Vio770, Human (Cat Number: 130-113-357)
- Miltenyi CD95 (FAS) – APC, Human (Cat Number: 130-113-070)
- Miltenyi CD8 – APC-Vio770, Human (Cat Number: 130-110-681)
- Miltenyi 7AAD – PerCp-Vio770 Staining Solution (Cat Number: 130-111-568).

Table 53 Staining volumes for FMO Controls

Channel	Antigen Marker	VioBlue FMO	VioGreen FMO	FITC FMO	PE FMO	PE-Vio770 FMO	APC FMO	APC-Vio770 FMO
405 nm laser, 450/50 filter	CCR7 VioBlue	X						
405 nm laser, 525/50 filter	CD4 VioGreen		X					
488 nm laser, 525/50 filter	CD3 FITC			X				
488 nm laser, 585/40 filter	Dextramer PE				X			
488 nm laser, 750LP filter	CD45RA PE-Vio770					X		
635 nm laser, 655-730 filter	CD95 APC						X	
635 nm laser, 750LP filter	CD8 APC-Vio770							X

Table 54 Staining volumes for Fully Stained Samples

Channel	Antigen Marker	Fully Stained Sample 1	Fully Stained Sample 2	Fully Stained Sample 3
405 nm laser, 450/50 filter	CCR7 VioBlue	X	X	X
405 nm laser, 525/50 filter	CD4 VioGreen	X	X	X
488 nm laser, 525/50 filter	CD3 FITC	X	X	X
488 nm laser, 585/40 filter	Dextramer PE	X	X	X
488 nm laser, 655-730 filter	7AAD PerCp-Vio770	X	X	X
488 nm laser, 750LP filter	CD45RA PE-Vio770	X	X	X
635 nm laser, 655-730 filter	CD95 APC	X	X	X
635 nm laser, 750LP filter	CD8 APC-Vio770	X	X	X

Antigen markers CD3, CD4, CD8 and CD45RA have been previously described in Chapter 5. The naïve T-cell staining panel used has been built upon in this chapter to develop an 8-colour panel to identify engineered T-cells and monitor different product characteristics. C-C chemokine receptor type 7 (CCR7), also known as Cluster of Differentiation 197 (CD197), is expressed in lymphoid tissues and stem cell memory T-cells (derived from naïve T-cells) [192]. Cluster of Differentiation 95 (CD95), or Fas-receptors are indicators of programmed cell death within engineered T-cell products [193,194]. Dextramer is the marker to monitor the transduction efficiency of the viral vector delivering the gene to the T-cell. This is a measure of how many cells will make the functioning protein required for this cell and gene therapy [195]. Finally, the viability stain used is 7-aminoactinomycin D (7-AAD), a nucleic acid stain that emits fluorescent emission spectra when it binds with DNA. This indicates ruptured, dead cells that have exposed nuclei from live cells with intact cell membranes [196].

Cells were run through a Miltenyi MACSQuant Analyser 10 Flow Cytometer (3 lasers, 8 optical channels), once a daily calibration was completed.

Three fully stained sample files were acquired as 3 representative repeats, along with the FMO controls. Files were exported as fcs 3.0 version types for use in Flowjo version 10.0.8r1 third party analysis software [123] and saved as a workspace.

6.2.3 Flow Cytometry Study Organisation

A total of 22 participants from three separate centres (4 from an academic institution, 13 and 5 participants from separate industrial institutions) were invited to complete the study in a quiet analysis space, to avoid distraction and the possibility of others seeing the study content and analysis. Study sessions had a one-hour maximum duration, and participants were shown three Flowjo workspaces, which contained a series of fully stained engineered T-cell .fcs files. Identical files were included in each workspace, and participants were instructed to gate through an eight-plot sequence to identify single cells, target cells, live cells, CD3+ cells, CD4+ CD8- cells, CD45RA+ CCR7+ cells, CD95+ cells and finally to identify transduced engineered T-cells. Flowjo was the choice of platform due to access of the software across all three industrial and participant sites, meaning a higher number of participants were likely to be familiar with the platform.

Participants were also provided with FMO controls in each workspace to aid gate application and were allowed to use whatever manual gating tool on Flowjo they felt best to gate the population in hand. An overall schematic of the gating sequence they were asked to follow is shown in Figure 96, and participants gated each workspace of files separately to ensure a correct quantification of uncertainty through standard deviation calculation.

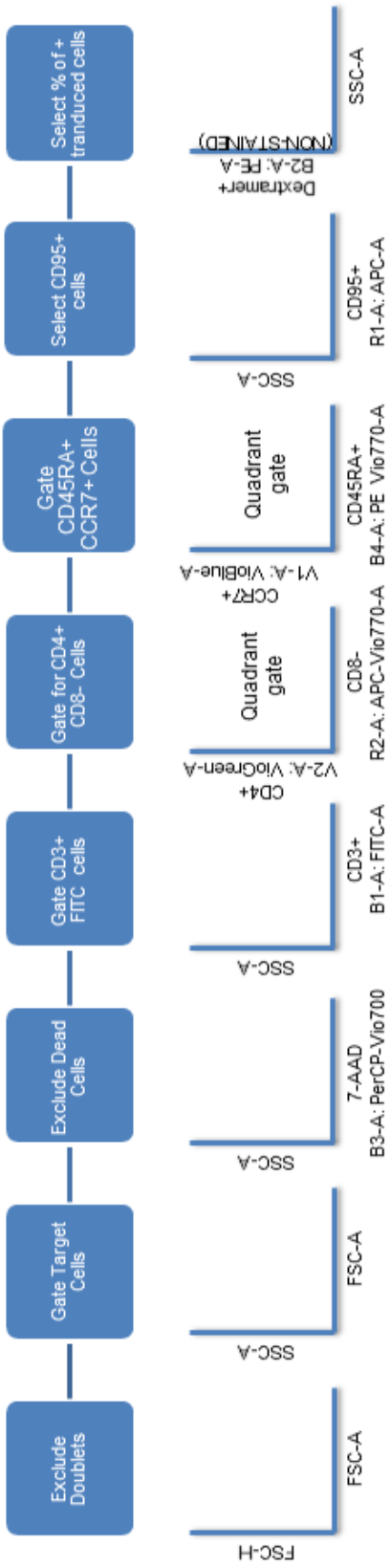


Figure 96 Gating sequence participants were asked to follow, to identify the transduced engineered T-cells.

Similar to the pre-study completed in Chapter 3 and intermediate model in Chapter 5, participants took part in a second phase, where they repeated the same gating process for the transduced engineered T-cell population but were asked to copy a diagrammatical protocol to apply gates instead of using their own judgement, shown in Figure 97. Participants were given the same three fully stained samples along with the FMO controls to aid them in applying the gates in each workspace, alongside the images given in the diagrammatical protocol. Participants followed the same gating sequence provided in Figure 96.

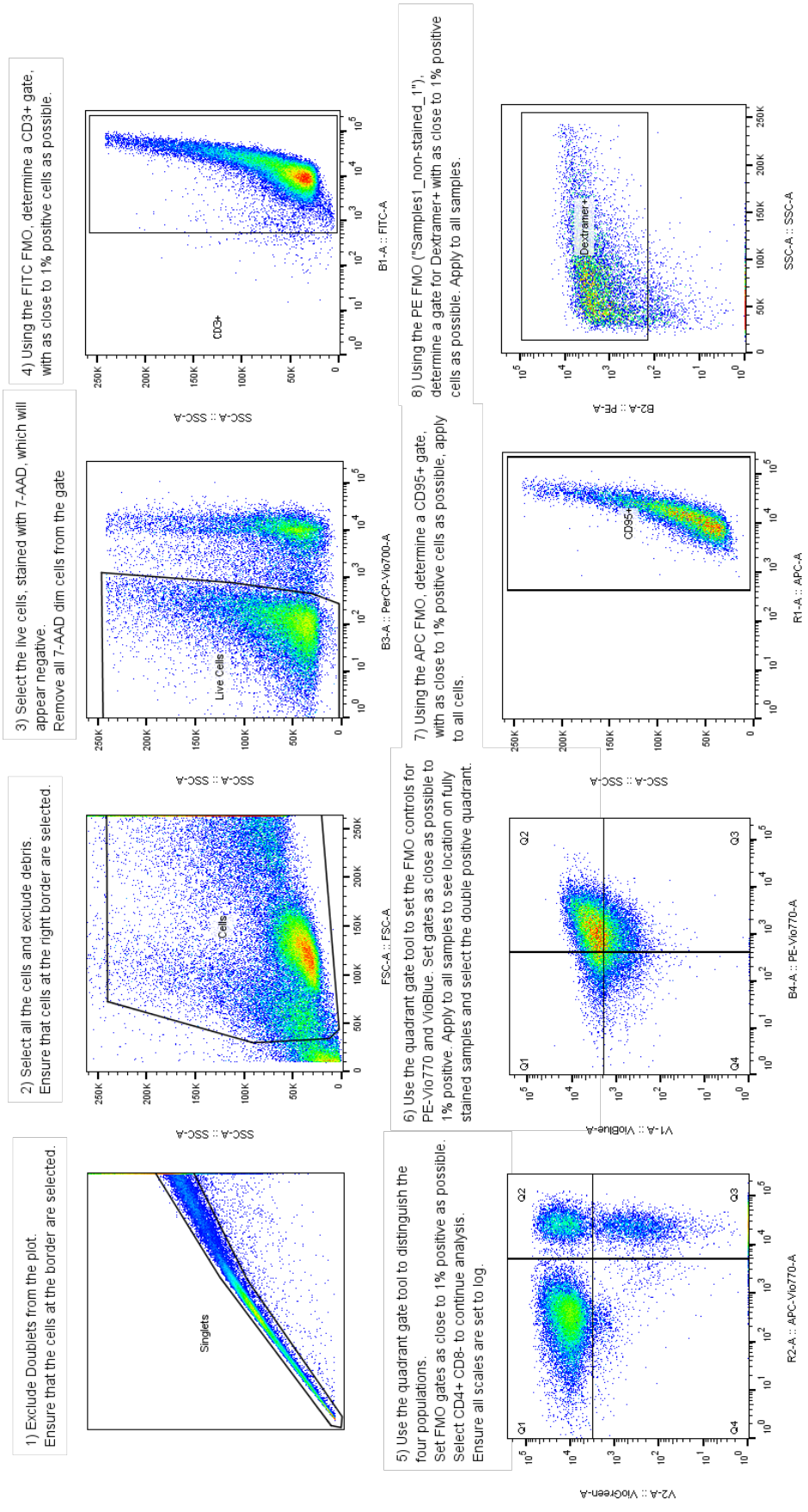


Figure 97 Diagrammatical protocol given to participants to copy gates instead of using personal judgement

This data exercise aimed to build upon previous work in Chapter 4 and 5 to identify if inter-participant variance does exist when reporting results from the same data and that uncertainty values can be calculated for participants to contribute towards an overall uncertainty estimation. It also aims to investigate if the participant range in uncertainty increases as the complexity of the gating exercise increases. This was designed to estimate a general absolute result median and uncertainty. Calculating significant differences between certain groups or testing conditions can be completed in this instance because participants took part in a two-phase analysis exercise, similar to the Pre-study in Chapter 3 and intermediate model in Chapter 5. Standard power calculations to determine appropriate sample size can be generated for future development reference, but also give an indication of current power from the number of participants that took part in the study.

As in the previous uncertainty exercises, participants made 3 repeat measures, within a 1-hour slot, because this was the maximum allowance given for each participant's time, agreed across the three institutions.

6.2.4 Uncertainty Calculation

Once studies had been completed the separate gated cell population metrics were extracted from the data, using the results from the identical repeated file situated in each Flowjo workspace, and were transformed into respective cell count percentages as a function of the original cell event number in the file. These were then used to calculate a mean cell count, *SD* and *CV* for each gating stage, per participant using Microsoft Excel software. Finally, a combined uncertainty (u_c) was calculated by combining these Type A uncertainties by summation in quadrature. The u_c value was expanded with a coverage factor of k = 2, representing a 95 % Confidence Interval for the uncertainty statement, which gave each participant a representative expanded uncertainty (U) figure, to show individual variance. The mathematical methodology used to calculate uncertainty metrics has been previously discussed in Section 2.5. An example of the data extraction through to calculation of metrics and uncertainty can be seen in Figure 98 for this complex model.

Total Cell Counts	
File 1	76012
File 2	72449
File 3	62475

COUNTS	Repeat 1			Repeat 2			Repeat 3		
	File 1	File 2	File 3	File 1	File 2	File 3	File 1	File 2	File 3
Exclude Doublets	54360	50538	43035	55340	51449	43786	55186	51328	43690
Target Cells	28107	25725	21182	27805	25458	20929	27493	25117	20679
Live Cells	26750	24393	19945	26662	24374	19907	26272	23922	19554
CD3+ Cells	26725	24367	19930	26642	24354	19894	26247	23896	19538
CD4+ CD8- Cells	19558	17760	14534	19544	17825	14430	19244	17466	14261
CD45RA+ CCR7+ Cells	6348	6211	4931	8362	8032	6376	7485	7198	5742
CD95+ Cells	6348	6211	4930	8361	8032	6375	7484	7198	5741
Dextramer+ Cells	4475	4330	3437	5931	5595	4430	5099	4809	3875

PERCENTAGES	Repeat 1			Repeat 2			Repeat 3		
	File 1	File 2	File 3	File 1	File 2	File 3	File 1	File 2	File 3
Exclude Doublets	71.52	69.76	68.88	72.80	71.01	70.09	72.60	70.85	69.93
Target Cells	36.98	35.51	33.90	36.58	35.14	33.50	36.17	34.67	33.10
Live Cells	35.19	33.67	31.92	35.08	33.64	31.86	34.56	33.02	31.30
CD3+ Cells	35.16	33.63	31.90	35.05	33.62	31.84	34.53	32.98	31.27
CD4+ CD8- Cells	25.73	24.51	23.26	25.71	24.60	23.10	25.32	24.11	22.83
CD45RA+ CCR7+ Cells	8.35	8.57	7.89	11.00	11.09	10.21	9.85	9.94	9.19
CD95+ Cells	8.35	8.57	7.89	11.00	11.09	10.20	9.85	9.94	9.19
Dextramer+ Cells	5.89	5.98	5.50	7.80	7.72	7.09	6.71	6.64	6.20

FILE 1	AVERAGE	STDEV	CV
Exclude Doublets	72.31	0.69	0.96
Target Cells	36.58	0.40	1.10
Live Cells	34.94	0.33	0.96
CD3+ Cells	34.91	0.34	0.96
CD4+ CD8- Cells	25.59	0.23	0.91
CD45RA+ CCR7+ Cells	9.73	1.33	13.65
CD95+ Cells	9.73	1.33	13.64
Dextramer+ Cells	6.80	0.96	14.13

Coverage Factor	
k	2

uc		2.32
U		4.64

Figure 98 Example of data extraction through to calculation of absolute results and uncertainty per participant

6.3 Results & Discussion

6.3.1 Flow Cytometry Complex Gating Exercise absolute results – Phase 1 Personal judgement

The absolute results reported here are the targeted cell population that participants were asked to identify using the gating sequence defined in Figure 96, during the first gating session where they used their own judgement to apply gates. These are akin to what would be reported in literature for specific cell types, in this instance it is transduced engineered T-cells. The uncertainty of the gating sequence will be discussed in subsequent sections of this Chapter.

Table 55 Measures of Location for the absolute results of the Complex Gating Study using personal judgement (%)

Arithmetic Mean	5.81
Median	5.45
Mode	N/A
Minimum	1.18
Maximum	10.50

Table 56 Measures of Spread for the absolute results of the Complex Gating Study using personal judgement (%)

Range	9.33
25th Percentile	4.04
75th Percentile	7.61
Interquartile Range	3.56
Standard Deviation	1.03
CV	12.10
Median Absolute Deviation	1.73

Table 57 Measures of Skew for the absolute results of the Complex Gating Study using personal judgement (%) (3dp for better resolution)

Skewness	0.168
Skewness standard Error	0.491
Skewness z-score	0.342
Kurtosis	-0.669
Kurtosis Standard Error	0.953
Kurtosis z-score	-0.702

Table 58 Measures of Normality for the absolute results of the Complex Gating Study using personal judgement (%) (3dp for better resolution)

Shapiro-Wilk statistic	0.980
Significance	0.911
Normal/Non-parametric	Normal

Using descriptive statistics to give a general report on the size and shape of the data, the distribution approximates to a normal shape because the mean and median are very close together, as identified in Table 55. This is supported by the skewness and kurtosis *z*-scores (0.342 % and -0.702 % respectively, Table 57) and the Shapiro-Wilk statistical test for normality (p = 0.911 %, Table 58) significantly concludes the distribution is normal. This normality definition is most probably indicated by the spread of the distribution, rather than any specific location parameters. Measures of spread (Table 56) show that the IQR of the participant data was approximately half the size of the range, again indicating normality. These participant average cell count values can be seen in Figure 99. Participant averages are more variable around the median than the previous model, and the error bars show ± 1*SD* from each participant's repeated measures.

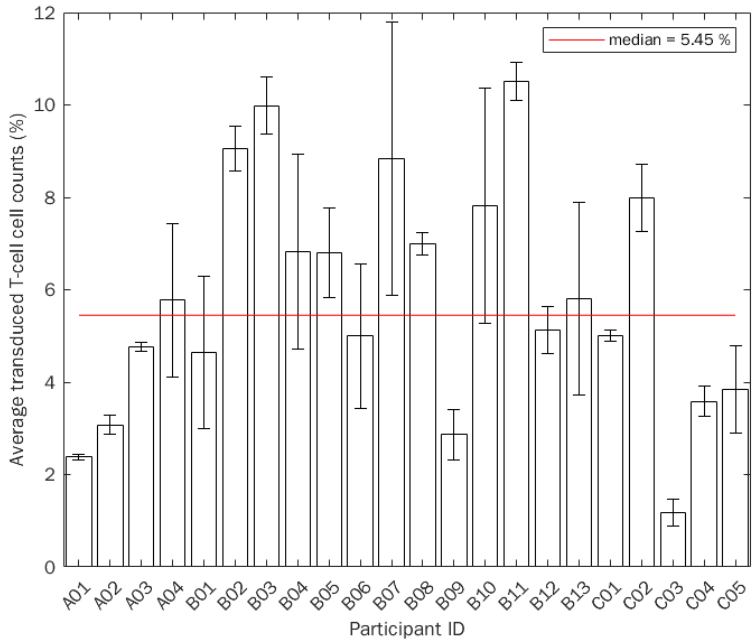


Figure 99 Absolute Results of Target Cell population, represented by each participant's average and ± SD.

Participant deviation from the median (residuals) has been more clearly visualised in Figure 100, with bars depicting each participant's average from the median group value (calculated by

subtracting participant averages from the group median). The *SD* of the total group has also been plotted, because these are most commonly used within traditional manufacturing boundaries to define out of control/out of specification limits. 59 % of participants are within 1*SD* of the median, showing good corroboration of final results. Of those who fell out of bounds one participant had a result below -2*SD*, four participants below -1*SD* and two participants above +1*SD*.

To compare the different types of error boundary estimator that can be used, the histogram of participant average cell counts has been plotted in Figure 101. This confirms that nearly all participants fall within Paxton's Criterion, mean ± 3*SD* and mean ± 2*SD*. Almost all fall inside the trimmed mean ± 2*SD*. A further 3 participants fell outside of the Median ± 2*MAD* range, whereas only 7 participants were contained within the 95 % Confidence Interval applied to this data set.

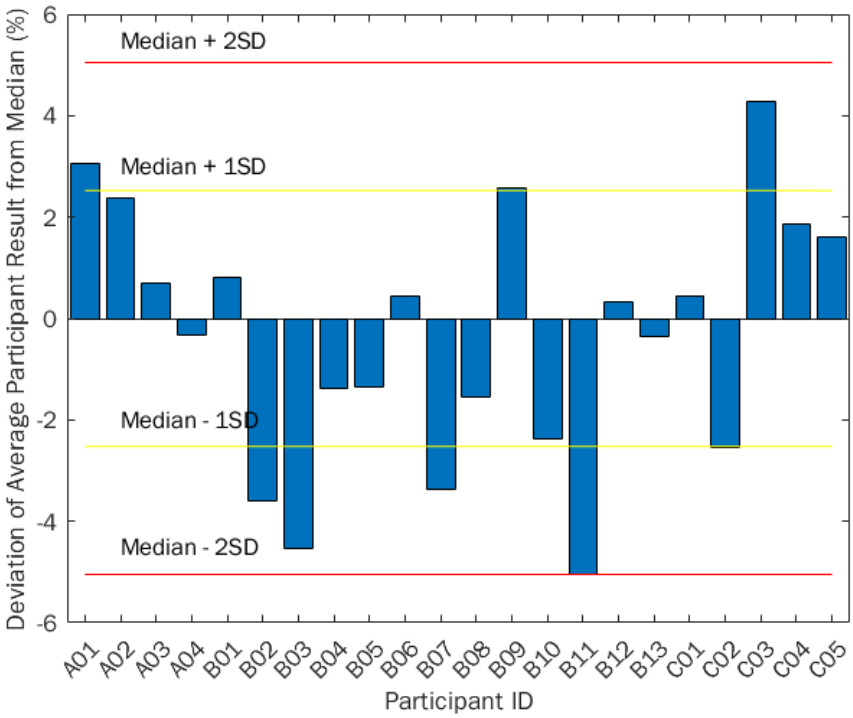


Figure 100 Participant average result deviations from overall group median.

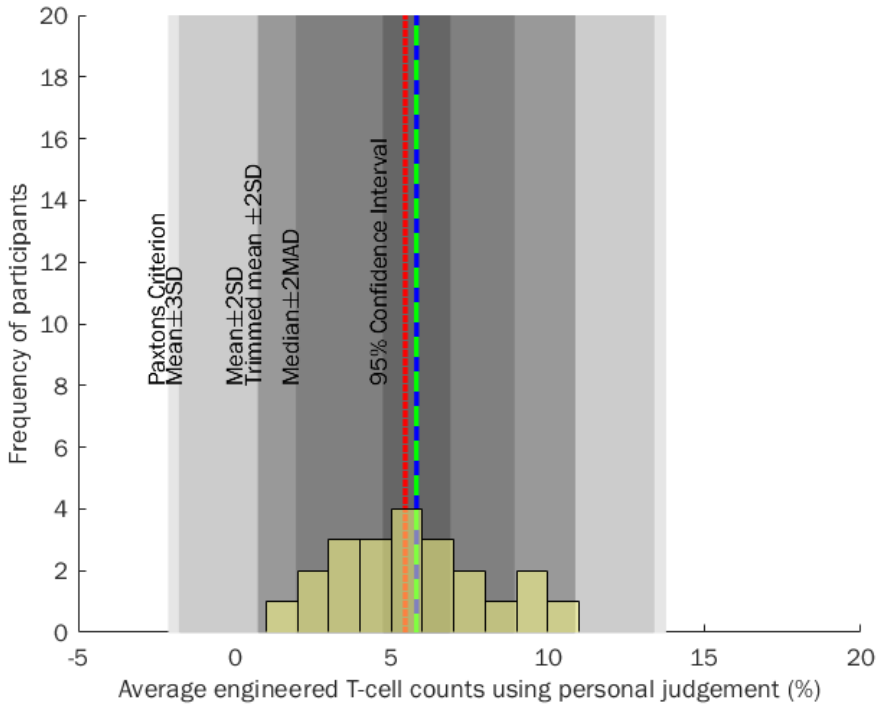


Figure 101 Comparison of average engineered T-cell counts to additional error boundaries

To compare both positive and negative deviation data extremes, Figure 102 shows participant B12's gating strategy, one of the participants very close to the group median value. By comparing the extreme participants to a median participant, there are obvious differences between participants when identifying populations based on the visualised density. This is also coupled with personal preferences on inclusion or exclusion of data points to further refine the data set in search of a particular target. The count average comes from the final gate applied only, but further back in the gating sequence gates have been applied to capture most of the relevant populations. Variation in the transduced T-cell population does not impact the results because there is a very low cell count across this bandwidth (gate 8). The layouts used in this Chapter to visualise the gating strategies are all uniform, with the top row showing gates 1-4 (left to right) and the bottom row showing gates 5-8 (left to right), also shown in Figure 97.

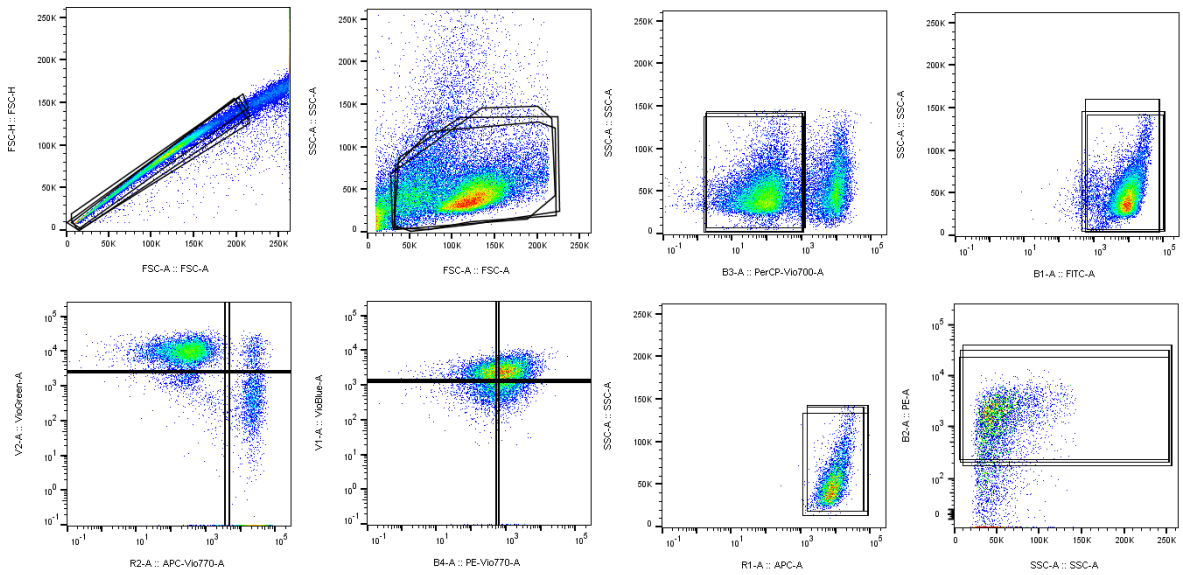


Figure 102 Participant B12 Gating Strategy interpretation, close to median result.

Focusing on those who fell outside of 2*SD*, only one participant (B11) had this much deviation from the median, with an overall percentage cell count of 10.00 %. This gating strategy shown in Figure 103, identifies how over-constraining the cell populations leads to an overall higher cell count than the median. The eight images show the gating sequence steps used to define the final population cell count, with B11's three repeats collated onto each sequence step image. Qualitatively reviewing the participant's gating strategy has shown that most of this bias is due to the gate applied to the lymphocytes (gate 2) and live cells (PerCp-Vio700+, gate 3) which has been applied to a restrained proportion of the population. B11 has applied this gate closer around the main T-cell population in the second gate and applied the separation boundary between the live and dead cells closer to the live cell population in the third gate, to not include the dying cells. Additionally, within the CD45RA+ CCR7+ gate (gate 6) the quadrant includes slightly more of the population than the median user has, which has increased the average cell count overall.

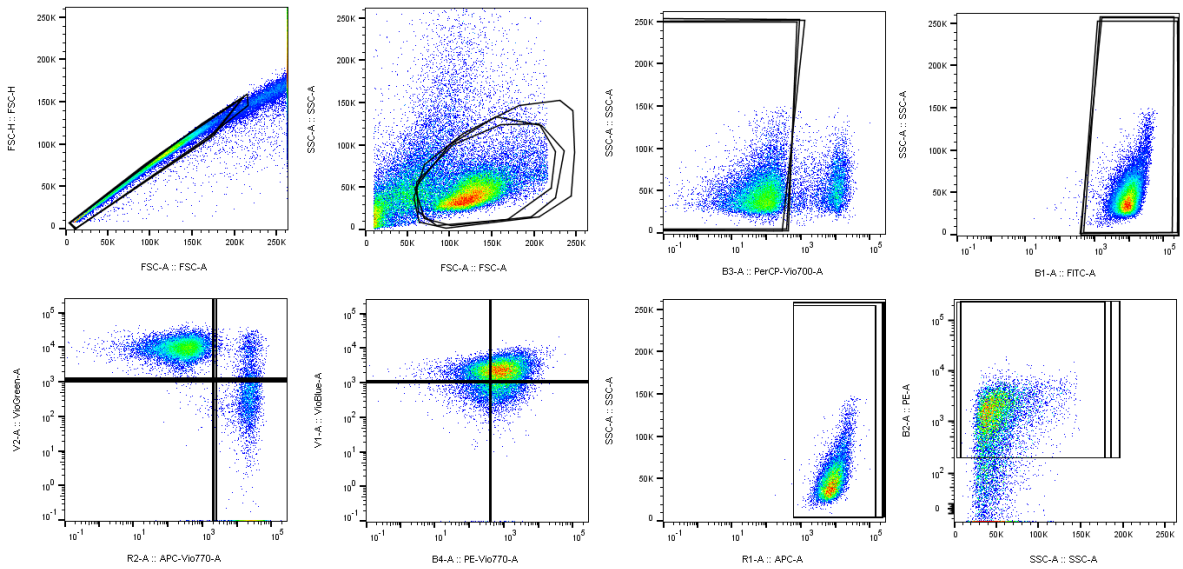


Figure 103 Participant B11 Gating Strategy interpretation

Participants A01, C03 and B09 fell outside of the ± 1*SD* boundary, because these participants have average cell counts lower than the median value. Figure 104 and Figure 105 show the respective gating strategies for the largest deviators, A01 and C03.

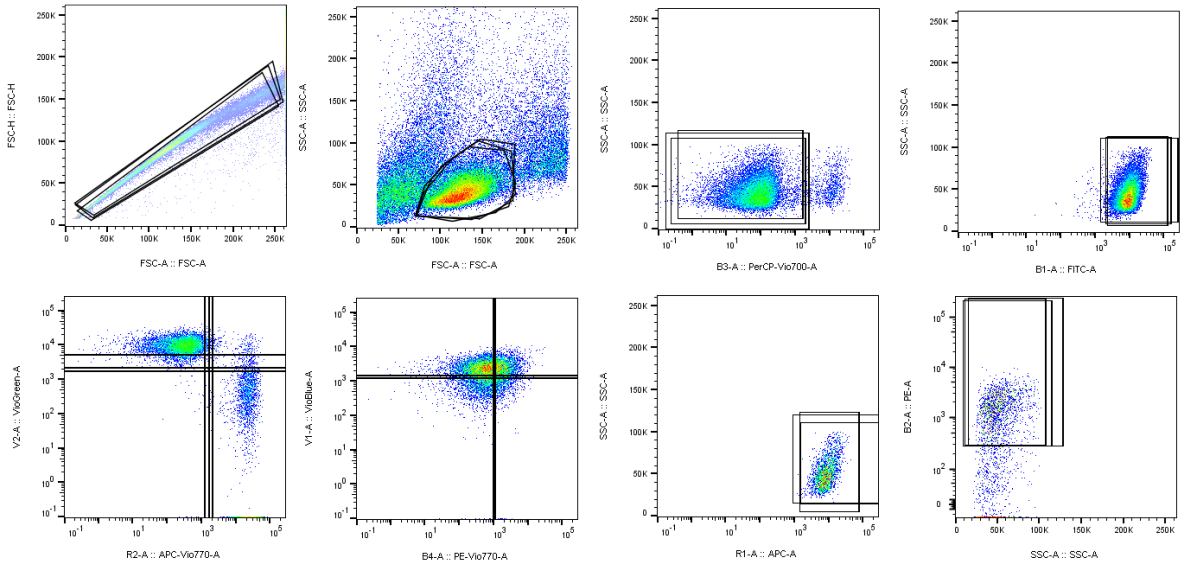


Figure 104 Participant A01 Gating Strategy interpretation

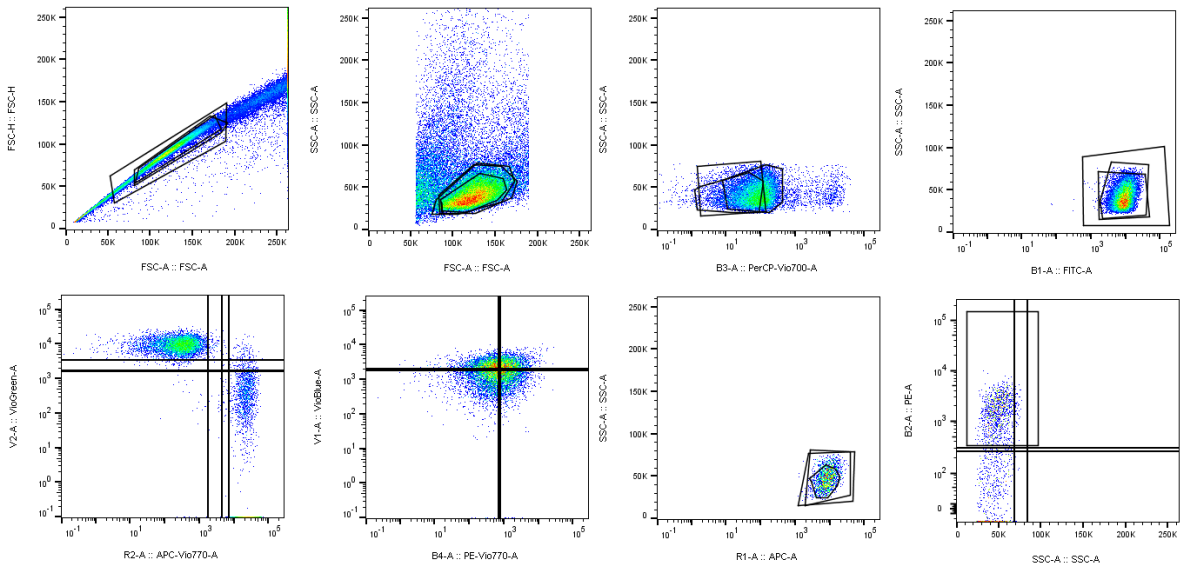


Figure 105 Participant C03 Gating Strategy interpretation

Participant A01 and C03 have lower average results than the median inter-participant value because throughout their gating process they have included less cells than the median user. This is specifically due to the second gate applied in the sequence, which has been more tightly constrained around the central lymphocyte population than the median user. In addition, the sixth gate applied to identify CCR7+ CD45RA+ cells, required participants to use FMO controls to intersect the dense cluster of cells to define those which had double positive expression. The vertical line placed by participants A01 and C03 have been placed to the right of the most dense region of this cluster, therefore selecting a smaller population that are expressing the desired markers, in comparison to the median user, who placed the gate more to the left of the most dense region. In this instance, different applications of gating 'clean-up' procedures and use of FMO controls has potentially caused this variance, alongside the participant perception of density.

Participants B02, B03, B07 and C02 all had average cell counts that fell outside of the median – 1*SD* boundary. To illustrate examples of these gating strategies, Participants B02 and B03 (who

had the highest cell counts within this boundary) can be seen in Figure 106 and Figure 107 respectively.

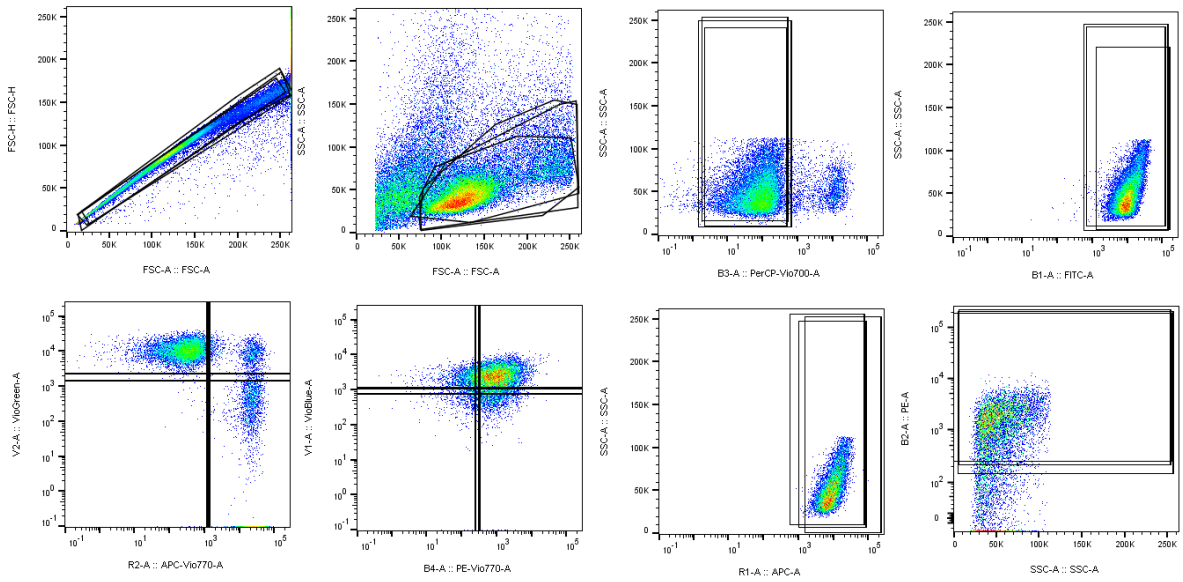


Figure 106 Participant B02 Gating Strategy Interpretation

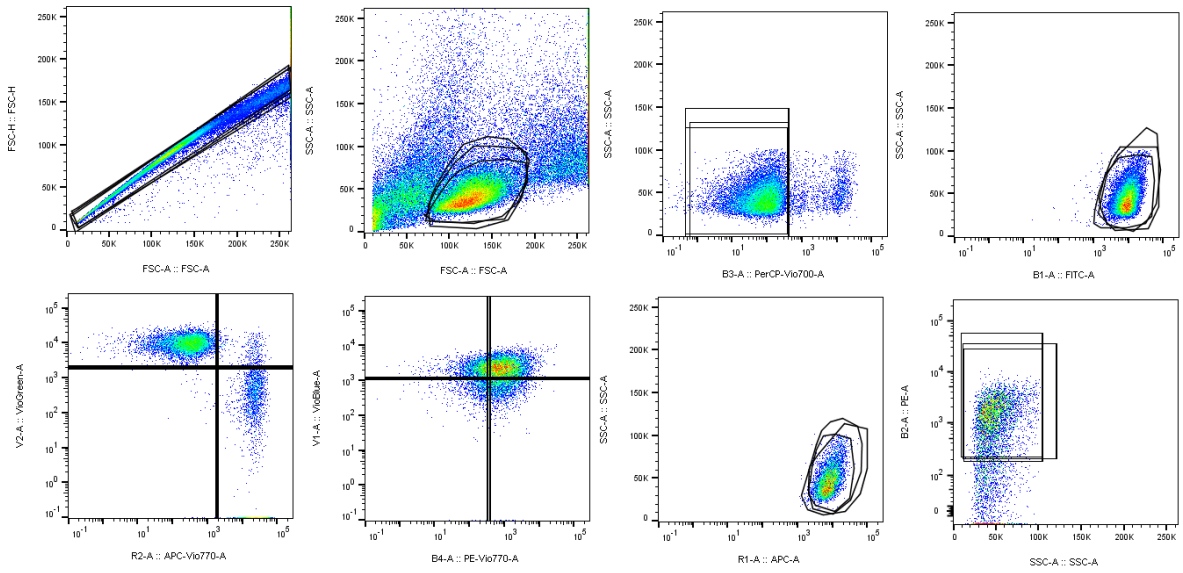


Figure 107 Participant B03 Gating Strategy Interpretation

Participant B02 has a higher cell count than the median participant, which can be seen from their second gate applied, and the sixth gate. The second gate applied includes a second larger cell

population to the right of the cell population in question. The larger overall cell count is also due to the sixth gate applied to identify the double positive CCR7+ CD45RA+ cell population. In contrast to participant A01 and C03, the vertical line has been placed to the left of the densest region of the cell population, including the majority of the population within the final average cell count for engineered T-cells. Participant B03 has also applied the sixth gate in this manner, to end up with the final average cell count population within this boundary, the only difference being they did not gate an additional population of cells within the second gate applied. The difference between these participants within the median +1*SD* boundary and participant B11 who fell outside of the median +2*SD* boundary is that B11 included a greater number of cells within gate 2. This carried forward through the rest of the gates applied, so the average cell count is greater as a result.

This analysis of the absolute results used to represent cell populations shows a 9.32 % cell count range between participants when determining final engineered T-cell population percentages. The further qualitative analysis of the extremes identifies 1 participant (B11) who falls outside of initial control limits. This extreme participant accounts for one third of the cell population percentage range. If this extreme participant value was removed, the range would fall to 7.60 % between participants (minimum value of 2.38 % and maximum value of 9.99 %).

Variability of absolute results is commonly assessed using the *CV*, which combines the average and standard deviation of final cell count measurand (as defined in Chapter 2). The distribution histogram of participant *CV* of reported results can be seen in Figure 108, plotted on top of 3 specification limits derived from the ICSH boundaries used within Chapter 4 and 5. Therefore, this amount of variance is not ideal in this exemplar. The optimal scenario would have all 22 participants (*n*) with < 1 % *CV*, which sets the total height at *y*-intercept of the graph.

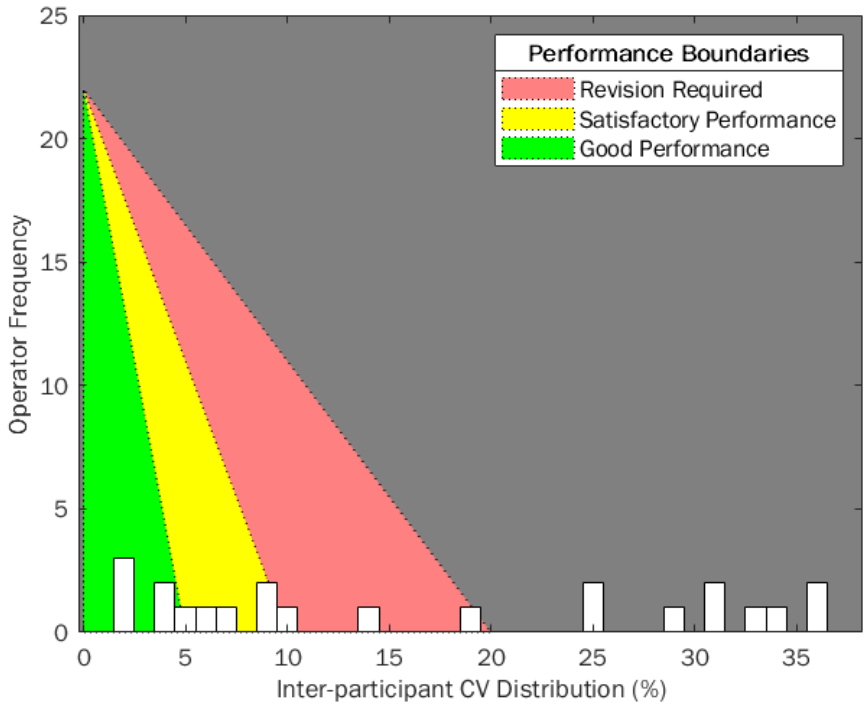


Figure 108 CV Performance of Participant absolute results when using personal judgement to gate complex cell model

Using these guidelines as boundaries, nine participants fell outside of the 'revision required' region due to very high *CV* and an additional two participants fall outside of the 'good' and 'satisfactory' performance regions. These extreme outliers had more variation within the final quadrant gate they drew when identifying the final naïve T-cell cell population. Participant B13 has been used as an example here, because they had the highest *CV*. Participant B04 had a *CV* of 36 %, and their final gates can be seen in Figure 109. B13 intersects the CCR7+ and CD45RA+ population much more on one of the gating repeats, creating a lower value for the final average cell count population of one repeat, carried forwards through the remaining gates. Therefore, there is more variation in the result, shown by the *CV* of the three repeats. This is the case for the other participants who fall outside of this region.

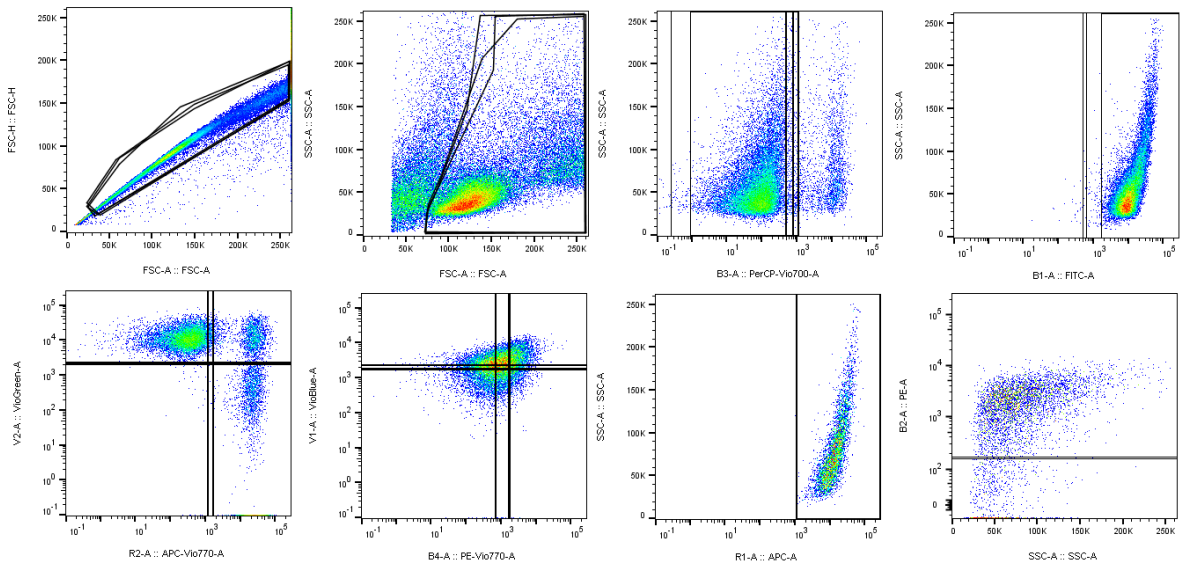


Figure 109 Participant B13 Complex Model Study Gating interpretation

Participants A04, B01, B06, B07, B10, B13, C03 and C05 also had *CV* values that fell outside the 'Revision Required' region of the graph. Similarly, the variability in the final result is caused by the placement of the sixth gate, separating the CCR7+ CD45RA+ population from the rest of the cluster. It appears that variability caused using FMOs at this phase has a knock-on effect for the final cell count, rather than being caused by the final gate applied in the sequence. In some instances, the variability between repeats is carried over from earlier in the gating sequence (typically the second or third gates).

Measurement uncertainty provides a way of combining variability measures (*SD*s) of each gate within the sequence, to provide a measurement that is more representative of the components of the gating sequence. When extremes in measurement uncertainty arise, uncertainty values can be easily deconstructed to identify which part of the gating sequence is responsible for causing variation within the measurement. Measurement uncertainty results for this phase are discussed in Section 6.3.4, once absolute results for phase 1 and 2 of this study have been reviewed and compared.

6.3.2 Flow Cytometry Complex Gating Exercise absolute results – Phase 2 Following Protocol

The absolute results reported here are the targeted cell population that participants were asked to identify using the gating sequence defined in Figure 96, during the second gating session where they used the diagrammatical protocol in Figure 97 and FMO controls to apply gates. A comparison of these results to Phase 1 and the uncertainty of the gating sequence will be discussed in the next section of this Chapter.

Table 59 Measures of Location for the absolute results of the Complex Gating Study using a diagrammatical protocol (%)

Arithmetic Mean	11.98
Median	12.23
Mode	N/A
Minimum	3.36
Maximum	16.60

Table 60 Measures of Spread for the absolute results of the Complex Gating Study using a diagrammatical protocol (%)

Range	13.24
25th Percentile	9.78
75th Percentile	15.06
Interquartile Range	5.28
Standard Deviation	1.47
CV	12.27
Median Absolute Deviation	2.66

Table 61 Measures of Skew for the absolute results of the Complex Gating Study using a diagrammatical protocol (%) (3dp for better resolution)

Skewness	-0.859
Skewness standard Error	0.491
Skewness z-score	1.749
Kurtosis	0.626
Kurtosis Standard Error	0.953
Kurtosis z-score	0.657

Table 62 Measures of Normality for the absolute results of the Intermediate Gating Study using a diagrammatical protocol (%) (3dp for better resolution)

Shapiro-Wilk statistic	0.936
Significance	0.165
Normal/Non-parametric	Normal

Using descriptive statistics to give a general report on the size and shape of the data, the distribution approximates to a normal shape because the mean and median fairly close together (11.98 % and 12.23 % respectively), as quoted in Table 59, however, these values are over double the mean and median for Phase 1. The skewness *z*-score (1.749 %,Table 61), and the low kurtosis *z*-score (0.657 %) also indicate a more normal distribution, supporting these measures of location. This is further supported by the Shapiro-Wilk statistical test for normality (Table 62) which indicates the distribution is normal. The measures of spread (Table 60) indicate that the IQR is approximately half the size of the total range of inter-participant results, which again alludes to a normal distribution shape of the data.

Participant average cell counts and *SD*s when gating following the protocol can be seen in Figure 110. Participant averages are more variable in Phase 2 of this complex study in comparison to the intermediate study, however, qualitatively, the inter-participant average results do appear to lie closer to the median, aside from a couple of extreme values.

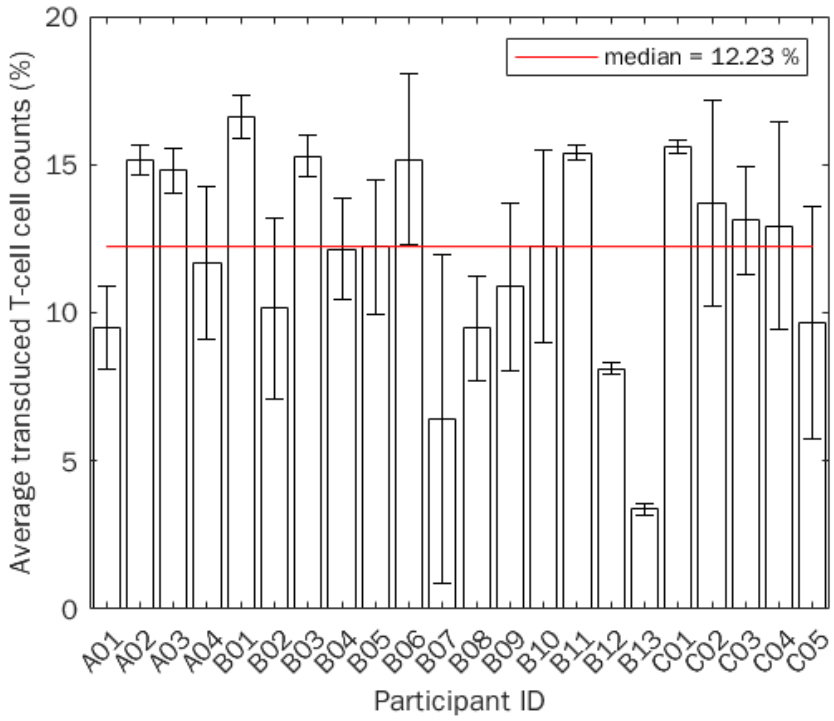


Figure 110 Absolute Results of Target Cell population when following a protocol, represented by each participant's average and ± SD.

Participant deviation from the median (residuals) has been more clearly visualised in Figure 111, with bars depicting each participant's average from the median group value (calculated by subtracting participant averages from the group median). The standard deviation limits have also been plotted, because these are most commonly used within traditional manufacturing to define out of control/out of specification limits. 82 % of participants are within 1*SD* of the median, showing good corroboration of final results. Of those who fell out of bounds one participant had a result above +2*SD*, two participants are above +1*SD* and one participant below -1*SD*.

The application of the Phase 1 error boundaries to the histogram of Phase 2 cell counts, created when participants followed the protocol can be seen in Figure 112. A shift in the whole population can be seen, with more participants returning a higher cell count when following the protocol. This is primarily due to a greater inclusion area specified in the second gate applied in the sequence, which overall has increased the cell count from what most participants highlighted when using their own judgement. Whilst there is still some variability around the final cell count achieved using a protocol, it shows that protocols can be used to improve the assumed accuracy of a gating strategy, to ensure the target cell populations are reproducibly selected. In this instance we cannot have metrological accuracy because a true value is unknown, however, if a protocol is used as a benchmark, then a known value can be taken from the benchmark to aim towards, creating an 'experimental accuracy' rather than a theoretical or traceable one.

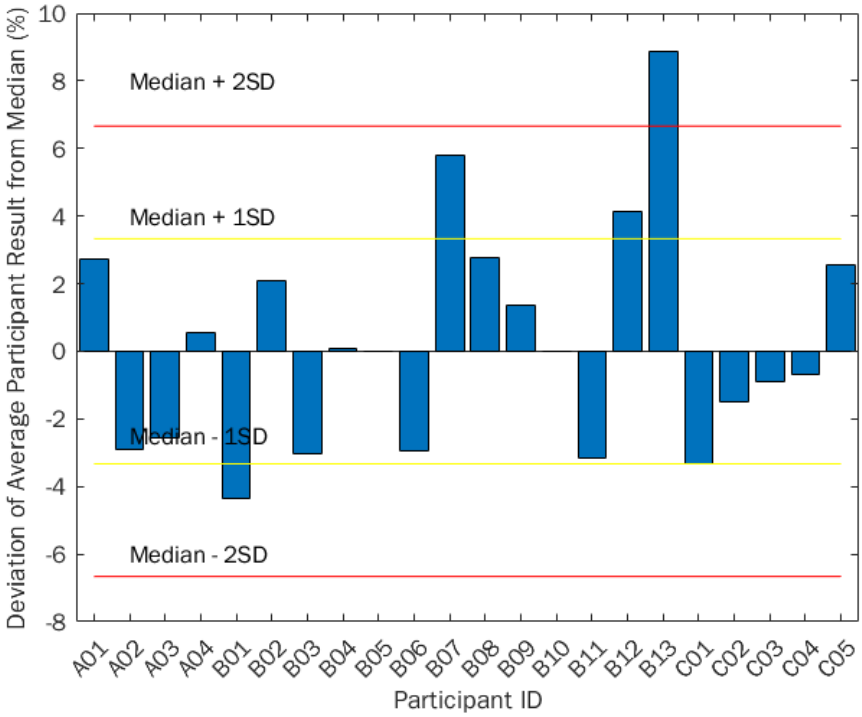


Figure 111 Participant average result deviations from overall group median.

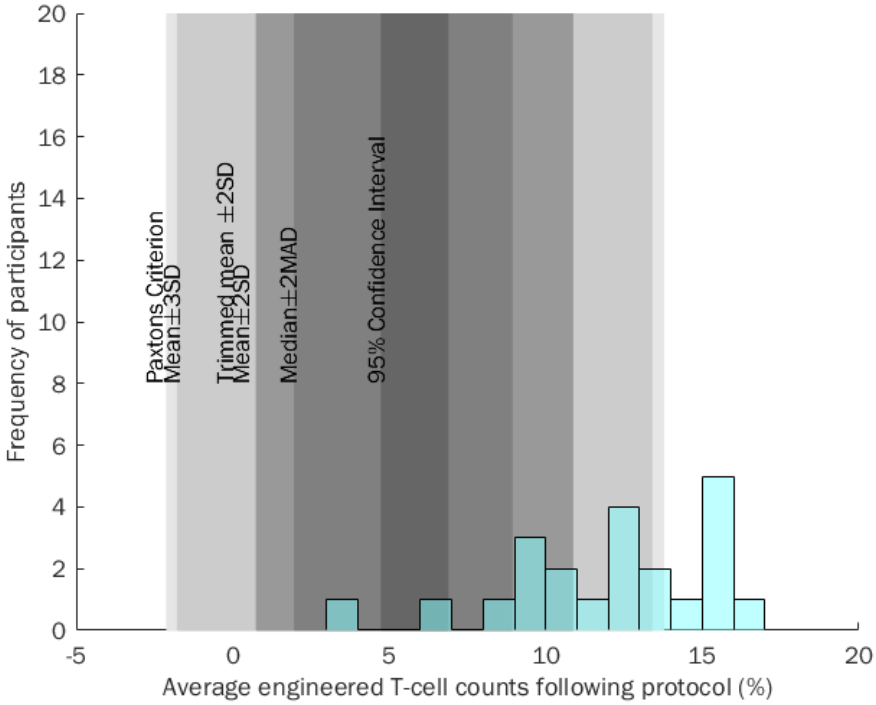


Figure 112 Application of Phase 1 error boundaries to engineered T-cell counts when participants follow a protocol

Qualitatively reviewing the extreme participant (B13) gating strategy outside of the Median ± 2*SD*, highlighted from the residual diagram in Figure 113, has shown that most of this bias could be due

to the fifth and sixth gates applied to the engineered T-cells, because these gates cut off a portion of the CD4+ CD8- cells in gate 5, and the quadrant placed on the sixth gate sits to the right of the densest region of the cell cluster. This cut across the dense region of this population, keeps a lower proportion of cells within the gate boundary. This gating strategy shows how over-constraining the cell population leads to an overall lower cell count than the median, with Participant B13 returning a final cell count of 3.56 % in comparison to the interparticipant median of 12.23 %.

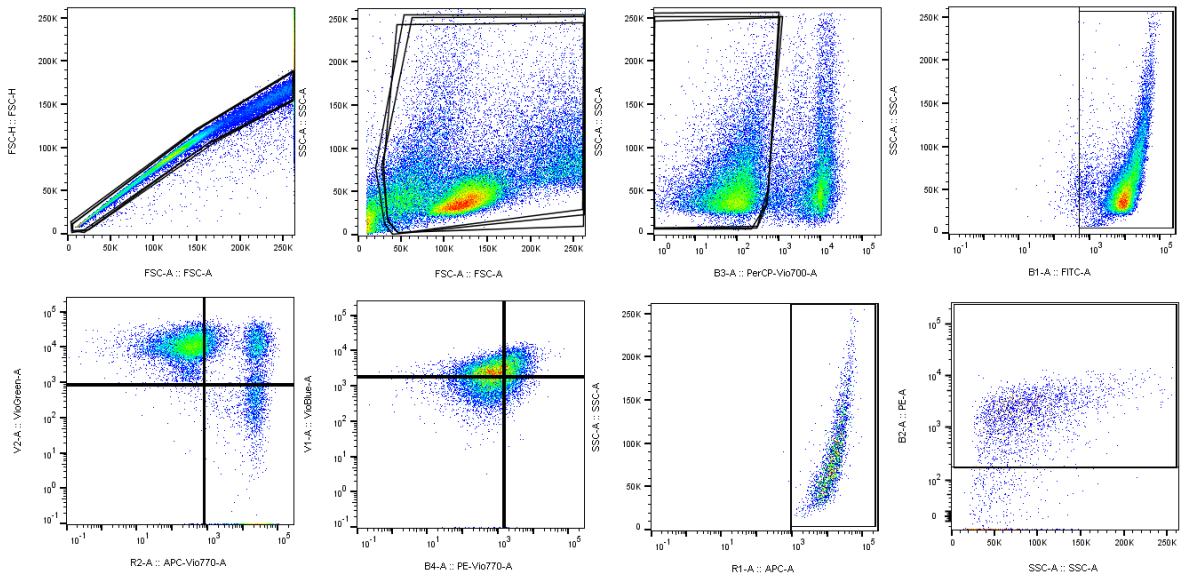


Figure 113 Participant B13 Gating Strategy interpretation when following a protocol

The distribution histogram of participant *CV* of reported results can be seen in Figure 114, plotted on top of 3 specification limits derived from the ICSH boundaries used throughout this thesis. All 22 participants took part in this second gating session so the optimal scenario would have all 22 participants (*n*) with < 1 % *CV*, which sets the total height at *y*-intercept of the graph.

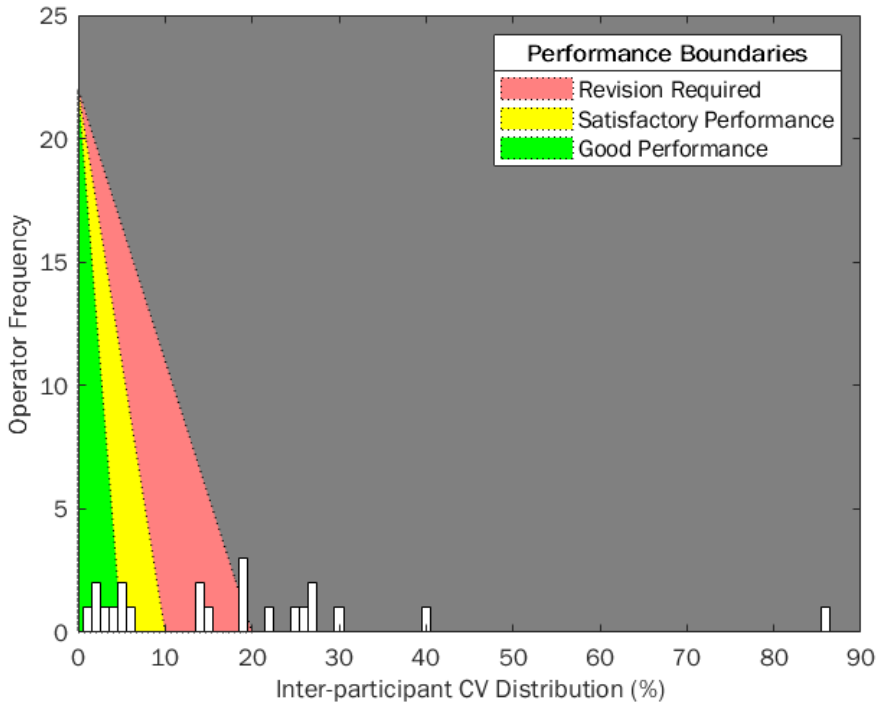


Figure 114 CV Performance of Participant Absolute results when following a protocol during the complex cell model

Using these guidelines as boundaries, eight participants fell outside of the 'revision required' region due to very high *CV* and a further five participants fell just outside of the 'good' and 'satisfactory' performance regions. These extreme outliers had more variation within the final quadrant gate they drew when identifying the final naïve T-cell cell population. Participant B07 had the highest *CV* when gating using the protocol, and their final gate can be seen in Figure 115. This variation comes from gate 6 applied to separate the CCR7+ CD45RA+ population from the remainder of the cell population. The second repeat did not intersect this population in the same way as repeat one and three, because B07 did not include any of the positive population within this gate, causing the large *CV*. The small amount of remaining cells is passed onto the remaining two gates, resulting in a small overall cell count for repeat two, a skewed average cell count and a larger *CV* due to the large *SD* between repeats.

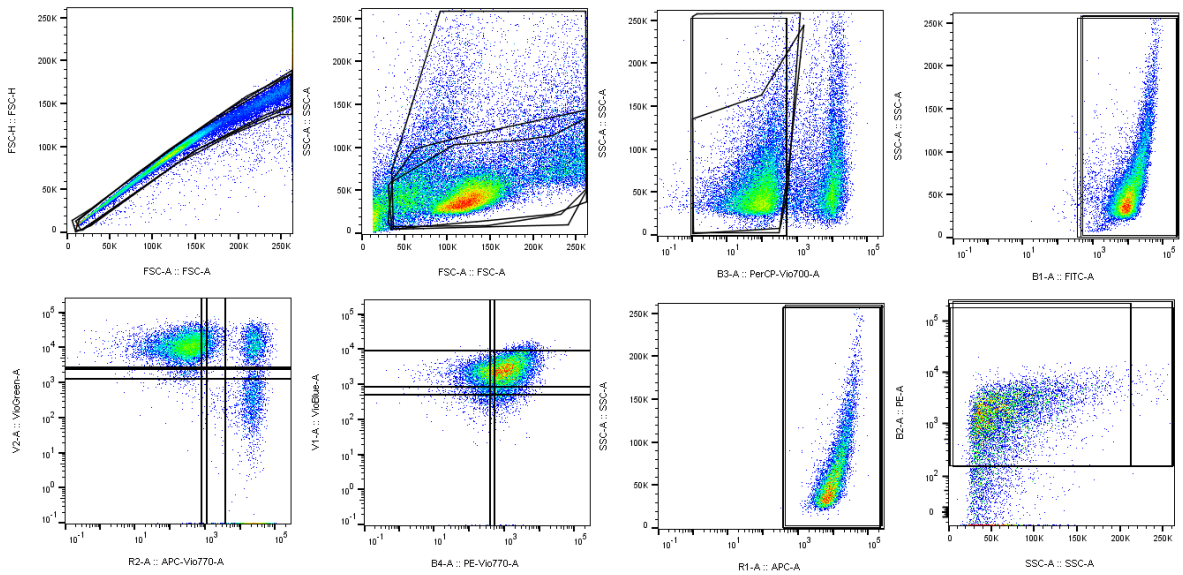


Figure 115 Participant B07 Complex Model Study Gating interpretation

The most extreme participant has been included here as an example of outlying *CV* results. Participant B07 had a *CV* of 86 % due to disparity between the three repeated gates applied to the data. The bulk of this deviation comes from the quadrant gates applied to the data to identify CD4+ CD8- populations (gate 5) and CCR7+ CD45RA+ populations (gate 6). One gate has been applied differently from the other two repeats. This is very noticeable in gate 6, where B07 has removed all of the same double positive population from one of the gates applied. This dramatically reduced the population count carried forward, causing the large *CV* result. The deviated gate was applied because B07 did not use both FMO controls to set the gate limits. One control was set for CCR7+ but B07 failed to cross-reference with the CD45RA+ control, causing the lower population count carried forward.

6.3.3 Flow Cytometry Complex Gating Exercise absolute results – Comparison of Phase 1 and Phase 2

The average cell counts for each participant when they gated following their own judgement and then a protocol have been compiled into the histograms in Figure 116. Any dark orange areas are overlap of the two respective histograms. The range of cell counts has increased by 6.18 % when participants followed the protocol, indicating that protocols may not be as helpful to participants when trying to conform to more reproducible cell counts, or the protocol specified gating areas that were wider than what participants would have personally drawn. Both phases of data have normally distributed populations, so in this instance the protocol has not made the population more kurtosed and repeatable to a specific cell count value.

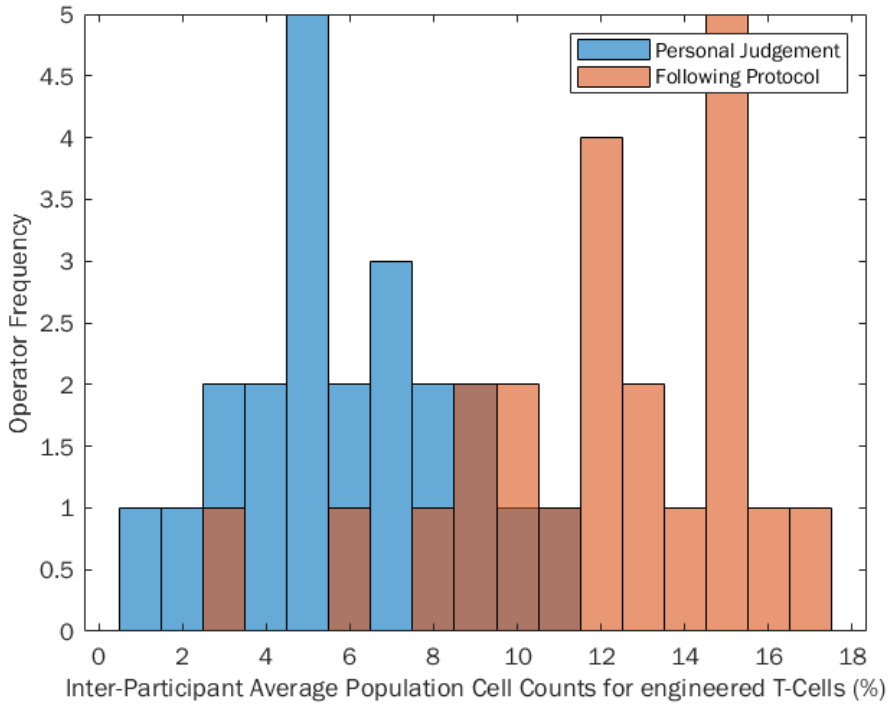


Figure 116 Comparison of Participant Absolute cell counts when gating using their own judgement and when following a protocol (brown areas indicate overlap between the two distributions)

Further comparison of inter-participant data when gating using their own judgement and following a protocol has been completed with a Sign statistical test (Table 63), to compare equality of medians between each test condition. The Sign test results reject the null hypothesis of no

difference in medians, confirming the qualitative histogram conclusion that there is a significant difference between the two medians of the two phases.

Table 63 Sign test results for comparison of Intermediate Gating Study Stages

Null Hypothesis	Test	Sig.	Decision
Median difference of P1 & P2 cell counts = 0	Related-samples Sign test	0.000	Reject null hypothesis

Comparison of these averages has been visualised in Figure 117, plotting the differences of participants average cell counts from the actual cell counts from the protocol gates they were provided to copy in Phase 2. Whilst this only considers the cell count values, not the size, shape or area of the gates, this potentially indicates that when participants copy the protocol, they are more likely to over constrain their gates, than when using their own judgement to place gates.

When applied to a CGT context, this could become dangerous when releasing products due to potential false positives. If an operator includes more cells due to their interpretation of a protocol, this indicates more cells of interest have grown within the therapy product, meeting the required threshold for filtration and patient infusion. If the desired cell count is not met, the therapy would either need more time to sufficiently expand, or it should be rejected when analysed in QC. If not given this time, the product would be filtered and prepared for the patient, but not actually provide the correct therapeutic dose of treatment. This could lengthen the treatment time if more starting material needs to be taken from the patient, causing them more distress, and more inefficiencies within the manufacturing process which in turn increase the costs of cell therapy treatments.

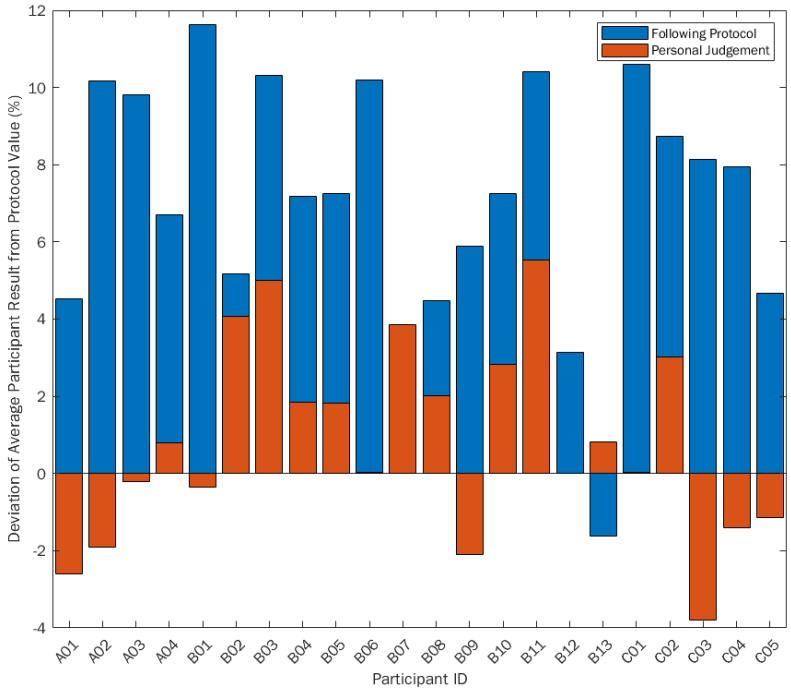


Figure 117 Comparison of Participant averages to the benchmark protocol given to follow in Phase 2

To further compare these two testing conditions the A Priori and Post Hoc power were calculated again (Table 64) to identify whether a suitable number of participants had been gathered based upon the difference in variance of absolute cell counts in each test condition. The use of these power analysis variables has been discussed within the methodology in Chapter 2.

Table 64 A Priori and Post Hoc Power analysis for Complex Study absolute cell counts

Variance Phase 1	Variance Phase 2	A-priori power	Sample size required	Actual power
6.364	11.092	0.800	82	0.343

The A Priori and Post Hoc power analyses in Table 64 show that for the variances achieved between the two test condition average cell counts, 82 participants would have been required to show this difference, to the required minimum desired power of 0.80. The actual power achieved through this study is 0.343, indicating a 34 % probability of the differences in variance of absolute cell counts being due to the two test conditions used and no other underlying factors.

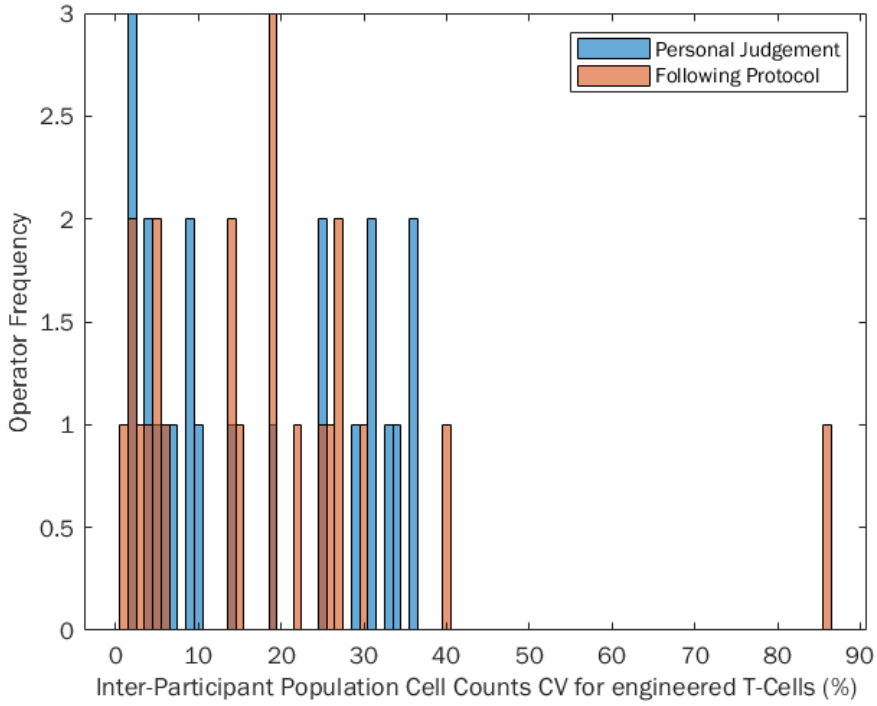


Figure 118 Comparison of participant absolute cell count CVs when gating using their own judgement and when following a protocol (brown areas indicate overlap between the two distributions)

The *CV* of cell counts is commonly reported alongside average cell counts within Flow Cytometry data and in this instance the range of participant *CV*s was greater when participants used the protocol to apply gates to the cell population, shown in Figure 118. The extreme value obtained here is due to FMO controls not being used properly during one repeat, causing a large overall *CV*. Aside from this extreme value the protocol appeared to then reduce the range of results *CV* between participants when they followed this, but this only really considers the final gate applied within the repeats. Measurement uncertainty provides a better way to compile and monitor variation over the whole gating process.

6.3.4 Flow Cytometry Complex Gating Exercise uncertainty results – Phase 1: Personal Judgement

The uncertainty results reported here are a combination of the eight gating stages defined in Figure 96 for gates applied when participants use their own judgement. The uncertainty values have been quantified following the prescribed methodology in Chapter 2, Section 2.5. The uncertainty would better represent variance of measurements with greater confidence, because this combines variability from all gates applied in the sequence, not just the variance of the final gate applied.

Table 65 Measures of Location for uncertainty of the Complex Gating Study using personal judgement (%)

Arithmetic Mean	10.6
Median	6.2
Mode	N/A
Minimum	0.8
Maximum	34.9

Table 66 Measures of Spread for uncertainty of the Complex Gating Study using personal judgement (%)

Range	34.0
25th Percentile	4.0
75th Percentile	13.2
Interquartile Range	9.1
Standard Deviation	10.8
Median Absolute Deviation	3.0

Table 67 Measures of Skew for uncertainty of the Complex Gating Study using personal judgement (%) (3dp for resolution)

Skewness	1.375
Skewness standard Error	0.491
Skewness z-score	2.800
Kurtosis	0.468
Kurtosis Standard Error	0.953
Kurtosis z-score	0.491

Table 68 Shapiro-Wilk test for normality for uncertainty of the Complex Gating Study using personal judgement (%) (3dp for resolution)

Shapiro-Wilk statistic	0.758
Significance	0.000
Normal/Non-parametric	Non-parametric

Unlike the descriptive statistics for absolute results for this study phase, the mean and the median are not close together (10.6 % and 6.2 % respectively), indicating a more skewed distribution, as monitored in Table 65. The median is less than the mean, indicating a slight positive skew to the data. This is further supported by the Shapiro-Wilk test for normality ($p < 0.0005$ %), shown in Table 68, indicating that the distribution is non-parametric in shape, indicating skewness.

There is a wide range (34.0 %) between minimum and maximum participant uncertainties. Table 66 also shows the interquartile range as 9.1 %, indicating a high kurtosis, because half of the data lies within 27 % of the total distribution range. This is supported by the skewness value in Table 67 (2.800 %) with skewness *z*-score falling outside of the ±2.58 boundaries specified for normality in Chapter 2. The raw data for each participant can be seen in Figure 119 with various extremes within the dataset. This distribution shape can be observed within Figure 120, showing the positive skew with a possible bimodal split and 4 larger uncertainty extremes.

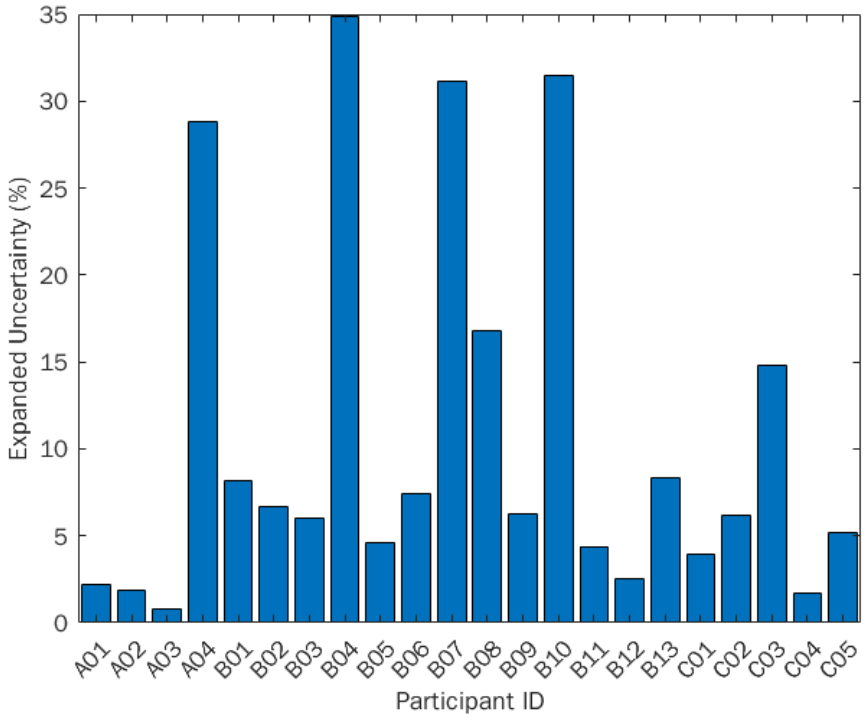


Figure 119 Expanded Uncertainty of all Participant Gating within the Intermediate Model Study

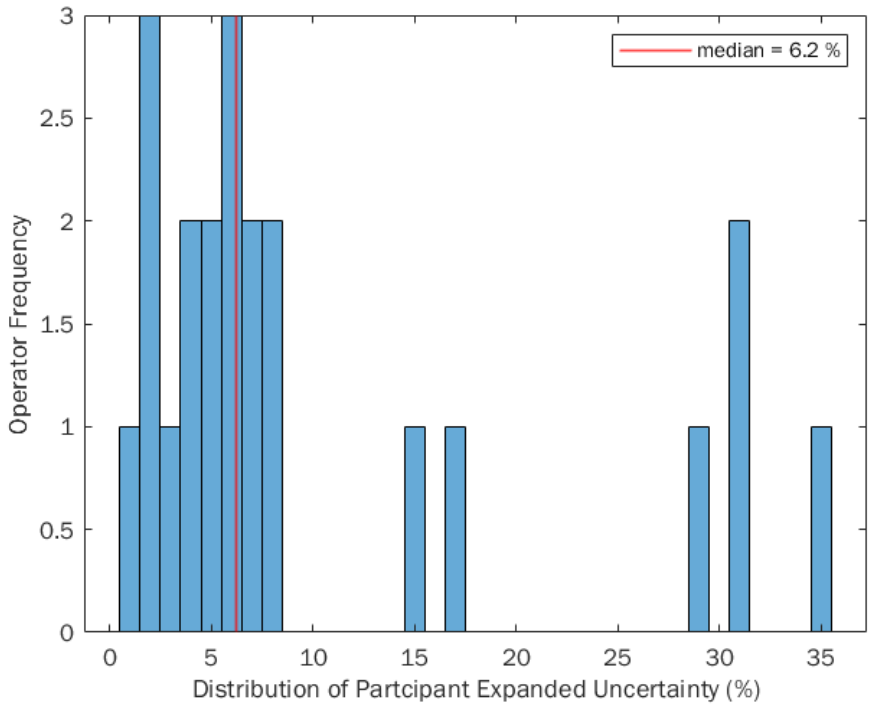


Figure 120 Histogram of Participant Expanded Uncertainty from repeats of Intermediate Model Study

Whilst deviation from a median can help to explain the distribution parameters, when analysing variance, a positively skewed distribution towards zero is preferred. Comparing uncertainty in its size order allows boundaries to be set for permissible specification limits for product release/laboratory quality that increase in value.

Again, in this instance the *CV* specification limits have been substituted for Uncertainty (Figure 122), and no other uncertainty specifications have been defined in the public body of knowledge from research or industry. However, unlike the previous model, this correlation is not strong because of the six participants with extreme uncertainty values, which will be explored within the extremes identified within Figure 121. Even when these extremes are removed the gradient is still low (0.1457*x*).

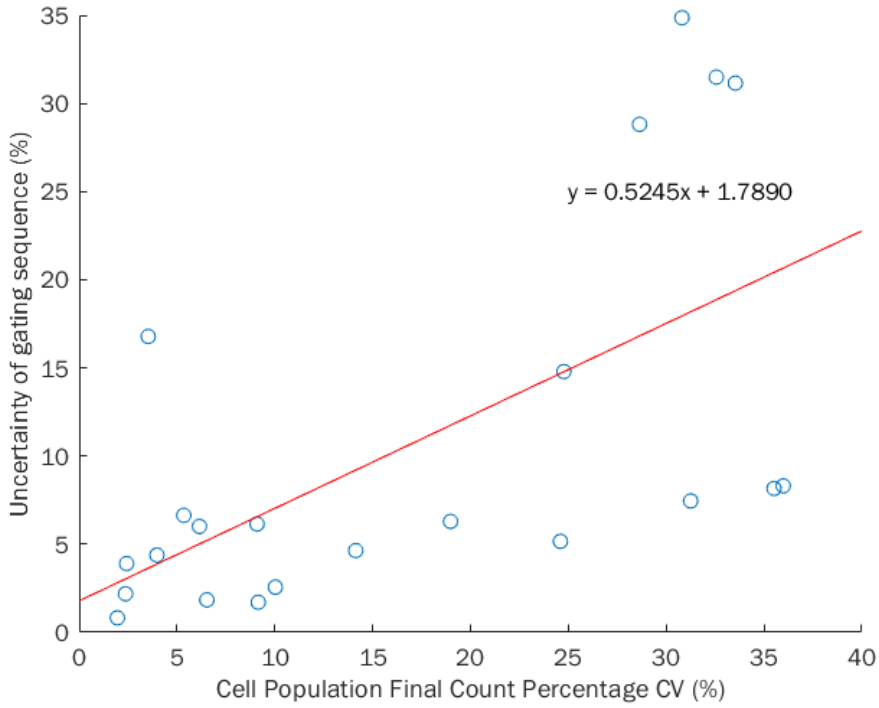


Figure 121 Final Cell Count Population Percentage CV versus Gating Uncertainty for participants

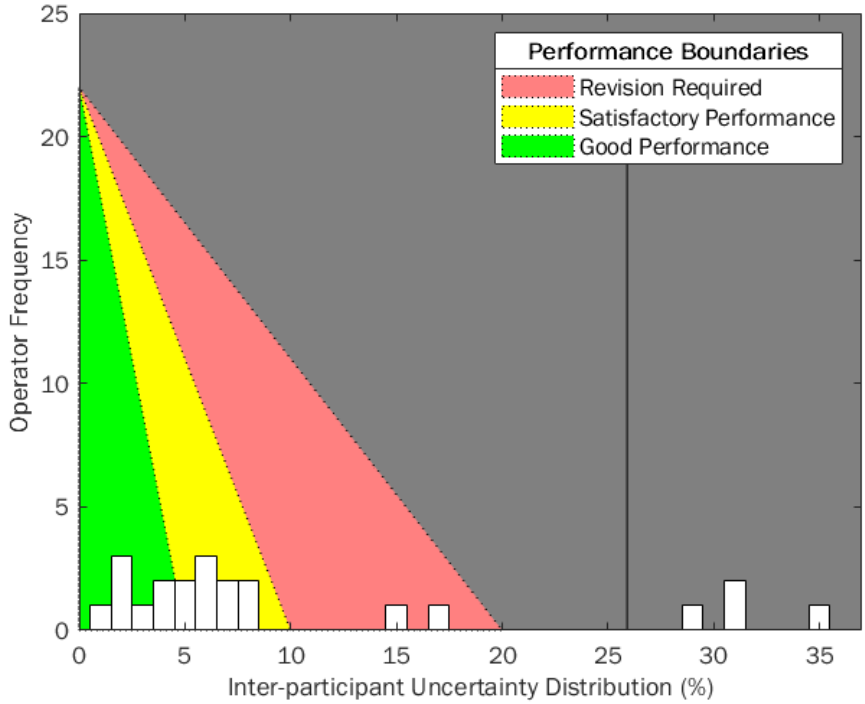


Figure 122 Participant Uncertainty performance monitoring diagram when using their own judgement during the complex model

Participants B04, B10, B07 and A04 had the highest overall uncertainties of the population, which fall outside of the 'Revision required' boundary, and are also above the limit of permissible uncertainty, shown as the black line on the graph in Figure 122.

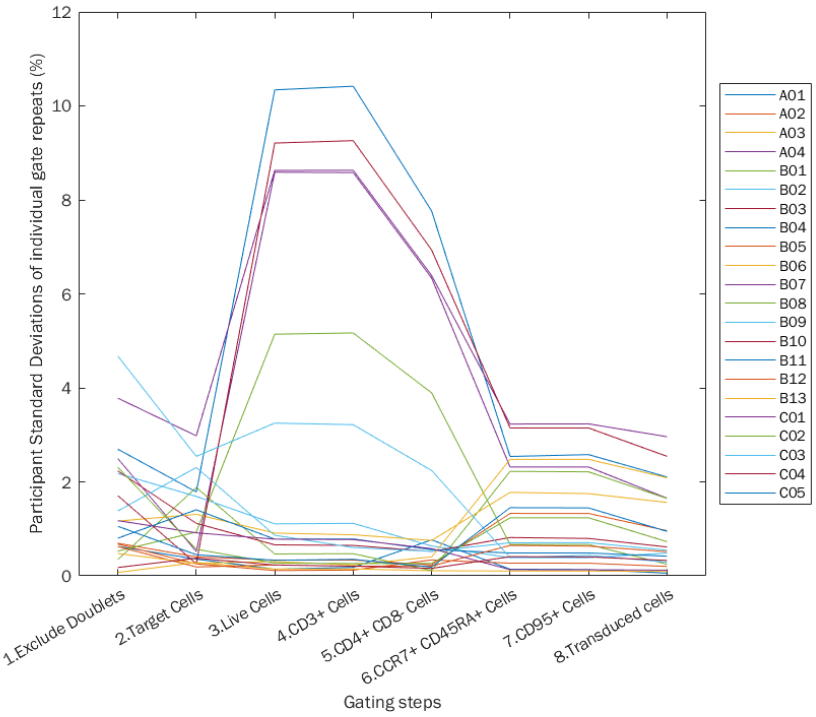


Figure 123 Standard Deviation of each gate applied in the sequence by each participant

The distribution of all participant *SD*s from each gate applied in the sequence can be seen in Figure 123. These standard deviations are combined in quadrature to create the total expanded uncertainty value, used to identify extreme participants. Participants B04, B10, B07 and A04, who are the uncertainty extremes in this instance can be identified as the top four lines of this standard deviation breakdown graph. Visualising the data in this way shows that all four participants had a large variation due the live cell gate (gate 3) applied in the sequence. This had a knock-on effect for the following two gates, which then lowered when gating the CCR7+ and CD45RA+ cells. Interestingly, this was not anticipated because of the difficulty of separation of the population, it was thought that there would be more variability here. However, because of the difficulty of separation, participants adhere to the FMO controls and instructions a lot more, causing the low

variation. As examples, the gating sequence for Participant B10 can be seen in Figure 124, in addition to Participant B04 (Figure 125), because these participants had the largest uncertainties.

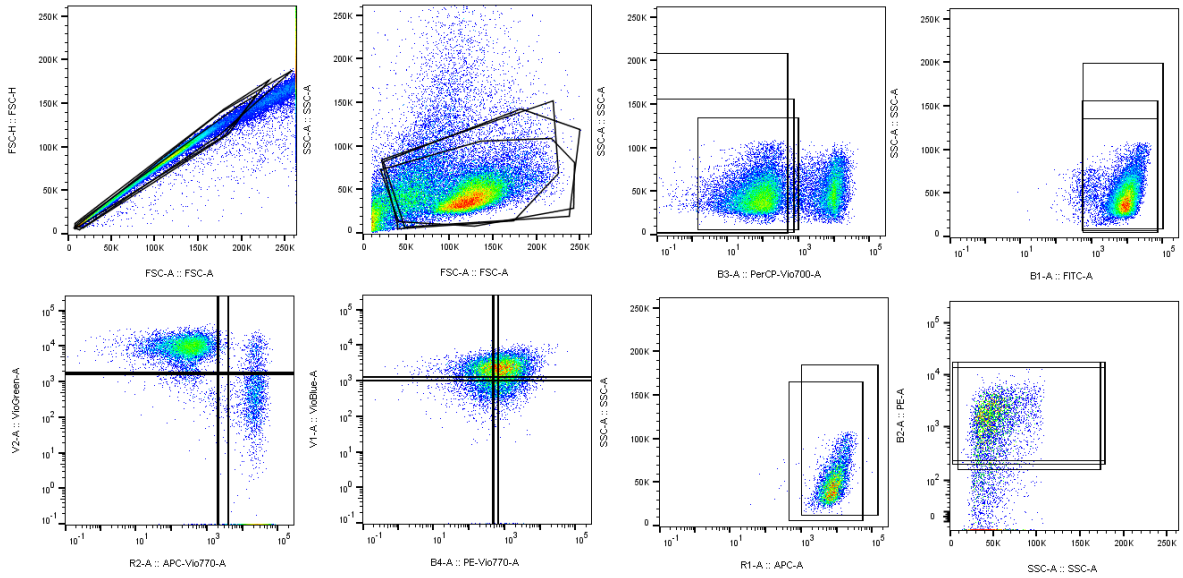


Figure 124 Participant B10 Complex Model Study Gating interpretation

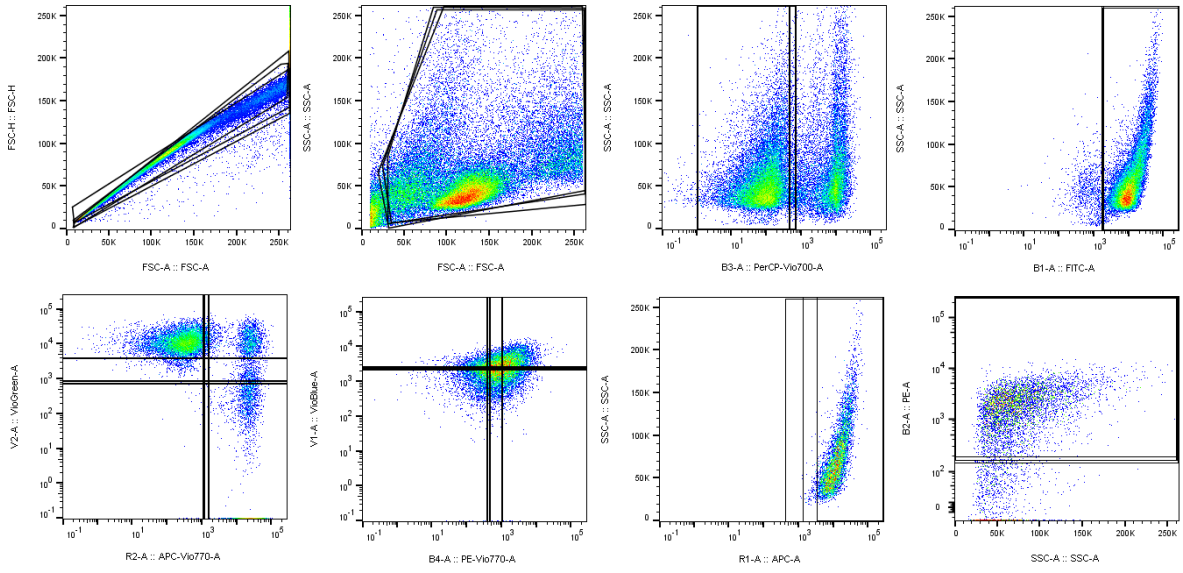


Figure 125 Participant B04 Complex Model Study Gating interpretation

In all instances there is one repeated gate that is significantly larger or smaller than the remaining two. Upon closer inspection, the larger gates all sit against the right, lower edge of the vertical axis and another pseudocolour stripe can be seen running along the edge of the boundary. Inclusion or

exclusion of this boundary effect is causing this variability, which is then carried forward into the CD3+ gate due to the population count. The remainder of the population had very small Standard Deviations, aside from Participants B08 and C03, who had mid-range variabilities. These variabilities are due to one gate being different from the rest, however there are differences between these two participants. Participant C03's gates have already been visualised in Figure 105, and Participant B08's gate can be seen in Figure 126. Participant C03 has variation due to the live gate (gate 3), with two repeats cutting through the population and one repeat placed around the entire population. This was due to the participant's inexperience with using control files to place gates, and a lack of final checking at each level of the gating sequence.

The difference with Participant B08, is that the largest gate applied to the live cell population cuts through the boundary effect seen on the left side, causing the higher cell population and larger standard deviation for that gate. This boundary effect has been investigated in further detail in Chapter 5, when additional cells were compiled together on the right edge of the plot. In this instance, smaller live cells that are not in the fluorescence range have been compiled on the left side. Where this has been cut through to including the lower density region, a spike in the population has been carried forward.

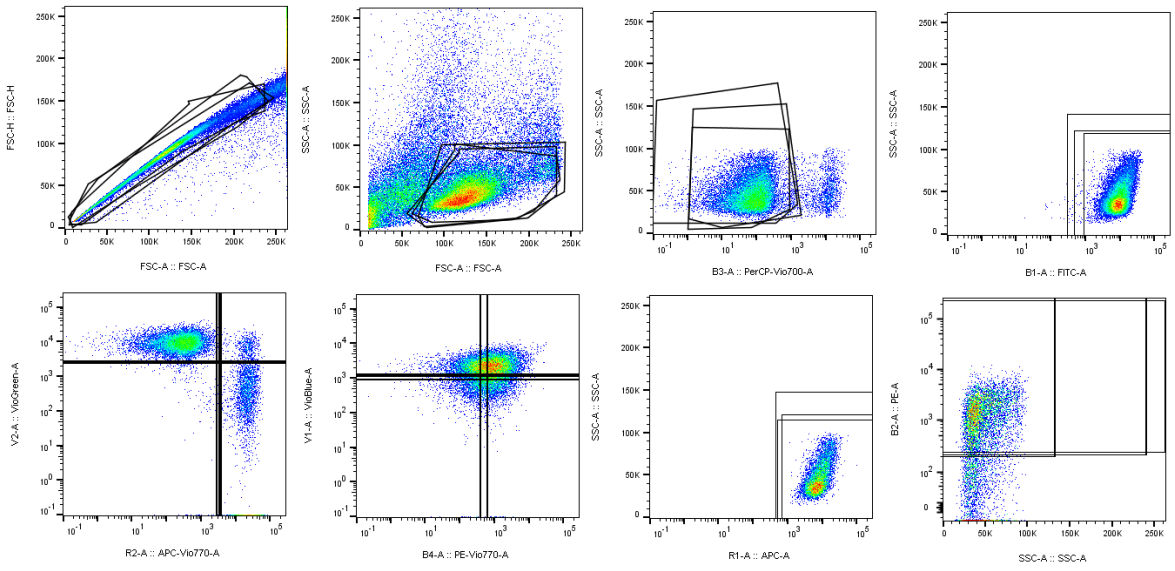


Figure 126 Participant B08 Complex Model Study Gating interpretation

6.3.5 Flow Cytometry Complex Gating Exercise uncertainty results – Phase 2: Following Protocol

The uncertainty results reported here are a combination of the eight gating stages defined in Figure 96 using the gating protocol provided, shown in Figure 97 and FMO controls. The uncertainty values have been quantified following the prescribed methodology in Chapter 2.5. The uncertainty would better represent variance of measurements with greater confidence, because this combines variability from all gates applied in the sequence, not just the variance of the final gate applied.

Table 69 Measures of Location for uncertainty of the Intermediate Gating Study when following a protocol (%)

Arithmetic Mean	22.0
Median	19.3
Mode	N/A
Minimum	2.1
Maximum	43.9

Table 70 Measures of Spread for uncertainty of the Intermediate Gating Study when following a protocol (%)

Range	41.8
25th Percentile	4.4
75th Percentile	38.8
Interquartile Range	34.4
Standard Deviation	16.6
Median Absolute Deviation	17.1

Table 71 Measures of Skew for uncertainty of the Intermediate Gating Study when following a protocol (%) (3dp for resolution)

Skewness	0.083
Skewness standard Error	0.491
Skewness z-score	0.169
Kurtosis	-1.800
Kurtosis Standard Error	0.953
Kurtosis z-score	-1.889

Table 72 S-W test for normality for uncertainty of the Intermediate Gating Study when following a protocol (%) (3dp for resolution)

Shapiro-Wilk statistic	0.836
Significance	0.002
Normal/Non-parametric	Non-parametric

Similar to the descriptive statistics for absolute results for this study, the mean and the median are close together (22.0 % and 19.3 % respectively), indicating a less skewed distribution, as monitored in Table 69. The median is less than the mean, indicating a slight positive skew to the data, which is preferable for uncertainty results. However, the Shapiro-Wilk test for normality, shown in Table 72, indicating that the distribution is non-parametric in shape (p = 0.002 %), indicating more skewness than the mean-median difference shows.

There is a wide range (41.8 %) between minimum and maximum participant uncertainties. Table 70 also shows the interquartile range as 34.4 %, which indicates a low kurtosis, because half of the data lies within 83 % of the total distribution. This is supported by the skewness and kurtosis values (0.169 % and -1.889 % respectively) in Table 71 with skewness and kurtosis *z*-scores inside of the ±2.58 boundaries specified for normality in Chapter 2. The raw data for each participant can be seen in Figure 127 with various extremes within the dataset. This ordered distribution shape can be observed within Figure 128, showing the positive skew with a bimodal split and 9 larger uncertainty extremes. Qualitatively assessing the shape of the distribution shows a different shape to the data than the descriptive statistics, because the mean calculated does not reflect central tendency of the raw data as would be assumed, which will be investigated further.

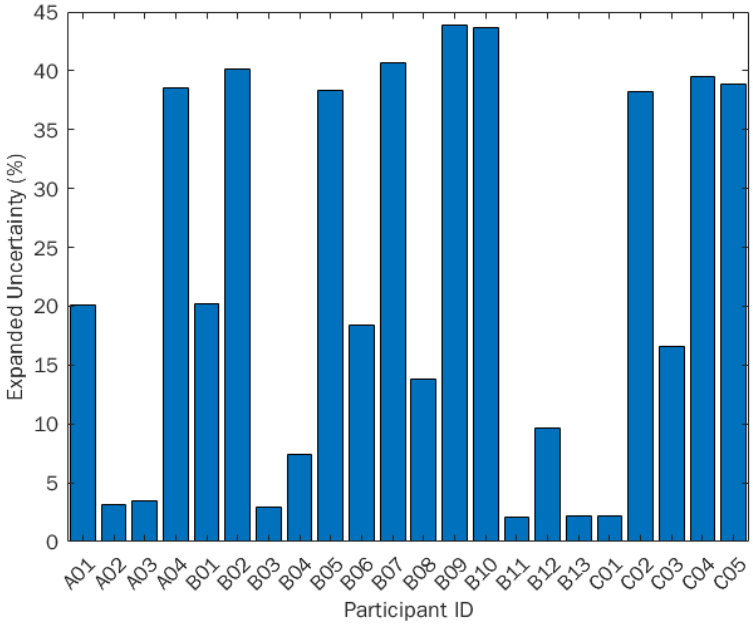


Figure 127 Expanded Uncertainty of all Participant Gating within the Complex Model Study when participants followed a protocol

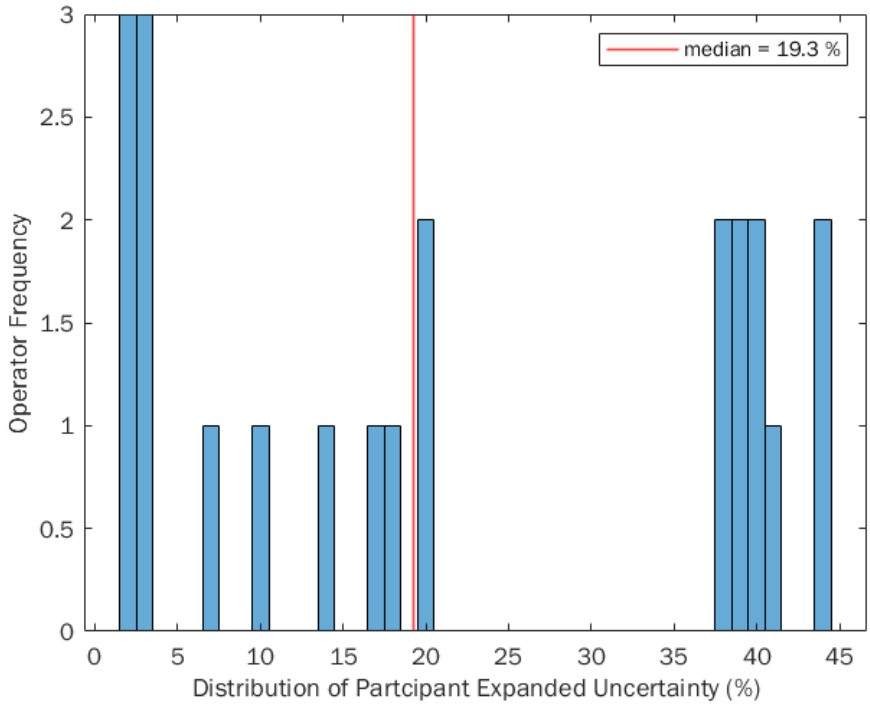


Figure 128 Histogram of Participant uncertainty from repeats of Intermediate Model Study when participants followed a protocol

Whilst deviation from a median can help to explain the distribution parameters, when analysing variance, a positively skewed distribution towards zero is preferred. Comparing uncertainty in its size order allows boundaries to be set for permissible specification limits for product release/laboratory quality that increase in value.

The ICCH and ICCS imprecision values described earlier for measurement *CV* and in previous Chapters have also been used here to define example specification limits if monitoring participant uncertainty. Again, in this instance the *CV* specification limits have been substituted for uncertainty from this gating phase (Figure 129), and no other uncertainty specifications have been defined in the public body of knowledge from research or industry.

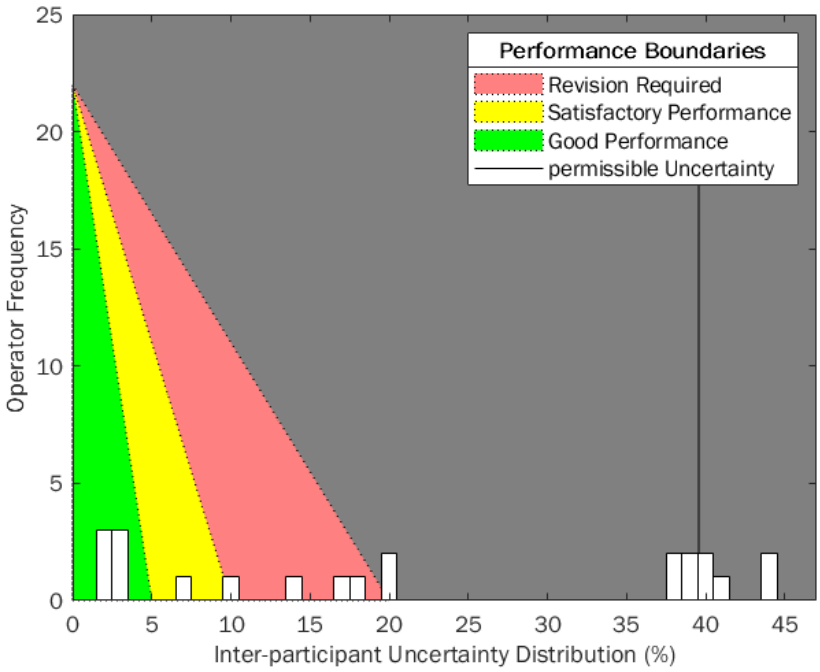


Figure 129 Participant uncertainty performance monitoring diagram when using their own judgement during the complex cell model

There were nine outliers that exceeded the higher control limits with five of these also exceeding the limit of permissible uncertainty. Participants B09 and B10 had the highest overall uncertainty, followed by Participants B07 and B02, who all exceeded the limit of permissible uncertainty, as well as exceeding the defined control limits. However, again there qualitatively appears to be a high variance participant cluster which has its own maxima at 39 % uncertainty.

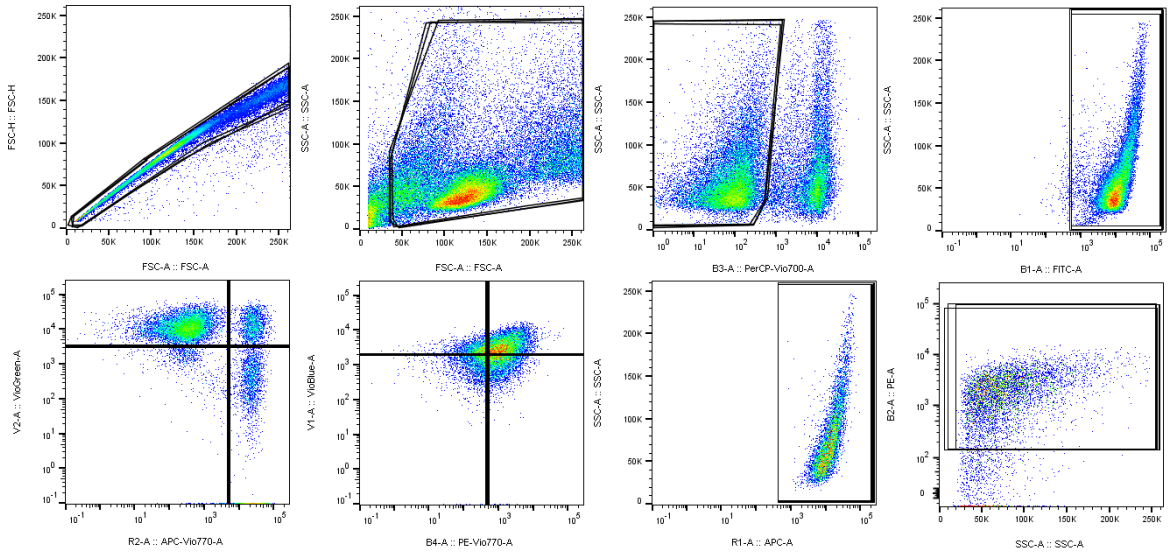


Figure 130 Participant B09 Complex Model Study Gating interpretation

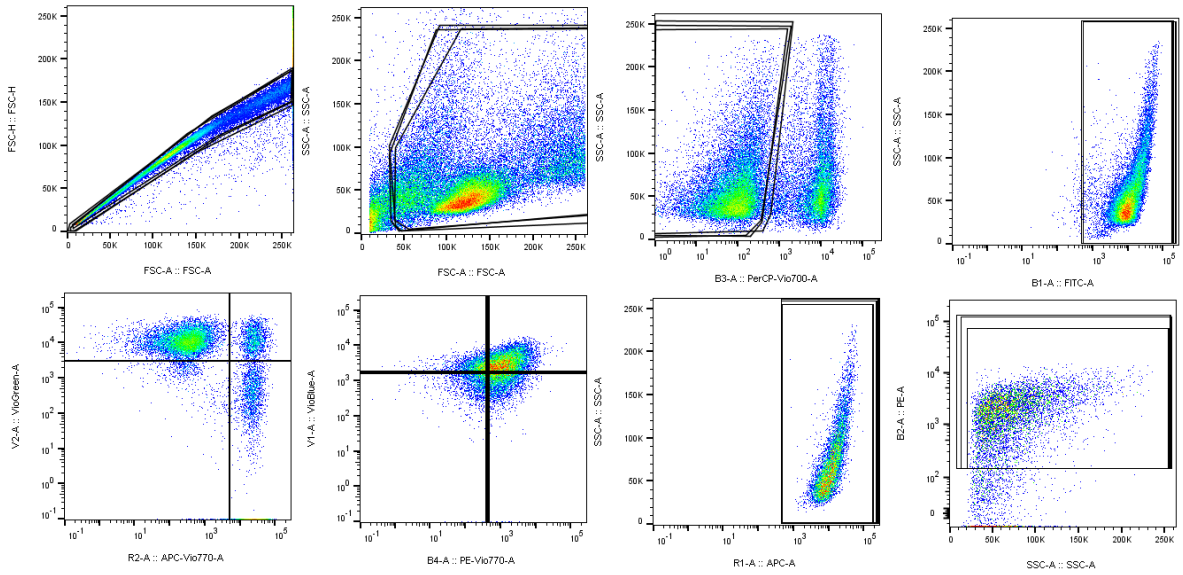


Figure 131 Participant B10 Complex Model Study Gating interpretation

Participant B09 (Figure 130) and B10's (Figure 131) most variable gates were the on the third plot in the sequence. The repeats are very uniform in size and structure, with a repeatable cut-off separating the primary population from the dead and dying cells. The variability appears to come from differences in the left edge of the gate, where the gates applied cut through a data spike that sits on the boundary of the plot, as first discussed in Chapter 5. In this instance, the axes of the plot could not be scaled up to better see this spike, because this spike always sits on the left

boundary, at the lowest scale point of 10^{-3}. This data spike could be an amalgamation of all the data points that exceed the plot limits, so they have been complied and added to the boundary, however, there is no information from Flowjo on this visualisation effect.

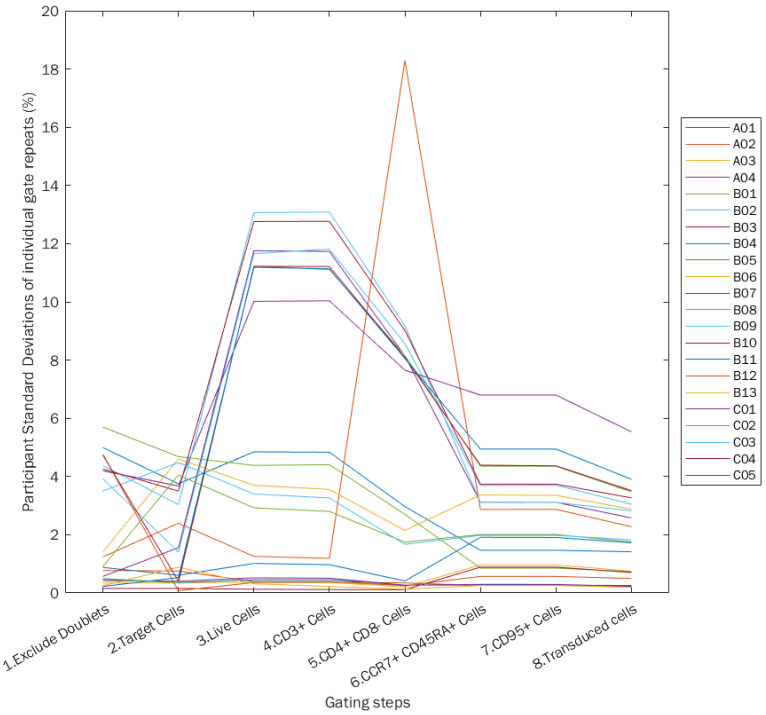


Figure 132 Standard Deviation of each gate applied in the sequence by each participant

The distribution of all participant standard deviations from each gate applied in the sequence can be seen in Figure 132. These standard deviations are combined in quadrature to create the total expanded uncertainty value, used to identify extreme participants. Participants B09, B10, B07, B05 and B02, who are the uncertainty extremes in this instance can be identified as the top lines of this standard deviation breakdown graph. Visualising the data in this way shows that these participants had a large variation due the live cell gate applied in the sequence, like the previous session where no protocol was used. This had a knock-on effect for the following two gates, which then lowered when gating the CCR7+ and CD45RA+ cells.

In a manner similar to Phase 1 uncertainty, there is a mid-range uncertainty group, populated by Participants A01, B01, B06, B08 and C03. These participants had a mid-range standard deviation

when gating the live cell population in gate 3. Examples of these can be seen in Figure 133 and Figure 134 for participants A01 and B01.

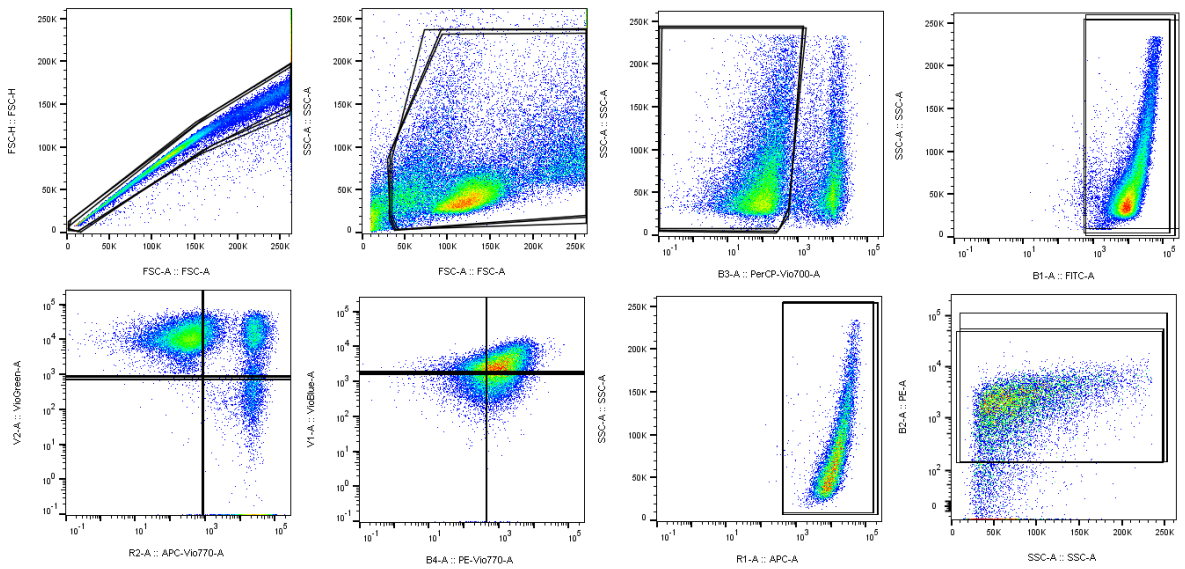


Figure 133 Participant A01 Complex Model Study Gating interpretation

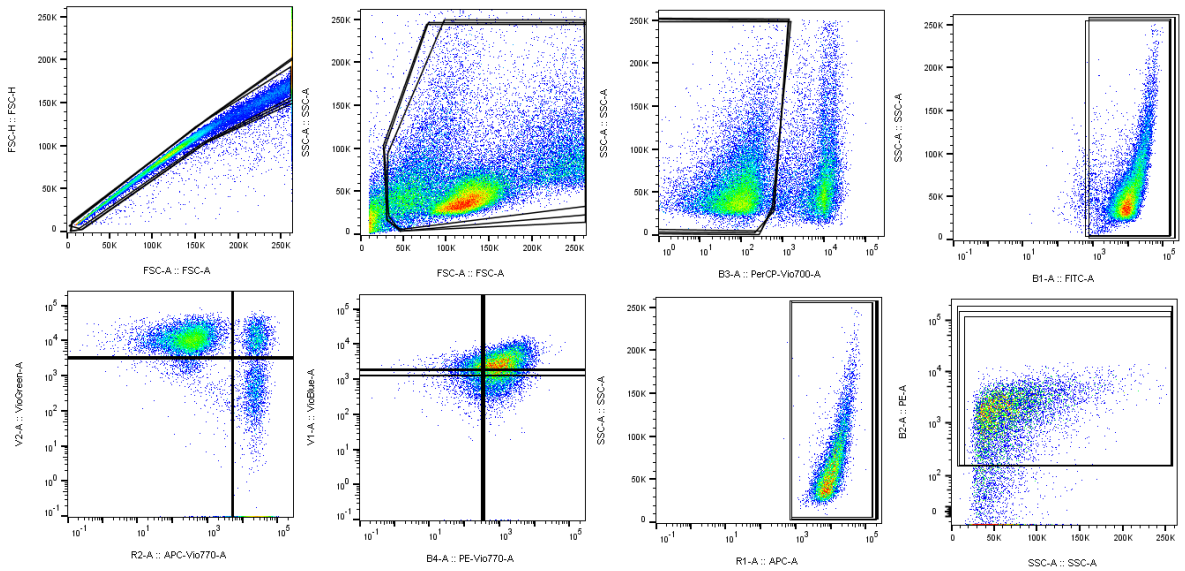
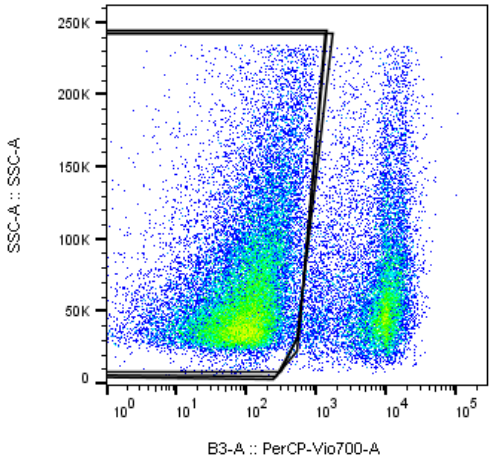


Figure 134 Participant B01 Complex Model Study Gating interpretation

To further investigate this, gate 3 was investigated for A01 and B01 as representatives of this mid-range group. Figure 135 and Figure 136 show the gates applied here for A01 and B01 respectively. The left images show the original scaling used for these gates, and the right images show the extended log scale to better see the left edge of the gates drawn.

a)



b)

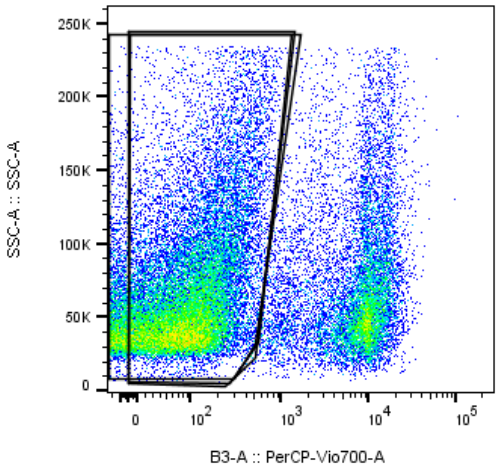
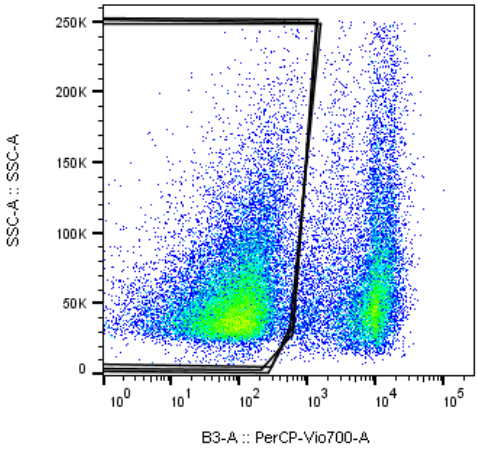


Figure 135 Participant A01 gate 3. a) Standard gates drawn with logarithmic scaling. b) Gates drawn to biexponential scaling

a)



b)

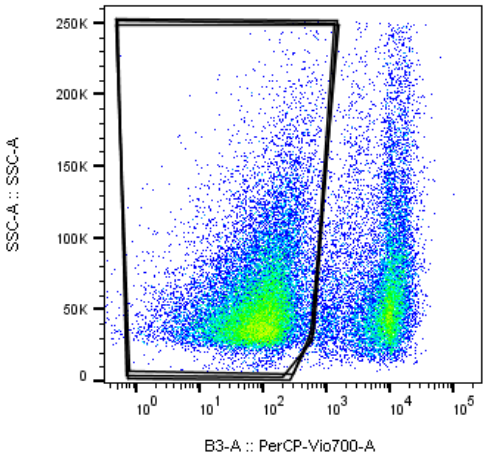


Figure 136 Participant B01 gate 3. a) Original gates drawn with logarithmic scaling. b) Scale extended to show shape of left edge

It is possible from looking at the two examples here that the mid-range standard deviations for this gate is due to the location of where the left edge is dragged to past the original logarithmic axis limit, and what angle this ends up being at. A01 gates are not placed in line with the 10^0 mark that the scale was set to, at which the boundary effect would have been displayed. Upon review of their

session recording, Participant A01 changed the axis scaling to biexponential, which is causing this variation, shown in Figure 135 (b), in comparison to the correct logarithmic layout in Figure 135 (a). However, B01 has a gate that is angled across this scale point, possibly causing the mid-range variability Both participants are consistent with the size and shape of the overall gates applied, and consistent with the cut off between the live, dead and dying cell populations.

To further investigate if the cause of this cluster variability in the third plot in the sequence, a selection of participants with low uncertainties were examined to see if there is a difference or obvious understanding if this boundary effect has been avoided during their analysis. Participants B03 and B13 had very low uncertainties from their total gating process and both fall within the lower variance cluster. Their initial plots can be respectively seen in Figure 137 and Figure 113. Further detailed plots of gate 3 can be seen in Figure 138 and Figure 139.

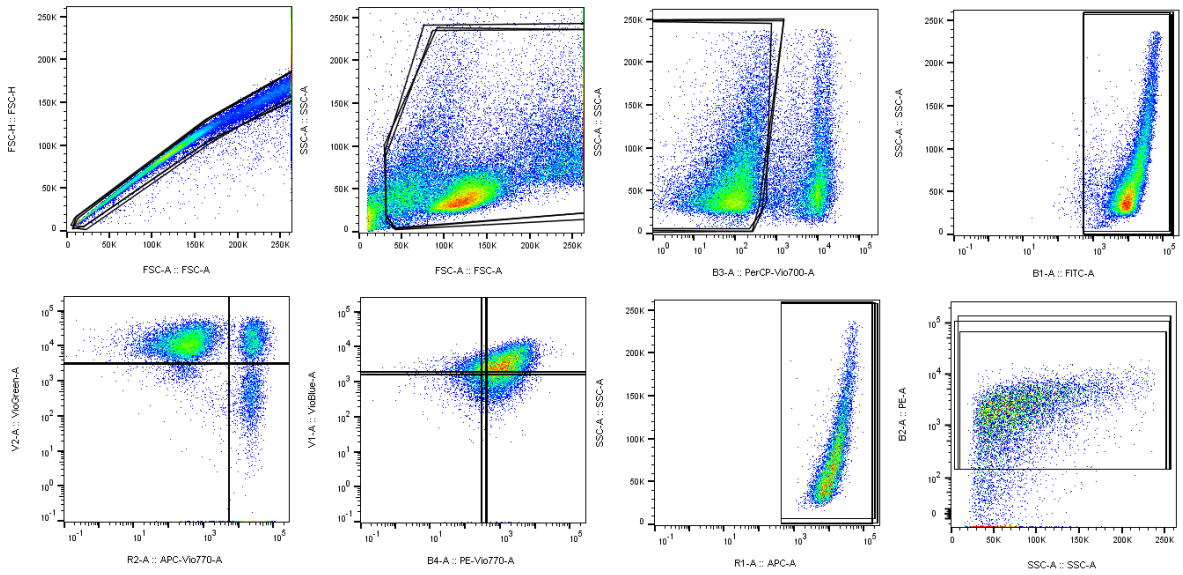


Figure 137 Participant B03 Complex Model Study Gating interpretation

Participants B03 and B13 have applied gates close to this boundary effect, however they are very repeatable in size and shape. Upon closer inspection B03 has drawn gates to all include the boundary effect. These are repeatable so the standard deviation is low, although this causes the

overall cell count to be higher. Participant B13 has consistently gated just inside the boundary effect, causing a lower variance and lower cell count due to the consistency of this edge.

a) b)

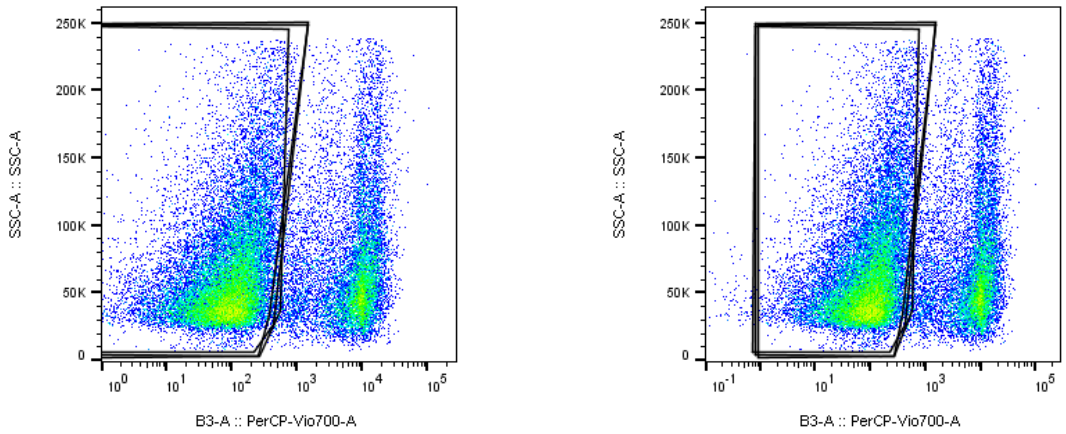


Figure 138 Participant B03 gate 3. a) Original gates drawn with logarithmic scaling. b) Scale extended to show left edge

a) b)

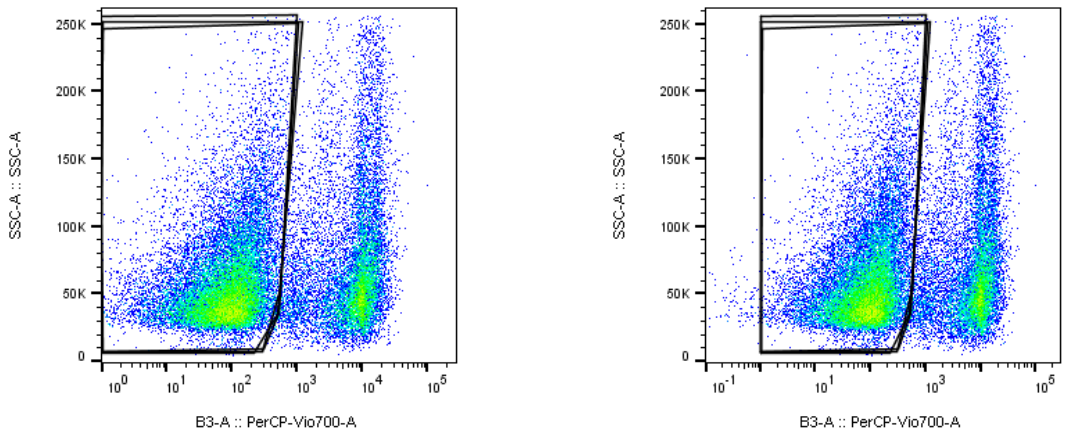


Figure 139 Participant B13 gate 3. a) Original gates drawn with logarithmic scaling. b) Scale extended to show left edge

Throughout this complex study, the variability has consistently arisen from a different gate in the series, not the first gate or quadrant gates which have been previously identified. This was due to the presence of a boundary effect appearing again as in Chapter 5, but in this instance it has appeared on the low fluorescence edge, due to the required cell population having low fluorescence

staining. This visual artefact has been considered again within different software platforms when it appears on a different edge. Figure 140 shows these outputs across different visualisation software, with FlowLogic removing this effect in the top and left edges of the plot, and Flowjo also removing it when figures are moved into the layout editor for exporting. Within the analysis window the boundary effect can still be seen, hence its effect on the data analysis, and the same can be said for FCS express, although the axes are defined and scaled differently to Flowjo and FlowLogic Logarithmic scaling.

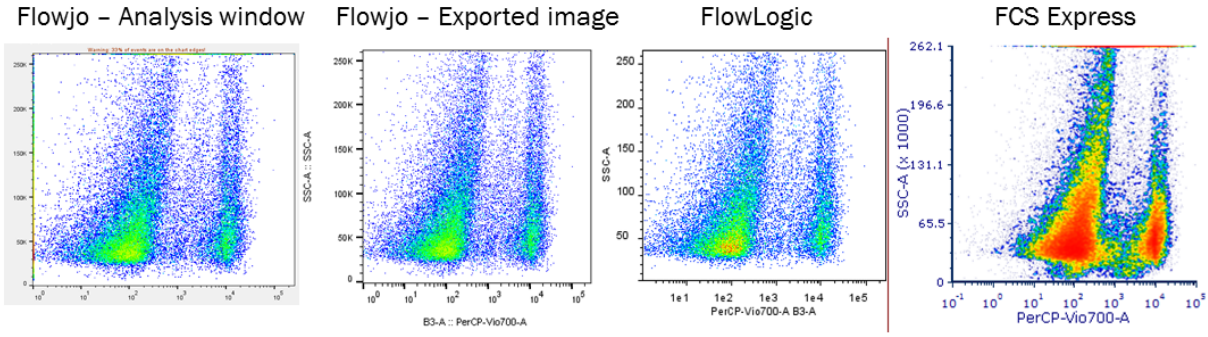


Figure 140 Software visualisations of live cell gate (gate 3) to identify boundary effect

6.3.6 Flow Cytometry Complex Gating Exercise uncertainty results – Comparison of Phase 1 and Phase 2

The uncertainties for each participant when they gated following their own judgement and then a protocol have been compiled into the histograms in Figure 141. Any dark orange areas are overlap of the two respective histograms. The range of cell counts has increased by 7.8 %, indicating that protocols may not help participants conform to reproducible cell counts. However, a bimodal/trimodal distribution has appeared so more needs to be completed to understand how subjectivity and interpretation of a protocol and visual images impact the final uncertainty calculated. An improvement in training to highlight boundary effects discussed could have a positive impact when trying to reduce inter-participant uncertainty in Flow Cytometry analysis. The skewness and kurtosis z-scores have reduced when participants use a protocol, however this is showing a tendency towards normality, which the distribution shape does not support in either instance.

When participants used their own judgement to apply gates, a split between higher and lower uncertainty groups appears, but the shape of the overall distribution is positively skewed, with only a few extremes exhibiting high variance. Use of a protocol has created further clustering effects, seen in Chapter 5 for the intermediate model, potentially caused by boundary effects within the software. This requires future investigation to understand what contributions training and awareness can do to remove this source of variation, as well as finding ways to remove this from the data. In addition, some software platforms remove these data spikes from the data, which can change the values operators specify for cell counts across different programs.

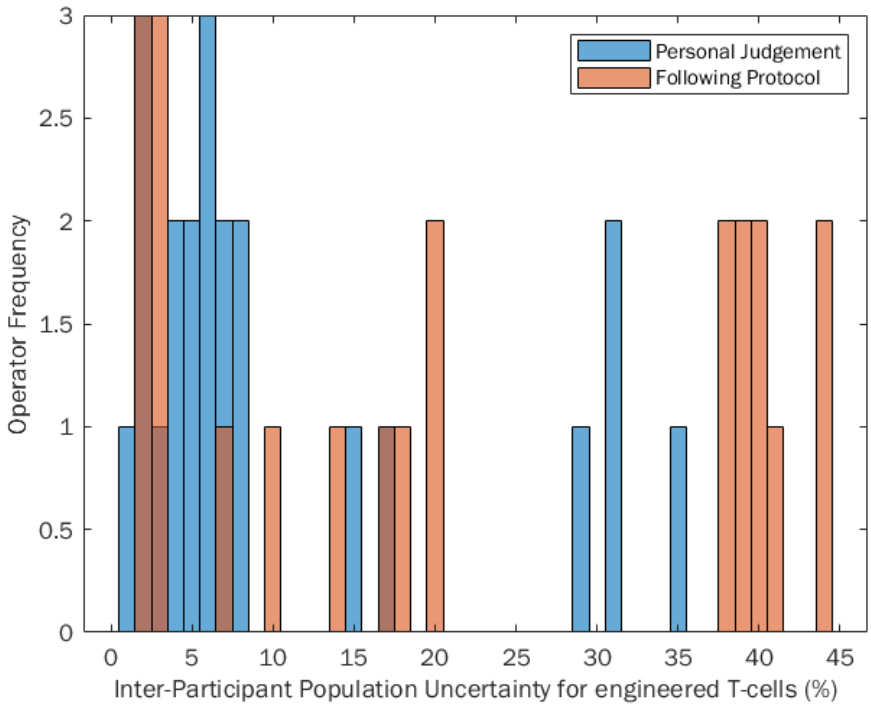


Figure 141 Overlaid histograms of participant uncertainties when gating the engineered T-cell population following their own judgment and then using a protocol (brown areas indicate overlap between the two distributions)

Further comparison of inter-participant data when gating using their own judgement and following a protocol has been completed with a Sign statistical test, to compare equality of median uncertainties between each test condition, shown in Table 73. The Sign test results reject the alternative hypothesis, because the uncertainty medians of the two testing conditions are not statistically different. This does not statistically show that the uncertainties are smaller when

participants use a protocol, however this only considered location values, so other measures of variability should also be considered before deciding whether a protocol effectively reduces inter-participant variability, or not. This is especially true in this instance due to the bimodal/trimodal nature of one of the populations, meaning a robust statistic such as the median is not a reliable location measure to represent the population.

In this study, an increase in overall range has been seen, indicating that diagrammatical protocols alone are not enough to control inter-participant variability and uncertainty contributions to measurements. Future considerations would need to look at incorporating training of noise parameters around data sets to ensure participants are aware of this effect on the variability of their subjectivity and interpretation, especially when cleaning the data at the start of a gating sequence. The use of this protocol is causing subjective behaviour to divide the population into high, medium and low variance clusters, which requires further testing to confirm this effect, and the effect of identifying boundary effects to participants before they apply gates to data.

Table 73 Sign test for median difference between Phase 1 and Phase 2 uncertainties

Null Hypothesis	Test	Sig.	Decision
Median difference of P1 & P2 uncertainties = 0	Related-samples Sign test	0.520	Retain null hypothesis

Table 74 A Priori and Post Hoc Power for Phase 1 and 2 uncertainties

Variance Phase 1	Variance Phase 2	A-priori power	Sample size required	Actual power
117.422	273.981	0.807	37	0.601

The A Priori and Post Hoc power analyses in Table 74 show that for the variances achieved between the two test conditions uncertainties, 37 participants would have been required to show this difference, to the required minimum power of 0.80. The actual power achieved through this study is 0.601, so less participants would be required if only uncertainty was being considered, not absolute cell counts (actual power of 0.343). This low power indicates that any differences seen from the data have a low probability of being just due to the two test conditions used and no other underlying factors present. Again, like the Sign test, this needs to be considered carefully. A greater number of participants in the study could always benefit and provide more confidence in the

results, however, the distributions are not normally distributed, so the variance calculated assumes a distribution with central tendency. In this instance, the range of data becomes more important to consider, due to the distribution shape and clusters appearing within the uncertainty data.

6.4 Chapter Conclusions

The engineered T-cell material used was a good model for the complex study because it provided an increase in complexity of the analysis pipeline participants were required to complete, and it also provided affinity towards current cell therapy treatments which are T-cell based. The studies run with the 3-workspace configuration for repeats continued to work well within the time available for participants, so this structure has shown to be a good working model for all studies in this experimental work. The eight-step process that each participant had to work through was also straight-forward to follow from the gating sequence protocol and the diagrammatical protocol, ensuring there was little deviation from the prescribed method.

When reporting the absolute cell count percentages for the results, the mean and median values for each Phase were very similar (mean = 5.8 % and median = 5.5 % for Phase 1, mean = 12.0 % and median = 12.2 % for Phase 2). These results both indicate a normal distribution within both data sets, however, the mean and median for Phase 2 is over double that of Phase 1. In this instance, there was an increase in the range of inter-participant cell counts when using a protocol. The protocol increased the absolute range of cell count results between-participants by 3.9 % of the overall cell count (2964 cell events) (which is a 42 % increase with respect to Phase 1 absolute range). This indicates protocols may not aid reproducible cell counts between participants in this instance, as initially identified in previous chapters.

The Sign test confirmed there was a significant difference between the medians of the two testing conditions (average cell counts), showing that protocols can potentially aid gating accuracy of the desired cell population. However, the Sign test to compare the medians of the uncertainties was not significant, indicating that the two phases did not have significant variabilities. Whilst this is a

definitive statistical test, the qualitative shape of the distributions was considered, and the bimodal/trimodal nature of the Phase 2 uncertainties indicates that a single location measure should not be used for sole comparison. The power analysis completed for both average cell counts and uncertainties indicated that more participants would be required to identify significant differences between the two testing conditions, requiring 82 and 38 participants for average cell counts and uncertainties respectively.

The use of the protocol made the distribution more bimodal, similar to the uncertainty for this Phase. Participants using their own judgement were more positively skewed towards 0 %, which is desirable for variation metrics, indicating that more participants were less variable (in their final cell count) when using their own judgement to apply gates. Reviewing the extremes in *CV* using the adjusted traffic light diagram shows variability in the final gate applied, but it also highlights that other variability seen upstream of this gate can have an impact on cell count, but the gate variability itself is not taken into consideration in the *CV* calculation, making measurement uncertainty a more suitable metric for accommodating variation throughout the whole gating sequence.

This complex model with an 8-step sequence has shown that calculating measurement uncertainty is possible for participants by using traditional measurement uncertainty methods. This was calculated successfully, by presenting participants with three repeated workspaces of data, and extracting one repeated file located in each workspace randomisation.

The uncertainties calculated to accompany the cell counts are more non-parametric than the absolute cell counts. The absolute mean and median uncertainties went from 10.6 % and 6.2 % respectively in Phase 1 to 22.0 % and 19.3 % respectively in Phase 2. The mean and medians in Phase 1 are not close together, indicating skewness. The mean and median in Phase 2 are closer together, indicating a normal distribution of uncertainty, although the central location measure has shifted to a higher uncertainty. The means are unsuitable metrics because they do not represent the peak maxima, especially in Phase 2, where the bimodal distribution causes the mean to sit

between two peaks. The range of participant uncertainty increased when using a protocol (absolute increase = 7.8 %, percentage increase with respect to Phase 1 = 23 %), indicating this may not reduce inter-participant uncertainty in the same way as observed for the basic and intermediate stages. Once more, the distribution shape of Phase 2 uncertainty separated into two bimodal peaks, indicating clusters of high and low variance participants, potentially due to understanding and bias when using the nominated software platform.

Further investigation has shown a high probability of this variation split coming from a boundary effect within the data visualisation software. Cells in the file that have a fluorescence signal lower than the visualisation axes are compiled on the boundaries. Inclusion of these in the repeated analysis can skew the cell counts and uncertainty significantly. Consistent inclusion or exclusion of the fluorescence spike gives a low variance, with high or low absolute cell counts respectively. Inconsistent inclusion/exclusion of the data spike leads to high variance. Those in the lower variance cluster have not included these cells in their analysis, or repeatedly have so the overall variation would be lower between repeats. Revision of extreme participants using the uncertainty boundary diagrams has shown most of the variation is contributed within the third gate applied, where the boundary effect is initially seen on the lower axis, increasing the possibility that the boundary effect causes this higher variation.

Overall, the structure of these analysis sessions and data extraction processes has worked well over the subsequent analysis phases, allowing for appropriate comparison between data in Chapter 7.

6.4.1 Consolidation of Objectives

- This study ran smoothly, acting as a good complex model for comparison of absolute reported results, *CV* and uncertainty measures. The session structures were suitable in time, and 3 repeats was suitable for participants to understand study context, but not become tired.

- Diagrammatical protocols used by participants during the second phase of this study have shown to increase the range in absolute results (3.9 %) reported and increase the range of participant absolute uncertainties (7.8 %), which calculates as a 23 % increase with respect to to Phase 1 counts). This is different to subsequent chapters and could possibly be due to the higher dimensionality of the data causing additional difficulty for participants.

- Extreme values in absolute reported results were due to participants either over constraining or under constraining the live cell population predominantly within the sixth gate, due to over- or under-constraining the quadrant around the desired double positive population. In some cases, lack of knowledge of using controls to set gates led to variance in population metrics.

- Extreme values in uncertainty results were due to participant variability in applying a gate to separate the live cell population from the dead or dying cells, alongside additional variation caused in the third gate by boundary effects on the left edge of the visualisation plot, caused by concatenated data that would otherwise be outside the plot axes. In addition, separating the CD45RA+ CCR7+ (gate 6) population from the remaining cells has caused some extremes in absolute cell counts due to placement of these quadrant gates. This gate was thought to be more variable, however, because of the difficulty of the gate, participants have been more repeatable when using FMO controls to place gates.

- The performance monitoring diagrams visualised continue to provide a straightforward way to monitor uncertainty performance with respect to the number of people in the study and defined quality satisfaction limits. These will be used in the subsequent chapter to monitor uncertainty performance in a more complex gating scenario.

- This study defines participant uncertainty for a more complex, industrially relevant, 8-colour panel cell model, which can be used as a complex model compare potential growth of inter-participant uncertainty through more difficult analysis scenarios.

Chapter 7: Comparison of Models

7.0 Introduction to the Chapter

Chapter 7 provides a comparison of key metrics monitored throughout the different complexity models discussed in Chapters 4 to 6. Basic statistics as well as more complex statistical tests such as the Friedman test have been used within this chapter to critically evaluate the different complexity models and their effect on cell counts and participant uncertainties.

7.1 Chapter Aims

This Chapter compares complexity models to identify whether the core hypothesis of this research has been met, showing an increase in range of *CV* and uncertainty with increased data complexity. This Chapter aims to compare the ranges of absolute cell counts, *CV*s and uncertainties, to identify a potential in case in between-participant variability. Only data from personal judgement has been compared, because a protocol phase was not conducted within the basic model (Chapter 4).

7.1.1 Chapter Aims & Objectives

The Aims and Objectives of this Chapter can be defined as follows:

- Identify potential changes in the range of absolute cell counts with complexity of data. An increase in range of reported cell counts would indicate greater variability between participants when reporting Flow Cytometry results.
- Identify potential changes in the range of absolute cell count *CV*s with complexity of data, because this is the most common variability reporting methods within the Flow Cytometry method so an indication of how this changes with data complexity could aid the community with a representative value for analyst contributions.
- Identify potential changes in the range of uncertainties calculated with complexity of data. This is an alternative measure of variation explored within this research and will be compared to *CV* to identify its suitability within Flow Cytometry measurements.

7.2 Methodology

To compare the three models across Chapter 4 to 6, basic statistical reporting has been considered, as well as more specific statistical tests to define differences between the models. This has been completed for all absolute cell counts and *CV*s reported, as well as respective uncertainties. This methodology section explains why these comparison methods have been chosen along with justification. Both IBM SPSS Statistics Version 24 and Matlab R2019a have been used to complete the more advanced statistical analysis and visualisation.

7.2.1 Basic Statistical Reporting

Basic statistical reports used within Chapter 4 to 6, with statistical definitions from Chapter 2, have been compared and discussed here. In addition to the compilation of these data sets, box plots have also been produced to visually compare the distributions of the separate data sets. A box plot shows the 25^{th}, 50^{th} (median) and 75^{th} percentiles (%iles) to visualise the distribution of the Interquartile Range (IQR). Whiskers have also been added to the box plot to show the distance of 1.5 x IQR. This is a common distribution marker for outliers. Anything within the whiskers is an inlier, or acceptable measurement. Any data point marked as a cross outside of the whiskers is defined as an outlier.

7.2.2 Further statistical testing for differences

Despite conclusions drawn from the basic statistics, more extensive statistical testing to confirm there are statistically significant differences between the three models has been completed for absolute cell counts and uncertainties.

A Friedman test has been used in this instance, because the data is non-parametric in distribution shape and three different groups or testing conditions have been considered [197]. This is the robust alternative to a one-way repeated measures ANOVA, which required the data to be normally

distributed. A Friedman test has also been used in place of a Kruskal Wallis H test, because it compares related data. Mostly the same participants were present in each testing model, making the Friedman test more suitable for analysis, whereas a Kruskal Wallis H test requires independence between participants in the separate testing conditions.

The Friedman test requires the same participants in each group, and each group represents repeated measures on the same dependent variable. In this instance there were not the same number of participants in each group, and some participants were not present in each group, which could compromise the power of the analysis. However, this was not in the control of the experimental studies. The Friedman test is an extension of the Sign test, which only compares two groups and has been used to compare test conditions within Chapters 5 and 6. The hypothesis of the Friedman test is as follows:

H_0 = the distribution of results in each group are the same

H_A = at least two distributions differ

If the null hypothesis is rejected because at least two distributions differ, Post Hoc tests (Pairwise comparisons) are used to identify similar distributions. These tests are like the Wilcoxon rank tests used to compare two non-parametric distributions [197].

7.3 Comparison of Results

The results reported here are split between the absolute cell count results and the uncertainties calculated from the gating sequences, as previously defined within Chapter 4 to 6. The results reported within these chapters have been repeated here for easier comparison.

7.3.1 Comparison of Absolute Cell Count Results

The statistical report results from the absolute cell counts generated from personal judgement of each complexity model can be found in Table 75, Table 76, Table 77 and Table 78. Only personal judgement results have been compared because the basic model does not have a 'protocol' phase, so only judgement has been assessed. Box plots to compare distributions of each complexity model are shown in Figure 142. It should be noted that the absolute results have been compared for continuity of the thesis structure, although the difference in cell type between the stages cannot be compared fairly, so these results are purely an indication of possible differences. Further testing with the same cell type at different gating stages could better investigate this.

Table 75 Measures of Location for the absolute results of the complexity models (%)

	Basic (Chapter 4)	Intermediate (Chapter 5)	Complex (Chapter 6)
Arithmetic Mean (%)	32.1	6.3	5.8
Median (%)	32.5	6.0	5.5
Mode (%)	N/A	N/A	N/A
Minimum (%)	19.7	3.5	1.2
Maximum (%)	51.3	8.0	10.5

Table 76 Measures of Spread for the absolute results of the complexity models (%)

Range	31.6	4.5	9.3
25th Percentile (%)	30.6	5.7	4.0
75th Percentile (%)	33.9	7.3	7.6
Interquartile Range (%)	3.3	1.6	3.6
Standard Deviation (%)	5.7	1.1	1.0
***CV* (%)**	17.8	18.2	12.1
Median Absolute Deviation (%)	1.9	6.0	1.7

Table 77 Measures of Skew for the absolute results of the complexity models (%) (3dp for better resolution)

Skewness (%)	0.492	-0.351	0.168
Skewness standard Error (%)	0.383	0.481	0.491
Skewness z-score (%)	1.280	-0.730	0.342
Kurtosis (%)	3.271	0.042	-0.669
Kurtosis Standard Error (%)	0.750	0.935	0.953
Kurtosis z-score (%)	4.560	0.045	-0.702

Table 78 Measures of Normality for the absolute results of the complexity models (%) (3dp for better resolution)

Shapiro-Wilk statistic	0.904	0.945	0.980
Significance	0.003	0.231	0.911
Normal/Non-parametric	Non-Parametric	Normal	Normal

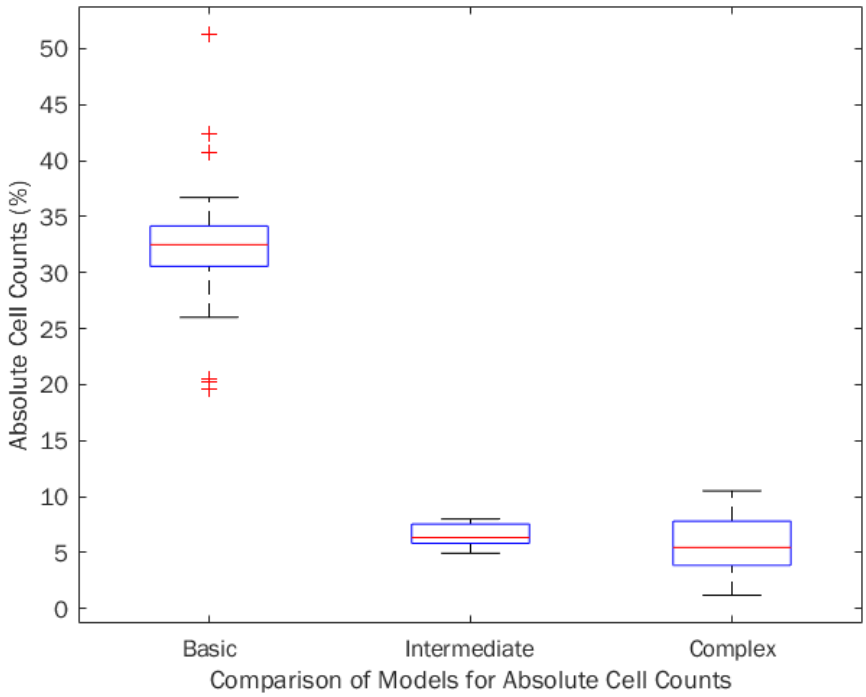
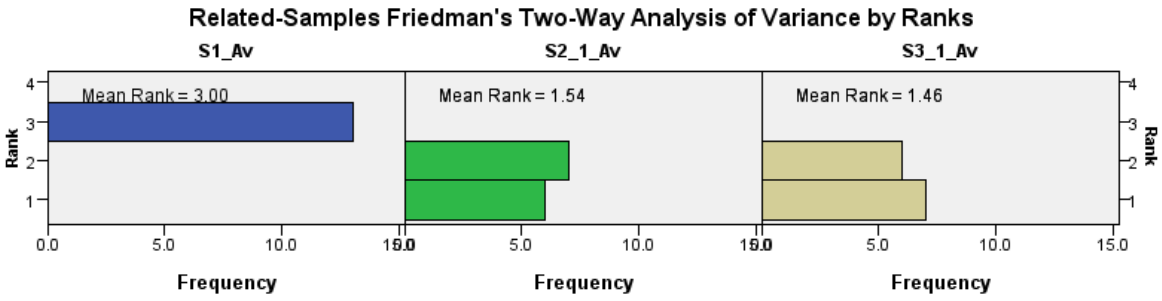


Figure 142 Absolute cell count distributions for each cell complexity model

The locations of the basic model distributions are qualitatively significantly different due to no overlap with distributions of the more complex models. This is quantified by the basic model median (32.5 %), compared to the medians of the intermediate and complex model (6.0 % and 5.5 % respectively). The shape of the distributions are significantly different according to the results of

the Friedman test, as shown in Figure 143 (Basic model = S1_Av, Intermediate model = S2_1_Av, Complex model = S3_1_Av). Absolute cell counts between models were statistically different, according to the related samples Friedman two-way analysis of variance by ranks, $\chi^2(2) = 19.538$, $p < 0.0005$, but this could be expected because of the difference in cell type.

Pairwise comparisons were performed as Post Hoc tests to further identify whether the significant differences were between some or all the models, shown in Figure 144. A Bonferroni correction was applied to adjust the significance levels [197,198]. Multiple comparisons increase the risk of a Type 1 error, which is why adjusted significance was used, and has been used for all subsequent Friedman tests in this Chapter. Absolute cell counts were statistically different between the basic and intermediate models ($p = 0.001$) and the basic and complex models ($p < 0.0005$). This statistical significance confirms the qualitative differences observed, but also because the basic model was based upon a different cell type (Embryonal Carcinoma cell line) in comparison to the intermediate and complex models which look at T-cell subsets.



Total N	13
Test Statistic	19.538
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.000

Figure 143 IBM SPSS results for the Friedman test comparing absolute cell counts for complexity models

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
S3_1_Av-S2_1_Av	.077	.392	.196	.845	1.000
S3_1_Av-S1_Av	1.538	.392	3.922	.000	.000
S2_1_Av-S1_Av	1.462	.392	3.726	.000	.001

Figure 144 IBM SPSS pairwise comparisons for the Friedman test comparing absolute cell counts for complexity models

The range of absolute results is largest in the basic model, which possibly could be because of the lower level of stratification required for this dataset, potentially suggesting that the larger the cell population cluster, the greater the range of participant cell counts. However, this is a different cell type, so this could confound this analysis. The range of absolute cell count results decreased for the intermediate model (4.5 %) and then increased again for the complex model (9.3 %). This may possibly have been due to the cell event number in the final model, which contained 76,012 cell events in comparison to 30,000 cell events gathered in the previous two stages. All cell counts were compared as percentages of the original cell count number to have better comparison between the files, to try and remove the inconsistent cell event numbers.

However, this would require further experimental clarification to formally design and test this possibility. For further comparison the ranges of the third gate applied in the intermediate and complex model have been considered, because these are both larger cell populations, comparable in size to the basic model as just shown. The range of results of the third gate applied in the intermediate and complex models are 25.8 % and 33.1 % respectively, which are similar to the basic model range (31.6 %). This potentially suggests that as further stratification of the data through gating steps occurs, there is a smaller range of inter-participant absolute cell counts, due to a smaller stratified cell population.

Due to the nature of gating, smaller cell counts through each gating step are logical. However, the ranges reported across the three models contradict this because the range of the complex model results is greater than the intermediate model. This is also seen in the IQR, so it is not just extreme

values that have an effect. More cell events were acquired in the complex model file than the previous two models, however, percentages of cell count populations have always been reported to standardise this. Gate 5 in the complex model has a range of 23.1 %, so it could be that because this model is more complex it is causing more deviation throughout the sequence, which could be expected.

This is not supported by other methods of variation measurement, such as the *SD*, which does decrease with every data set (5.7 %, 1.1 %, 1.0 %) respectively. However, consideration of the data shape is required before drawing these conclusions because the basic model is more skewed and kurtosed than the other model, shown by skewness and kurtosis *z*-scores and Shapiro Wilk test results for normality. This could suggest that *SD* values are not valid for use in this context, however they are still suitable to consider because of their use within general statistical reporting.

A more robust measure to use instead of *SD* is Median Absolute Deviation (*MAD*), but this shows no obvious trend in the data, with *MAD* starting at 1.9 % for the basic model, increasing to 6.0 % for the intermediate model and decreasing to 1.7 % for the complex model. Finally, the most common metric of variation in the Flow Cytometry community, Coefficient of Variation (*CV*) is fairly constant between each test model, reporting 17.8 %, 18.2 % and 12.1 % for each model increase. This *CV* will be known as inter-participant *CV*, because it is the *CV* of the absolute cell counts reported by each participant. This shows that if *CV* is to be continually used in the field, an average inter-participant *CV* of 16 % could be considered as an operator analysis component, taken from inter-participant absolute cell counts reported. This falls within the allowable *CV* specified by the ICSH guidelines, and also NHS KPIs for uncertainty, for Flow Cytometry discussed in Chapter 1 [72].

Intra-participant *CV* has also been investigated throughout this research and is defined as the *CV* across the repeated measures of the three repeats each participant completed. The difference between inter-*CV* and intra-*CV* has been depicted in Figure 145, re-imagined from the earlier Gauge

R&R explanations in Chapter 2. In this instance, intra-*CV* is the repeatability of each participant, and the inter-*CV* is *CV* of absolute cell counts between participants.

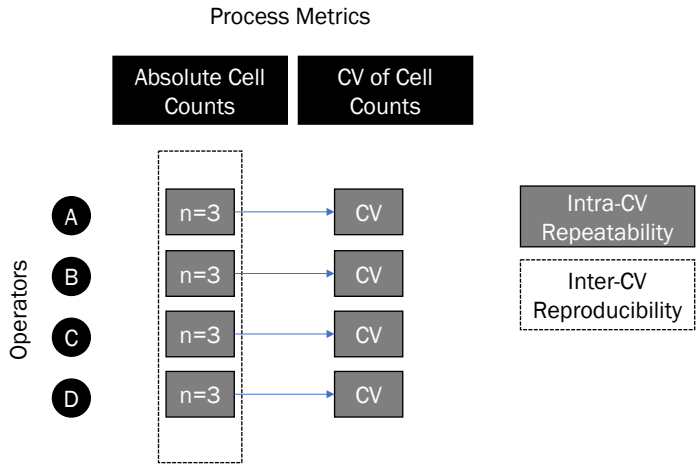


Figure 145 Diagram of Inter-CV and Intra-CV for subjectivity comparison

Figure 146 shows boxplot distributions of intra-participant *CV* results, from their 3 repeats acquired in each cell complexity model.

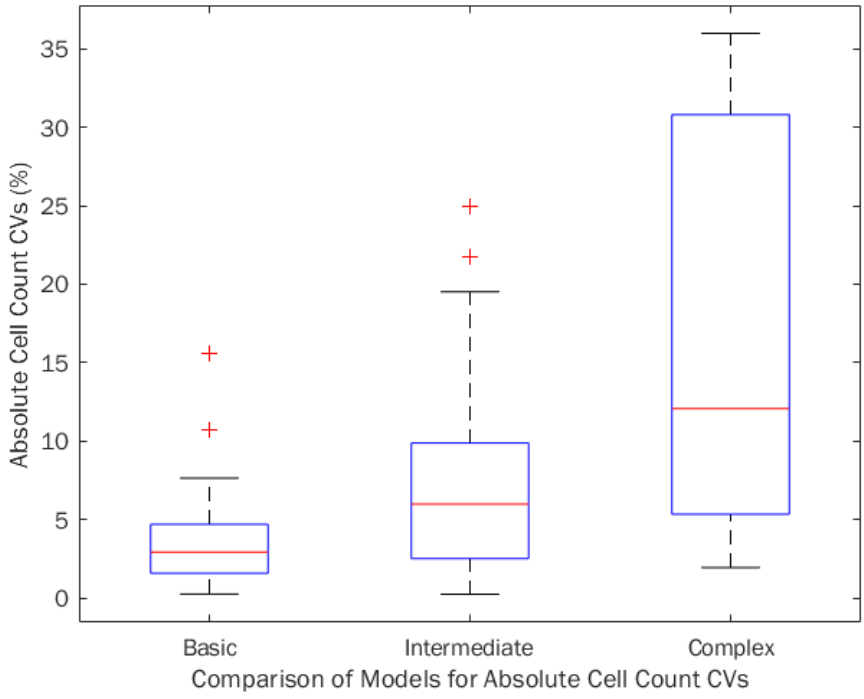
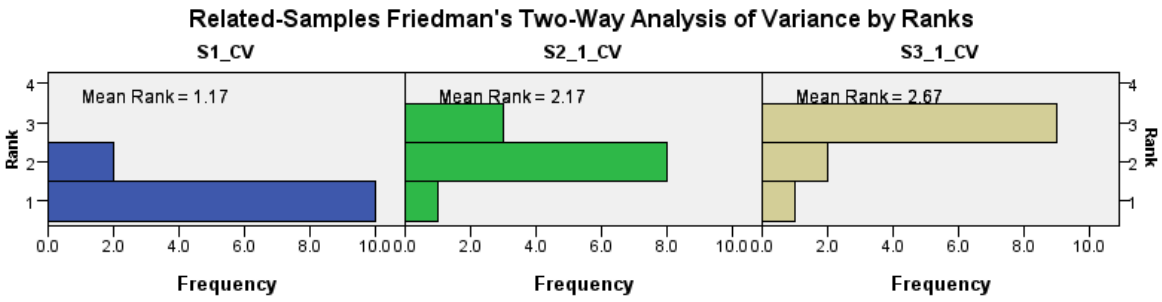


Figure 146 Absolute cell count CV distributions for each cell complexity model

It is clear that the range of intra-participant *CV*s increase with cell complexity, with gradual increase in the median of each group. A related-samples Friedman rank test was completed to further confirm a significant difference between the test conditions. Cell count *CV*s were statistically different through the different models, $\chi^2(2) = 14.000$, $p < 0.001$, shown in Figure 147. Post Hoc pairwise comparison tests were then conducted to identify which pairs were significant, summarised in Figure 148. These was a significant difference between the basic and intermediate models ($p = 0.043$) and the basic and complex model ($p = 0.001$), but not between the intermediate and complex model as their distribution shape was deemed similar.



Total N	12
Test Statistic	14.000
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.001

Figure 147 IBM SPSS results for the Friedman test comparing absolute cell counts CVs for complexity models

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
S1_CV-S2_1_CV	-1.000	.408	-2.449	.014	.043
S1_CV-S3_1_CV	-1.500	.408	-3.674	.000	.001
S2_1_CV-S3_1_CV	-.500	.408	-1.225	.221	.662

Figure 148 IBM SPSS pairwise comparisons for the Friedman test comparing absolute cell count CVs for complexity models

All distributions have a positive skew, aligning with the traffic light diagrams in Chapter 4 to 6, used to monitor the individual *CV* distributions. Even though the distribution of the complex model is the largest, it contains no outliers specified by the whisker limitations (Figure 146). Therefore, the data is less likely to by kurtosed because it may have a more evenly distributed shape than the other two test conditions.

Overall, this indicates that as the complexity increases, the *CV* of the participant is likely to increase, confirming the thesis hypothesis if intra-participant *CV* is the variation metric of choice. If a general *CV* of absolute cell counts of a population is used (inter-participant *CV*), an increase has not been observed as cell models become more complex. An average of 16 % *CV* was achieved across the three models, so this could possibly be used as a rule of thumb when considering general operator variation around a measurement.

7.3.2 Comparison of Cell Count Uncertainty Results

Currently *CV* is commonly used within Flow Cytometry communities as a measure of variability along with the absolute reported result. Uncertainty has been explored as a potential alternative to *CV*, because of the better resolution it provides when monitoring variability between participants through the gating sequences. Uncertainty has also been explored as an alternative to *CV* because uncertainties for equipment are required to be calculated for ISO 15189 to show competency of medical testing laboratories [82]. This replaced Clinical Pathology Accreditation (CPA) in the United Kingdom, where every pathology laboratory in the National Health Service (NHS) must be accredited to this new standard. If interpretation forms part of a measurement, this should also be monitored and included in the uncertainty budget, stated alongside the final measurement reported.

The statistical report for the uncertainty results across all three complexity models can be found in Table 79, Table 80, Table 81 and Table 82. These results have been taken from 'personal judgement' phases through each stage for comparison. Boxplots representing each of these populations has been visualised in Figure 149.

Table 79 Measures of Location for the uncertainty results of the complexity models (%)

	Basic (Chapter 4)	Intermediate (Chapter 5)	Complex (Chapter 6)
Arithmetic Mean (%)	4.0	3.8	10.6
Median (%)	3.6	2.1	6.2
Mode (%)	N/A	N/A	N/A
Minimum (%)	0.7	0.4	0.8
Maximum (%)	13.1	16.1	34.9

Table 80 Measures of Spread for the uncertainty results of the complexity models (%)

Range (%)	12.4	15.7	34.0
25th Percentile (%)	2.0	1.4	4.0
75th Percentile (%)	5.6	3.8	13.2
Interquartile Range (%)	3.6	2.3	9.1
Standard Deviation (%)	2.7	4.3	10.8
Median Absolute Deviation (%)	2.0	0.8	3.0

Table 81 Measures of Skew for the uncertainty results of the complexity models (%) (3dp for better resolution)

Skewness (%)	1.288	1.942	1.375
Skewness standard Error (%)	0.388	0.481	0.491
Skewness z-score (%)	3.320	4.037	2.800
Kurtosis (%)	2.311	2.899	0.468
Kurtosis Standard Error (%)	0.759	0.935	0.953
Kurtosis z-score (%)	3.045	3.101	0.491

Table 82 Measures of Normality for the uncertainty results of the complexity models (%) (3dp for better resolution)

Shapiro-Wilk statistic	0.900	0.692	0.758
Significance	0.003	0.000	0.000
Normal/Non-parametric	Non-parametric	Non-parametric	Non-parametric

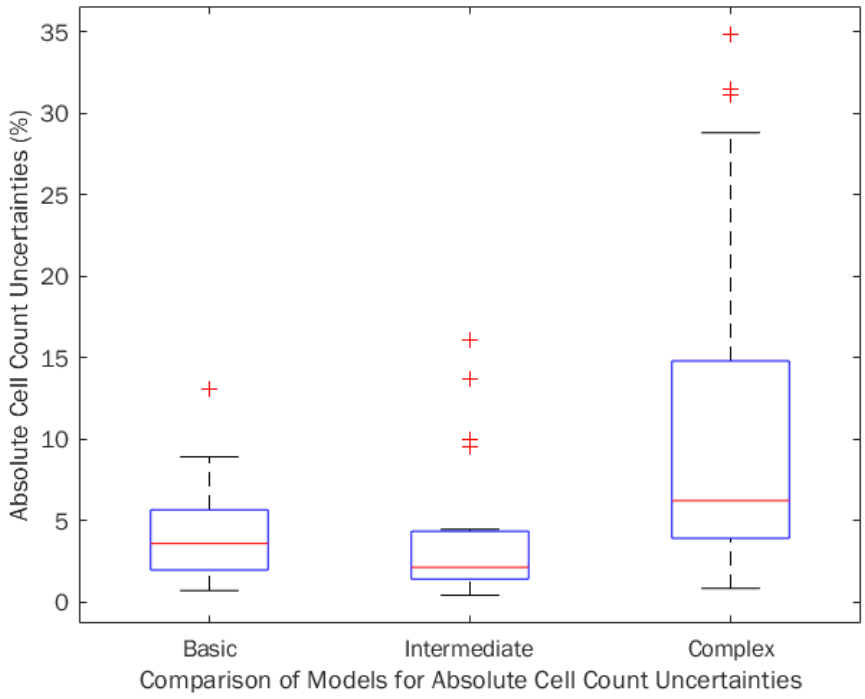
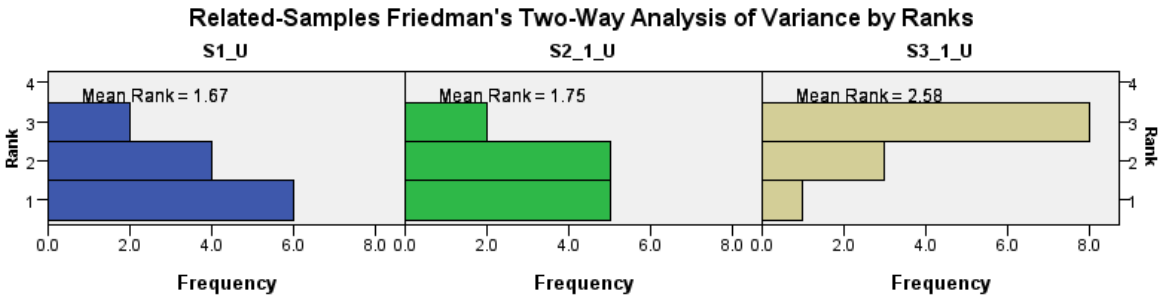


Figure 149 Absolute cell count uncertainty distributions for each cell complexity model

There is no steady increase of mean or median between the analysis stages because these location measures for the intermediate model are lower. However, as identified in Table 80 there is a steady increase in the range of participant uncertainties from basic model (12.4 %) through the intermediate model (15.7 %) to the complex model (34.0 %). This indicates more participant variability throughout the entire gating process as the data they analyse becomes more complex. Unlike *CV*, this combines variability from each gate applied, rather than just the final cell counts, giving better resolution and traceability to the variability.

The shape of the distributions were significantly different according to the Friedman test, $\chi^2(2) = 6.167$, $p < 0.046$, shown in Figure 150. However, when conducting a Post Hoc examination using pairwise comparisons, no significant differences between individual pairings were reported, shown in Figure 151. The Friedman test assesses the distribution shape, so the positively skewed nature of all three models may have caused this outcome. All models returned positive skewness z-scores (3.320 %, 4.037 %, 2.800 % respectively) for the increase in complexity. All are outside of the 2.58

boundary limits for normality, but some are more skewed than others. This could be why the Friedman test is initially significant and the pairwise comparisons show no significant differences.



Total N	12
Test Statistic	6.167
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.046

Figure 150 IBM SPSS results for the Friedman test comparing absolute cell counts uncertainties for complexity models

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
S1_U-S2_1_U	-.083	.408	-.204	.838	1.000
S1_U-S3_1_U	-.917	.408	-2.245	.025	.074
S2_1_U-S3_1_U	-.833	.408	-2.041	.041	.124

Figure 151 IBM SPSS pairwise comparisons for the Friedman test comparing absolute cell count uncertainties for complexity models

The kurtosis of the basic and intermediate models were also high (3.045 % and 3.101 % respectively), whereas the complex model has a much lower kurtosis *z*-score (0.491 %), indicating this distribution is less affected by outliers. This can be further supported by the boxplots in Figure 149 showing outliers across a smaller range in the basic and intermediate models rather than in the complex model.

The comparison of variation between the stages again comes down to how the data is presented and what variation metrics are chosen to describe the data. In this instance, the *SD* shows a similar trajectory to the range, increasing in size as the complexity of data increases. However, the *MAD* score decreases from basic to intermediate models and then increases from intermediate to complex models again.

7.3.3 Comparison of *CV* and Uncertainty Results as variation metrics

Although uncertainty was not significantly different with regards to the distribution shapes of the separate models, the range of the intra-participant uncertainty increased with complexity, comparable to *CV* in this instance, because *CV* also increased in range as the complexity of the data increased, although the range is larger for *CV* than uncertainty.

Rather than replacing *CV* with uncertainty because it is more specific, both metrics could be used when training and reporting FC measurements. Uncertainty gives much more resolution to the variability within data, shown when analysing uncertainty components throughout Chapters 4 to 6. The method shown to obtain participant uncertainties can be used to combine other sources of uncertainty within the FC measurement, to obtain a representative combined and expanded uncertainty value that meets required standards.

CV can also be used to obtain a quick point-in-time measure of variability used alongside uncertainty which is a lot more specific and takes longer to achieve. It can be used within training and refresher exercises to quickly monitor an analyst's variation on a specific instrument or analysis pipeline. Ultimately, both variation measures can be used, however, it cannot be assumed that an analyst's uncertainty could be judged from their *CV* or vice versa, so one should not be used to provide an indication or estimation of other metrics and variability. The '*CV* versus uncertainty' scatter graphs created in Chapter 4 to 6 show very poor correlations, so *CV* should not be used to

assume a participant's uncertainty. If an analyst's *CV* is high, their uncertainty may not also be high. Uncertainty requires much more structured methodology and analysis in order to calculate a combined uncertainty figure.

Overall, the core thesis hypothesis has been proven for both *CV* and uncertainty of participant Flow Cytometry results. As the complexity of Flow Cytometry data increases, the range of participant *CV* of results will also increase (Figure 152), as well as the range of participant measurement uncertainty (Figure 153).

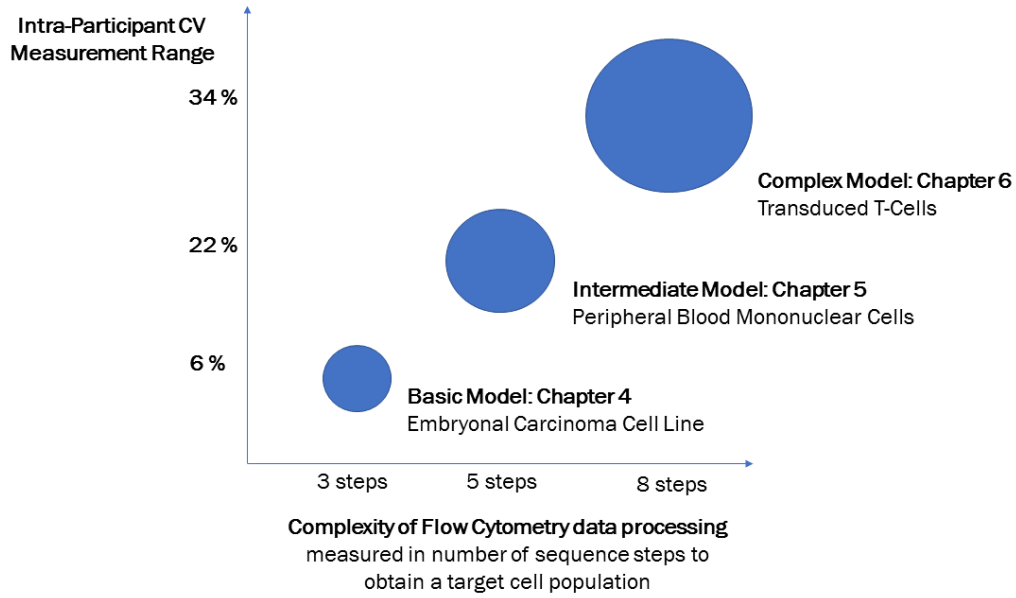


Figure 152 Core hypothesis of thesis, showing an increased range of intra-participant CV with FC data complexity

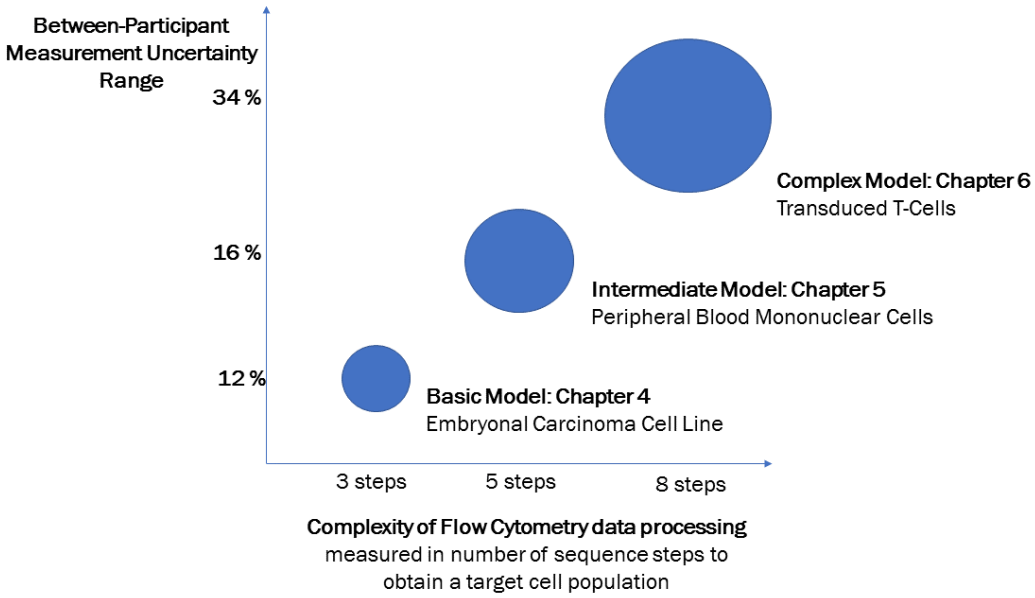


Figure 153 Core hypothesis of thesis, showing increased range of inter-participant uncertainty with FC data complexity

7.4 Chapter Conclusions

When considering reporting of absolute cell counts (means and medians), the lower cell counts reported as the complexity increased could be a function of the number of analysis steps used within the gating analysis sequence. If so, this could mean rare cell events are lost through gating, or the true cell count of populations is skewed. However, this is a comparison across different cell types, so this could confound results. The more sequence steps completed, the lower the mean or median cell count could be due to increased stratification of the cell populations. This could intensify when 18 colour panels are used to consider specific markers, although all 18 markers would not necessarily all be used to investigate or monitor one specific function, they may monitor three or four targets at a time, requiring less processing steps for each target.

Although the location metrics reduces as the complexity of the models increased, this cannot be assumed for the range or inter-participant absolute results reported for each model. As the complexity increased, the range of results between participants did not decrease. As the target becomes smaller and perhaps more specific, this does not necessarily mean that the range of results reported by participants will become smaller or more focused.

The absolute range of results for the basic model was the largest (31.6 %) when three gating steps were completed, which reduced to 4.5 % for the intermediate model (5 steps), however, this increased again when 8 steps were completed in the complex model, giving an inter-participant range of 9.3 %. All cell counts were compared as percentages of the original cell count number to have better comparison between the files, to try and remove the inconsistent cell event numbers between files used in each model.

All cell model distributions were deemed to be statistically different from each other using the related-measures Friedman test, and when investigated further using Post Hoc pairwise comparisons between each pair of models, it was found that the intermediate and complex models

were statistically different from the basic model, but not different from each other. These differences may be due to the lower number of gating steps used in the basic model, but also that the exemplar used in the basic model was an immortalised cell line, where pluripotent stem cell markers were identified. However, in the intermediate and complex model, T-cell lineages were investigated, so this change in cell type could have impacted the results because of different sized sub-populations. This was investigated by looking at the ranges for the third gate applied within the intermediate and complex models, returning absolute ranges of 25.8 % and 33.1 % respectively between participants. These values are comparable to the range reported for the basic model, so this higher range is more likely due to the lower number of stratification steps to identify the target cells, at this point.

Using the basic statistical reports to compare the three models has shown further inconsistency between different statistical variability metrics, building upon the findings in Chapter 3, where different boundary estimators were tested. When considering the variability of the absolute cell counts, there were inconsistencies in trend between the total range, Standard Deviation and Median Absolute Deviation of all three cell models. This further supports the conclusion that the data distributions need to be properly understood before choosing the most appropriate metric to define and monitor variance.

Inter-participant *CV* was stable across all three models (17.8 %, 18.2 % and 12.1 %, with an average of 16.0 % respectively), however, it does not concur with other manual gating analyses conducted on Flow Cytometry data which show higher *CV*s for individual gating or centralised gating [1,2]. This *CV* metric is indicative of what is normally reported in FC literature, as a calculation from repeated absolute cell counts reported. Therefore, the stability across the three models of this inter-participant *CV*, could suggest that an overall value for participant *CV* contributions is 16.0 %. This could be used when considering variations of measurements as well as within quality control documentation and training purposes. However, if this were to be considered alongside the ICSH guidelines, this indicates that participants do not fall within satisfactory variation limits, because

having < 10 % *CV* is highlighted as being satisfactory, and the cell types and populations used within this research were not rare [71]. This satisfactory value was quoted for the overall measurement, but it could also be used for individual measurement components such as analyst variation or biological variation.

When considering the intra-participant *CV*s, the absolute range of results increased each time with an increase in complexity (6 %, 22 %, 34 % respectively with each complexity model), showing that as the data becomes more difficult to analyse, the range in reported *CV*s increases, showing more variability as a result. If *CV* is used as a variance metric, it was confirmed using the Friedman test that there is a significant difference between the stages, showing that there is a difference in variance between the complexity stages. When considering using *CV* these ranges can be used effectively within training exercises and when considering repeatability of analysts on the same test. The greater range of variability reported for more complex models indicates that a wider range of *CV*s should perhaps be anticipated when running FC assays with larger panels, although they may not be initially acceptable.

Continuous improvement efforts to reduce variability can potentially be monitored using *CV* as the dependent variable measured, to identify how these small changes impact the variability of analysts and therefore data over time. This has been effectively shown through many proficiency testing schemes, using *CV* as the comparable measure between participating laboratories to indicate good and poor performance [94,95].

A similar trend in range was also observed for uncertainty across each of the complexity phases. An increase in range was seen as the complexity of the model increased (12 % to 16 % to 34 % respectively with each complexity model), which was further confirmed as a significant difference by the Friedman test, to show differences in shape. However, when further investigated, there were no significant differences between the pairs of models, possibly due to the smaller ranges.

The mean and median uncertainties did not go up continually compared to the range, there was a decrease from the basic model (mean = 4.0 %, median = 3.6 %) to the intermediate model (mean = 3.8 %, median = 2.1 %) and an increase from the intermediate model to the complex model (mean = 10.6 %, 6.2 %). This shows that if location metrics were only used to describe each population, the increase in the whole population would not be seen. Although the mean and median values have been considered throughout the whole research, the core hypothesis of the thesis investigates the variability and spread of the uncertainty of the data, rather than location metrics, so range is considered as a more important marker of comparison at this stage.

The core thesis hypothesis has been confirmed, because the range of inter-participant uncertainties has increased with more complex data. This further consolidates using uncertainty as a variance metric within FC, specifically in a manufacturing and quality control environment, not just a clinical one. This can also benefit research and development communities, because a full understanding of uncertainty can provide more specific root cause analysis when high variance is detected. This can then help deliver better training and continuous improvement as a result.

Uncertainties reported alongside measurements gives a much more informed judgement on the cell therapy product through all stages of manufacture, sorting and release. If an uncertainty reported alongside a cell count result is high, it enables the manufacturer, inspector or clinician to really question the product, measurement equipment, process and reagents involved, before making a more conclusive decision. Uncertainty reporting provides better resolution to the whole measurement, with quantitative traceability back through the components, to identify significant contributions of variance. This overall provides better awareness and control to all products and processes contributing to the measurement. If a measurement is close to acceptance limits for signing off the product or approval for the next manufacturing phase and the uncertainty boundary causes concern for a specific batch, this could provide more solid grounds to ask for further testing of the product. This would be preferred over potentially providing a false positive or negative result, which when passed onto the patient could have costly impacts.

7.4.1 Consolidation of Objectives

- There was no increase in range of absolute cell counts reported as the complexity of the data increased. The target population medians got smaller as the complexity increased, due to further stratification of the data, however, the range of participant reported cell counts did not decrease in line with this.

- Inter-participant *CV*, calculated from all the participant cell counts across each complexity phase was very stable, with an average of 16 % *CV*. This could possibly be used as an indication of variation within the Flow Cytometry community, to highlight the significant impact analysts can have on the analysis and conclusion of results.

- Intra-participant *CV* range increased significantly with each complexity model (6 %, 22 %, 34 % respectively), showing that as the complexity of the data increases, a greater range of *CV* results can be expected between participants.

- There was a significant difference in range between the cell complexity models when measurement uncertainty was calculated. As the complexity of the data increased, the range of uncertainties between participants also increased (12 %, 16 %, 34 % respectively). This confirms the core hypothesis of this thesis, that as the complexity of the data models increases, the uncertainty contributed from participants also increases.

- The use of *CV* is common within Flow Cytometry communities, and it still has significant strengths as a reporting and training tool when looking at FC data. It shows that as operators are trained to analyse more complex data, initially larger ranges of results are to be expected from more difficult analysis panels. This is also the case for uncertainty, which allows laboratories to better conform to ISO 15189 requirements for competency within medical testing laboratories, by showing more specificity of uncertainty calculations by including quantified subjective elements.

Chapter 8: Participant Surveys

8.0 Introduction to the Chapter

Chapter 8 introduces the participant experience surveys, which participants completed at the start and end of the study period (2 years apart) to capture training state and issues observed over time. This gave a better understanding of the participants and allowed certain experience metrics to be investigated in relation to uncertainty study outcomes. This provides some human factor context to the quantitative data explored in Chapters 4 - 6. This also stratified common issues participants faced when gating their own Flow Cytometry data, or when interpreting it from another source, to identify areas for improvement within protocols and standardisation.

8.1 Chapter Aims

This Chapter aims to give a better understanding of the participants (and their training background) who were involved in some or all stages of this research. Not only will this identify dichotomous indicators of preference and experience, but ordinal metrics such as use frequency are investigated alongside uncertainty results from previous Chapters. This indicates whether correlations can be drawn between use frequency metrics, experience and the results achieved as well as the measurement uncertainty, which were so commonly assumed by participants.

8.1.1 Chapter Aims & Objectives

The Aims and Objectives of this Chapter can be defined as follows:

- Stratify the results of the initial and end surveys, to compare key dichotomous markers that define the population for various personal, experience and motivational factors.
- Investigate the preferences for manual or automated gating over time, stratifying the key reasons why, to see if these can be addressed with data analysis, software or cultural changes.

- Investigate problems encountered by participants when they gate manually, and if these stratified issues change through the course of the research studies.
- Investigate problems encountered by participants when interpreting Flow Cytometry gating data from literature, and what issue this may cause for reproducibility.
- Investigate possible correlations between continuous participant experience metrics and the range of results from the previous uncertainty studies (Chapter 4-6), focusing on the range of absolute cell counts, *CV* and expanded uncertainty results obtained.

8.2 Methodology

A total of 43 participants provided answers to the initial questionnaire. These participants took part in either one or all of the uncertainty studies, and if they joined part way through the total experimental phase, they were given this initial questionnaire to complete before starting the respective study. Sixteen participants provided answers to the follow-up questionnaire. This difference between participant involvement is due to a staffing change at one of the centres just before the intermediate uncertainty model. Once these new participants had completed the first questionnaire, it was deemed too close in time to send the follow-up questionnaire, so they were excluded from this final survey exercise.

The timescale and delivery of the starting and ending questionnaire will be defined throughout this section, moving on to provide further detail on the structure of each questionnaire provided to participants.

8.2.1 Questionnaire timescales and delivery

To ensure background information of Flow Cytometry participants could be captured, participants answered an initial questionnaire. This was hosted electronically using Google Forms to minimise disruption to participants and provide automatic stratification into Microsoft Excel file formats (.xls). Once participants had provided consent to take part in the research (as discussed in Chapter 3) and 24 hours had elapsed (cooling off period, to ensure all participants were happy to take part), the questionnaire was sent to participants, with instructions to complete it before their first analysis session took place. This progression is shown in Figure 154, depicting the sequence of questionnaires and the main analysis models. The pre-study analysis discussed in Chapter 3 has not been included, because the results obtained from this study are not comparable with the absolute cell counts, *CV*s and uncertainties investigated in Chapters 4 to 6.

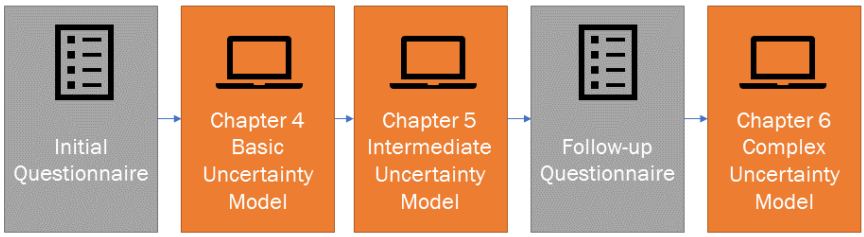


Figure 154 Questionnaire timescale in conjunction with analysis models

Following the Intermediate Uncertainty model (Chapter 5) and the investigation of the boundary effects present in the software visualisation, a follow-up questionnaire was introduced to identify additional participant understanding of Flow Cytometry. This was given to participants before the complex uncertainty model (Chapter 6) took place, to ensure a good rate of response. Online Surveys (previously Bristol Online Surveys) was utilised for this questionnaire [199], due to institutional access and migration away from Google platforms, as well as the ability to download the data into different analysis platforms previously used, such as Microsoft Excel (.xls) and IBM SPSS Statistics (.sav), for more streamlined analysis. A series of questions were repeated from the initial questionnaire, to monitor experience gained over time, for example, to see whether collectively the experience distribution changed, or whether specific participant experiences affected results. New questions were introduced to ask participants about what noise parameters they identified and how there were dealt with. This was to identify underlying understanding about the boundary effects, without asking a leading question.

8.2.2 Initial Questionnaire Outline

A series of dichotomous, scalar and ordinal questions were asked to participants throughout this survey, stratified into groups to identify personal Flow Cytometry qualifications, experience with FC, any visual impairment that could affect their judgement during the studies and motivations factors to identify enjoyment. Each of these sessions will be discussed throughout the methodology, with results discussed later in this chapter. A full template of the initial questionnaire given can be found in Appendix D.

8.2.2.1 Personal Questions

These questions were simple identifiers, to log who the participants were (name) and their responsible manager or Principal Investigator. This was to group people into teams, with the understanding that they had all received similar training, and potentially may have similar biases, if splits appeared in the data. Upon analysis participants were all anonymised and given alphabetical codes for further analysis.

Participants were also asked if they had any formal Flow Cytometry Qualifications. This refers to being a Chartered Cytometrist (C.Cy), recently updated to Specialist in Cytometry (SCYM). This is a qualification required for clinical pathologists using Flow Cytometry to perform diagnostic assays, so this question aimed to investigate whether anyone in a research, manufacturing or process development setting in the 3 centres used in this study had this level of externally approved qualification.

8.2.2.1 Experience Questions

The experience questions covered personal experience with Flow Cytometry, with respect to the length of time they have used Flow Cytometry for, and how often they use the equipment. This was to avoid situations where participants may have completed an FC experiment 5 years ago, but not completed any analytical work since. This section also asked participants to highlight what they predominantly use FC for; Cell counting, Immunophenotyping, Cell Sorting (FACS) or other assays. This was also coupled with what cell types participants were most familiar with when running FC assays, to identify what cells types were most common across the three centres.

Training was also investigated, to identify whether participants had been taught to gate internally, externally (by training courses or manufacturers) or were self-taught using text/online resources. To expand on the training questions, participants were also asked if they used protocols to apply

gates to their data and whether these were listed within internal Standard Operating Procedures (SOPs) or from external publications.

Participants were asked to identify their preference between manual and automated gating. In this instance, manual gating refers to software platforms that allow users to manually draw or identify cell populations of interest, whilst also giving more flexibility over the scaling and axes used. Automated gating platforms refer to machine learning algorithms and statistical tools that cluster data to find common sub-groups due to different dimensionality factors. Populations are then selected by the user, and no multi-step sequence of gates is required as found in manual gating. Participants were also asked to identify why they chose their preferred gating style (manual or automated), which was given as a written text answer. These text answers were further coded and analysed (Section 8.2.4.5).

Finally, participants were asked to identify whether they had experience using the Flowjo software platform, or not, to identify additional training that may be required to complete the studies. Participants were also asked if they had taken part in any External Quality Assessment Schemes (EQAS) throughout their career, or likewise, submitted data files for central processing.

8.2.2.1 Vision Questions

Participants were asked if they had any visual impairment that could affect their ability to participant in the study or use a laptop computer for a maximum duration of one hour. If participants had issues such as colour blindness or mentioned other visual impairments that could affect their perception of shape or density, they were not able to take part in the further studies, due to known differences in observation of the data that could skew results [200,201]. In Chapter 3, this was explored with a small number of participants, however, no participants had colour blindness according to the test used, so this is something that could be investigated in future with two distinctive test groups to quantify potential differences.

8.2.2.1 Motivational Factors Questions

Motivational Factors focused on identifying situations that were preferential for participants to complete FC gating during their day-to-day tasks, and what problems were caused through their own experience of gating data or interpreting data.

Participants were asked to list problems they encountered when they manually gated, and what solutions they have to rectify or deal with these issues. They were also asked to identify problems that occur when trying to interpret published Flow Cytometry data from another source, to investigate whether better reporting standards may be required throughout the community. Both questions required written text responses, which were coded manually, as detailed in section 8.2.4.5.

Participants were then asked to provide an ordinal response, to specify how much they enjoyed gating (scale of 1- to 5, 1 = strongly dislike, 5 = strongly like) and dichotomous results (yes/no) to identify whether they had to be in a certain mood to gate, a certain environment to complete the analysis, and whether they preferred to complete their analysis at a certain time of day. If participants responded 'yes' to preferring to gate at a certain time of day, they were then asked to specify further and results were stratified into morning or afternoon bins.

8.2.3 Follow-up Questionnaire Outline

A further series of dichotomous and scalar questions were asked to participants throughout the follow up questionnaire, to follow-up on factors from the previous questionnaire, and to ask new questions that arose during Chapter 4 and 5 data analysis. This questionnaire was divided into sections to cover FC usage and FC gating preferences. Each of these sessions will be discussed throughout the methodology, with results discussed later in this chapter. A full template of the follow-up questionnaire given can be found in Appendix D.

8.2.3.1 FC Usage Questions

In a similar manner to the personal questions, FC usage questions were simple identifiers, to log who the participants were (name) and their responsible manager or Principal Investigator. Again, this was to group people into teams, with the understanding that they had all received similar training, and potentially may have similar biases, if splits appeared in the data. Also, over time participants may have left the company or changed reporting roles, so this change in participant number was also captured.

More specificity was asked about what Flow Cytometers and analysis software packages participants had the most experience with. Further to stratifying into research groups, it became apparent through the different analysis models that different teams/centres used different brands of Flow Cytometer and different software. These biases could also cause bias in results analysis so this information was captured for reference if necessary.

The experience questions covered personal experience with Flow Cytometry, with respect to the length of time they have used Flow Cytometry for, and how often they use the equipment. Participants were also asked to identify whether their usage was consistent over time, or on and off with a project, and to highlight if they were currently working on a project that required FC. To further diversify FC usage, participants were also asked to identify how often they analyse FC data, as well as use the equipment.

8.2.3.1 FC Gating Preferences Questions

Participants were again asked to identify whether they preferred manual or automated gating platforms, and to list why. This was to monitor changes over time with usage, hence the staggered questionnaires, because automated gating software and plug-ins are becoming more apparent throughout FC literature.

Similarly, participants were asked to identify problems they had when gating FC data and when interpreting it from other sources, to monitor whether the initial issues identified from the first questionnaire were addressed over time, or whether they are still an issue in the community.

Participants were asked to identify (on a scale) how much they actively read or research FC gating techniques to stay aware of current findings, to possibly indicate motivation around completing 'good' analysis. This could possibly suggest that those who do more self-directed reading may have lower variation because they are more aware of the different controls and variables to aid their analysis. Finally, participants were asked to list what noise parameters they consider affecting FC gating and how these are dealt with when analysing data. This new question was determined by the boundary effect observed within Chapter 5, to see if participants are aware of this issue, without being asked a leading question.

8.2.4 Questionnaire Analysis and Coding

After both questionnaires were completed by participants, the results were downloaded and coded where necessary. Data was stored in accordance with GDPR requirements, although no sensitive personal information was gathered through these surveys. This section will address how each type of question was coded (where necessary) for further analysis. Questions could be stratified into five types: Dichotomous (Yes/No), Continuous (numerical measures), Nominal (pick from a selection of options), Ordinal (choose an option from a scale of answers), and written text responses. Three software platforms were used to store, analyse and visualise the data: Microsoft Excel 2016, Matlab R2018a and IBM SPSS Statistics Version 24.

8.2.4.1 Dichotomous Questions

Questions where participants had to select from two options, for example either a 'Yes' or 'No' were coded within the respective platforms, and then visualised using either pie or bar charts to represent the data split of the population.

8.2.4.2 Continuous Questions

Continuous questions were questions where participants had to list a numerical measurement factor. For example, total experience duration with FC were formatted to ensure all results were displayed in terms of months of use (standard unit). Some participants had specified their answers in years, so for further stratification, this data was transformed to the same time scale.

8.2.4.3 Nominal Questions

Nominal questions were questions where participants had to select one of a series of different options that nearest matched their personal experience. For example, when asked how often they used FC, they were asked to choose from: 'Everyday', 'More than once a week', 'Once a week', 'Once a month' or 'Less than once a month'. This data could then be easily visualised in the form of a bar chart to compare the population.

8.2.4.4 Ordinal Questions

Ordinal questions were questions where participants had to select one of a series of scalar options that nearest matched their personal opinion of a particular topic. For example, when asked how much they enjoyed gating on a scale of 1 to 5, they were asked to choose from: '1 – Strong Dislike', '2 – Dislike', '3 – Neutral', '4 – Like' and '5 – Strong Like'. These are commonly used to monitor satisfaction levels and hence were used in the Motivational Factors section of the initial questionnaire.

8.2.4.5 Written Text Responses

These were questions where participants had to expand on choices made or identify particular problems they encountered, by providing a written text answer. This did not limit them to preconceived options that could potentially display bias from the questionnaire author. These

responses were qualitatively coded using manual analysis protocols [179]. Comments were added to stratify responses into one- or two-word answers to represent the problem identified. These results were tallied, and a second round of coding was completed to further align and stratify the responses into similar bins. For example, if a participant mentioned 'fluorescence spill over into different channels' and 'accounting for fluorescence of the data set' were mentioned, these would be coded into a 'Compensation' tally, which then logged how many times particular issues were mentioned to build up a picture of the landscape. Full definitions for the stratified text responses for all questions, along with examples of inclusion and exclusion criteria can be found in Appendix D.

8.3 Survey Results

8.3.1 Experience Results

Participants surveyed predominantly used Flow Cytometry for immunophenotyping purposes, followed by Cell counting. In this instance participants were able to select multiple options, because of the diverse assays that can be run with a Flow Cytometer. Most participants were focused on development of immune therapies, shown by the high use of Hematopoietic lineage cells defined within the survey, shown in Figure 155. Use of pluripotent stem cells for other cell therapy research was common, as well as the use of HEK293 cells which are used within the development of viral vectors for autologous treatments.

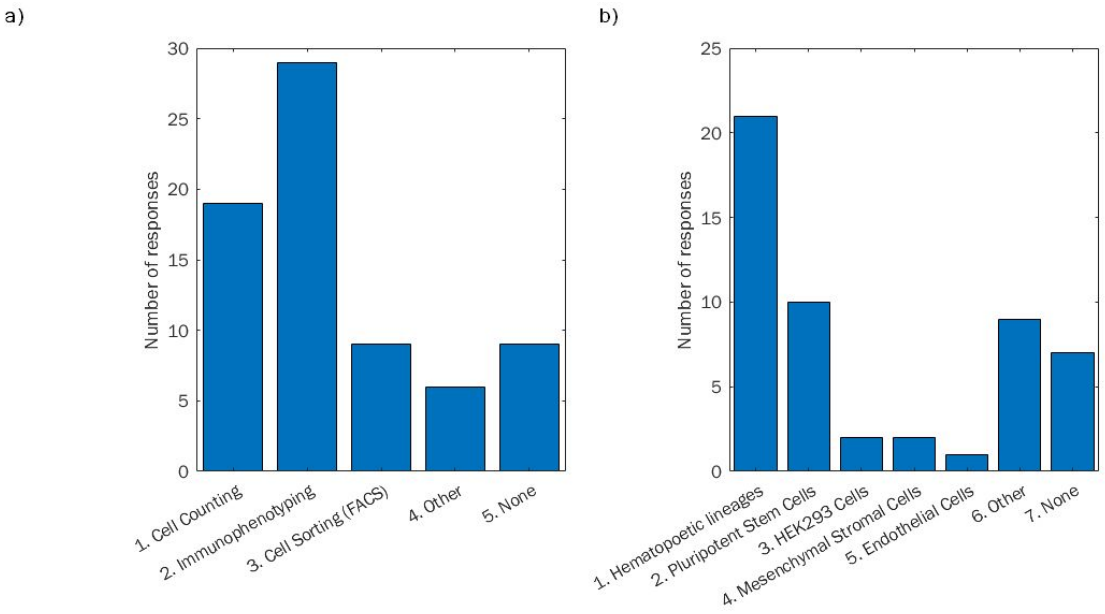


Figure 155, a) Common uses for Flow Cytometry, b) Common cell types analysed using Flow Cytometry

Participants were asked about what they most commonly used, indicating current usage for both Flow Cytometry equipment and software (Figure 156). The most common Flow Cytometry equipment used was a MACSQuant Analyser 10 [202], followed by BD instruments [203,204]. This is a core piece of equipment for participants located at the cell therapy product development centre, hence the popular results. Flow Cytometry software used was a lot more variable and based on user preference. Flowjo [123] was the most popular software because participants found it easy to use and it gave them more control to complete more complex analysis. However, MACSQuantify [205] was popular due to use in cell therapy process development at one site. It is not as flexible for the user; however, it was mentioned on multiple occasions that the automated sample gating it provided was beneficial.

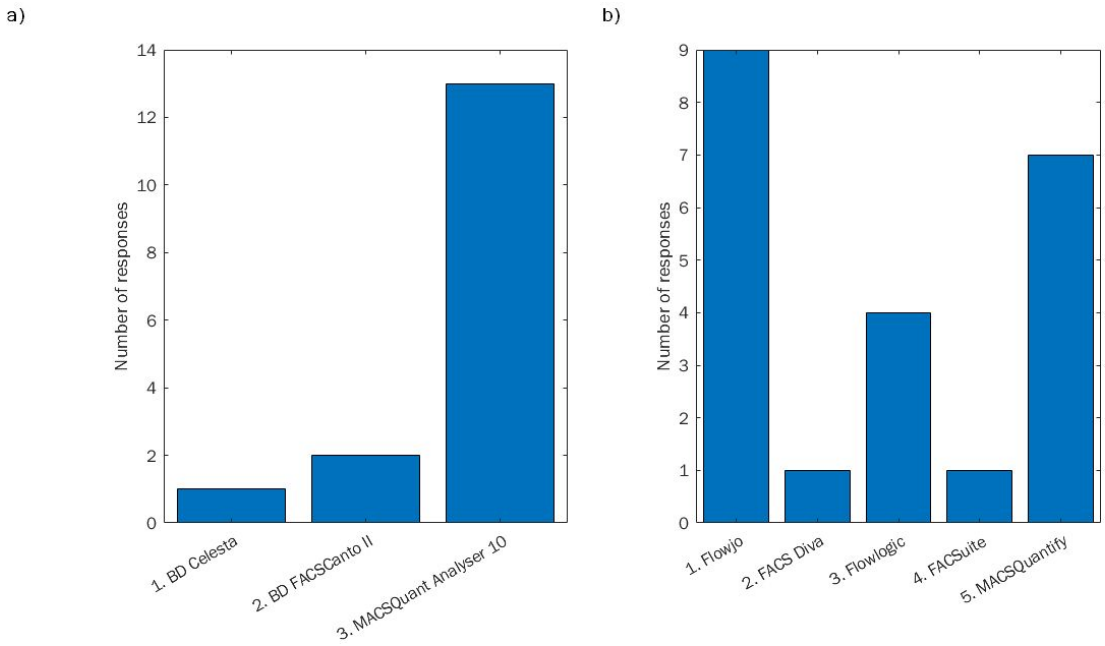


Figure 156, a) Flow Cytometers used by participants, b) Flow Cytometry analysis platforms used by participants

Participant experience (total time using a FC) for the initial and follow-up questionnaire has been overlaid using the histograms in Figure 157. More participants took part in the initial survey, hence the wider range of results. Experience was captured alongside use frequency of a Flow Cytometer.

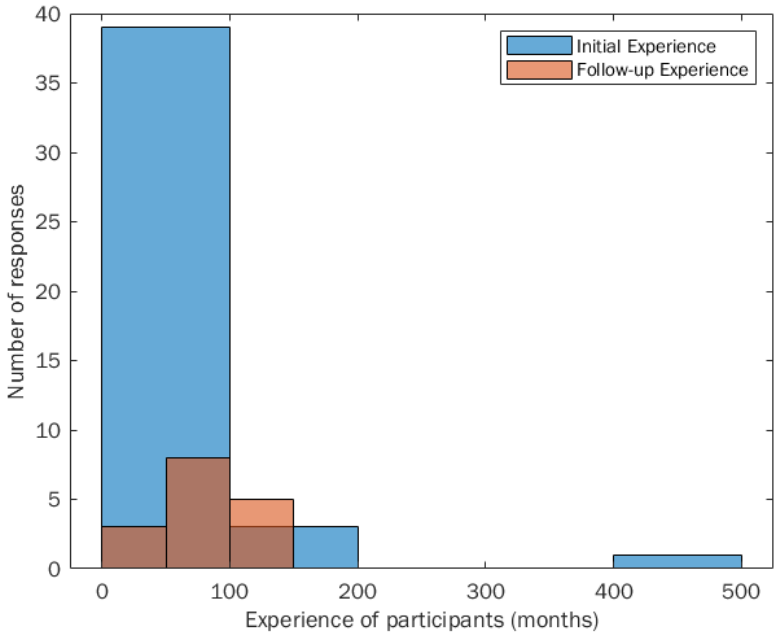


Figure 157 Overlaid histograms of participant experience with Flow Cytometry

The initial survey experience values have been plotted against results from all three uncertainty studies to show the range of absolute cell counts (Figure 158a), *CV*s (Figure 158b) and

uncertainties (Figure 158c) in comparison to the experience of the operator. These results, as well as results for Figure 159, have been taken from the 'personal judgement' phases of the uncertainty models, because this was the only phase tested for the basic model and is thus comparable. This was to identify if there was a potential convergence of results as the experience of an operator increased.

Experience of operators has been considered in different contexts throughout Flow Cytometry literature. Often for comparison studies, a selection of 'experienced users' are chosen to complete analysis, to ensure participants are all familiar with analysis pipelines [1,206], because it has been noted that differences between operators can cause variance in final data analysis, among other steps [19].

Correlations between analyst experience and variation has not been conducted within Flow Cytometry literature, so a formal, wide-range study could aid this information. However, many comparisons have been made in other healthcare and screening fields. For instance, variability of dentin adhesion and resin cements have been explored between students and fully qualified dentists [207,208], with one study identifying qualified dentists having a lower *CV*, however, the other study found no difference in adhesive cement performance between student and dentist application. The impact of training and practice has been monitored for carotid artery procedures between students and registrars, highlighting the improvements simulation training can provide to aid novices [209]. Comparisons between human and automated, novice and expert judgement of luggage screening has also been investigated, with monitoring of advice given how much this is subjectively trusted. 'Novice' automation showed more reliability than novice humans, however, 'expert' automation had lower compliance relative to expert humans. Perhaps stratified experience or particular job roles are a more suitable way to consider experience going forwards, however the distribution of experience with results was required here, so stratification has not taken place [210].

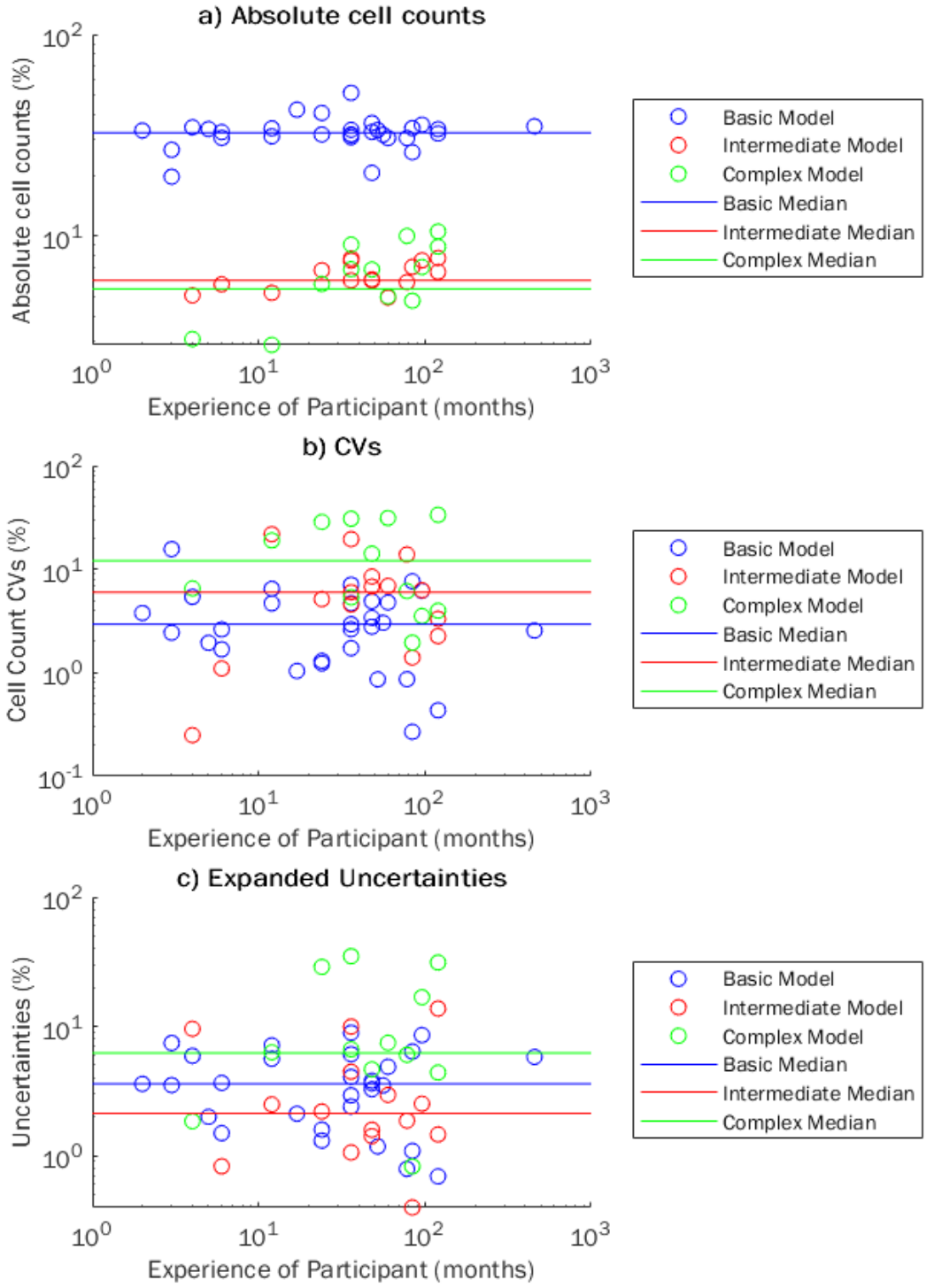


Figure 158 Comparison of participant experiences with a) Absolute cell counts, b) Reported CVs, c) Reported expanded uncertainties

There is no apparent correlation between participant experience and the absolute cell count achieved in any of the cell complexity models (Figure 158 scatter diagrams shown on log scales for clearer identification of population shapes). This is the same for *CV* and uncertainties for all three cell complexity models. The absolute cell count median for the basic model was higher than the remaining models, however this model required less stratified analysis than the others and a different cell type, so cell counts are higher because of these reasons.

Experienced operators are more knowledgeable on the topic having spent longer using equipment (assuming constant use). However, these results have shown that knowledge should not be confused with result accuracy or repeatability. More 'experienced' users are not necessarily more accurate, shown by the scatter plot in Figure 158a, because experienced users are not more convergent towards the median value, assuming this to be the 'most true' cell count. If the true value of a cell count was known, and more experienced users were more accurate, there would be a higher probability of achieving this result.

This is also the case for precision, whether it is represented by *CV* or uncertainty. There appears to be no convergence of *CV* or uncertainty results for any of the gating models, as shown by Figure 158b and Figure 158c respectively. Again, it should be noted that more 'experience' does not necessarily mean operators are more repeatable, although they may be more knowledgeable of the sources of variation. Continuous improvement efforts for reducing variation within FC gating are required, to constantly identify sources of variation that impact inter-operator reproducibility. Throughout this research, boundary effects have been identified to be potential influencers on both the accuracy and precision of data, depending on the software platform in use. A new training session or information provided to participants within future studies should identify this issue, to ensure it is a variable removed from analysis. Software platforms could also make a unified effort to align visualisation methods.

The use frequencies that participants specified in the initial questionnaire have also been plotted against the absolute cell counts, *CVs* and expanded uncertainties for each model, visualised in Figure 159 (y axis scales are linear to ensure easy comparability across different plots). The use frequencies have been coded as follows: 1 = Use every day, 2 = Mode than once a week, 3 = Once a week, 4 = Once a month, 5 = Less than once a month.

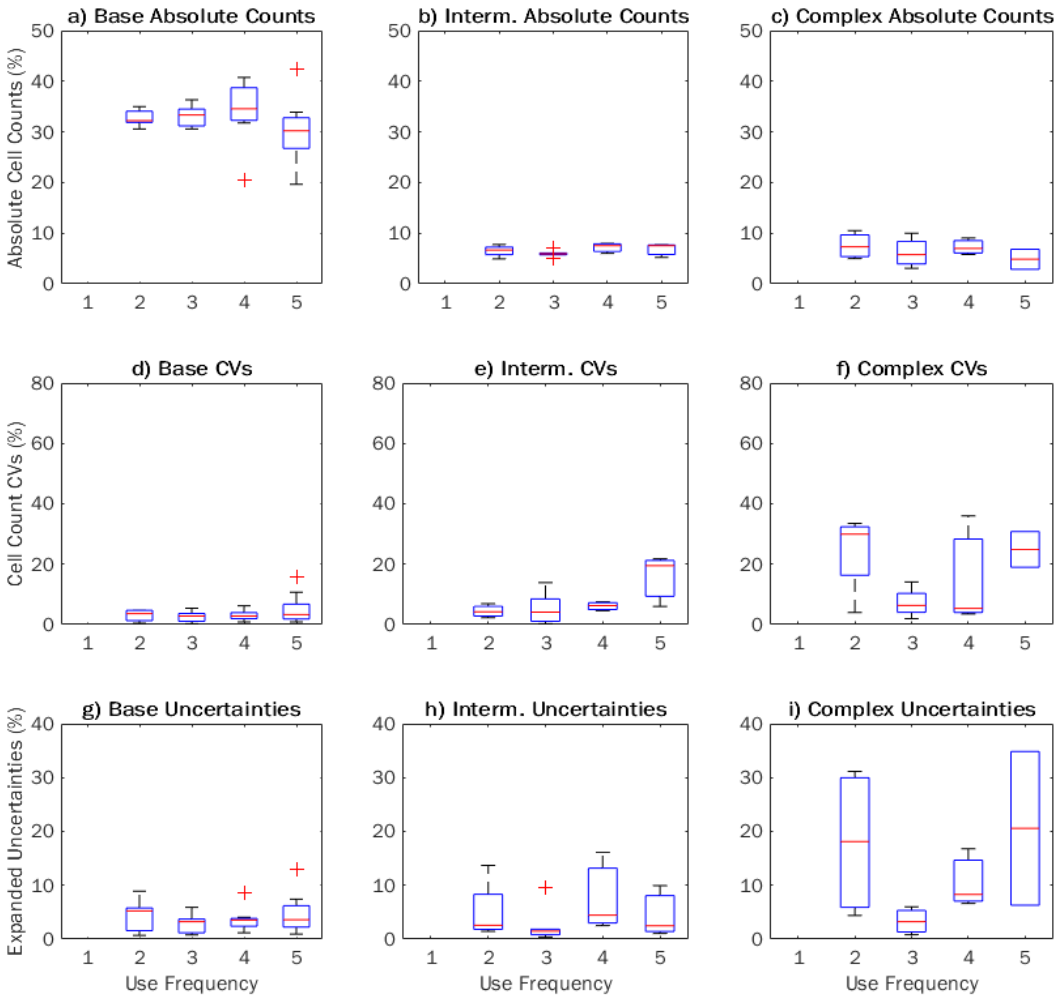


Figure 159 Use frequency compared to a-c) Absolute cell counts per model, d-f) Cell Count CVs, g-i) Expanded Uncertainties

There is no significant difference between the different use frequencies with respect to absolute cell counts across any of the cell models (Figure 159a-c). However, in the basic model (a) there is a greater range in those who use a Flow Cytometer the least (5). Those who used a FC less than

once a month also have the greatest interquartile range (IQR) of inter-participant *CV* in each model, however, this is not the case for uncertainties. Those who use a Flow Cytometer more than once a week have the largest IQR within the base and intermediate model, and the second largest IQR in the complex model, after those who use FC most infrequently.

In a similar manner to the experience data, there is no significant evidence to suggest that those who use a flow cytometer more frequently are more accurate or more precise. This is especially the case with the uncertainty data (g-i), because those who use it the most frequently and the most infrequently have the greatest ranges of uncertainties. This could be due to under- and over-familiarity with the process. However, further studies to achieve statistical power would be required to confirm this. Those who are using FC regularly will probably commit protocols to memory and over time possibly find shortcuts or ways to speed up the process. This could cause deviations in measurement, impacting accuracy and repeatability. Those who are unfamiliar with the process will probably take longer to identify the necessary populations and may not be aware of all the factors that could affect their subjectivity of the final population.

Further work could investigate these factors in relation to the range of cell counts and variability between participants, gathering a wider range of experience and use frequency of participants. In addition, covariance between the experience of a participant and their use frequency could also be explored, to identify whether a combination of total experience with FC and how often it is used could better identify sub-populations of participants who contribute to greater inter-participant variation.

This could help within Cell and Gene Therapy manufacturing scenarios, to better inform training protocols, and also identify suitable intervals for ‘refresher’ or ‘retraining’ periods for operators. These would have to be carefully integrated into the work environment, with an emphasis on fostering a good working culture, to ensure participants see it as an opportunity to improve

themselves and the quality of their product, rather than it being seen as a test or reprimand if they are not within the manufacturer-defined control limits.

8.3.2 Training and Proficiency Results

Participants were predominantly trained internally by another member of staff when they were required to use Flow Cytometry (Figure 160). Some had attended training courses instead, but the courses listed were operational Flow Cytometry training, ensuring all participants could turn on, calibrate and run FC experiments with correct compensation, but none covered how to gate. Arguably, there are so many different cell types to gate that there is no 'one size fits all' policy for applying gates to data, hence why internal supervision or learning was more popular. Some participants supported their training with self-led activities such as watching online tutorials or videos to show how others applied gates to similar data or cellular subpopulations. This highlights the need for some uniform gating templates or a good practice guide to possibly be created for particular cell types, which gives a cheap and effective point for users to tailor to their needs.

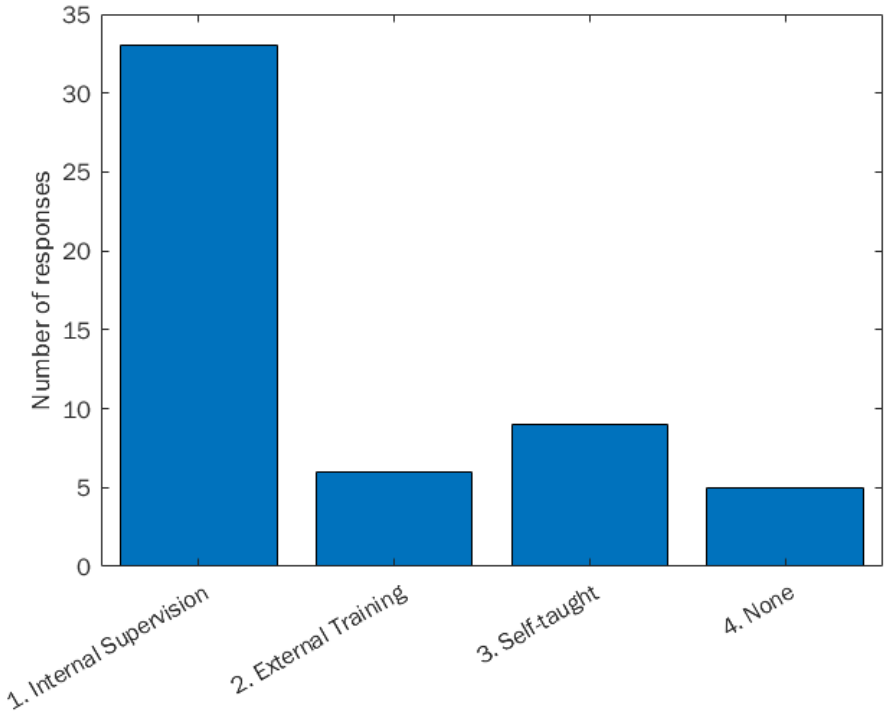


Figure 160 Flow Cytometry training

Participants were also asked if they used Standard Operating Procedures (SOPs) to apply gates to their data, as well as using protocols to gather the data, with the analysis shown in Figure 161. Most participants did not use a protocol to apply gates to their data, however, of those who did the majority used an internal SOP to apply gates. When asked about this in the gating sessions, most indicated they had a rule of thumb when applying gates to control files. For example, they would get as close as possible to 1 % positive cells in the FMO control, rather than use features of the visualisation such as axis scale points or images of protocols. This highlighted that individuals completed analysis separate of one another, so a general rule worked better for alignment of protocols, because there was no guarantee that the compensation, voltages used, or scaling would be the same between participants running the same assay.

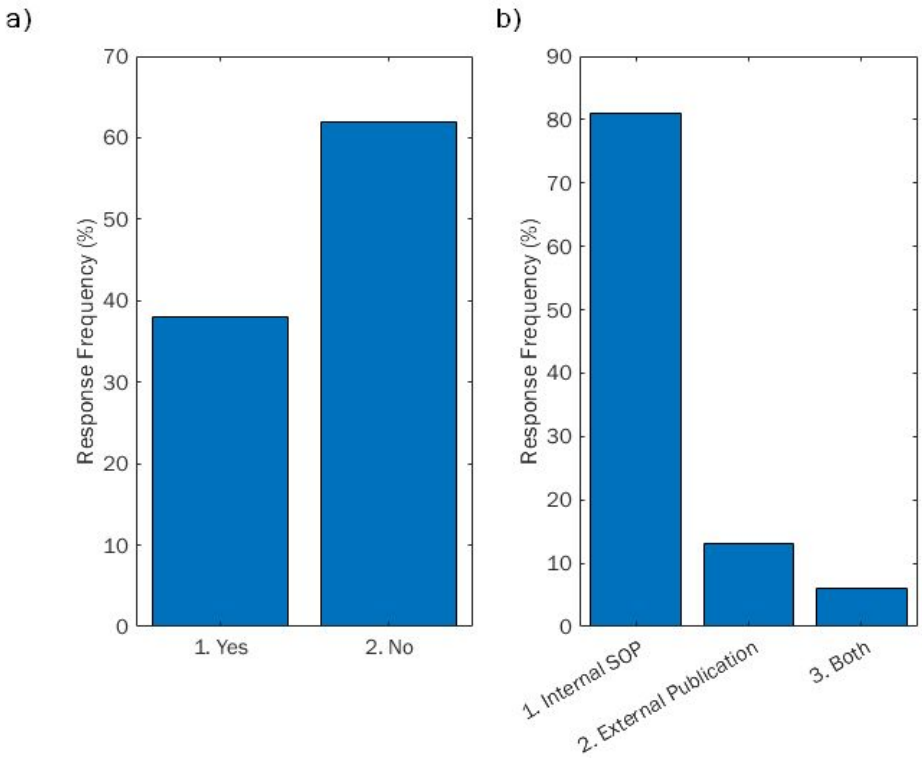


Figure 161, a) SOPs used to apply gates, b) Types of SOPs used

In addition, no participants had taken part in an External Quality Assessment Scheme (EQAS) or submitted files for central processing. These schemes or file transfers are common within clinical cytometry to monitor laboratory proficiency and reduce inter-participant variation. It remains to be

seen whether these are schemes that the wider manufacturing community could also benefit from, to reduce subjectivity, and continually identify sources of variations contributed within analysis. This questionnaire was only given to the three collaborator sites, so if this were to be repeated it could be sent to participants across a diversity of centre and job roles. No participants had formal FC qualifications unlike their clinical counterparts, so this could be something that may be required in future for operators within manufacturing environments.

8.3.3 Manual and Automated Gating

Participants were asked in both the initial and follow-up questionnaires to identify their preference for manual or automated gating platforms. Manual gating allowed the user to completely identify where they want to place the gates, and automated uses a machine learning algorithm to create separate population clusters. Figure 162 shows the split between participant preference for manual or automated gating, within the initial questionnaire and the follow-up questionnaire. The results are shown as response frequency percentages, because the follow-up questionnaire contained less responses. The follow-up questionnaire asked this question to identify whether the increased prevalence of automation algorithms within analysis software were becoming more popular to use than manual gating. In both instances, manual gating was greatly preferred to automated gating, with stratified reasons presented in Figure 163.

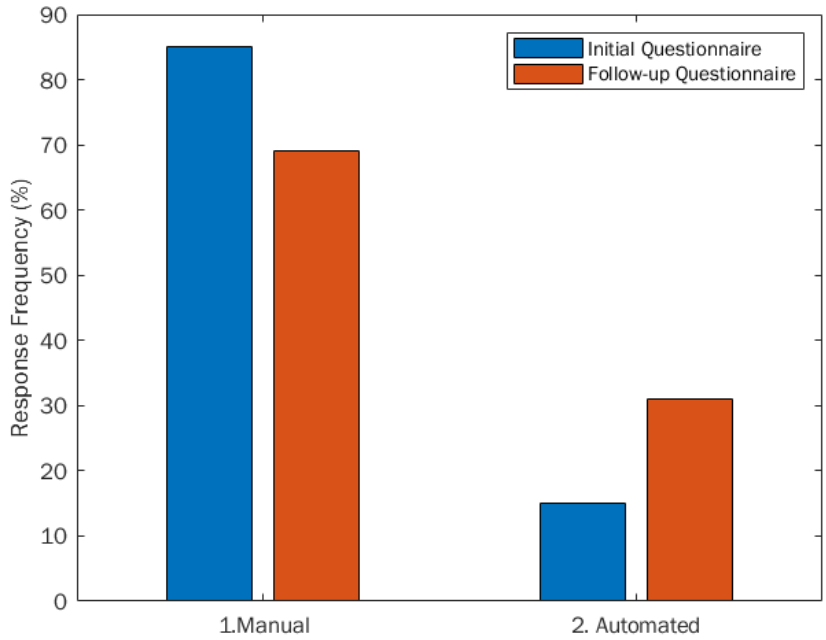


Figure 162 Preference for manual or automated gating within the two questionnaires

From the initial questionnaire responses, the only beneficial reason given for using automated platforms was due to the speed of analysis, with more dimensional data being more cumbersome for an analyst to review. Even though manual gating was preferred within both questionnaires, a greater amount of reasoning for using automated platforms was identified within the follow-up questionnaire, suggesting that the benefits of automation are starting to make traction within the community.

To further monitor this, the results from one site were compared between the two questionnaires. This site had the least changeover of staff, for better consistency and tracking of automation popularity. 3/20 participants preferred automation to start with, however two of these operators did not take part in the follow-up survey. The remaining participant still felt automation was better in the second survey, with a total of 5/13 participants preferring automated algorithms at the time of the follow-up survey. Only one of these participants was a new member of staff, so there had been a few preferences changed to automated gating platforms over the course of the research. It was expressed that algorithms can be trained to identify particular subsets, hence can potentially

be accurate, but cannot identify or deal with biological variation as well as a manual analyst, so it lacks precision and reproducibility.

Both questionnaires identified users wanting much more analytical control, for visualisation purposes as well as double-checking findings. This, as well as legacy of use ('*I prefer manual gating because it's what I've always known*'), identify a potential resistance to change to new techniques, especially where removal of specialist human analysis is concerned. Many participants identify that they don't trust automated platforms, and they would have less control with exploration of the data once it had been clustered using a machine-learning algorithm.

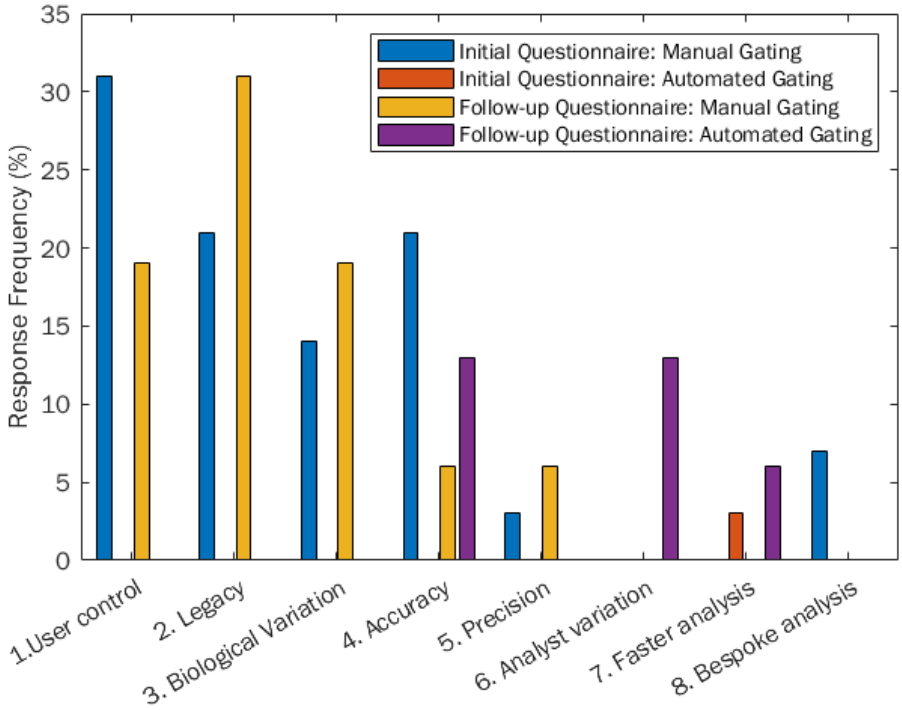


Figure 163 Preferential reasons for choosing manual or automated gating across both questionnaires.

Automated platforms need to improve in accuracy, precision and handling of biological variation to allow participants to move away from manual analysis. Equally, an awareness of a participants own variation when gating would give them a personal benchmark when testing and potentially validating new algorithms for use in population identification, which is partially what this research aims to show.

Consideration of automated gating requires a user to be working with Flow Cytometry very frequently, due to the time taken for the algorithms to learn from historical data or training sets, which is suitable for those in process development or manufacturing, but not those in a research environment, who may use bespoke panels at infrequent time points, making manual analysis more convenient. At all developmental stages of a cell therapy product, from research through to approved manufacture, an analyst should have an awareness of their variability, to ensure false positives or false negatives are not declared, which could become more costly to patients, research groups and manufacturers if carried forwards.

8.3.4 Problems identified when gating

Participants were asked to identify what problems they encountered when gating FC data. This was asked in both questionnaires to see if the same issues were prevalent after two years had passed, in case more had been done in the community to resolve well-known issues. Figure 164 identifies the stratified results from both questionnaires, with the biggest issue being reliable cell cluster separation. The ability to separate clusters was a key issue for many participants, even when proper titration and voltage setup had taken place, and it was still the main problem identified within the follow-up questionnaire. More cell separation and clustering algorithms are present for the Flow Cytometry community, however, they often require knowledge of an additional software package such as R or Matlab to process the data, which many analysts may find difficult if this is not a main component of their job role [1].

Participants were asked to detail general problems they deal with when gating, outside of the constraints of the studies they took part in. Analyst variation was noted as much more of a problem in the initial questionnaire, however, inclusion in these experimental studies could have potentially skewed the follow-up analysis because this is the focus of the research, so they listed other issues. Dealing with biological variation was also noted as a difficulty when analysing lots of donor

information together, calling into question the validity of the gates, and where definitive cut-offs should be places for positive and negative discrimination. Staining quality can potentially help here, if all the upstream process steps have been followed to minimise fluorescence spillover, however, differences between analysts could cause these differences to have knock on effects within the final analysis. The introduction of automation could be a good way to potentially alleviate these issues to remove differences in staining times and temperatures, due to standardised, repeatable processing.

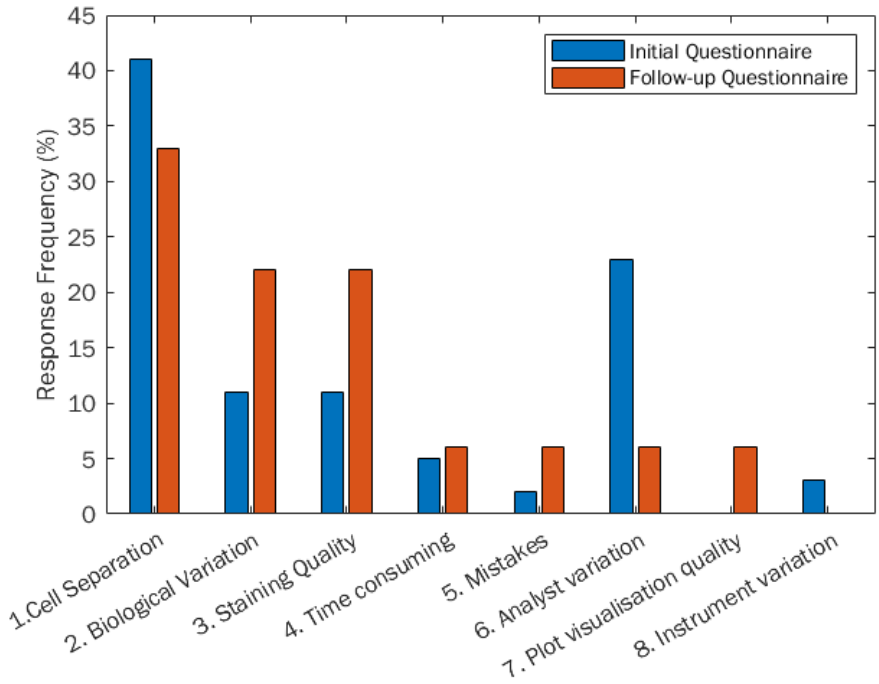


Figure 164 Identified problems when gating

After observing boundary effects during the participant analysis for the intermediate model, an additional question was added to the follow-up questionnaire, to identify whether participants were already aware of this effect, without asking leading questions. This boundary effect was not listed among the noise parameters specified by participants, stratified in the bar chart in Figure 165. The noise parameters identified mostly concerned upstream analysis of the sample itself, either due to the sample preparation or instrumental variables during file generation and acquisition of data. Fluorescence spillover can affect the noise within a file and subsequent gating of the data, however,

participants always discussed this issue as a feature of poor compensation, which occurs before the file generation, rather than the post-analytical gating itself.

Cell debris and doublets were the only noise issues that occur at the post-analytical gating phase. Removal of these is usually completed during the first two gates, where the desired population must be separated from the smaller broken pieces of dead cells (debris) and the cells which stick together through the laser interaction point during sample acquisition (doublets). During the basic and intermediate models, the cell debris and doublet gates were the gates which caused the most variation within the uncertainty budget, so there is a need to focus on these to reduce variation. In addition to this, FC has many uses outside of cellular measurement, once of which is bacterial measurement. Bacteria are much smaller than cells, which would cause the cut-off for debris to be even closer to the required population, making subjectivity of gate application more important [211].

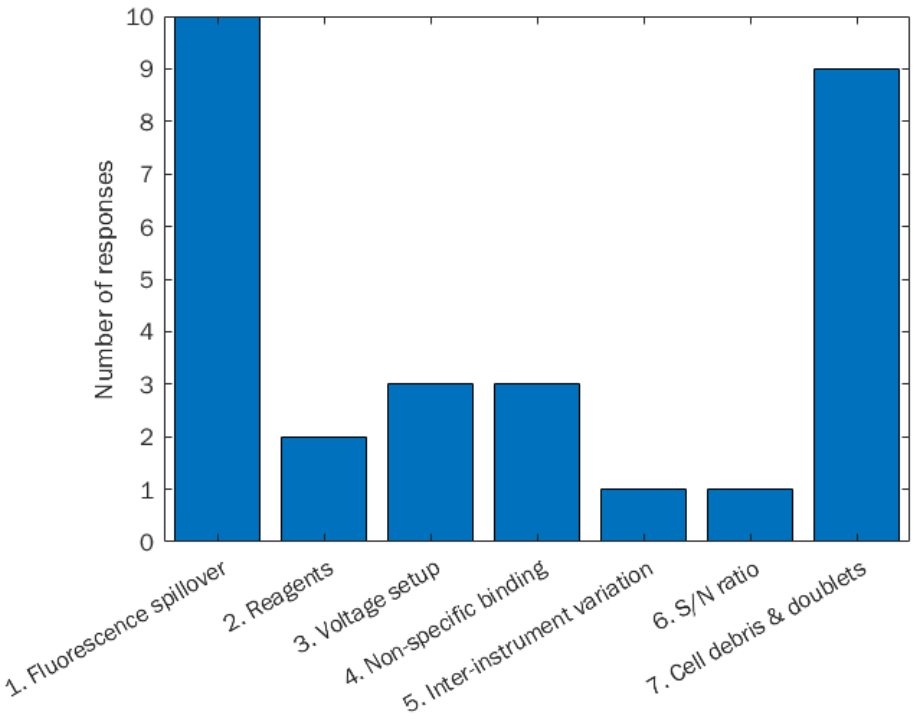


Figure 165 Noise parameter issues identified by participants

8.3.5 Interpretation problems from literature

Participants were asked in both the initial and follow-up questionnaire to identify what issues they encountered when they interpreted FC data or gating methods from published literature. This gave an indication of the initial problems faced, and whether standardisation in the community had improved between the questionnaires or whether the sample problems were still being faced.

Gating standardisation was the most common problem faced within both questionnaires (Figure 166), officially defined in the codebook located in Appendix D. Participants identified that more information on the gating hierarchy completed and standardised gating sequences for well-known marker panels were required for reproducibility. This was followed (in both questionnaires) by a lack of control file images provided by authors, to show the reader how the gates were applied to the fully stained sample. This can guide the reader to show what percentage positive to work to if not specified, to aid reproducibility of FC data. Better reporting of reagents (and staining methodology) was also a common issue, as well as poor visualisation of the Flow Cytometry data itself.

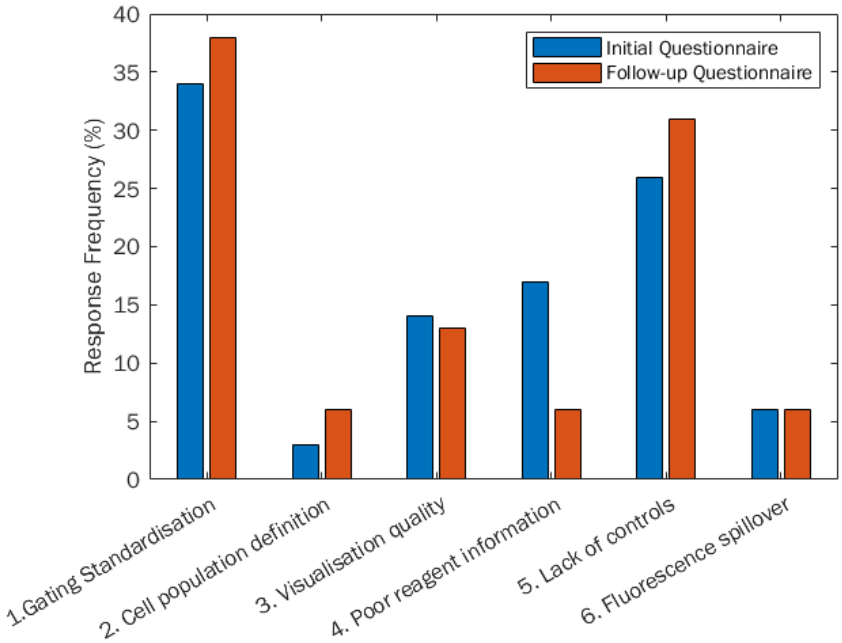


Figure 166 Identified problems when interpreting FC gating from literature

Minimum Information for a Flow Cytometry experiment (MIFlowCyt), is a reporting standard generated by the International Society for Advancement of Cytometry (ISAC), which lists the minimum requirements for a Flow Cytometry experiment, from sample preparation information through to how the sample has to be acquired and gated [49]. Despite these excellent efforts in the right direction by ISAC, many participants were not aware of these sorts of standard reporting formats, because they do not follow the Flow Cytometry community journals (such as Cytometry A & B) and relevant publications from ISAC. When asked if they undertook self-driven reading to stay up-to-date on current methods, most participants were only familiar with what was listed in the training protocols provided at their institution/company. Any additional reading completed (if a participant wanted to) was mostly focused around specific cell types they worked with, rather than Flow Cytometry as a technique. Therefore, if this information is not provided to wider communities, or they do not actively seek it, it will always be an achievable unknown, hence the need for Good Practice Guides (GPGs) that can be universally accessed to address general issues or training tips.

8.4 Chapter Conclusions

Providing two questionnaires throughout the process of the experimental research, has delivered potential correlations between experience metrics and study results to be investigated. In addition, key issues participants have identified as problems when gating are consistent over time, indicating key areas of work for the Flow Cytometry community.

The structure of both questionnaires has been successful and sending the surveys as online links for participants to complete was the most convenient way to gather the data for participants as well as providing secure storage of the results for quick download and analysis. Manual coding of written text responses was completed by the author, following manual coding protocols, with two analysis cycles for effective stratification and codebooks written to ensure complete definitions of the code were provided for all.

Most participants used Flow Cytometry for immunophenotyping haematopoietic lineages or pluripotent stem cells. This relates to the three centres that took part, which are all focused on the research, development and manufacture of cell therapies, most of which are currently haematopoietic-based and relevant to the wider Cell and Gene Therapy community. Immunophenotyping allows participants to build panels to monitor the progress of cell-based treatment for different antigen markers that are features of cells or engineered additions. A MACSQuant Analyser 10 Flow Cytometer was the most popular piece of equipment, due to most participants involved being located at the centre where these are the validated FCs for research and development. These participants therefore were all trained to use the MACSQuantify analysis software that comes with the instrument, as well as most participants using Flowjo to complete their analysis.

Flowjo was the chosen platform for the uncertainty studies, due to all participants and centres being familiar with the software, requiring less pre-participation training, and it was straight forward to transport as a plug-in dongle. Participants preferred this platform due to easy of use and the amount of control it gave them, further identified when asked about preference of manual or automated gating platforms. Most participants preferred to use manual gating rather than automation within both questionnaires 85 % and 70 % respectively for the initial and follow-up questionnaire), because of the control it gave them to investigate and explore the data with different visualisation methods. Legacy was also a prominent factor, with participants resistant to change because they are familiar with this style of analysis, and sceptical of more automated methods being introduced to try and help more highly dimensional datasets. Possible culture changes are needed to help address this *'I've always done it this way*' attitude, by empowering employees with training opportunities and easy access to validated software (if proven to be suitable for use).

Most participants within the initial questionnaire had less than 12 months experience with FC, hence why this was followed-up in the second questionnaire. Experience was continually identified

by participants through their training experiences because they had been trained internally by someone who had worked on the equipment longer (suggesting they were better because they had done it for longer), in addition to also attending just a few external training courses delivered by instrument manufacturers. This training mostly covers 'operational' FC, so participants are signed off to show competency when turning the machine on and off, running calibrations and compensation for experiments. These training courses generally do not include gating principles because of the diversity of FC application, but there is an opportunity for training to include this for client specificity. Therefore, this gating training is completed internally by observing others, reading protocols or a combination of both. Most participants did not use SOPs to apply their gates, but those who did used an internal SOP to complete this rather than using external references or a self-made protocol.

No obvious correlations were identified between the amount of experience or use frequency of a participant in relation to the range of absolute cell counts reported, *CV*s from cell counts or respective expanded uncertainties across any of the cell models investigated. The amount of experience a participant has does not mean they are more precise or accurate than a participant with no experience at all. Further research is required into this to ensure a wide range of participants are gathered to represent the complete diversity of experience and use frequency for statistically significant, powered experimental design. Experience can indicate the number of variables a participant is aware of that can affect the final outcome of the data, due to greater repeats of the experiments.

In addition, participants were made aware of the aims of the study to meet ethical requirements, but this informed them that their data would be subsequently investigated to anonymously quantify their variability through uncertainty calculations. Because this was not completely blind, this could have inferred Hawthorne's effect because participants knew they were being reviewed, so changed their behaviour and gating accordingly during the experimental studies. Some participants mentioned this during their study session, making remarks such as '*I'll gate this properly for you,*

not how I usually do it', indicating they would have done things a different way to how they want a third-party observer to see. This could indicate that the results identified throughout the studies could be conservative, not just because they were tightly controlled but because participants knew the data would be reviewed, so tried to be as uniform as possible, defeating the objectives of the research. To address this over time, routine analysis could be taken form participants, but it would be difficult to do blindly due to ethical consent and how the reporting and feedback will be handled.

EQAS can be used to monitor proficiency of the laboratory and its participants, but it requires the right culture and participant behaviour to support its introduction and integration, to ensure that all behave professionally and use it as an opportunity to learn rather than becoming a big brother environment, fostering blame culture in the workplace. Employees need to be empowered through training and given the opportunity to speak openly and safely about the problems they have faced to ensure a team-led dynamic can enable effective solutions to improve the quality of a product. Equally, management also needs to handle the data responsibly, to deliver continuous improvement professionally and anonymously, identifying potential issues that can affect employees without making those involved feel responsible for issues which brought these opportunities to their attention.

This is a community wide issue, not just one that could reside within one company or site. Participants identified many problems when they gate data which could cause variation, with key difficulties being cell separation of multiple populations and biological variation of samples. This was identified at the start and end of the research, indicating these issues are still prevalent within the community. Many separate initiatives have tried to address cell separation issues using automation platforms, especially where high-dimensionality data is concerned [1,118,212]. However, these are all still research-focused and not validated for mass manufactured products. As well as this validation requirement for use in CGT manufacturing, this questionnaire has highlighted a resistance to using these tools, indicating that providers need to demonstrate

usability of the platforms as well as very good accuracy and precision, and the ability to handle biological variation between donors.

There also needs to be a community moves towards gating standardisation and better identification of the use of controls used to set gates, according to these questionnaires. When interpreting data from literature, many participants struggled to identify key factors from the text of figures that enabled good reproducibility of the data. Participants also failed to mention key efforts that have taken place in the FC field to standardise FC experimental reporting. Better cross-over between assay-focused and cell-based journals is needed to enforce effective reporting standards to the wider community. Additionally, within teams or departments, nominated employees could be responsible for staying abreast of the latest updates in certain development journals and disseminating this information within the team, to ensure all are aware of current methods to improve reproducibility and good practice.

8.4.1 Consolidation of Objectives

- These questionnaires were implemented and managed well, allowing effective analysis of a series of dichotomous, ordinal, nominal and written text responses for personal, experience and motivational factors.
- No observed correlations between participant experience and FC use frequency was seen with respect to the range of absolute cell counts, *CV*s or uncertainties. It could be possible that experience, or overall knowledge on the topic infers that people are better at gating if they are more knowledgeable about FC. However, more knowledge on the topic does not necessarily mean they are more repeatable and precise in their analysis.
- Both questionnaires showed a majority of participants preferring manual analysis of data (85 % and 70 % respectively) using platforms such as Flowjo, because of the amount of control they had of the data, as well as a resistance to change to new methods and platforms being addressed in the literature and from companies. However, those

preferring automated methods indicated it is much quicker than manual analysis of large data sets, setting a precedent for future development.

- Problems identified from manual gating were the same at both time points, with Cell separation and Biological variation being the biggest issues participants faced when analysing lots of data. These were also things that automated platforms struggled with, suggesting a focused effort is needed in this area.
- Reproducibility of data from literature is a big concern for participants, with key issues for improvement in the community being standard experimental and visual gate reporting as well as including images of the control files used to help others repeat the analysis when potentially using different instrumentation. Wider knowledge of standards such as ISAC MIFlowCyt reporting is needed across different cell-based communities, to maximise the standardisation of reported data.

Chapter 9: Translational Outlook

9.0 Introduction to the Chapter

Chapter 9 investigates whether subjective issues seen within Flow Cytometry processing and post-analytical data analysis, such as cell population area selection are also present within other assays and platforms used by Cell and Gene Therapy Process Development teams. 8 interviews were conducted with industrial contacts who work with the Center for Biomedical Innovation at MIT, during a 6-week secondment, hosted by Jacqueline Wolfrum and Anthony Sinksey. The platforms investigated as part of this translation analysis were Flow Cytometry, Imaging Platforms and quantitative Polymerase Chain Reaction (qPCR) assays. Therefore, this Chapter captures some subjectivity issues already addressed and discussed from the main uncertainty models in Chapter 3-7, and from the questionnaire results in Chapter 8. This Chapter aims to give a high-level review of translational similarities to human factor variables seen within Flow Cytometry but does not aim to deep-dive into specific factors at this stage.

Imaging platforms were considered for the interviews because of the prevalence of imaging-based technology within CGT manufacturing. qPCR was also considered in one case where a participant had qualified this assay against a viral plaque imaging assay. This was considered across a range of exemplar CGT products, such as viral vectors, CAR-T suspension culture and Mesenchymal Stromal Cell (MSC) adherent cell culture, so was suitable for a wide range of potential interviewees.

Operator subjectivity has been investigated across a wide variety of applications within medical imaging [213–215], however, with the recent use of imaging platforms to monitor to progress of cell therapy growth [216], differentiation and quality of products [217], the variability of analysts needs to be understood. A lot has been done to develop automated analysis for cell imaging platforms, to monitor cell morphology in relation to cell therapy development [218–220], because

it has the potential to be used as an in-line measurement tool. This is an attractive alternative to current invasive or off-line sampling methods to monitor CGT expansion. The interviews conducted here aim to identify how imaging platforms can be enhanced to make them more standardised measurement tools for monitoring CGT products throughout manufacture and release measurements. This translational outlook also investigated whether there was a possibility to quantify subjectivity across other platforms, using measurement uncertainty principles exemplified throughout this thesis. This would show a broad applicability of the technique within biomanufacturing, to improve measurement variation, resolution and control in manufacturing.

9.1 Chapter Aims

This Chapter aims to highlight subjective and cultural issues when analysing data across other instrumental platforms used within Cell and Gene Therapy Process Development teams, that are similar to subjectivity seen within FC assays. This focuses on the measurement process, to identify subjective human factor influences within measurements. The second half of the interviews focused on training to identify what training processes are followed and what cultural differences could affect the variability of results between analysts.

9.1.1 Chapter Aims & Objectives

The Aims and Objectives of this Chapter can be defined as follows:

- Interview a selection of industrialists who use cell imaging platforms within CGT process development and research teams to identify subjectivity within the measurement process.
- Identify common causes for subjectivity across these imaging platforms and assays, for future work to address process improvements.
- Identify similarities and differences between training, and attitude to training and using Standard Operating Procedures to remove variation between analysts.

9.2 Methodology

These imaging exemplars were chosen in collaboration with MIT hosts through a review of core measurement techniques used across different CGT products. A lot of optical spectroscopy techniques are used (Flow Cytometry falls into this category), but because of the heavily automated nature of these instruments, it was decided that there was only small scope for application here. Imaging platforms were identified as being heavily used, along with qPCR, so it was decided these techniques would be the core focus. The MIT Center for Biomedical Innovation host a network of international CGT industrial and academic contacts, so this provided a suitable base for this six-week long translational study, ensuring all relevant contacts could be accessed.

9.2.1 Interviewee selection

Relevant industrial contacts known to be working with one of the three exemplars were contacted through the MIT Centre for Biomedical Innovation. An additional conference call with each individual contact was made after verbal acceptance of the initial interview scope to further explain the scope and boundaries of the exercise, to confirm the most suitable interviewees could be contacted and the imaging platform they used was discussed in more detail. This ensured the interview would be suitable for their assay application and if they wanted to discuss a different platform, which they felt had subjectivity issues (other than Optical Spectrometry methods, because this has been largely standardised for Spectrophotometers, and Flow Cytometry has been previously investigated within this research), this was identified and confirmed with the research team. Interviews were held with 8 delegates from across 7 large USA biomanufacturing and standards companies.

9.2.2 Ethical Approval

Ethical Approval was granted from the Loughborough University Human Participants Ethical Sub-committee and the MIT Committee on the Use of Humans as Experimental Subjects (COUHES). All approval documents and consent templates can be found in Appendix E. Participants were required

to consent to the interview at least 24 hours before it was scheduled to take place and forms were stored securely throughout the duration of the project according to UK GDPR regulations. Participants were all anonymised at the point of transcription, so any further analysis was completed blindly by the author.

9.2.3 Interview Questions

Interview questions were split into sections, detailed in the following subsections. An interview template of all the questions asked can be found in Appendix E. One hour was allocated for these interviews, so that all questions could be answered. The interview questions were a mix of continuous, nominal, scalar, and text response questions, as defined and coded previously in Chapter 8. Throughout these interviews, various questions aimed to identify where subjectivity was present in the measurement processes, to identify where translational subjective variation could also be measured, as shown throughout this thesis using measurement uncertainty within FC.

9.2.3.1 Technique Use Questions

In a manner similar to the questionnaires described in Chapter 8, participants identified how frequently they used the respective measurement platform, what the specific measurand used was and whether it was used for quantitative or qualitative analysis. This aimed to identify whether participants used imaging platforms for purely qualitative analysis, or if quantitative analysis was used (viral plaque counting, for example), and what level of automation was used for quantification.

9.2.3.2 Process mapping exercise

Participants detailed the full measurement process for the respective assay, going through pre-process, in-process and post-process sections, described as follows:

- Pre-Process: Any sample preparation required to tag or label the biological product for analysis, as well as preparing substrates or master mixes for analysis.
- In-Process: Instrumental preparation required to run the sample, such as instrumental calibration and background measurements if participants conducted these steps.
- Post-Process: Once the analysis file or exported measurand has been obtained, any further image processing, selection or calculation steps that were required for final reporting.

Once these process steps were defined, participants went through each specified step and identified what human contact time was required, and what elements of each step could be affected by operator subjectivity. Once all interviews were completed, results for similar processes were compiled and transcripts were coded for further stratification, using coding methods previously described [179]. These stratified variables were then tallied for the three process sections to identify subjectivity issues prevalent across the measurement processes. This identified similarities and differences to FC post-analytical variation specifically, to further identify other biomanufacturing platforms that could benefit from harmonisation and subjective variation quantification.

9.2.3.4 Training and Standard Operating Procedure (SOP) Questions

Participants were asked whether they were responsible for training new users on the respective platforms and if so, whether this was purely operational or whether theoretical training was also provided. Participants also identified whether they felt there was a difference between new and experienced users, and why. Finally, participants indicated whether additional reference materials were used for the platform and whether SOPs for data analysis were provided. If yes, they described whether these were internal or from external literature sources.

9.3 Survey Results

Fifty percent of interviewees spoke about their experiences with viral infectivity assays, used in the production and release of viral vector products used in other therapies, so this has been used as an exemplar throughout this results section. Only 8 interviews could be conducted within the secondment time-frame (due to preparation time, geography and interviewee availability), which is a limitation of this study, but it was only designed to be an initial pilot investigation to highlight issues for future clarification with more structured experimental work. This will provide a translation of the FC variation analysis techniques and evidence for application of measurement uncertainty techniques to further quantify operator subjectivity within other biometrology platforms.

9.3.1 Training and Proficiency Results

In a manner similar to the questionnaires in Chapter 8, interviewees were asked how often they actually completed the respective assay themselves. Most participants aimed to complete assays at least once a week, to ensure they stayed familiar with the technology and processes. Those who did not use the assays as regularly were in more managerial positions. Within the CGT space there is a lot to explore with respect to experience and use frequency, however, it is difficult to complete within this research, because participants are based across different companies. Different company structures can possibly indicate different levels of interaction with equipment, which are perhaps better understood from an internal perspective before comparing between different sites and structures.

Table 83 indicates whether these assays were used for quantitative or qualitative purposes by interviewees (or both). It shows results for all interviews, not just those focused on Viral Infectivity. If participants stated it was a qualitative test, they explained what they used as a 'determination of success' to satisfy a decision on the product. Only one process stated was purely qualitative, where Fluorescence Microscopy was used to monitor cell growth over time by a visual increase in fluorescence. Other imaging platforms have been used for qualitative analysis as well as

quantitative analysis. In these instances, the qualitative is used to support the quantitative results if further clarity is required. A qPCR experiment had been validated for viral plaque counting against an imaging platform, so this quantitative method was discussed by the relevant interviewee.

Table 83 Quantitative or Qualitative nature of assays

Assay Measurement	Quantitative/Qualitative	Qualitative: Determination of Success	Quantitative: Final Metric	Numeric ?
Cell Count/Viability	Both	Images to match control	Cell Count	Yes
			Viability	Yes
			Cell Size (µm)	Yes
Cell Growth	Qualitative	Visual increase in fluorescence		
Viral Infectivity	Both	Images for sanity checking	Transducing Units / mL	Yes
Viral Infectivity	Quantitative		Transducing Units / mL	Yes
Viral Infectivity	Both	Manual subjective counting of plaques	Transducing Units / mL	Yes
Viral Infectivity	Quantitative		Transducing Units / mL	Yes
Cell Count/Viability	Quantitative		Viability	Yes
Cell Shape Quantification	Quantitative		Differentiation score	Yes
Immunophenotyping	Quantitative		Cell Count	Yes

Any interviewee who stated a quantitative metric was used was asked to describe the final reported metric. This identified whether different metrics could be reported from the same assay or platform. Viral Infectivity assays all produced a Transducing Units / mL measure, from counts of viral plaques or through qPCR. Despite the methods of achieving these results being different (described in the subsequent process mapping exercise results section), the reported result is the same. This requires further quantitative experimentation, because differences between platforms and methods could return consistently different measures. Counts were provided for any cell counting method for MSC or CAR-T exemplars as well as novel scoring methods being created to monitor cell shape with respect to differentiation.

9.3.3 Process mapping exercise results

Each interviewee gave a detailed process breakdown of their respective assay, from the start of pre-processing to the final measurement and/or decision. Viral infectivity maps have been exemplified here because they were most commonly discussed, have the same reported metric (Transducing Units / mL), yet all methods of obtaining this metric were different between interviewees, as shown in process maps in Figure 167 to Figure 170.

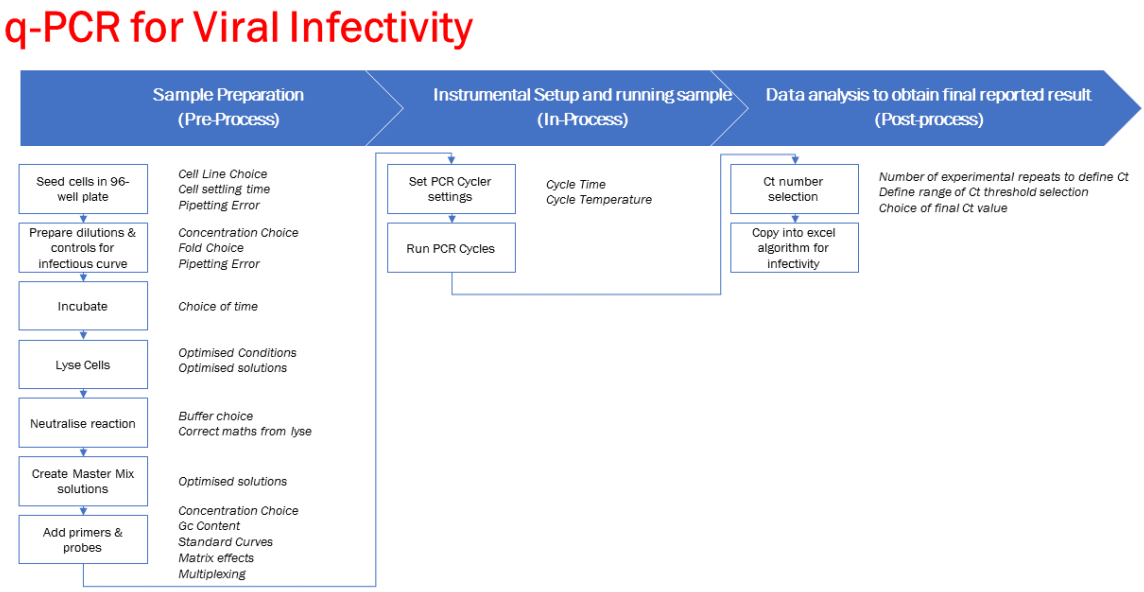


Figure 167 Process Flow Map for qPCR Viral Infectivity measurements

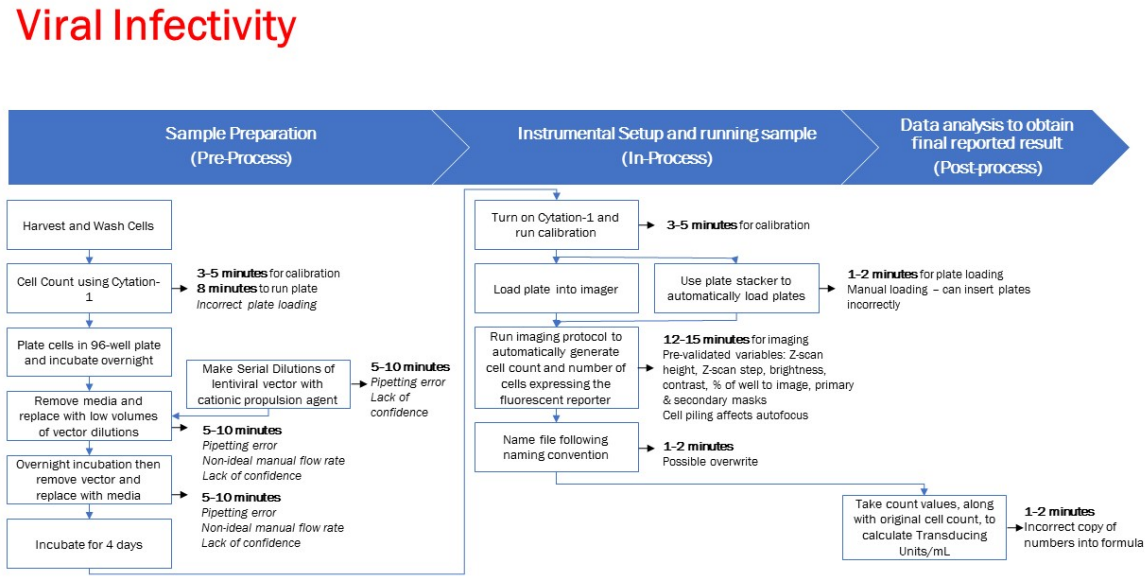


Figure 168 Process Flow Map for imager Viral Infectivity measurements

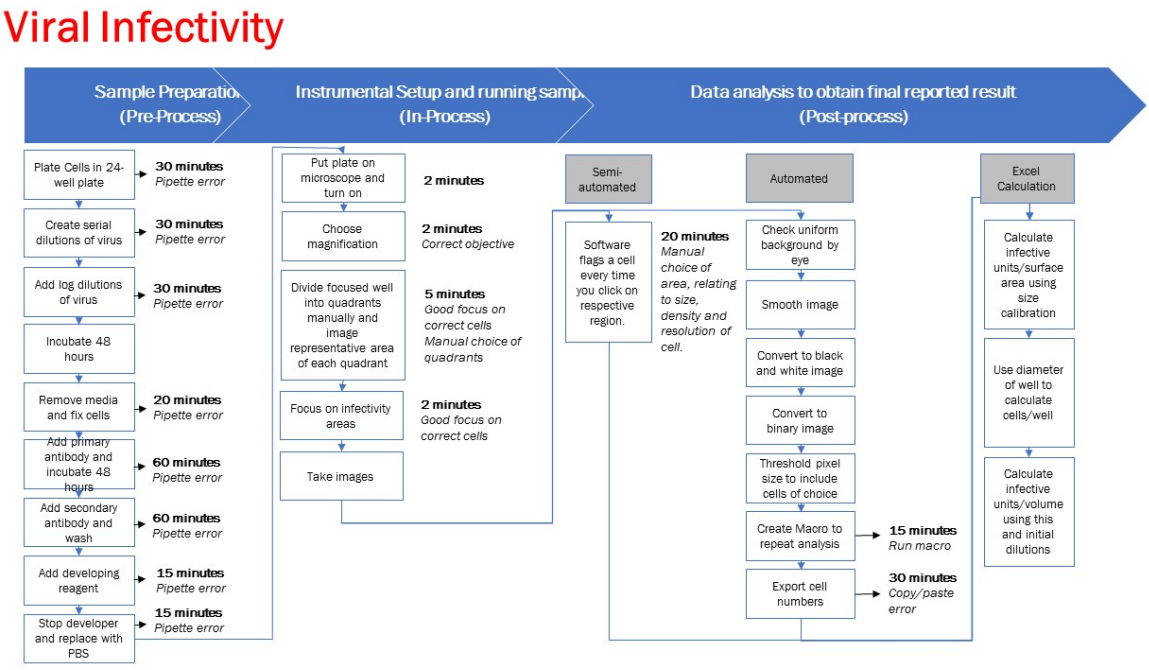


Figure 169 Process Flow Map for manual & imageJ Viral Infectivity measurements

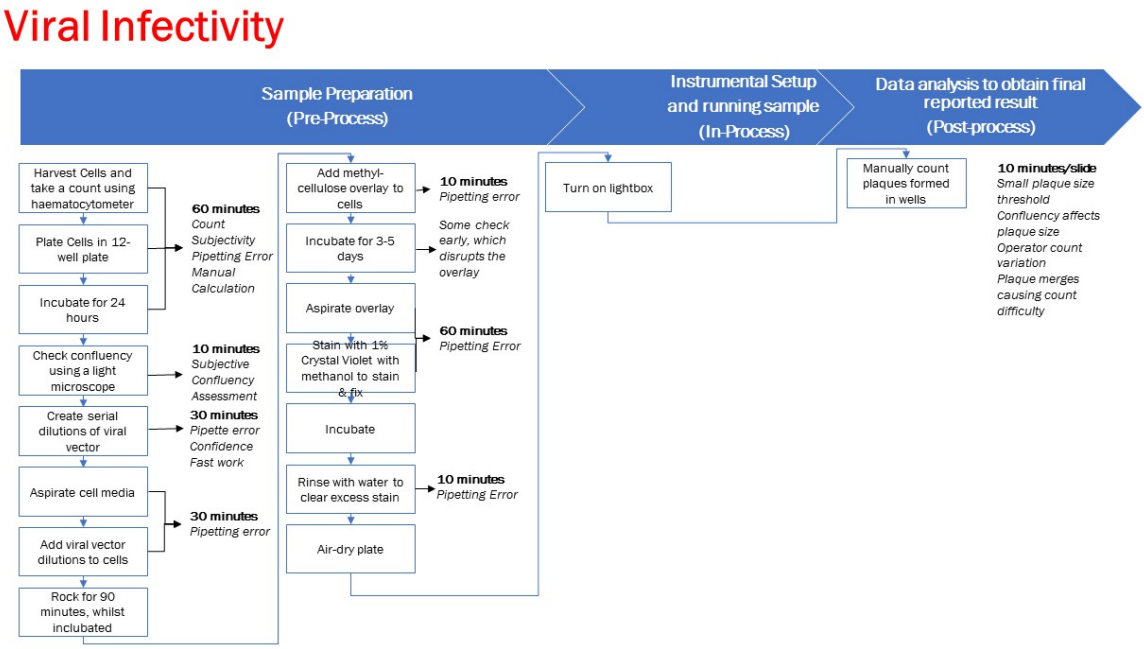


Figure 170 Process Flow Map for manual Viral Infectivity measurements

The four Viral Infectivity process maps have been organised into the pre-process, in-process and post-process columns defined in Section 9.2.3.2. Participants were also asked to identify human factor or subjectivity issues that arose at each step in the process, listed in italics next to the respective step. Some participants also gave an indication of contact time for each step. Where possible, this is listed next to the respective step in bold. These maps have been listed in reverse order of measurement automation, i.e., the more automated procedures are first, with the most manual procedures last. The qPCR method has been included although it is not a cell imaging method, because the relevant company and interviewee had validated this against a cell imaging assay and used qPCR as a more robust method.

It became evident that the presence and use of automation made operators believe any human subjectivity is removed from the process. In Figure 168, where a Cytation-1 cell analyser imaging platform was used, the addition of an automated plate loading machine also made the interviewee strongly believe there were no subjective components, because the operator had been replaced with robotics. However, when discussing the setup of imaging platforms, there was significant

operator involvement to obtain good image for further use. Imaging platforms can be affected by many factors such as digital image parameters (brightness and contrast), as well as spatial issues such as selecting quadrants of wells for imaging or choosing a specific population by eye. Interviewees did not identify these as potential variables, however, human factor or subjectivity variables they initially provided were all concerned with sample preparation, mainly pipetting error, speed of work and operator confidence.

All subjectivity variables listed through the interview process were stratified into core themes, listed in the bar charts in Figure 171. Full definitions of these bins can be found in Appendix E. This bar chart is stacked to show how variables featured across the three process sections. The responses are frequency of results as a function of the total number of subjectivities listed across all three sections for all assay types.

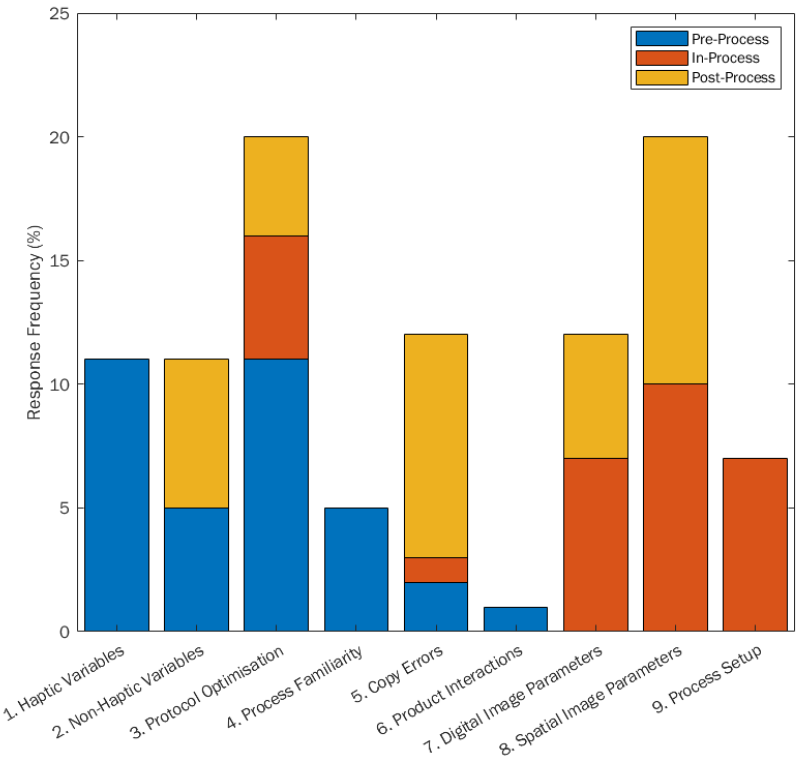


Figure 171 Stratified subjectivity issues within imaging platforms

Protocol Optimisation and Spatial Image parameters were the most frequent variables identified as subjectivity concerns for imaging platform. Both are defined, but protocol optimisation concerned parameters that would have been investigated before the actual measurement process, during previous validation experiments, so are discounted here. Therefore, Spatial Image Parameters were the most prevalent subjectivity factor, followed by Copy Errors and Digital Image Parameters. These variables contrasted with interviewee conception of subjectivity only being present during sample preparation, because this is where there was a lot of human contact with the physical sample. Statements like 'pipetting error' have been stratified under 'Haptic Variables' because they were all physical interactions between analyst and sample. 'Haptic Variables' are affected by the operator completing a physical task or manual manipulation of the product, whereas 'Non-Haptic Variables' are affected by the operator but are not physical in completion. These often involve software-based tasks where the operator is using a computer to select and test a variety of parameters. Full coding definitions are located in Appendix E.

Most of the subjectivity variables physically listed by participants were located within in-process and post-process sections, contrary to initial verbal interviewee opinions. Although Spatial Image Parameters may be difficult to standardise to image plaques or cells, participants did also discuss ways they have achieved this. One participant used Haematocytometer grids to select imaging areas, to remove subjectivity of camera placement. Further validation runs with camera magnifications, fields of view and other digital image parameters can all be standardised for uniform image quality. This can also aim to standardise process parameters upstream, such as the thickness of viral plaque substrates, to ensure imaging, depth and focus can be uniform each time. There are many subjectivity variables listed across these process stages which can be categorised accordingly, but these cannot begin to be addressed or standardised if the entire process is not uniform. It may be acceptable to have different processes to achieve the same metric, but validation is first required to ensure there are no significant differences of results obtained.

9.3.4 Training and SOP results

Interviewees were asked whether training provided for the respective assay was operational, theoretical or a combination of both. Figure 172 shows that no one delivered purely theoretical training because this was not industrially pragmatic. However, 45 % delivered theoretical understanding where relevant alongside operational training. The remaining 55 % provided purely operational training in how to run the assay and equipment. If employees wanted further information, it was up to the individual to research it.

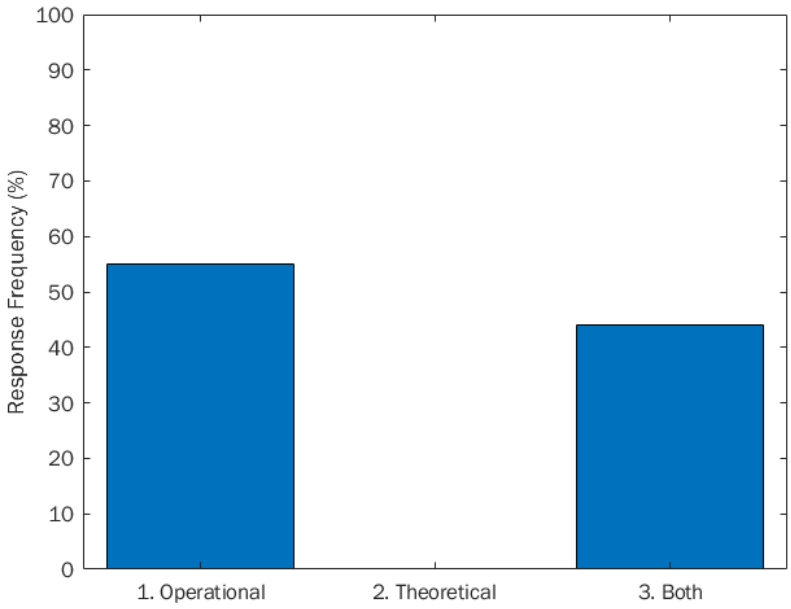


Figure 172 Types of training provided on respective assays

All training discussed was very front-end in an employees' use of the equipment. Employees would have to read relevant risk assessments, SOPs and documentation, then some employees also calculated *CV* of training repeats, so they were only signed off when their *CV* was below a certain level of acceptance. In the case of manual plaque counting, this was ≤ 40 % *CV*. This supports the use of intra-participant *CV* calculated as a training tool, as discussed in Chapter 7. No mention of measurement uncertainty or combined errors was made amongst the interviewees, indicating a significant opportunity for uncertainty quantification across other biometrology platforms to monitor subjectivity. Interviewees discussed variability predominantly in terms of *CV* indicating an

awareness of precision, which is very similar to the FC community who also heavily use *CV* as a key variation metric, as discussed previously in Chapter 1 and throughout Chapters 4 to 6.

All participants stored training documents appropriately to meet auditing requirements, so documents were always accessible. Despite the general training processes being similar, there were differences observed that could affect subjectivity at the learning and training stage. A cultural attitude needs addressing once training has been completed, as many interviewees felt it was inappropriate to tell others what they did was incorrect and why, if they returned abnormal results. Rather than addressing employee differences in cell culture or image acquisition, employees were often asked to repeat the assay and learn from their own mistakes. Whilst this is a good way to learn, differences between operators should be able to be confidently and confidentially addressed in the workplace, to minimise variance on the CGT product.

One participant explained how they tackled this issue in the workplace, by going back and addressing training from SOPs. For each step in an SOP, images and written content was provided to define good and bad process steps, consequences and further actions. Using this in future training enabled new employees to better understand why they completed certain tasks, giving them ownership and confidence in their work. Consequently, they had a very open team culture, so whenever a problem arose, it was comfortable to address and inform the wider team to work together to find the root cause and develop a robust solution, rather than foster a blame culture.

88 % of interviewees used SOPs for post-process data analysis and when questioned further, these were mostly internal documents. SOPs for data analysis mostly provided images of endpoints or good data for relevant steps. Most did not include examples of bad data or consequences of poor variable selections but mentioned it would be good to include in future SOP versions.

Overall, a culture change to the whole measurement process is required, in order for other process changes to have maximum benefit. Standardisation of processes can be achieved, if there is an

open culture to allow employees and managers to discuss continuous improvement efforts, without people feeling subjected to blame or 'big brother' ideology. This has been identified as a function of the broader remit of the interviews here, which could also be further investigated within Flow Cytometry, because the core research within this thesis does not address work culture effects.

9.4 Chapter Conclusions

This translational exercise has shown that human subjectivity issues exist within other measurement platforms used across a variety of CGT products, not just in Flow Cytometry. Imaging and qPCR are highly utilised techniques across many CGT processing stages, so a better understanding of operator subjectivity may have a significant impact on standardisation efforts.

The interviewees have further demonstrated the diversity of processes to obtain 'comparable' results, with many levels of autonomy used to achieve a quantifiable result. Even though only 8 interviewees were included, it still demonstrates diversity of results, which could be investigated more rigorously in future analysis. It became evident that most interviewees believe operator subjectivity issues were found only in pre-processing steps. However, from the actual variables the interviewees listed, most subjectivity was identified as being in the later processing stages. This reflects the variability seen within the Flow Cytometry uncertainty models in Chapters 4 to 6. The comparison in Chapter 7 has shown that variation exists in operator analysis in the post-analytical phase also. Whilst regulatory frameworks exist for the use of validation of software within medical measurement, more needs to be done to ensure user variability is reduced as a function of the platform, or to ensure that its use by different operators does not induce more uncertainty into the measurement.

The lack of participant recognition on the subjective spatial and digital image parameters indicates that further education is required on this topic, especially when the general perception is that automation removes operators and therefore any subjectivity. Automation only transforms the

subjectivity, from manual handling, to manual selection and determination of manufacturing and measurement parameters. When combined with the additional issues of maintenance and accreditation, there are a large number of subjective elements that an operator must understand to make an informed, confident decision when manufacturing a CGT product using automated platforms, especially if measures are close to defined performance limits, which currently causes issues for automation, as discussed in Chapter 1. There is significant subjectivity that resides within the post-analytical processing for imaging platforms, which could greatly benefit from similar measurement uncertainty application that has been successfully demonstrated throughout this thesis. Well defined imaging assays could use the subjectivity breakdowns within this thesis to isolate specific processes for quantification of manual and automated analysis uncertainty where relevant, showing translation of this research across a wide application of biometrology CGT assays.

Whilst considerable effort is put towards training to ensure all new employees receive it and the documentation is stored correctly, a cultural change may be needed to empower employees to foster a more open working culture to speak openly about issues and problem solving. Interviewees stated that additions to SOPs and training to teach more about good and bad data, consequences and root causes could foster a healthier attitude to problem solving when differences arise. In a similar manner to Flow Cytometry, all imaging users will use platforms for diverse measurements, so it could be difficult to deliver standardised training in analytical image processing. However, more internal validation could be used to ensure better standardisation and therefore quality of images used for qualification of CGT products.

9.4.1 Consolidation of Objectives

- 8 participants were interviewed around imaging platforms they used to monitor CGT products they worked with their respective research and process development roles. Throughout the process mapping exercise and other elements of the interview, they

identified various subjectivity elements affecting imaging of CGT products, throughout pre-process, in-process and post-process phases.

- Common causes for subjectivity were spatial image parameters, that occur during in-process and post-process analysis phases. These are variables used to select areas of a well or plate to image, as well as *Z*-height where three dimensions are required.
- Most interviewees believed subjectivity was only based in the pre-process phase during sample preparation, however, most of the subjectivity elements listed were later in the analysis process.
- All interviewees identified that there was a procedure in place for training and documentation, however, participants agreed this could be enhanced by a protocol that contained more information on good and bad data for each step.
- A cultural change is possibly needed in some instances to ensure employees and managers can speak openly about issues that need addressing, or problems they identify, to ensure continuous improvement can be instigated.
- This high-level review of alternative CGT analysis techniques has identified the potential of translation of the novel uncertainty-based analysis methods developed for FC, within this thesis.

Chapter 10: Conclusion & Future Work

10.0 Introduction to the Chapter

The aim of this thesis was to investigate the following research hypothesis as defined in the Prelude (along with the key objectives):

"*As the complexity of Flow Cytometry data increases, the variability contributed to the measurement (either CV or measurement uncertainty) from operator subjective analysis also increases*."

10.1 Thesis Conclusions

With reference to the aims of this thesis key conclusions based upon the results of this research are identified and discussed as follows:

- A succinct review of External Quality Assessment Schemes (EQAS) (Chapter 1) has shown that there are opportunities for translation from the clinical environment into Cell and Gene Therapy manufacturing environments. Current and relevant schemes in clinical Flow Cytometry (FC) have direct application to CGT manufacturers, to ensure measurement quality is passed onto the patient in the form of a correct decision and appropriate treatment. Clinical Chemistry EQAS have also shown how integration of measurement uncertainty into EQAS is achievable, providing further measurement resolution, control and potential alignment to ISO 17025 and ICH Q7 standards for CGT manufacture.

- Measurement uncertainty principles (described in Chapter 2 and applied in Chapters 4 to 6) have been shown to be very applicable to FC inter-participant data, giving an indication of overall participant variation, but also identify specific gates that cause issues. This shows

its' worth as a continuous improvement tool within CGT measurement and manufacturing. One drawback of using measurement uncertainty is the number of repeats recommended for *SD* calculations (between 4 to 10 according to the GUM). No more than 3 repeats could be taken for this data (due to participant fatigue), which could affect the results, but the benefits of using measurement uncertainty outweigh this. Further repeats could be conducted in future work, because it was not practical as part of this experimentation, to indicate an optimum number of repeats for human participant variation quantification.

- To model an increase in complexity throughout the uncertainty models, three cellular profiles were chosen for fcs file generation. The Embryonal Carcinoma 2102 Ep immortalised cell line was used for the basic model (Chapter 3) because it had been shown to remain pluripotent over 10 passages, contributing less biological variation to the analysis. This ensured the basic model could be used as a baseline. The intermediate model files were created from Peripheral Blood Mononuclear Cells (PBMCs) because this population contained further sub-populations which provided greater depth and complexity. The 2102 Ep line did not have this capability, and the PBMC material was more comparable to current autologous therapies available to patients, that collaborators were also focused on. This PBMC model enabled further complexity enhancement for the final model, which used engineered T-cells derived from PBMCs to measure transduction efficiency. This is representative of assays used in current CGT expansion, and was provided by one of the collaborators, giving affinity to relevant therapies and their manufacture.

- Throughout the complexity models in Chapters 4 to 6, a standard reporting structure of metrics has been used to report; the absolute results obtained by participants, respective Coefficients of Variation (*CV*) and measurement uncertainty. Only *CV* and measurement uncertainty are considered within these overall conclusions, because the absolute results were not stated within the initial hypothesis to monitor variability, and they are not comparable to a metrologically 'true value' due to lack of traceability through Flow Cytometry standards.

- It is possible that Hawthorne's effect could have impacted the results of this human participant research. This effect can be attributed when participants are aware they are being watched, monitored or evaluated within their work. Participants were aware of their work being analysed due to the information and consent provided to take part in the research. It is also very difficult to conduct Gauge R&R and measurement uncertainty analysis session without this impact, because of the experimental specificity required for calculation, which informs measurement participants.

- During the pre-study investigation in Chapter 3 (in which participants applied gates to histograms), there was a range of 8 % *rCV* (for optimal 500 *V* file) between participants when analysing a data file that had the correct instrument setup. The range increased up to 400 % *rCV* (250 *V*– 450 *V*) when the instrument was poorly setup, clarifying the need for appropriate instrument optimisation to reduce downstream measurement variance between analysts.

- Throughout this research, many non-parametric distributions are identified when participants gate with or without the use of protocols. Many of the basic statistical tools such as skewness and kurtosis define non-parametric distributions by their discordance with normal distribution metrics. The in-built assumption of desirable normal distribution shape is challenged here. When trying to reduce variation, conformance to specific performance criteria is required, which can mean distributions should become more kurtosed or concentrated within a smaller distribution range. This questions how data is processed, to identify core metrics of success to ensure that all statistics used are relevant.

- Chapter 3 has also highlighted the differences between performance criteria (as explained in Chapters 1 and 2) used across various industries and respective outliers of the data set. There is no harmonisation between these methods, although the correct choice should be informed from thorough normality assessment of the distribution, and data should not

necessarily be log transformed, due to discrepancies between outlier definitions. The analysis reported here has therefore been kept closer to the raw data, making analysis and decisions easier to interpret for manufacturers and regulators.

- Experimental studies were well designed (as discussed in Chapter 2) because the relevant amount of data was captured within the allotted participant time frames for each complexity model (Chapters 4 to 6). Data was stored in accordance to ethical guidelines, once participant consent was obtained. These have provided a suitable methodology for future uncertainty calculations where subjectivity estimations are required, to obtain at least three repeats within 1 hour of participant processing time.

- Measurement uncertainty was quantified successfully across all three complexity models, by calculating the *SD* of three repeats of each gate applied in the gating sequences. These were successfully combined in quadrature using the GUM principles to create combined uncertainties and expanded uncertainties to show a 95 % Confidence Interval of data (k=2). Median uncertainties of the basic to complex model were 3.6 %, 2.1 % and 6.2 % respectively, indicating location parameters do not show any relation to increase variation.

- Inter- and intra-*CV* have been defined as; *CV* of a participant population of results, and *CV,* of a participant's individual repeats, respectively. An inter-*CV* mean of 16 % has been found across all three complexity models for participant judgement when gating data. Inter-*CV*s across the three complexity models were 17.8 %, 18.2 % and 12.1 % respectively as complexity increased. 16 % could be used as a rule of thumb for participant variability, although it falls outside of ICSH satisfactory criteria of < 10 % *CV*.

- An increase in the intra-participant range of *CV* and measurement uncertainty was seen as the complexity increased, shown in Table 84, and illustrated in Figure 173 and Figure 174 (rounded to the nearest integer). This is the primary novel result that confirms the hypothesis of the thesis, showing more variability is contributed to the measurement from

the operator conducting the data analysis. This has significant implications within CGT, because FC panels can increase in dimensionality to 18 colours in some instances, contributing more variation to the final measurement and interpretation. This becomes more complicated when monitoring rare cell events and minimal residual diseases, although these elements were not a feature of the study. As cell emission spectra get closer to the limit of sensitivity of the instrument, it can increase the probability of false positives / false negatives as a function of operator subjectivity and thresholds, causing a therapy to be inappropriately given, or a product discarded when it could have provided a patient a treatment.

Table 84 Comparison of intra-participant CV and uncertainty ranges across the cell complexity models

	Basic	Intermediate	Complex
CV (%)	6	22	34
Uncertainty (%)	12	16	34

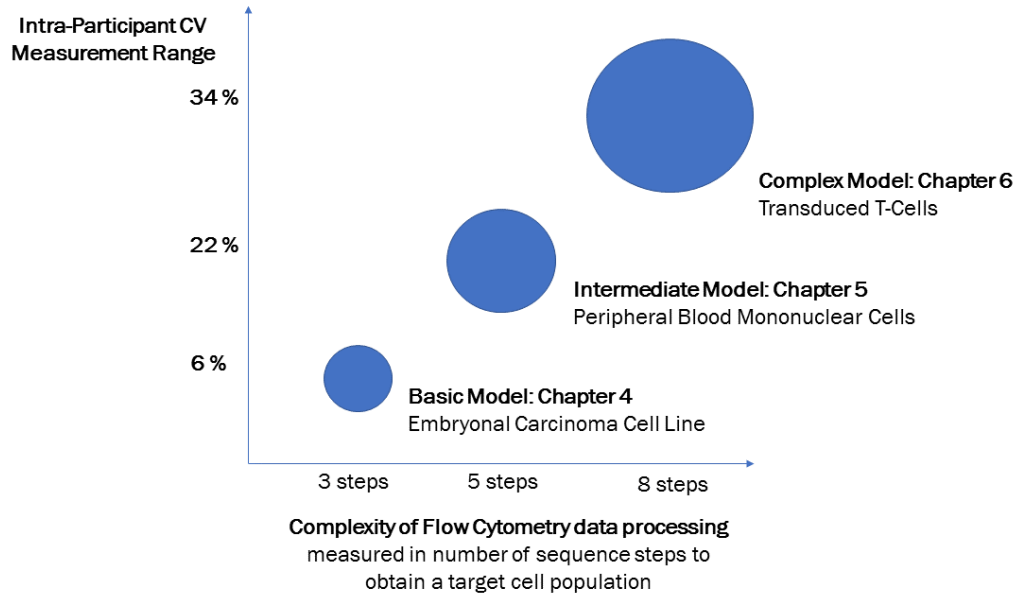


Figure 173 Core hypothesis of thesis, showing an increased range of intra-participant CV with FC data complexity.

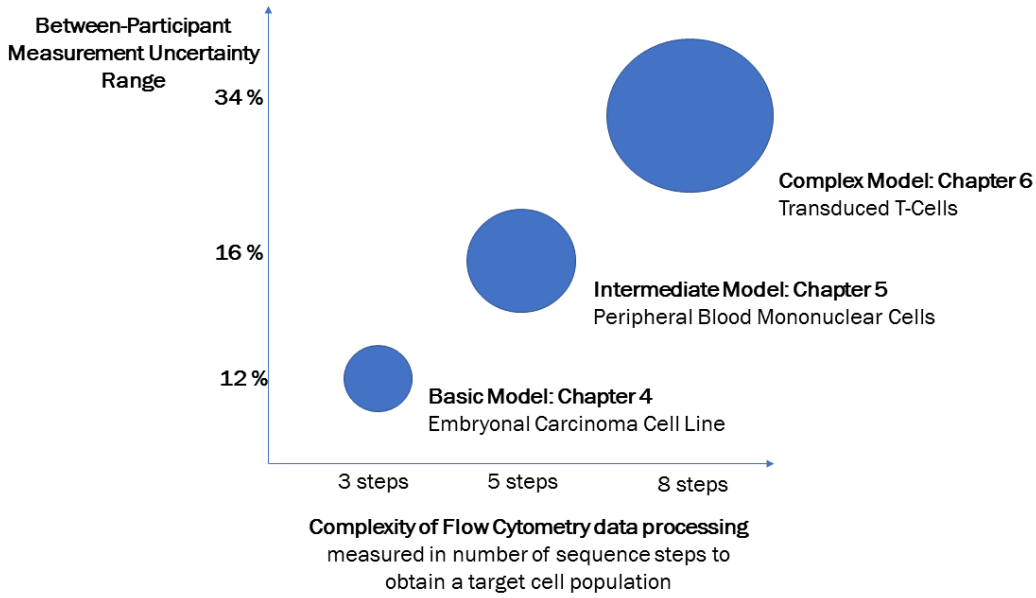


Figure 174 Core hypothesis of thesis, showing increased range of inter-participant uncertainty with FC data complexity.

- *CV* is a commonly used variability metric within FC and other biological assays. Throughout these analyses, it has shown it is more difficult to identify sources of variance from *CV* extremes, because this metric only accounts for variation from repeats of the final gate measurements. Measurement uncertainty calculated variability using a bottom-up approach, combining uncertainties from individual gates applied within the FC analysis sequence (in this instance). It facilitates easier root cause analysis of variation, to better identify process elements or operators that are more variable. This continuous improvement loop provides opportunity for appropriate optimisation of the process, and appropriate training of staff. On this note, *CV* is a useful measure of initial competency, because it is quick to calculate, whereas uncertainty requires well controlled experiments for quantification that are more time-consuming. Throughout each model (Chapters 4 to 6) it has been shown there is no strong correlation between CV and uncertainty, so these metrics should not be used to estimate each other.

- Using the ICSH guideline criteria for acceptable *CV*, a performance diagram was created by adapting these guidelines, exemplified here in Figure 175. The satisfactory limit (10 %) was

halved to 5 % to show 'good' performance and doubled to 20% to show where revision is required, with actual histogram data overlaid. This was used effectively to monitor intra-participant *CV* throughout Chapters 4 to 6, as well as substituting these values to monitor uncertainty. No defined limits for uncertainty have been provided in the Flow Cytometry literature, so these were used alongside permissible uncertainty defined for clinical chemistry values. This diagram was effective at monitoring extreme data and showed how it could be used as an indication within a manufacturing facility to easily monitor quality towards a common goal of lower variance.

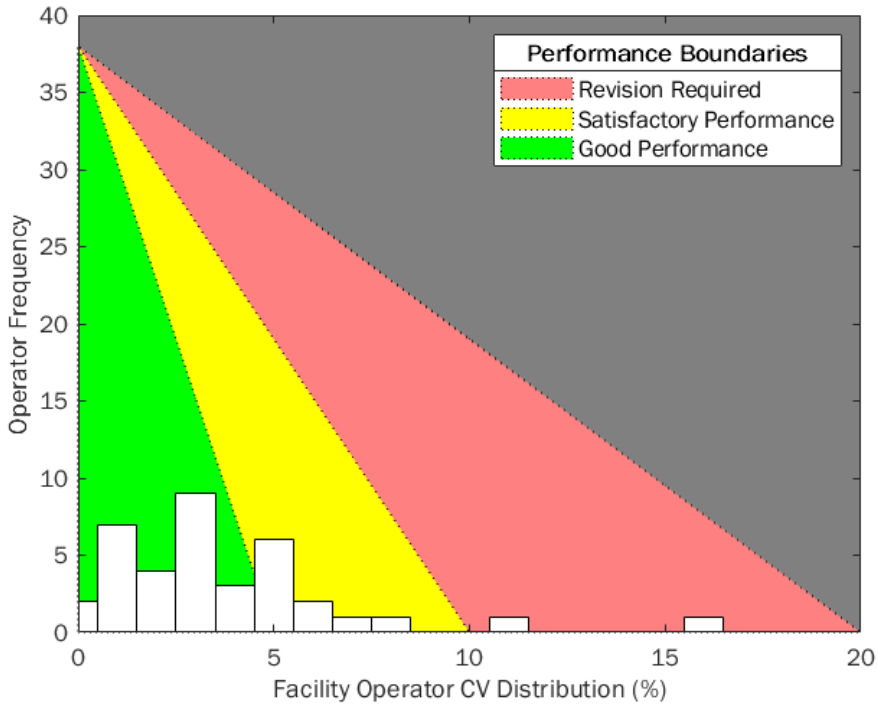


Figure 175 Example of the novel performance monitoring diagram, highlighting good, satisfactory and revision regions based upon ICSH guidelines

- The use of a protocol showed initial promise within the pre-study, reducing operator variability (*CV*) when copying histogram gates by up to 90 % in optimal conditions (Chapter 3). There was also a 57 % reduction in range seen during the intermediate model when participants used the protocol to identify naïve T-cells (Chapter 5). However, there was an increase in range of count results (42 % increase) and uncertainties (23 % increase) when participants used the protocol in the complex model (Chapter 6), indicating that copying a

protocol in lengthy gating sequences may not be optimal to reduce variation between participants, or it requires redesigning and investigating protocol usage sensitivities.

- Bimodal distributions were observed within Chapters 5 and 6, although only through visual observation. Basic statistics used did not identify this shape, often placing mean values where no data maxima were present. This highlights the importance of visual representation of the data, to further identify extreme or unexpected deviances. These high and low distributions were due to boundary effects identified on the edges of plots, and how participant gates were affected by these.

- Throughout the complexity models, it was noticed that the same gates in each sequence were causing high variability between participants. Upon further investigation, participants with higher uncertainties due to these gates were mostly affected by the boundary effect on the respective edges of the high variance gates. Improper use of controls was the other main cause of variance between participants.

- A difference in high and low uncertainty clusters has been identified during the intermediate and complex model, due to the boundary effects observed on the axes limits during these respective studies. Repeatable inclusion or exclusion of this data gives a low uncertainty, but low or high respective cell counts. Inconsistent gating on edges where repeats do not all contain this effect have a high uncertainty. This has shown that these boundary artefacts require future investigation to identify the impact they have on measurement variability when removed. Boundary effects were also observed when the data was visualised in some software platforms, but not others, highlighting a significant issue for reproducibility of data and analysis across Flow Cytometry measurements.

- Power analysis was conducted for all analysis studies where two test conditions were evaluated (Chapters 3, 5, 6), although these studies were conducted without initial sample size limitations to ensure an initial variance could be quantified. Power was used to define

the appropriate number of participants across the models, to identify whether a suitable number of participants were acquired to determine differences between test conditions. A-Priori and Post Hoc power indicated ideal sample sizes from variances gathered from data and the actual power of the models respectively. Power analysis could not be conducted for singular test conditions (Chapter 4), because there was no comparison to another test condition or a hypothesised ideal value. Whilst some power analysis indicated further participants were required, this could not have been determined before the studies took place, because initial variation was unknown. These values can now be used to inform further validation studies.

- The results of this questionnaire (Chapter 8) indicated that there was no correlation between experience of a participant and their respective uncertainty, in any of the gating models. This was also true of use frequency of a Flow Cytometer, with no correlation to measurement uncertainty as the equipment is used more regularly. This potentially requires more investigation; however, it indicates that more knowledge of Flow Cytometry does not identify if participants are going to be more precise when applying their gates. Further work on how training is implemented could further expand upon this.

- The questionnaires also showed the majority of participants prefer manual gating to automated algorithms for cluster identification, due to scepticism of the automated methods (relating to literature in Chapter 1). This indicates a cultural barrier needs to be addressed in order for these methods to be fully adopted, in addition to better quantification and validation of automation precision. Participants also indicated that cell cluster separation was the biggest issue facing FC data analysis, despite many automation efforts in this area. Many participants identified the need for better reporting standards of FC experiments, although typically did not note the MIFlowCyt standard, suggesting that these reporting standards (discussed in Chapter 1) need to cross boundaries into different cell-based communities.

- It was evident from the FC questionnaires for gating (Chapter 8) and the translational work for imaging and qPCR (Chapter 9), that the training provided for assay specificity is all internally driven. Further focus on gating training or uniform guidelines that could aid general analysis are required to ensure internal training is reproducible.

- Chapter 9 has shown the range of subjectivity issues prevalent across imaging and qPCR platforms used within CGT process development and manufacturing, with many people not realising that subjectivity also appears within data analysis, not just sample handling and upstream pipetting error. This translation exercise has identified a possible need for training across multiple platforms, to highlight further sources of variation that can impact a measurement during post-analytical analysis of images and assays. Even when automation is used to aid manufacture, this does not necessarily remove the operator subjectivity, it translates it to other areas, such as validation and setup options, as well as the additional issues of maintenance and calibration.

- This identification of subjectivity in downstream measurement across other platforms demonstrates the applicability to quantify subjectivity in other measurement techniques, where quantitative measurements are derived from interpreted images, plots or qualitative data.

10.2 Thesis Novelty

The perceived novelty of this research was originally identified in the Prelude of this thesis. Table 85 identifies how well these objectives have been met, and the perceived novelty of each element of the work, as well as other novel contributions that have developed during the research.

Table 85 Comparison of Novelty

Perceived Novelty	Relevant Chapter	Novelty
This thesis provides a critical review of current External Quality Assessment Schemes (EQAS), to identify opportunities for integration into CGT manufacturing.	1	**Low/ Medium** This review has shown current EQAS that are used for clinical FC, and also current EQAS for clinical chemistry that have quantified measurement uncertainty. This highlights a potential for CGT manufacturing, to learn and adapt these principles to a new focus but is not novel in terms of uncertainty consideration into EQAS.
Relevance of application of manufacturing outlier definitions to define process control limits.	3	**Low/Medium** A review of different performance criteria definitions across various industries has shown that there are many different control limits that can be applied, but a lot of these require normal data, so are unsuitable for skewed data. More robust methods should be used for non-parametric data, with various suggestions made in this research.
Application of Gauge Repeatability & Reproducibility techniques to Flow Cytometry post-analytical variation.	4	**High** This is a novel approach to determining post-analytical FC variation, purely isolating the gating from the rest of the FC measurement process. Gauge R&R is an effective way to quantify variation using

		electronic data files, making the application of this easy to implement across different scenarios.
Use of measurement uncertainty for Flow Cytometry post-analytical variation.	4, 5, 6	**High** Measurement uncertainty has for the first time been demonstrated and used to quantify purely the post-analytical variation stage of FC, and importantly for the first time within a CGT manufacturing context.
Use of measurement uncertainty for better measurement resolution and control with CGT manufacturing.	4, 5, 6	**High** The bottom-up uncertainty calculated for each respective gate provides better root cause analysis for extreme values and shows how this can be utilised for precision, accreditation and continuous improvement.
Quantification of participant subjectivity as a function of cell model complexity.	4, 5, 6	**High** This is the first-time intra-participant and inter-participant subjectivity from the gating process has been quantified within a FC context and within CGT analysis scenarios. This can be used in conjunction with other uncertainty estimates to create a more confident uncertainty budget for FC measurements and their interpretation.
Increased variability as a function of cell complexity.	4, 5, 6	**High** This is the first time that increase variation has been shown with respect to cell complexity, using intra-participant CV or uncertainty.
Development of a new performance monitoring	4, 5, 6	**High** This diagram and respective code was developed to allow centres to define their own performance limits in

diagram to aid continuous improvement of variation.		relation to the number of people in the facility, to easily monitor conformance and performance to highlight extreme data and subsequent learning opportunities. This challenges the central tendency seen as 'ideal' for normal distributions, because variation always aims to be as low as possible, indicating a strong positive skew, which used equation of a straight line to inform this shape from the input performance parameters. This has been a very significant tool to identify extremes in this research and shows applicability in many other quality and monitoring situations.
Comparison of measurement variability metrics suitable for precision of FC measurements within CGT manufacturing contexts.	7	**Medium/ High** *CV* and measurement uncertainty represent different levels of controlled and quantified variation, with uncertainty giving more resolution, but *CV* being more pragmatic. These have been combined for KPIs for the NHS (Chapter 1), showing how they could be utilised within CGT manufacture for better manufacturing control.

10.3 Further work

This research has already successfully shown there is an increase in variability input to Flow Cytometry measurements, when data being analysed becomes more complex. However, further time and effort are still needed to expand and investigate areas of this work. Further work has been stratified into 'general' and 'detailed' further work, to identify larger experimental studies and focused validation efforts in relation to this work respectively.

10.3.1 General Further work

- A top priority for this work, is translation of the precision of human operators in comparison to machine learning algorithms for cluster definition. This work can be used as a benchmark for automated analysis of the same files, to compare variability of clustering methods. Even though this complex model precision is high, it can be used to identify performance criteria for automation: for example, automation precision to be 10 % of the human counterpart.

- Further measurement uncertainty calculations of other FC variation sources could be completed to estimate expanded uncertainty for a particular FC assay. This can indicate the amount of variation present for the whole assay, as well as identifying sources of variation to reduce.

- Further experimental repeats could be taken, to strengthen the validity of this work. This would require significant experimental design to choose suitable repeats and obtain them without causing participant fatigue or enabling them to learn the sequence. However, this would need to be well defined due to Hawthorne's effect and recognition of the data.

- Further human participant work would focus on measurement/identification of rare cell events and residual disease, because these measurements will be needed to monitor

participant treatment and progression, so measurement uncertainty becomes of utmost importance to ensure results are not reported or interpreted incorrectly.

- Development of EQAS for CGT FC measurements is required, to ensure standardised instrumental setup and analysis between sites and manufacturers, to improve reproducibility of data.

- Application of the novel performance monitoring diagram created in this research to other biomanufacturing environments.

- Extension of the complexity versus measurement uncertainty research to capture more complex panels, such as 12- and 18-colour panels, because these are becoming more common within Flow Cytometry analysis.

10.3.2 Detailed Further work

- Further benchmarking of the Intermediate Model (Chapter 5) could be completed, to monitor participant results in relation to the cell count values extracted from the diagrammatical protocol provided to participants in Phase 2, similar to benchmarking for the Complex model (Chapter 6).

- Comparison of machine learning parameter repeatability (similar to t-SNE parameters discussed in Chapter 1) needs to be completed for each analysis file used for the Basic, Intermediate and Complex model. This would provide a comparable precision measure against the participant results calculated in this research, to define where improvements in precision are required.

- Conduct similar uncertainty studies for alternate cell types that are common within CGTs, such as Mesenchymal Stromal Cells (MSCs) to build a database of uncertainty ranges of post-analytical variation.

- Recruit more participants to take part in further studies, to provide more confidence in the results, and possibly repeat some of the current models that had bimodal distributions due to boundary effects within the files. Educating participants about this issue before the study (or identifying how to remove this effect from the data), could potentially remove the difference in variance seen, to get a better understanding of gating the data with or without a protocol.

- Detailed analysis of the prevalence and influence of boundary effects in different software platforms.

- Further studies to test the use of a protocol need to be conducted, to identify how these can be useful and provide training, without negatively impacting variance of results. A focus on culture change to ensure operators speak openly about their analysis and difference would help harmonisation efforts.

- Protocol design sensitivity exercise to develop protocols that helps at the higher dimensional analysis stages, to identify where higher or lower variance between participants is introduced as a function of interpretation.

References

[1] Pedersen NW, Chandran PA, Qian Y, Rebhahn J, Petersen NV, Hoff MD, et al. Automated analysis of flow cytometry data to reduce inter-lab variation in the detection of major histocompatibility complex multimer-binding T cells. Front Immunol 2017;8:1–12. doi:10.3389/fimmu.2017.00858.

[2] Gouttefangeas C, Chan C, Attig S, Køllgaard TT, Rammensee H-G, Stevanović S, et al. Data analysis as a source of variability of the HLA-peptide multimer assay: from manual gating to automated recognition of cell clusters. Cancer Immunol Immunother 2015;64:585–98. doi:110.1016/j.bbi.2017.04.008.

[3] De la Salle B. Pre- and postanalytical errors in haematology. Int J Lab Hematol 2019;41:170–6. doi:10.1111/ijlh.13007.

[4] Hourd P, Ginty P, Chandra A, Williams DJ. Manufacturing models permitting roll out/scale out of clinically led autologous cell therapies: Regulatory and scientific challenges for comparability. Cytotherapy 2014. doi:10.1016/j.jcyt.2014.03.005.

[5] Mason C, Brindley DA, Culme-Seymour EJ, Davie NL. Cell therapy industry: Billion dollar global business with unlimited potential. Regen Med 2011;6:265–72. doi:10.2217/rme.11.28.

[6] Minutolo NG, Hollander EE, Powell DJ. The emergence of universal immune receptor T-cell therapy for cancer. Front Oncol 2019;9. doi:10.3389/fonc.2019.00176.

[7] Riegler LL, Jones GP, Lee DW. Current approaches in the grading and management of cytokine release syndrome after chimeric antigen receptor T-cell therapy. Ther Clin Risk Manag 2019;15:323–35. doi:10.2147/TCRM.S150524.

[8] Eyles JE, Vessillier S, Jones A, Stacey G, Schneider CK, Price J. Cell therapy products: focus on issues with manufacturing and quality control of chimeric antigen receptor T-cell therapies. J Chem Technol Biotechnol 2018;94:1008–16. doi:10.1002/jctb.5829.

[9] Harrison RP, Ruck S, Medcalf N, Rafiq QA. Decentralized manufacturing of cell and gene therapies: Overcoming challenges and identifying opportunities. Cytotherapy 2017;19:1140–51. doi:10.1016/j.jcyt.2017.07.005.

[10] Simon CG, Lin-Gibson S, Elliott JT, Sarkar S, Plant AL. Strategies for achieving Measurement Assurance for Cell Therapy Products. Stem Cells Transl Med 2016;5:705–8.

[11] Verschoor CP, Lelic A, Bramson JL, Bowdish DME. An introduction to automated flow cytometry gating tools and their implementation. Front Immunol 2015;6. doi:10.3389/fimmu.2015.00380.

[12] Thurman-Newell JA, Petzing JN, Williams DJ. Quantification of biological variation in blood-based therapy - a summary of a meta-analysis to inform manufacturing in the clinic. Vox Sang 2015;109:394–402. doi:10.1111/vox.12288.

[13] BS EN ISO / IEC 17025 : 2017 BSI Standards Publication General requirements for the competence of testing and calibration laboratories 2017.

[14] ICH Expert Working Group. Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients Q7. ICH Harmon Tripart Guidel 2000:49.

[15] van Dongen JJM, Orfao A. EuroFlow: Resetting leukemia and lymphoma immunophenotyping. Basis for companion diagnostics and personalized medicine. Leukemia 2012;26:1899–907. doi:10.1038/leu.2012.121.

[16] Wang L, Hoffman RA. Standardization, calibration, and control in flow cytometry. Curr Protoc Cytom 2017;2017:1.3.1-1.3.27. doi:10.1002/cpcy.14.

[17] Kalina T, Flores-Montero J, van der Velden VHJ, Martin-Ayuso M, Böttcher S, Ritgen M, et al. EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. Leukemia 2012;26:1986–2010. doi:10.1038/leu.2012.122.

[18] van Dongen JJM, Lhermitte L, Böttcher S, Almeida J, van der Velden VHJ, Flores-Montero J, et al. EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. Leukemia 2012;26:1908–75. doi:10.1038/leu.2012.120.

[19] Mizrahi O, Ish Shalom E, Baniyash M, Klieger Y. Quantitative Flow Cytometry: concerns and recommendations in clinic and research. Cytom Part B - Clin Cytom 2018;94:211–8. doi:10.1002/cyto.b.21515.

[20] Finak G, Langweiler M, Jaimes M, Malek M, Taghiyar J, Korin Y, et al. Standardizing Flow Cytometry immunophenotyping analysis from the human ImmunoPhenotyping Consortium. Sci Rep 2016;6:1–11. doi:10.1038/srep20686.

[21] Aghaeepour N, Finak G, Hoos H, Mosmann TR, Brinkman R, Gottardo R, et al. Critical assessment of automated flow cytometry data analysis techniques. Nat Methods 2013. doi:10.1016/j.jfma.2017.09.008.

[22] Krishnan A, Krishnamurthy H, Totey S. Applications of Flow Cytometry in stem cell research and tissue regeneration. New Jersey: Wiley; 2010.

[23] Krishnamurthy H, Cram L. Applications of Flow Cytometry in stem cell research and tissues regeneration. In: Krishnan A, Krishnamurthy H, Totey S, editors. Appl. Flow Cytom. Stem Cell Res. Tissues Regen., New Jersey: Wiley; 2010, p. 1–12.

[24] Gelderman M, Simak K. Flow Cytometric analysis of cell membrane particles. Funct. Proteomics - Methods Protoc., vol. 53, 2013, p. 79–97. doi:10.1017/CBO9781107415324.004.

[25] Braga F, Pasqualetti S, Panteghini M. The role of external quality assessment in the verification of in vitro medical diagnostics in the traceability era. Clin Biochem 2018;57:23–8. doi:10.1016/j.clinbiochem.2018.02.004.

[26] Braga F, Panteghini M. Verification of in vitro medical diagnostics (IVD) metrological traceability: Responsibilities and strategies. Clin Chim Acta 2014;432:55–61. doi:10.1016/j.cca.2013.11.022.

[27] Ibrahim SF, van den Engh G. Flow cytometry and cell sorting. Cell Sep., Berlin: Springer; 2007, p. 19–39.

[28] Shapiro HM. The evolution of cytometers. Cytometry 2004;58A:13–20. doi:10.1002/cyto.a.10111.

[29] Shapiro HM. Practical Flow Cytometry. 4th ed. Hoboken, NJ: John Wiley & Sons; 2003.

[30] Spidlen J, Moore W, Parks D, Goldberg M, Bray C, Bierre P, Gorombey P, Hyun B, Hubbard M, Lange S, Lefebvre R, Leif R, Novo D, Ostruszka L, Treister A, Wood J, Murphy RF, Roederer M, Sudar D, Zigon R BR. Data File Standard for Flow Cytometry, Version FCS 3.1. Cytom A 2010;23:97–100. doi:10.1038/jid.2014.371.

[31] Maciorowski Z, Chattopadhyay PK, Jain P. Basic multicolour Flow Cytometry. Curr Protoc Immunol 2017;117:5.4.1-5.4.38.

[32] Wood B. 9-Color and 10-Color Flow Cytometry in the clinical laboratory. Arch Pathol Lab Med 2006;130:680–90.

[33] Roederer M. Spectral compensation for flow cytometry: Visualization artifacts, limitations, and caveats. Cytom Part A 2001;205:194–205. doi:10.1002/1097-0320(20011101)45:3<194::AID-CYTO1163>3.0.CO;2-C.

[34] Maecker HT, Trotter J. Flow Cytometry controls, instrument setup, and the determination of positivity. Cytom Part A 2006;69:1037–42. doi:10.1002/cyto.a.

[35] Flowjo. Concatenation and tSNE in FlowJo Webinar. 2017.

[36] Van Der Maaten L, Postma E, Van Den Herik J. Dimensionality Reduction : a comparative review. October 2009:1–35. doi:10.1080/13506280444000102.

[37] Van Der Maaten LJP, Hinton GE. Visualizing high-dimensional data using t-SNE. J Mach Learn Res 2008;9:2579–605. doi:10.1007/s10479-011-0841-3.

[38] Flowjo. tSNE | FlowJo v10 Documentation 2018. http://docs.flowjo.com/d2/plugins/tsne/ (accessed May 18, 2018).

[39] Wattenberg M, Viégas F, Johnson I. How to use t-SNE effectively. Distill 2016;1. doi:10.23915/distill.00002.

[40] Kaiser AD, Assenmacher M, Schröder B, Meyer M, Orentas R, Bethke U, et al. Towards a commercial process for the manufacture of genetically modified T cells for therapy. Cancer Gene Ther 2015;22:72–8. doi:10.1038/cgt.2014.78.

[41] Lutwama F, Serwadda R, Mayanja-Kizza H, Shihab HM, Ronald A, Kamya MR, et al. Evaluation of dynabeads and cytospheres compared with flow cytometry to enumerate CD4+ T cells in HIV-infected ugandans on antiretroviral therapy. J Acquir Immune Defic Syndr 2008;48:297–303. doi:10.1097/QAI.0b013e31817bbc3a.

[42] CHMP. European Medicines Agency. Committee for Medicinal Products for Human Use (CHMP) Assessment report. KYMRIAH 2018;44.

[43] Schwartz A, Marti GE, Poon R, Gratama JW, Fernández-Repollet E. Standardizing flow cytometry: A classification system of fluorescence standards used for flow cytometry. Cytometry 1998;33:106–14. doi:10.1002/(SICI)1097-0320(19981001)33:2<106::AID-CYTO4>3.0.CO;2-H.

[44] Davis BH, Dasgupta A, Kussick S, Han JY, Estrellado A. Validation of cell-based fluorescence assays: Practice guidelines from the ICSH and ICCS - Part II - Preanalytical issues. Cytom Part B - Clin Cytom 2013;84:286–90. doi:10.1002/cyto.b.21105.

[45] Barnett D, Louzao R, Gambell P, De J, Oldaker T, Hanson CA. Validation of cell-based fluorescence assays: Practice guidelines from the ICSH and ICCS - Part IV - Postanalytical Considerations. Cytom Part B - Clin Cytom 2013;84:309–14. doi:10.1002/cyto.b.21105.

[46] BD Biosciences. SPHERO ™ Rainbow Calibration Particles 2015:7–8.

[47] NIBSC. Flow Cytometry Reference Materials 2019. https://www.nibsc.org/science_and_research/biotherapeutics/cellular_immunology/flow_cytometry_reference_materials.aspx#Antibody and cell ref materials (accessed July 31, 2019).

[48] FDA. Guidance for Industry Part 11, Electronic Records; Electronic Signatures — Scope and Application. 2003.

[49] Lee JA, Spidlen J, Boyce K, Cai J, Crosbie N, Dalphin M, et al. MIFlowCyt: The minimum information about a flow cytometry experiment. Cytom Part A 2008;73:926–30. doi:10.1002/cyto.a.20623.

[50] Gasparetto M, Spidlen J, Brinkman RR. Minimum Information about a Flow Cytometry Experiment Experiment annotation example. MIFlowcyt 2008;40076.

[51] Program TI. MIATA Project 2019. http://miataproject.org/ (accessed July 27, 2019).

[52] Roederer M, Tárnok A. OMIPs - Orchestrating multiplexity in polychromatic science. Cytom Part A 2010;77:811–2. doi:10.1002/cyto.a.20959.

[53] Connelly MC, Knight M, Giorgi J V., Kagan J, Landay AL, Parker JW, et al. Standardization of absolute CD4+ lymphocyte counts across laboratories: An evaluation of the ortho CytoronAbsolute flow cytometry system on normal donors. Cytometry 1995;22:200–10. doi:10.1002/cyto.990220307.

[54] Reimann KA, O'Gorman MRG, Spritzler J, Wilkening CL, Sabath DE, Helm K, et al. Multisite Comparison of CD4 and CD8 T-lymphocyte counting by single- versus multiple-platform methodologies: evaluation of Beckman Coulter Flow-Count Fluorospheres and the tetraONE System. Clin Vaccine Immunol 2000;7:344–51. doi:10.1128/CDLI.7.3.344-351.2000.

[55] Baker M. Is there a reproducibility crisis? Nature 2016;533:452–4.

doi:10.1038/533452a.

[56] Baker M. Quality time. Nature 2016;529:456–8.

[57] Rivière I, Roy K. Perspectives on manufacturing of high-quality Cell Therapies. Mol Ther 2017;25:1067–8. doi:10.1016/j.ymthe.2017.04.010.

[58] Carpenter M, Couture L. Regulatory considerations for the development of autologous induced pluripotent stem cell therapies. Regen Med 2010;5:569–79.

[59] Archer R, Williams DJ. Why tissue engineering needs process engineering. Nat Biotechnol 2005;23:1353–5. doi:10.1038/nbt1105-1353.

[60] Williams DJ, Thomas RJ, Hourd PC, Chandra A, Ratcliffe E, Liu Y, et al. Precision manufacturing for clinical-quality regenerative medicines. Philos Trans R Soc A Math Phys Eng Sci 2012;370:3924–49. doi:10.1098/rsta.2011.0049.

[61] Daniels JT, Secker G a, Shortt AJ, Tuft SJ, Seetharaman S. Stem cell therapy delivery: treading the regulatory tightrope. Regen Med 2006;1:715–9. doi:10.2217/17460751.1.5.715.

[62] der Strate B van, Longdin R, Geerlings M, Bachmayer N, Cavallin M, Litwin V, et al. Best practices in performing flow cytometry in a regulated environment: feedback from experience within the European Bioanalysis Forum. Bioanalysis 2017;9:1253–64. doi:10.4155/bio-2017-0093.

[63] Hourd P, Chandra A, Medcalf N, Williams DJ. Regulatory challenges for the manufacture and scale-out of autologous cell therapies. StemBook 2008. doi:10.3824/stembook.1.96.1.1.

[64] European Medicines Agency (EMA). ICH Guideline Q9 on quality risk management 2014;44:1–20.

[65] Plant AL, Hanisch RJ. Reproducibility and Replicability in Science , A Metrology Perspective. Washington DC: 2018.

[66] Beck SC, Lock RJ. Uncertainty of measurement: an immunology laboratory perspective. Ann Clin Biochem 2015;52:7–17. doi:10.1177/0004563214551066.

[67] Lindmo T, Steen HB. Characteristics of a simple, high-resolution flow cytometer based o a new flow configuration. Biophys J 1979;28:33–44. doi:10.1016/S0006-3495(79)85157-7.

[68] Daly J, Tiersch T. Sources of variation in flow cytometric analysis of aquatic species sperm: The effect of cryoprotectants on flow cytometry scatter plots and subsequent population gating. Aquaculture 2012;11:179–88. doi:10.1038/jid.2014.371.

[69] Davis BH, Wood B, Oldaker T, Barnett D. Validation of cell-based fluorescence assays: Practice guidelines from the ICSH and ICCS - Part i - Rationale and aims. Cytom Part B - Clin Cytom 2013;84:282–5. doi:10.1002/cyto.b.21104.

[70] Tanqri S, Vall H, Kaplan D, Hoffman B, Purvis N, Porwit A, et al. Validation of cell-based

fluorescence assays: Practice guidelines from the ICSH and ICCS - Part III - Analytical issues. Cytom Part B - Clin Cytom 2013;84:291–308. doi:10.1002/cyto.b.21105.

[71] Wood B, Jevremovic D, Béné MC, Yan M, Jacobs P, Litwin V. Validation of cell-based fluorescence assays: Practice guidelines from the ICSH and ICCS - Part V - assay performance criteria. Cytom Part B Clin Cytom 2013;84:315–23. doi:10.1002/cyto.b.21108.

[72] Nightingale M. Measurement uncertainty for FMH , antibody titration and antibody quantification. Uncertain It All 2015.

[73] Peeling RW, Sollis KA, Glover S, Crowe SM, Landay AL, Cheng B, et al. CD4 enumeration technologies: A systematic review of test performance for determining eligibility for antiretroviral therapy. PLoS One 2015;10:1–26. doi:10.1371/journal.pone.0115019.

[74] Baradez MO, Lekishvili T, Marshall D. Rapid phenotypic fingerprinting of cell products by robust measurement of ubiquitous surface markers. Cytom Part A 2015;87:624–35. doi:10.1002/cyto.a.22637.

[75] Carraro P, Plebani M. Errors in a stat laboratory: Types and frequencies 10 years later. Clin Chem 2007;53:1338–42. doi:10.1373/clinchem.2007.088344.

[76] Epner PL, Gans JE, Graber ML. When diagnostic testing leads to harm: A new outcomes-based approach for laboratory medicine. BMJ Qual Saf 2013;22:6–10. doi:10.1136/bmjqs-2012-001621.

[77] Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care: Estimations from three large observational studies involving US adult populations. BMJ Qual Saf 2014;23:727–31. doi:10.1136/bmjqs-2013-002627.

[78] ISAC. Specialist in Cytometry, SCYM 2017. https://www.ascp.org/content/docs/default-source/boc-pdfs/boc-us-guidelines/scym_content_outline.pdf?sfvrsn=4 (accessed July 24, 2019).

[79] Ceriotti F. The role of external quality assessment schemes in monitoring and improving the standardization process. Clin Chim Acta 2014;432:77–81. doi:10.1016/j.cca.2013.12.032.

[80] Badrick T, Punyalack W, Graham P. Commutability and traceability in EQA programs. Clin Biochem 2018;56:102–4. doi:10.1016/j.clinbiochem.2018.04.018.

[81] International Organization for Standardization IEC. Conformity assessment — General requirements for proficiency testing. ISO/IEC 17043:2010 (E) 2010;1:1–46.

[82] ISO. ISO BSI 15189: Medical laboratories — Requirements for quality and competence. 2014.

[83] ISO. ISO 17511 In vitro diagnostic medical devices - Measurement of quantities in biological samples - Metrological traceability of values assigned to calibrators and control materials. ISO/TC 212 2003;3:1–10.

[84] Braga F, Infusino I, Panteghini M. Performance criteria for combined uncertainty budget in the implementation of metrological traceability. Clin Chem Lab Med 2015;53:905–12. doi:10.1515/cclm-2014-1240.

[85] BIPM. Joint Committee for Traceability in Laboratory Medicine (JCTLM) 2019. https://www.bipm.org/en/committees/jc/jctlm/ (accessed July 24, 2019).

[86] Li K, Donaldson B, Young V, Ward V, Jackson C, Baird M, et al. Adoptive cell therapy with CD4+ T helper 1 cells and CD8+ cytotoxic T cells enhances complete rejection of an established tumour, leading to generation of endogenous memory responses to non-targeted tumour epitopes. Clin Transl Immunol 2017;6:e160. doi:10.1038/cti.2017.37.

[87] Zanetti M. Tapping CD4 T Cells for Cancer Immunotherapy: The Choice of Personalized Genomics. J Immunol 2015;194:2049–56. doi:10.4049/jimmunol.1402669.

[88] Brando B, Gatti A, Chianese R, Gratama JW. Twenty years of external quality assurance in clinical cell analysis - A tribute to Jean-Luc D'Hautcourt. Cytom Part B - Clin Cytom 2007;72:2–7. doi:10.1002/cyto.b.20154.

[89] Waxdal M, Monical M, Fleisher T, Marti G. Inter-laboratory survey of lymphocyte immunophenotyping. Pathol Immunopathol Res 1988;7:345–56.

[90] Schnizlein-Bick CT, Spritzler J, Wilkening CL, Nicholson JKA, O'Gorman MRG, Investigators S. Evaluation of TruCount Absolute-Count tubes for determining CD4 and CD8 cell numbers in Human Immunodeficiency Virus-positive adults. Clin Vaccine Immunol 2000;7:336–43. doi:10.1128/cdli.7.3.336-343.2000.

[91] Barnett D, Granger V, Whitby L, Storie I, Reilly JT. Absolute CD4+ T-lymphocyte and CD34+ stem cell counts by single-platform flow cytometry: the way forward. Br J Haematol 1999;106:1059–62. doi:bjh1632 [pii].

[92] Reilly JT, Barnett D. UK NEQAS for leucocyte immunophenotyping: The first 10 years. J Clin Pathol 2001;54:508–11. doi:10.1136/jcp.54.7.508.

[93] UK NEQAS. UK NEQAS 2019. https://ukneqas.org.uk/ (accessed July 24, 2019).

[94] Whitby L, Granger V, Storie I, Goodfellow K, Sawle A, Reilly JT, et al. Quality control of CD4+ T-lymphocyte enumeration: Results from the last 9 years of the United Kingdom National External Quality Assessment Scheme for Immune Monitoring (1993-2001). Clin Cytom 2002;50:102–10. doi:10.1002/cyto.10094.

[95] Whitby L, Whitby A, Fletcher M, Helbert M, Reilly JT, Barnett D. Comparison of methodological data measurement limits in CD4+T lymphocyte flow cytometric enumeration and their clinical impact on HIV management. Cytom Part B - Clin Cytom 2013;84:248–54. doi:10.1002/cyto.b.21094.

[96] Barnett D, Whitby L, Wong J, Louzao R, Reilly JT, Denny TN. VERITAS?: A Time for VERIQAS™ and a new approach to training, education, and the quality assessment of CD4 + T lymphocyte counting (I). Cytom Part B - Clin Cytom 2012;82 B:93–100.

doi:10.1002/cyto.b.20624.

[97] Whitby L, Whitby A, Fletcher M, Barnett D. Current laboratory practices in flow cytometry for the enumeration of CD 4+ T-lymphocyte subsets. Cytom Part B - Clin Cytom 2015;88:305–11. doi:10.1002/cyto.b.21241.

[98] Bainbridge J, Rountree W, Louzao R, Wong J, Whitby L, Denny TN, et al. Laboratory accuracy improvement in the uk neqas leucocyte immunophenotyping immune monitoring program: An eleven-year review via longitudinal mixed effects modeling. Cytom Part B - Clin Cytom 2017:1–7. doi:10.1002/cyto.b.21531.

[99] Kluin-Nelemans JC, Van Wering ER, Van Der Schoot CE, Adriaansen HJ, Van't Veer MB, Van Dongen JJM, et al. SIHONSCORE: A scoring system for external quality control of leukaemia/lymphoma immunophenotyping measuring all analytical phases of laboratory performance. Br J Haematol 2001;112:337–43. doi:10.1046/j.1365-2141.2001.02500.x.

[100] Levering WHBM, Van Wieringen WN, Kraan J, Van Beers WAM, Sintnicolaas K, Van Rhenen DJ, et al. Flow cytometric lymphocyte subset enumeration: 10 Years of external quality assessment in the Benelux countries. Cytom Part B - Clin Cytom 2008;74:79–90. doi:10.1002/cyto.b.20370.

[101] Vesely R, Barths J, Vanlangendonck F, Hannet I, Stmuss K. Initial results of Central European Immunophenotyping Quality Control Program (CEQUAL). Cytometry 1996;26:108–12.

[102] Brando B, Sommaruga E. Nationwide quality control trial on lymphocyte immunophenotyping and flow cytometer performance in Italy. Cytometry 1993;14:294–306. doi:10.1002/cyto.990140310.

[103] Bergeron M, Faucher S, Minkus T, Lacroix F, Ding T, Phaneuf S, et al. Impact of unified procedures as implemented in the Canadian Quality Assurance Program for T lymphocyte subset enumeration and the participating Flow Cytometry laboratories of the Canadian Clinical Trials Network for HIV/AIDS Therapies. Cytometry 1998;33:146–55.

[104] Bergeron M, Ding T, Houle G, Arès L, Chabot C, Soucy N, et al. QASI, an international quality management system for CD4 T-cell enumeration focused to make a global difference. Cytom Part B - Clin Cytom 2010;78:41–8. doi:10.1002/cyto.b.20487.

[105] Pattanapanyasat K, Shain H, Noulsri E, Lerdwana S, Thepthai C, Prasertsilpa V, et al. A multicenter evaluation of the PanLeucogating method and the use of generic monoclonal antibody reagents for CD4 enumeration in HIV-infected patients in Thailand. Cytom Part B - Clin Cytom 2005;65:29–36. doi:10.1002/cyto.b.20052.

[106] Noulsri E, Lerdwana S, Pattanapanyasat K. Long-term external quality assessment program for CD4+ T-lymphocyte enumeration in Thailand. Accredit Qual Assur 2016. doi:10.1007/s00769-016-1225-9.

[107] Kalina T, Flores-Montero J, Lecrevisse Q, Pedreira CE, van der Velden VHJ, Novakova M, et al. Quality assessment program for EuroFlow protocols: Summary results of four-year (2010-2013) quality assurance rounds. Cytom Part A 2015;87:145–56. doi:10.1002/cyto.a.22581.

[108] Glencross DK, Aggett HM, Stevens WS, Mandy F. African regional external quality assessment for CD4 T-cell enumeration: Development, outcomes, and performance of laboratories. Cytom Part B - Clin Cytom 2008;74:S69–79. doi:10.1002/cyto.b.20397.

[109] Kunkl A, Risso D, Terranova MP, Girotto M, Brando B, Mortara L, et al. Grading of laboratories on CD4+ T-lymphocyte evaluations based on acceptable data boundaries defined by the measurement error. Clin Cytom 2002;50:117–26. doi:10.1002/cyto.10069.

[110] Iso 13528. Statistical methods for use in proficiency testing by interlaboratory comparison 2015;2015:1–16.

[111] Granger V, Barnett D, Reilly J, Mayr P, Fay S. US Patent 6,197,540: Preparation and stabilization of cells using aged transition metal solutions. 6197540, 2001.

[112] Whitby A, Whitby L, Fletcher M, Reilly JT, Sutherland DR, Keeney M, et al. ISHAGE protocol: Are we doing it correctly? Cytom Part B - Clin Cytom 2012;82 B:9–17. doi:10.1002/cyto.b.20612.

[113] Sutherland R, Anderson L, Keeney M, Nayar R, Chin-Yee I. The ISHAGE guidelines for CD34+ cell determination by Flow Cytometry. J Hematother 1996;5:213–26.

[114] Nomura LE, Walker JM, Maecker HT. Optimization of whole blood antigen-specific cytokine assays for CD4+ T cells. Cytometry 2000;40:60–8. doi:10.1002/(SICI)1097-0320(20000501)40:1<60::AID-CYTO8>3.0.CO;2-J.

[115] Maecker HT, Rinfret A, D'Souza P, Darden J, Roig E, Landry C, et al. Standardization of cytokine flow cytometry assays. BMC Immunol 2005;6:1–18. doi:10.1186/1471-2172-6-13.

[116] Rawstron AC, Villamor N, Ritgen M, Böttcher S, Ghia P, Zehnder JL, et al. International standardized approach for flow cytometric residual disease monitoring in chronic lymphocytic leukaemia. Leukemia 2007;21:956–64. doi:10.1038/sj.leu.2404584.

[117] Rawstron AC, Böttcher S, Letestu R, Villamor N, Fazi C, Kartsios H, et al. Improving efficiency and sensitivity: European research Initiative in CLL (ERIC) update on the international harmonised approach for flow cytometric residual disease monitoring in CLL. Leukemia 2013;27:142–9. doi:10.1038/leu.2012.216.

[118] O'Neill K, Aghaeepour N, Špidlen J, Brinkman R. Flow Cytometry Bioinformatics. PLoS Comput Biol 2013;9:1–10. doi:10.1371/journal.pcbi.1003365.

[119] Brinkman RR, Aghaeepour N, Finak G, Gottardo R, Mosmann T, Scheuermann RH. Automated analysis of flow cytometry data comes of age. Cytom Part A 2016;89:13–5.

doi:10.1002/cyto.a.22810.

[120] Bashashati A, Brinkman RR. A survey of Flow Cytometry Data analysis methods. Adv Bioinformatics 2009;2009:1–19. doi:10.1155/2009/584603.

[121] Lee H, Sun Y, Patti-Diaz L, Hedrick M, Ehrhardt AG. High-throughput analysis of clinical Flow Cytometry data by automated gating. Bioinform Biol Insights 2019;13:117793221983885. doi:10.1177/1177932219838851.

[122] Kvistborg P, Gouttefangeas C, Aghaeepour N, Cazaly A, Chattopadhyay PK, Chan C, et al. Thinking outside the gate: single-cell assessments in multiple dimensions. Immunity 2015;42:591–2. doi:10.1016/j.immuni.2015.04.006.

[123] LLC. Home | FlowJo, LLC 2018. https://www.flowjo.com (accessed January 10, 2019).

[124] Jimenez-Carretero D, Ligos JM, Martínez-López M, Sancho D, Montoya MC. Flow Cytometry data preparation guidelines for improved automated phenotypic analysis. J Immunol 2018;200:3319–31. doi:10.4049/jimmunol.1800446.

[125] Hourd P, Williams DJ. Scanning the horizon for high value-add manufacturing science: Accelerating manufacturing readiness for the next generation of disruptive, high-value curative cell therapeutics. Cytotherapy 2018;20:759–67. doi:10.1016/j.jcyt.2018.01.007.

[126] NEQAS U. Digital Blood Film Morphology for CPD 2019. https://ukneqas.org.uk/programmes/result/?programme=digital-blood-film-morphology-for-cpd (accessed July 27, 2019).

[127] MMIP. Advanced Therapies Manufacturing Action Plan. ABPI Taskforce Rep 2016.

[128] Tate JR, Johnson R, Barth J, Panteghini M. Harmonization of laboratory testing - Current achievements and future strategies. Clin Chim Acta 2014;432:4–7. doi:10.1016/j.cca.2013.08.021.

[129] Legg M, Swanepoel C. The Australian pathology units and terminology standardisation project - An overview. Clin Biochem Rev 2012;33:103–8.

[130] Plebani M, Sciacovelli L, Bernardi D, Aita A, Antonelli G, Padoan A. What information on measurement uncertainty should be communicated to clinicians, and how? Clin Biochem 2018;57:18–22. doi:10.1016/j.clinbiochem.2018.01.017.

[131] Padoan A, Sciacovelli L, Aita A, Antonelli G, Plebani M. Measurement uncertainty in laboratory reports: A tool for improving the interpretation of test results. Clin Biochem 2018;57:41–7. doi:10.1016/j.clinbiochem.2018.03.009.

[132] Infusino I, Panteghini M. Measurement uncertainty: Friend or foe? Clin Biochem 2018;57:3–6. doi:10.1016/j.clinbiochem.2018.01.025.

[133] Down M, Czubak F, Gruska G, Stahley S, Benham D. Measurement System Analysis. 2010. doi:10.1002/9780470997482.ch11.

[134] Montgomery DC. Introduction to statistical quality control. 6th ed. Hoboken, N.J.: Wiley;

2008.

[135] Ficalora J, Cohen L. Quality function deployment and Six Sigma : a QFD handbook. Upper Saddle River, N. J.: Pearson Education; 2010.

[136] Nicolay CR, Purkayastha S, Greenhalgh A, Benn J, Chaturvedi S, Phillips N, et al. Systematic review of the application of quality improvement methodologies from the manufacturing industry to surgical healthcare. Br J Surg 2012;99:324–35. doi:10.1002/bjs.7803.

[137] Williams T, Howe R. W. Edwards Deming and Total Quality Management: An Interpretation for Nursing Practice. J Healthc Qual 1992;14:36–9.

[138] Smart NJ. Lean Biomanufacturing: Creating value through innovative bioprocessing approaches. Cambridge, JK<: Woodhead Publishing Ltd; 2013.

[139] European Commission. Eudralex EU Guidelines for Good Manufacturing Practice for Medicinal Products for Human and Veterinary Use, Chapter 1: Pharmaceutical Quality System 2013;4:1–11. doi:10.2903/j.efsa.2015.4206.0J.

[140] Parkash V, Kumar D, Rajoria R. Statistical Process Control. Process Autom Handb 2007:837–48. doi:10.1007/978-1-84628-282-9_102.

[141] Joint Committee For Guides In Metrology. Vocabulaire International de Métrologie. vol. 3. 2012. doi:10.1016/0263-2241(85)90006-5.

[142] Bell S. Good Practice Guide No. 11 Issue 2 - A Beginners Guide to Uncertainty of Measurement. NPL; 2001.

[143] Statistics L. Measures of Spread | How and when to use measures of spread | Laerd Statistics n.d. https://statistics.laerd.com/statistical-guides/measures-of-spread-range-quartiles.php (accessed January 30, 2019).

[144] Field A. Discovering Statistics using IBM SPSS Statistics. Fifth edit. Los Angeles: Sage; 2018.

[145] Ghasemi A, Zahediasl S. Normality tests for statistical analysis: A guide for non-statisticians. Int J Endocrinol Metab 2012;10:486–9. doi:10.5812/ijem.3505.

[146] Yap BW, Sim CH. Comparisons of various types of normality tests. J Stat Comput Simul 2011;81:2141–55. doi:10.1080/00949655.2010.520163.

[147] Cramer D, Howitt D. The Sage dictionary of statistics : a practical resource for students in the social sciences. London: Sage; 2004.

[148] Montgomery DC. Statistical Quality Control. Sixth. Hoboken, N.J.: 2009.

[149] Cohen J. Statistical Power Analysis for the Behavoural Sciences. New York: Academic Press; 1977.

[150] Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39:175–91. doi:10.1109/ISIT.2013.6620417.

[151] Mayr S, Buchner A, Erdfeld E, Faul F. A short tutorial of G Power. Tutorials Quant Methods Psychol 2007;3:51–9.

[152] ISO. Evaluation of measurement data — Guide to the expression of uncertainty in measurement. Int Organ Stand Geneva ISBN 2008;50:134. doi:10.1373/clinchem.2003.030528.

[153] Kimothi S. The Uncertainty of Measurements. Milwaukee: American Society for Quality; 2002.

[154] Grant R, Coopman K, Medcalf N, Silva-Gomes S, Campbell JJ, Kara B, et al. Understanding The Contribution Of Operator Measurement Variability Within Flow Cytometry Data Analysis For Quality Control Of Cell And Gene Therapy Manufacturing. Measurement 2020;150:106998. doi:10.1016/j.measurement.2019.106998.

[155] BD Bioscience. BD ™ CS & T Beads 2016. http://www.bdbiosciences.com/ds/europe/tds/23-14666.pdf (accessed April 11, 2018).

[156] BD Bioscience. BD Human and Mouse Pluripotent Stem Cell Analysis Kit 2014. http://www.bdbiosciences.com/ds/pm/others/23-10865.pdf (accessed August 11, 2019).

[157] FIRM Symposium. Previous Symposia | FIRM Symposium 2017. https://www.firmsymposium.com/copy-of-past-events (accessed April 11, 2018).

[158] Thermofisher Scientific. Optics of a Flow Cytometer 2019. https://www.thermofisher.com/uk/en/home/life-science/cell-analysis/cell-analysis-learning-center/molecular-probes-school-of-fluorescence/flow-cytometry-basics/flow-cytometry-fundamentals/optics-flow-cytometer.html (accessed June 2, 2019).

[159] Bainbridge J, Rountree W, Louzao R, Wong J, Whitby L, Denny TN, et al. Laboratory Accuracy Improvement in the UK NEQAS Leucocyte Immunophenotyping Immune Monitoring Program: An Eleven-Year Review via Longitudinal Mixed Effects Modeling. Cytom Part B - Clin Cytom 2018;94:250–6. doi:10.1002/cyto.b.21531.

[160] Mount NM, Ward SJ, Kefalas P, Hyllner J. Cell-based therapy technology classifications and translational challenges. Philos Trans R Soc B Biol Sci 2015;370. doi:10.1098/rstb.2015.0017.

[161] Best M, Neuhauser D. Walter A Shewhart, 1924, and the Hawthorne factory. Qual Saf Heal Care 2006;15:142–3. doi:10.1136/qshc.2006.018093.

[162] Christensen SL, Anglov JTB, Christensen JM, Olsen E, Poulsen OM. Application of a new AMIQAS computer program for integrated quality control, method evaluation and proficiency testing. Anal Bioanal Chem 1993;345:343–50.

[163] Healy MJR. Outliers in clinical chemistry quality-control schemes. Clin Chem 1979;25:675–7.

[164] Goodfellow KJ, Storie I, Granger V, Whitby L, Antcliffe J, Reilly JT, et al. The United Kingdom

National External Quality Assessment Scheme gating and standardization strategy for use in residual WBC counting of WBC-reduced blood components. Transfusion 2002;42:738–46. doi:10.1046/j.1537-2995.2002.00116.x.

[165] Leys C, Ley C, Klein O, Bernard P, Licata L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. J Exp Soc Psychol 2013;49:764–6. doi:10.1016/j.jesp.2013.03.013.

[166] Paxton H, Kidd P, Landay A, Giorgi J V, Flomenberg N, Walker E, et al. Results of the flow cytometry AGTG quality control program: Analysis and findings. Clin Immunol Immunopathol 1989;52:68–84.

[167] Gramatica P. Principles of QSAR models validation: Internal and external. QSAR Comb Sci 2007;26:694–701. doi:10.1002/qsar.200610151.

[168] Coucke W, China B, Delattre I, Lenga Y, Van Blerk M, Van Campenhout C, et al. Comparison of different approaches to evaluate External Quality Assessment Data. Clin Chim Acta 2012;413:582–6. doi:10.1016/j.cca.2011.11.030.

[169] Bainbridge J, Wilkening CL, Rountree W, Louzao R, Wong J, Perza N, et al. The Immunology Quality Assessment Proficiency Testing Program for CD3+4+ and CD3+8+ lymphocyte subsets: A ten year review via longitudinal mixed effects modeling. J Immunol Methods 2014;409:82–90. doi:10.1016/j.jim.2014.05.017.

[170] Hoo KA, Tvarlapati KJ, Piovoso MJ, Hajare R. A method of robust multivariate outlier replacement. Comput Chem Eng 2002;26:17–39. doi:10.1016/S0098-1354(01)00734-7.

[171] Nieboer D, Steyerberg EW, Soedamah-Muthu S, Vergouwe Y. Log transformation in biomedical research: (mis)use for covariates. Stat Med 2013;32:3770–1. doi:10.1002/sim.5793.

[172] Feng C, Wang H, Lu N, Chen T, He H, Lu Y, et al. Log-transformation and its implications for data analysis. Shanghai Arch Psychiatry 2014;26:105–9.

[173] Roederer M. Interpretation of cellular proliferation data: avoid the panglossian. Cytom Part A 2011;79:95–101.

[174] Marko NF, Weil RJ. Non-gaussian distributions affect identification of expression patterns, functional annotation, and prospective classification in human cancer genomes. PLoS One 2012;7:1–15. doi:10.1371/journal.pone.0046935.

[175] Josephson R, Ording CJ, Liu Y, Shin S, Lakshmipathy U, Toumadje A, et al. Qualification of embryonal carcinoma 2102 Ep as a reference for human embryonic stem cell research. Stem Cells 2007;25:437–46. doi:10.1634/stemcells.2006-0236.

[176] Zeineddine D, Papadimou E, Chebli K, Gineste M, Liu J, Grey C, et al. Oct-3/4 dose dependently regulates specification of Embryonic Stem Cells toward a cardiac lineage and early heart development. Dev Cell 2006;11:535–46. doi:10.1016/j.devcel.2006.07.013.

[177] Pazhanisamy S. Adult stem cell and embryonic stem cell markers. Mater Methods 2013;3:200.

[178] Draper JS, Pigott C, Thomson JA, Andrews PW. Surface antigens of human embryonic stem cells: changes upon differentiation in culture. J Anat 2002;200:249–58.

[179] Saldaña J. The Coding Manual for Qualitative Researchers. 3rd ed. London: Sage Publications Ltd; 2016.

[180] Holmes JL, Grant R, Petzing J. Quantifying operator variation to improve biomanufacturing process control (Restricted Access). Loughborough: 2017.

[181] Haeckel R, Wosniok W, Gurr E, Peil B. Permissible limits for uncertainty of measurement in laboratory medicine. Clin Chem Lab Med 2015;53:1161–71. doi:10.1515/cclm-2014-0874.

[182] HTA. Human Tissue Act 2004. HTA Website 2004:30: 1-63. doi:10.1258/rsmmlj.72.4.148.

[183] Perfetto SP, Chattopadhyay PK, Lamoreaux L, Nguyen R, Ambrozak D, Koup RA, et al. Amine-reactive dyes for dead cell discrimination in fixed samples. Curr Protoc Cytom 2010:1–20. doi:10.1002/0471142956.cy0934s53.

[184] National Institute of Health. CD3 Complex. 2019 2019. https://meshb.nlm.nih.gov/record/ui?name=CD3 Antigens (accessed June 10, 2019).

[185] Bell L. CD4 T cells. Manchester: 2019.

[186] van Hall T, van der Burg S. Chapter 3 - Mechanisms of Peptide Vaccination in Mouse Models: Tolerance, Immunity and Hyperreactivity. In: Melief C, editor. Adv. Immunol., 2012, p. 51–76.

[187] Wissinger E. CD8 T Cells. London: 2017. doi:10.1007/978-1-4419-0717-2_93.

[188] Andersen M, Schrama D, thor Straten P, Becker J. Cytotoxic T Cells. Compr Toxicol Second Ed 2006;126:32–41. doi:10.1016/B978-0-08-046884-6.00606-0.

[189] McBride JA, Striker R. Imbalance in the game of T cells: What can the CD4/CD8 T-cell ratio tell us about HIV and health? PLOS Pathog 2017;13:1–7. doi:10.1371/journal.ppat.1006624.

[190] Mahnke YD, Brodie TM, Sallusto F, Roederer M, Lugli E. The who's who of T-cell differentiation: Human memory T-cell subsets. Eur J Immunol 2013;43:2797–809. doi:10.1002/eji.201343751.

[191] Poli A, Michel T, Thérésine M, Andrès E, Hentges F, Zimmer J. CD56 bright natural killer (NK) cells: an important NK cell subset. Immunology 2009;126:458–65. doi:10.1111/j.1365-2567.2008.03027.x.

[192] Kershaw MH, Westwood JA, Darcy PK. Gene-engineered T cells for cancer therapy. Nat Rev Cancer 2013;13:525–41.

[193] Dhodapkar M V. Navigating the Fas lane to improved cellular therapy for cancer. J Clin

Invest 2019;129:1522–3.

[194] Tschumi BO, Dumauthioz N, Marti B, Zhang L, Schneider P, Mach JP, et al. CAR-T cells are prone to Fas- and DR5-mediated cell death. J Immunother Cancer 2018;6:1–9. doi:10.1186/s40425-018-0385-z.

[195] National Institute of Health. How does gene therapy work? 2019. https://ghr.nlm.nih.gov/primer/therapy/procedures (accessed June 28, 2019).

[196] Philpott NJ, Turner AJ, Scopes J, Westby M, Marsh JC, Gordon-Smith EC, et al. The use of 7-amino actinomycin D in identifying apoptosis: simplicity of use and broad spectrum of application compared with other techniques. Blood 1996;87:2244–51.

[197] Laerd Statistics. Friedman test using SPSS Statistics. Stat Tutorials Softw Guid 2015. https://statistics.laerd.com/ (accessed July 9, 2019).

[198] Armstrong RA. When to use the Bonferroni correction. Ophthalmic Physiol Opt 2014;34:502–8. doi:10.1111/opo.12131.

[199] Jisc. Online Surveys 2019. https://www.onlinesurveys.ac.uk/ (accessed July 21, 2019).

[200] Spalding JAB, Cole BL, Mir FA. Advice for medical students and practitioners with colour vision deficiency: A website resource. Clin Exp Optom 2010;93:39–41. doi:10.1111/j.1444-0938.2009.00434.x.

[201] Dargahi H, Einollahi N, Dashti N. Color blindness defect and medical laboratory technologists: Unnoticed problems and the care for screening. Acta Med Iran 2010;48:172–7.

[202] Miltenyi Biotec. MACSQuant Analyzer 10 2019. https://www.miltenyibiotec.com/GB-en/products/macs-flow-cytometry/flow-cytometers/macsquant-analyzer-10/macsquant-r-analyzer-10.html (accessed July 21, 2019).

[203] BD Bioscience. BD FACSCanto II 2019. http://m.bdbiosciences.com/us/instruments/research/cell-analyzers/bd-facscanto-ii/m/744810/features (accessed July 21, 2019).

[204] BD Bioscience. BD FACSCelesta 2019. http://www.bdbiosciences.com/en-us/instruments/research-instruments/research-cell-analyzers/facscelesta (accessed July 21, 2019).

[205] Miltenyi Biotec. MACSQuantify 2019. https://www.miltenyibiotec.com/CA-en/products/macs-flow-cytometry/software/macsquantify.html (accessed July 21, 2019).

[206] Gratama JW, Kraan J, Beemd V Den, Hooibrink B, Bockstaele DR Van, Hooijkaas H. Analysis of variation in results of Flow Cytometric lymphocyte immunophenotyping in a multicenter study 1997;30:166–77.

[207] Sano H, Kanemura N, Burrow M, Inai N, Yamada T, Tagami J. Effect of operator variability students vs. dentists on Dentin adhesion. Dent Mater J 1998;17:51–8.

[208] Gomes G, Gomes O, Reis A, Gomes J, Loguercio A, Calixto A. Effect of operator experience

on the outcome of fiber post C cementation with different resin cements. Oper Dent 2012;38:555–64. doi:10.2341/11-494-l.

[209] Gosling AF, Kendrick DE, Kim AH, Nagavalli A, Kimball ES, Liu NT, et al. Simulation of carotid artery stenting reduces training procedure and fluoroscopy times. J Vasc Surg 2017;66:298–306. doi:10.1016/j.jvs.2016.11.066.

[210] Madhavan P, Wiegmann DA. Effects of information source, pedigree, and reliability on operator interaction With decision support systems. Hum Factors J Hum Factors Ergon Soc 2007;49:773–85. doi:10.1518/001872007x230154.

[211] Ambriz-aviña V, Contreras-garduño JA, Pedraza-reyes M. Applications of Flow Cytometry to Characterize Bacterial Physiological Responses. Biomed Res Int 2014;2014:1–14.

[212] Bruggner R V, Bodenmiller B, Dill DL, Tibshirani RJ, Nolan GP. Automated identification of stratifying signatures in cellular subpopulations. Proc Natl Acad Sci 2014;111:E2770–7. doi:10.1073/pnas.1408792111.

[213] Reiner B. Redefining image quality analysis. Mycobact Dis 2014;4:2–3. doi:10.4172/2167-7964.1000e126.

[214] Wolstenhulme S, Davies AG, Keeble C, Moore S, Evans JA. Agreement between objective and subjective assessment of image quality in ultrasound abdominal aortic aneurism screening. Br J Radiol 2015;88. doi:10.1259/bjr.20140482.

[215] Suther KR, Hopp E, Smevik B, Fiane AE, Lindberg HL, Larsen S, et al. Can visual analogue scale be used in radiologic subjective image quality assessment? Pediatr Radiol 2018;48:1567–75. doi:http://dx.doi.org/10.1007/s00247-018-4187-8.

[216] Wang J, Jokerst J V. Stem Cell Imaging: Tools to Improve Cell Delivery and Viability. Stem Cells Int 2016;2016:1–16. doi:10.1155/2016/9240652.

[217] Lee Z, Dennis J, Alsberg E, D.Krebs M, Welter J, Caplan A. Imaging Stem Cell Differentiation for Cell-Based Tissue Repair. Methods Enzymol., 2012, p. 247–63.

[218] Nishida K, Hotta K. Robust cell particle detection to dense regions and subjective training samples based on prediction of particle center using convolutional neural network. PLoS One 2018;13:1–13. doi:10.1371/journal.pone.0203646.

[219] Chen D, Sarkar S, Candia J, Florczyk SJ, Bodhak S, Driscoll MK, et al. Machine learning based methodology to identify cell shape phenotypes associated with microenvironmental cues. Biomaterials 2016;104:104–18. doi:10.1016/j.biomaterials.2016.06.040.

[220] Smith D, Glen K, Thomas R. Automated image analysis with the potential for process quality control applications in stem cell maintenance and differentiation. Biotechnol Prog 2016;32:215–23. doi:10.1002/btpr.2199.

Appendix A

Ethics Approvals (Human Participants) Sub-Committee

Loughborough University

Ethical Clearance Checklist

Has the Investigator read the 'Guidance for completion of Ethical Clearance Checklist' before starting this form?	Yes

Does the study require NHS approval? *Please complete a copy of the checklist providing a brief project description in the additional information section. Please send this to the Secretary of the Ethics Approvals (HP) Sub-Committee before starting your NHS application.*	No

Project Details

1. Project Title: Understanding the effects of Operator Variation on Flow Cytometry Measurements

Investigator(s) Details

2. Name of Investigator 1: Rebecca Grant	10. Name of Investigator 2: Dr Jon Petzing
3. Status: PGR student	11. Status: Staff
4. School/Department: Wolfson School of MEME	12. School/Department: Wolfson School of MEME
5. Programme (if applicable): Click here to enter text.	13. Programme (if applicable): Click here to enter text.
6. Email address: r.grant@lboro.ac.uk	14. Email address: j.petzing@lboro.ac.uk
7a. Contact address: Healthcare Engineering Wolfson School of MEME Loughborough University, LE11 3TU	15a. Contact address: Wolfson School of MEME Loughborough University, LE11 3TU
7b. Telephone number: 01509 564890	15b. Telephone number: 01509 227617
8. Supervisor: No	16. Supervisor: Yes
9. Responsible Investigator: Yes	17. Responsible Investigator: No
List all other investigators (name/email address): Professor Nick Medcalf	

Participants

18. Does the project involve NHS patients from the National Centre for Sport and Exercise Medicine. *NHS approval may be required. Please complete a copy of the checklist providing a brief project description in the additional information section. Please send this to the Secretary of the Ethics Approvals (HP) Sub-Committee.*	No

Positions of Authority

19. Are investigators in a position of direct authority with regard to participants (e.g. academic staff using student participants, sports coaches using his/her athletes in training)?	No

Vulnerable groups

20. Will participants be knowingly recruited from one or more of the following vulnerable groups?	
Children under 18 years of age	No
Persons incapable of making an informed decision for themselves	No
‡ Pregnant women	No
Prisoners/Detained persons	No
Other vulnerable group Please specify: Click here to enter text	No
If Yes to any of question 20, please answer the following questions:	
21. Will participants be chaperoned by more than one investigator at all times?	Choose an item
22. Will at least one investigator of the same sex as the participant(s) be present throughout the investigation?	Choose an item
23. Will participants be visited at home?	Choose an item

Investigator Safety

24. Will the investigator be alone with participants at any time?	No
If Yes to question 24, please answer the following questions:	
24a. Will the investigator inform anyone else of when they will be alone with participants?	Choose an item

24b. Has the investigator read the Guidance Notes on 'Conducting Interviews Off-Campus and Working Alone' and will abide by the recommendations within?	Choose an item

Methodology and Procedures

25. Please indicate whether the proposed study:	
‡ Involves taking bodily samples (please refer to published guidelines)	No
Involves using bodily samples previously collected with consent for further research	No
Involves transporting Human Tissue Act relevant material to or from Loughborough (a materials transfer agreement is required)	No
Involves procedures which are likely to cause physical, psychological, social or emotional distress to participants	No
Is designed to be challenging **physically** or **psychologically** in any way (includes any study involving physical exercise)	No
Exposes participants to risks or distress greater than those encountered in their normal lifestyle	No
‡ Involves collection of body secretions by invasive methods	No
‡ Prescribes intake of compounds additional to daily diet or other dietary manipulation/supplementation	No
‡ Involves pharmaceutical drugs/medicines	No
Involves use of radiation	No
Involves use of hazardous materials	No
Assists/alters the process of conception in any way	No
Involves methods of contraception	No
Involves genetic engineering	No

‡ Involves testing new equipment	No
‡ Involves testing of medical equipment or devices	No

Observation/Recording

26. Does the study involve observation and/or recording of participants?	Yes
27. **If Yes to question 26,** will those being observed and/or recorded be informed that the observation and/or recording will take place?	Yes

Informed consent

28. Will participants give informed consent freely?	Yes
29. Will participants be fully informed of the objectives of the study	

and all details disclosed (preferably at the start of the study but, where this would interfere with the study, at the end)?	Yes
30. Will participants be fully informed of the use of the data collected (including, where applicable, any intellectual property arising from the research)?	Yes

31. For children under the age of 18 or participants who are incapable of making an informed decision for themselves:	
a. Will consent be obtained (either in writing or by some other means)?	Choose an item
b. Will consent be obtained from parents or other suitable person?	Choose an item
c. Will they be informed that they have the right to withdraw regardless of parental/guardian consent?	Choose an item
d. For studies conducted in schools, will approval be gained in advance from the Head-teacher and/or the Director of Education of the appropriate Local Education Authority?	Choose an item
e. For detained persons, members of the armed forces, employees, students and other persons judged to be under duress, will care be taken over gaining freely informed consent?	Choose an item

Deception

32. Does the study involve deception of participants (i.e. withholding of information or the misleading of participants) which could potentially harm or exploit participants?	No
If Yes to question 32, please answer the following questions:	
33. Is deception an unavoidable part of the study?	Choose an item
34. Will participants be de-briefed and the true object of the research revealed at the earliest stage upon completion of the study?	Choose an item
35. Will there be an increased physical or emotional risk to participants or investigators when participants are informed of the withholding of information or deliberate deception?	Choose an item

Withdrawal

36. Will participants be informed of their right to withdraw from the investigation at any time and to require their own data to be destroyed?	Yes

Storage of Data and Confidentiality

37. Will all information on participants be treated as confidential and not identifiable unless agreed otherwise in advance, and subject to the requirements of law?	Yes

38. Will storage of data comply with the Data Protection Act 1998 and the Guidance Note on 'Data Protection and Storage'?	Yes
39. Will any transcripts and video/audio recording of participants be kept in a secure place and not released for any use by third parties?	Yes
40. Will video/audio recordings be destroyed within ten years of the completion of the investigation or securely archived if required by funder?	N/A
41. Will full details regarding the storage and disposal of any human tissue samples be communicated to the participants?	N/A
42. Will research involve the sharing of data or confidential information beyond the initial consent given?	No
43. Will the research involve administrative or secure data that requires permission from the appropriate authorities before use?	No

Incentives

44. Will incentives be offered to the investigator to conduct the study?	No
45. Will incentives by offered to potential participants as an inducement to participate in the study?	No

Work Outside of the United Kingdom

46. Is research being conducted by investigators travelling outside of the United Kingdom?	No
If Yes to question 46, please answer the following questions:	
47. Country or countries researcher will travel to for the conduct of the research:	Click here to enter text
48. Is this the investigator's home country?	Choose an item
49. Has a risk assessment been carried out to ensure the physical, emotional and cultural safety of the investigator whilst working outside of the United Kingdom?	Choose an item
50. Have you considered the appropriateness of your research in the country you are travelling to and checked the FCO guidance: https://www.gov.uk/foreign-travel-advice?	Choose an item
51. Is there an increased physical, emotional or cultural risk to investigators outside of the United Kingdom as a result of your research study or has the FCO issued a travel warning?	Choose an item
52. Have you obtained any necessary ethical permission needed in the country you are travelling to?	Choose an item

53. Will any of the participants be outside of the United Kingdom?	No
54. **If Yes to 53**, is there an increased physical, emotional or cultural risk to participants who are outside of the United Kingdom as a result of taking part in your research study?	Choose an item

Risk Assessment

55. Has a risk assessment been carried out and approved by the School, to ensure the physical, emotional and cultural safety of the investigator and participants involved in the study?	Yes

Information and Declarations

‡ If you have selected answers marked with this symbol you should complete the additional INSURANCE FORM which is available on the Sub-Committee's website to ensure appropriate insurance cover.

Checklist Application Only:
If you have completed the checklist to the best of your knowledge, and not selected any answers marked with an *, # or †, your investigation is deemed to conform with the ethical checkpoints. Please sign the declaration and lodge the completed checklist with your Head of Department/School or his/her nominee.

† Checklist with Additional Information to the Secretary:
If you have completed the checklist and have only selected answers which require additional information to be submitted with the checklist (indicated by a †), please ensure that all the information is provided in detail below and send this signed checklist to the Secretary of the Sub-Committee.

Checklist with Generic Protocols Included:
If you have completed the checklist and selected one or more of the answers marked with this symbol # a full Research Proposal needs to be submitted to the Ethical Approvals (Human Participants) Sub-Committee unless you, or one of the investigators on this project, are a named investigator on an existing Generic Protocol which covers the procedure. Please download the Research Proposal form from the Sub-Committee's web page. **A signed copy of this Checklist should accompany the full proposal to the Sub-Committee.**

If you, or one of the investigators on this project, are using a procedure covered by a generic protocol, please ensure the relevant individuals are on the list of approved investigators for that Generic Protocol. Include the Generic Protocol reference number and a short description of how the proposal will be used at the end of the checklist in the space provided for additional information.

The completed checklist should be lodged with your Head of Department/School or his/her nominee.

* Full Application needed:
If on completion of the checklist you have selected one or more answers which require the submission of a full proposal (indicated by a *), please download the Research Proposal form from the Sub-Committee's web page. **A signed copy of this Checklist should accompany the full Research Proposal to the Sub-Committee.**

Space for Additional Information and/or Information on Generic Proposals as requested:

N/A

For completion by Supervisor

Please tick the appropriate boxes. The study should not begin until all boxes are ticked.

☑ The student has read the University's Code of Practice on investigations involving human participants

☑ The topic merits further research

☑ The student has the skills to carry out the research or is being trained in the required skills by the Supervisor

☑ The participant information sheet or leaflet is appropriate

☑ The procedures for recruitment and obtaining informed consent are appropriate

Comments from supervisor:

No further comments to be added.

Signature of Applicant:

Signature of Supervisor (if applicable):

Signature of Dean of School/Head of Department or his/her nominee:

Date: 7th November 2016

Appendix B

amsbio
www.amsbio.com
sales@amsbio.com
T: +44 (0) 1235 828200
T: +49 (0) 69 779099
T: +41 (0) 91 604 55 22
T: +1.949.265.7717

GlobalStem, Inc.
www.globalstem.com

Certificate of Analysis

Name: 2102Ep Cl.2/A6

Description: Human Embryonal Carcinoma, P48

Unit Size: ~5 x 10^6 cells/vial

MSDS: Available upon request

Shipping: Frozen on dry ice

Storage: Vapor phase of liquid nitrogen

Preparation: Dilute the vial contents 1:10 with media and plate the cells at the suggested density of 6.7 x 10^4 cells/cm^2. Refer to the Product Information Sheet provided for related information and protocols.

Catalog #: GSC-2001
Lot #: 61500006
Expiration Date: N/A

GlobalStem cell lines are prepared under aseptic conditions and each lot must pass screening for bacterial/fungal contamination, pathogens, and mycoplasma. Karyotype and cell line identity data available.

TEST PARAMETERS	SPECIFICATION	RESULT
Sterility[1]	no growth	no growth
Mycoplasma[2]	negative	negative
Human Pathogens[3]	negative	negative
Marker Panel[4]	acceptable	acceptable
Viability	> 50%	95-98 %

1. Direct inoculation and culture of blood agar plates, thioglycolate broth, tryptocase soy broth, and sabaoraud agar to check for bacterial and fungal contamination.
2. PCR-based assay sensitive to the detection of 25 species typically found as contaminants of cell culture.
3. Cells are tested for HIV-1, HIV-2, Human T-Cell Lymphotrophic Virus I/II, HSV1, HSV2, EBV, CMV, Hepatitis B Virus and Hepatitis C Virus.
4. Cells are stained with common markers for undifferentiated human embryonic stem cells.

NOTE: This product is intended for *Research Purposes* only. Utilization of this product apart from the intended use may be a violation of Federal Law.

Approved: _____ **Quality Systems**

Date: 30 Nov 2007

ATCC®

CERTIFICATE OF ANALYSIS

ATCC® Number: PCS-800-011™

Lot Number: 80628171

ATCC
10801 University Boulevard
Manassas, VA 20110-2209 USA
www.atcc.org

800-638-6597 or 703-365-2700
Fax: 703-365-2750
E-mail: tech@atcc.org
or contact your local distributor

ATCC®

CERTIFICATE OF ANALYSIS

ATCC® Number: PCS-800-011™

Lot Number: 80628171

Name:	**Primary Peripheral Blood Mononuclear Cells, Normal, Human**
Description:	**Mononuclear Cells**
Species:	**Human (*Homo sapiens*)**
Source:	**Human Peripheral Blood**
Age:	**36 years**
Gender:	**Male**
Ethnicity:	**Hispanic**
Blood Type:	**O Positive**
Volume/Ampule:	**Approximately 1 mL**
Product Format:	**Cells cryopreserved in the appropriate cryopreservation medium**
Expiration Date:	**Not applicable**
Storage Conditions:	**Vapor phase of liquid nitrogen**

Test / Method	Specification	Result
Average viable cells/ampule	≥ 2.5 x 10^7 (25 million) viable cells/ampule	3.2 x 10^7 viable cells/ampule
Post-freeze viability	≥ 70%	98.7%
Human pathogenic virus testing for HIV (I/II), HepB, HepC, and HTLV (I/II)	Negative	HIV (I/II) – Negative HepB – Negative HepC – Negative HTLV (I/II) – Negative
Characterization / cell specific staining (by flow cytometry)	CD45+: Positive (≥ 70%) CD3+, CD4+, CD8+, CD14+, CD19+, CD56+: Report results	CD45+: Positive, 94.13% CD3+: 53.59% CD4+: 51.78% CD8+: 23.92% CD14+: 22.41% CD19+: 10.10% CD56+: 12.41%

Jo Salisbury

Digitally signed by Jo Salisbury
DN: cn=Jo Salisbury, o=ATCC, ou=Quality Assurance, e=mailto:jsalisbury@atcc.org, c=US
Date: 2017-09-12 15:27:50 -04'00'

Quality Assurance Specialist; Quality Assurance

ATCC hereby represents and warrants that the material provided under this certificate is pure and has been subjected to the tests and procedures specified and that the results described, along with any other data provided in this certificate, are true and correct to the

ATCC
10801 University Boulevard
Manassas, VA 20110-2209 USA
www.atcc.org

800-638-6597 or 703-365-2700
Fax: 703-365-2750
E-mail: tech@atcc.org
or contact your local distributor

November 14, 2017

To,
Rebecca Grant MEng (Hons) DIS AMIMechE
PhD Researcher
IMechE Postgraduate Research Scholar
Metrology & Healthcare Engineering Groups
Wolfson School, Loughborough University
Email: r.grant@lboro.ac.uk
Tel: +44 (0) 1509 227 567

Re: Informed Consent for Primary Peripheral Blood Mononuclear Cells (ATCC® PCS-800-011™) and Primary Peripheral Blood CD14+ Monocytes, Normal, Human (ATCC® PCS-800-010™)

Dear Rebecca Grant,

ATCC primary cells are intended for research use only and are derived from source tissues that were obtained through organ procurement organizations (OPO) with the appropriate donor consent forms. Donor information is compliant with the U.S. Department of Health and Human Services, Health Insurance Portability and Accountability Act of 1996 (HIPAA) guidelines (http://www.hhs.gov/ocr/privacy).

Most of the human-derived materials which are distributed by ATCC are not derived at ATCC directly from donor specimens. Thus, ATCC was not directly involved in the donor consent process for these materials. ATCC does not request and does not have access to information about the identity of the tissue donors of the cell cultures. ATCC human-derived cell lines do not have identifiers or links back to specific individuals or patients.

For 90 years, ATCC has been a nonprofit supplier of biological materials to the scientific community, with a well-known reputation for the highest quality products. ATCC is committed to providing excellent products and services to our customers.

Regards,

Sharon Sequeira, MS., Ph.D
Supervisor, ATCC Customer Care – Technical Support

Certificate of Attendance

This is to certify that

Becky Grant

Has attended a

Centre for Biological Engineering
Local HTA Induction Training
On
16th November 2016

Carolyn Kavanagh
Departmental Quality Manager

Loughborough University

Centre for Biological Engineering

This is to Certify that

.....REBECCA...GRANT.....

Has attended the HTA Awareness Training session

On

.....Monday 19th September 2016.....

Signed

(Designated Individual - Loughborough University)

MRC Medical Research Council

This is to certify that

Rebecca Grant

completed the following e-learning with an assessment (England, Wales and Northern Ireland) score of

70%

Research and human tissue legislation

Overview of Human Tissue Act 2004 and Human Tissue (Scotland) Act 2006

When the Acts apply

What constitutes best practice

Top tips to support compliance

Where to find help

December 14, 2016 MRC Regulatory Support Centre

Loughborough University The Centre for Biological Engineering	HTA Training Record REBECCA GRANT			
Document Ref: HTA-RM-FORM/005	Version N°:	1.0	Issue Date:	24/10/2016

HTR No: CBE-HTA/TR/000	Working with HTA Licensable Material

Part 1A: Standard Operating Procedure (SOP) Sign-off

This section is to certify that you have:

- Read the SOPs relevant to your role
- Understand how to apply the SOPs to your work
- Know where to locate a copy of the current SOPs

SOP Number	Title	Version	Date Read	Signature
HTA-QS-SOP001	Production & control of SOPs applied to activities involving HTA licensable material	001	24/10/2016	RGrant.
HTA-QS-SOP002	Change control for work activities involving HTA licensable material	001	24/10/2016	RGrant.
HTA-RM-SOP003	Induction and training for HTA compliance	001	24/10/2016	RGrant.
HTA-PR-SOP004	Receipt and storage of HTA licensable material	001	24/10/2016	RGrant.
HTA-PR-SOP005	Setting up and editing projects in the Procuro database	001	24/10/2016	RGrant.
HTA-PR-SOP006	Acquisition, transfer and transport of HTA licensable material	001	09/11/2016	RGrant.
HTA-PR-SOP007	Disposal of HTA licensable material	001	09/11/2016	RGrant.
HTA-MI-SOP008	Reporting of adverse events relating to HTA licensable material	001	09/11/2016	RGrant.
HTA-MI-SOP009	Self-inspection audit of research activities involving HTA licensable material	001	09/11/2016	RGrant.
HTA-MI-SOP010	Risk management and contingency planning	001	09/11/2016	RGrant.

Part 1B: Relevant Reading Material Sign-off

Document Title	Version	Date Read	Signature
HTA Code of Practice for Research _Code 9	14.0	16/11/16	RGrant.
University HTA Compliance Quality Manual	004	16/11/16	RGrant.
CBE Quality Manual	1.0	16/11/16	RGrant.

Appendix D

D.1 Initial Questionnaire Outline

D.1.1 Personal Questions

1) What is your name?

2) Who is your Principal Investigator/ Team Leader (if applicable)?

3) Do you hold any Flow Cytometry qualifications (e.g. C.Cy)?

If you answered yes, please list what these qualifications are and when you attained them.

D.1.2 Experience Questions

4) Please tick what you have experience using Flow Cytometry for.

5) How much experience (in years and months) do you have with Flow Cytometry?

6) On average, how frequently do you use a Flow Cytometer?

7) How were you originally taught to apply gates?

8) Have you attended any Flow Cytometry Training courses? If so, when?

9) What cell types do you most commonly work with when using a Flow Cytometer?

10) Do you use a written/diagrammatical protocol to apply gates to your data?

10a) If yes, is this a protocol you have written, an internal SOP or an external publication?

11) Do you prefer to use automated or manual gating?

11a) Why do you prefer to use automated or manual gating? If you use automated gating, please list the software package used.

12) Do you have experience using Flowjo Software?

13) Have you ever participated in an External Quality Assessment Scheme (EQAS) for Flow Cytometry?

14) Have you ever submitted your files for central processing (e.g. quality scheme)?

D.1.3 Vision Questions

Do you have colour blindness, or any other visual impairment that can affect your judgement of colour and shape on a computer screen?

D.1.4 Motivational Factors Questions

15) What problems do you encounter when gating manually? How do you deal with these?

16) What problems do you encounter when interpreting published Flow Cytometry data from another source?

17) How much do you enjoy gating?

18) Do you feel you have to be in a certain mood to gate?

19) Do you have to be in a certain environment to gate?

20) Do you prefer to analyse your data by applying gates at a certain time of day? If so, when do you like to apply your gates?

D.2 Follow-up Questionnaire Outline

D.2.1 Usage Questions

1) What is your Forename?

2) What is your Surname?

3) Who is your Principal Investigator/Team leader (if applicable)?

4) What Flow Cytometer do you most commonly have experience with?

5) Which analysis software package(s) do you most commonly use to analyse Flow Cytometry data? If you use multiple software packages, please indicate why and what benefit each package brings.

6) How long have you been working with Flow Cytometry? Please indicate length of time in years and months.

7) Has your Flow Cytometry usage been consistent over time, or on and off with projects?

8) How often do you currently use a Flow Cytometer?

9) How often do you analyse Flow Cytometry data?

D.2.2. Gating Preferences

10) Do you prefer manual/automated gating?

11) What problems or frustrations do you encounter when gating flow data?

12) What problems or frustrations do you have when interpreting flow data from literature sources?

13) Do you actively read/research into Flow Cytometry gating techniques to stay aware of up to date literature and findings?

14) Please indicate what noise parameters you consider affecting Flow Cytometry gating and analysis and how you deal with these when analysing data.

Training			
Short description	Long description	Inclusion criteria	Exclusion criteria
Internal supervision	Internal training given by a supervisor or more senior team member within a team. This also includes directed reading for training and observation of other members of staff within the facility.	supervisor training or observation; studying internal protocols	external delegates providing in-house training, self-directed learning
Self-taught	Self-led learning about applying gates, not directed by a supervisor or internal protocols. This could be through the use of online material or literature.	Self-directed learning that includes videos, online material and FC literature	instructed learning or resources, internal supervision, external delegate training
External training course	External training provided by an instrument manufacturing or training company, delivered externally or internally within the company.	External delegates delivering training	internal supervision; self-led learning
None	No training delivered to participants because they are new to flow cytometry.	no training delivered	Internal supervision, external training or self-directed learning

Manual/Automated gating Preferences			
Short Description	**Long description**	**Inclusion criteria**	**Exclusion criteria**
User control	This means the platform gives the user more control over the analysis and specificity for targeting cell populations and using specific tools or functions to highlight certain features.	Familiarity with the workspace to	Experience of platform or use
Legacy	This means theplatform has always been used, and the operator is reluctant to change, try new methods or has not had the opportunity to do so within their analysis	I've always used this...'; stated experience with this platform	Useability of platform to further define features or populations
Biological Variation	In this instance, biological variation means how well each platform allows variation or adjustment due to biological or donor variability. Small adjustments often need to be made in these instances, which may be easier to adjust by hand if an automated system cannot detect changes.	Sample to sample variation';donor differences	Not differences due to the platform or staining panels
Accuracy	In this instance, accuracy refers to how well the desired cell population can be identified. It is not measurement accuracy, because this requires traceability to a primary unit.	Gating correctly'; 'identifies correct populations'	any mention of variability, consistency or reproducibility
Precision	In this instance, precision refers to how variable a platform can be in its analysis. This is defined using repeated measures, whether this participants, samples, donors or measurement repeats.	More consistent'; 'More repeatable';'Can deal with variation better'	Gating correctly'; 'identifies correct populations' as these are more accuracy terms
Less inter-analyst variation	This is a more specific precision term, relating specifically to differences between participants when completing Flow Cytometry analysis and the subjectivity input to measurements and interpretation.	Differences between people'; 'Discrepancy between users'	Setup variation between participants
Faster analysis	This relates to how quickly the analysis for the target cell population can be completed. With larger data-rich files it can take longer to analysefiles due to multiple parameters to consider.	Quicker to analyse'; 'XXX takes longer'	Does not include upstream preparation times or sample acquisition times

Problems gating			
Short Description	**Long description**	**Inclusion criteria**	**Exclusion criteria**
Cell separation	The ability to separate cell populations or clusters of cells within the file. This could be difficult due to how much overlap is present and what controls are present to help this split.	Discerning populations'; 'Hard to know where to gate'	Repeatability statements or discussion of variation of the cell source
Biological variation	In this instance, biological variation means how well each platform allows variation or adjustment due to biological or donor variability. Small adjustments often need to be made in these instances, which may be easier to adjust by hand if an automated system cannot detect changes.	Sample to sample variation';donor differences	Not differences due to the platform or staining panels
Staining quality	Staining quality relates to the quality of the cell staining process and how repeatably this was done, due to variation caused in measurement. This also relates to any variability caused by fluorescence or autofluorence of cells within the file.	Autofluorescence of cells; staining protocols discussed; spillover and compensation	Does not include voltages used to define population separations when setting up FC.
Time-consuming	Any reference to how long the gating process can take, using either manual or automated platforms to define a desirec cell population.	takes too long'; 'XXX is faster;	No reference to upstream processes and time variables associated with these.
Mistakes	Any manual mistakes that are made due to copy errors or dealing with high volumes of data or long analysis pipelines.	Mistakes'; 'Copy errors' ; 'Incorrect transfer'	Errors made uspteam of the gating process, unless they have knock-on effects.
Analyst variation	This is a more specific precision term, relating specifically to differences between participants when completing Flow Cytometry analysis and the subjectivity input to measurements and interpretation.	Differences between people'; 'Discrepancy between users'	Setup variation between participants
Plot visualisation quality	This refers to any way the data can be visualised and the resolution of the data points that are used for the analysis. This can include colour of plots, density mapping and aspet ratios, for example.	Colour of plots'; Ability to visualise with large layouts'; Discussion of colours, scales,axes clarity and aspecti ratios;	Nothing upstream of the gating that could influence this, such as voltages or stanining profiles

Noise parameters Codebook			
Short Description	**Long description**	**Inclusion criteria**	**Exclusion criteria**
Fluorescence spillover	This relates to any fluorescence spillover that can occur between optical channels in the FC. Any mention or correction or compensation to rectify this issue is included.	Compensation'; Fluorescence spillover'	No mention of the antigen itself of the biological conjugate, this is purely focused on the fluorophore.
Reagents	This relates to variability in the use of reagents upstream of the gating process. Antigen markers and fluoreochromes as well as staining buffers used can cause noise due to fluorescene spillover but also if the process of using this varies there can be differences in fluorescence measures.	Antibody lot variaitons'; Staining processes and repeatability'	Does not include the actual fluorescence spillover that occures due to the fluoreochromes, this has more concern with the impact on biological variation.
Voltage setup	This relates to setting the PMTs or PDs for the the acquision of data within each channel being used. The voltages allow suitable amplification of the fluorecence in order to see the required populations and discern from overalpping populations by using controls.	Setting up PMTs'; 'gain'; 'voltage setup'	Does not include the upstream staining and fluorochromes used, or spillover of the fluorochromes into different channels.
Non-specific binding	This relates to non-specific binding that occurs due to FC receptors wtthin immune cells and certain antigen markers that can bind to incorrect surface markers due to this.	Non-specific staining'; 'Fc blockers'; 'Non-specific binding	Does not inclue the fluorochromes themselves or anythign to do with Fluorescence spillover
Instrument to instrument variation	This relates to difference between different FC instruments made by different manufacturers or located at different sites.	Discussino of instrumental differences or specific manufacturer setting	Does not relate to laser lines used or voltages used for specific setup, nor reagents
S/N ratio	The Signal to Noise ratio relats to the limit of detection of the instrument and how this is defined with smaller cell types that are harder to detect doe to smaller size and potentially weaker signals, This can be difficult to separate from instrumental noise which could be over-amplified when looking for small or weak particles.	Limit of detection; Sensitivity; Signal to Noise; Small particle amplification	Does not include fluorescence particle signals or spillover, not specific voltages used for setup.
Cell Debris & Doublets	This relates to dead or dying cell populations and doublet cell populations that would normally be cut out of analysis in early 'clean up' gates applied/. These cells can often overlap required populations so their removal can be difficult when trying to fully etract the population.	Cell debris; Dead cells' Dying cells'; Doublets	Does not include control files or specific methods for defining gating boundaries.

Interpretation problems			
Short Description	**Long description**	**nclusion criteria**	**Exclusion criteria**
Gating standardisation	This includes the use of well defined and populat gating strategies,validated through the FC community, which also includes information provided on hwo the gating was completed, such as the hierarchy and process of gating steps to crate the gating images.This also includes difficulty to understand the analysis from descriptions given, along with understanding the conclusions drawn from the results in the literature.	Gating standards'; 'Ambiguous analysis'; 'Poor conclusions'; 'No gating hierarchy'	Does not include upstream standardisation such as reagent panels or voltage setup.
Cell population definition	This relates to the ability to understand how the target cell population is defined, but in attached descriptions. This also relates to the wider body of knowledge on how well certain sub-populations re determined and defined by specific markers and techniques.	Know what you are looking for'; Understanding of markers for populations	Does not include how markers bind to cells or gating standards used to define certain target populations.
Visualisation quality	This refers to any way the data can be visualised and the resolution of the data points that are used for the analysis. This can include colour of plots, density mapping and aspet ratios, for example.	Colour of plots'; Ability to visualise with	Nothing upstream of the gating that could
Poor reagent information	Lack of information surrounding reagents used, with respect to quantities and concentrations of each reagent, incubation times and titrations completed.	Lack of information on concentrations'; 'Lack of information on titrations used'	This does not relate to the cell number or event count within the file itself. Not does it relate to spillover with regards to the conjugated fluorophore.
Lack of controls	Lack of control files used to place the gates shown in the fully stained sample. These provide information on how the gate was placed, and how much separation can be seen between positive and negative populations.	FMOs'; 'Isotypes'; 'Live/dead stains'; 'Negative markers'	Does not include replicates of the sample itself or upstream optimisation such as titration.
Fluorescence spillover	This relates to any fluorescence spillover that can occur between optica channels in the FC. Any mention or correction or compensation to rectify this issue is included.	Compensation'; Fluorescence spillover'	No mention of the antigen itself of the biological conjugate, this is purely focused on the fluorophore.

☐ Current Proposal	Proposal #_____
Sponsor _____	
Title _____	
☐ Current Award	Account # _____
Sponsor _____	
Title _____	

B. Institutional Funding:

☐ Gift
☐ Departmental Resources
X☐ Other (explain) Visiting Student (Rebecca Grant) is funded by an Institute of Mechanical Engineers (IMechE) Research Scholarship
☐ No Funding

6. Statement of Financial Interest

Does the investigator, study personnel involved in the study or their Family have a financial interest in a company or other organization participating in or providing drugs, devices, biological agents, investigational medical devices, or any other tangible material or financial sponsorship for the research?

☐ Yes X☐ No

Does this study contemplate receiving/using any materials/data (data sets, confidential information) or making any purchases from or subawards to a company or other organizations in which you or a Family member hold a Financial Interest?

☐ Yes X☐ No

If yes was checked for any of the questions above, then attach a **Supplement for Disclosure of Financial Interest** for each individual with an interest. *This supplement, together with detailed guidance on this subject and definitions of the highlighted terms, is available in the COUHES site under Policies & Procedures in the Financial Conflicts of Interest section.*

7. Human Subjects Training. *All study personnel in research **MUST** take and pass a training course on human subjects research. MIT has a web-based course that can be accessed from the main menu of the COUHES web site. COUHES may accept proof of training from some other institutions. List the name, MIT or outside affiliation and emails of all study personnel and indicate if they have taken a human subjects training course.*

The personnel list is attached. All personnel will complete human subjects training and proof of completion of training will be submitted to COUHES prior to the start of the research.

8. Anticipated Dates of Research

Start Date: July 19, 2018	Completion Date: August 24, 2018

	Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects	**Application #** (assigned by COUHES)	
		Date	

APPLICATION FOR APPROVAL TO USE HUMANS AS EXPERIMENTAL SUBJECTS (EXEMPT STATUS FORM)

*Please answer every question. Positive answers should be amplified with details. You must mark N/A where the question does not pertain to your application. Any incomplete application will be **rejected and returned for completion.***

I. BASIC INFORMATION

1. Title of Study

A Horizon-Scanning Investigation to Explore and Identify Human Factors that Influence the Manufacture and Measurement of Cell and Gene Therapy Products

2. Investigator

Name: Anthony Sinskey	Building and Room #: 68-370A
Title: Professor	Email: asinskey@mit.edu
Department: Biology	Phone: **(office)** 617 253 6721; **(cell)** 781 354 3157

3. Faculty Sponsor. *If the investigator does not have PI Status (faculty, SRS or PRS) then a faculty sponsor must be identified and sign below.*

Name: same as above	Email: same as above
Title: same as above	Phone: same as above
Affiliation: same as above	

4. Collaborating Institutions. *If you are collaborating with another institution(s) then you must obtain approval from that institution's institutional review board, and forward copies of the approval to COUHES).*

The Visiting Student who will be working on this project is a PhD student at Loughborough University (Loughborough, Leicestershire, England). The project has been submitted to the Loughborough Human Participants Ethics Sub-Committee. Documents submitted to the committee are attached. A copy of their final approval will be forwarded to COUHES.

5. Funding. *If the research is funded by an outside sponsor, the investigator's department head must sign this form. Please enclose one copy of the research proposal (draft is acceptable) with your application. **Do not leave this section blank. If your project is not funded check No Funding.***

A. Sponsored Project Funding:

II. STUDY INFORMATION

1. Purpose of Study. *Please provide a brief statement of the background, nature and reasons for the proposed study. Use non-technical language.*

The purpose of this study is to identify specific human factors that can impact measurement of cell or gene therapy material. It has been previously that there is variation in Flow Cytometry results due to differences in how experienced operators isolate the required cell populations. Operator variation has an effect on many other areas of cell culture and therapy manufacture that can contribute to the final results. There is difficulty when reporting data, because different process parameters and gating strategies are used to analyse the results, which means findings are difficult to compare. The study hopes to identify these human factor issues that contribute to variability in measurement and manufacture of cell and gene therapy products across different technical platforms.

2. Study Protocol. *Please provide an outline of the proposed research. You should provide sufficient information for effective review by non-scientist members of COUHES. Define all abbreviations and use simple words. Unless justification is provided, this part of the application must not exceed 2 pages. Attaching sections of a grant application is* ***not*** *an acceptable substitute for the description requested here. Include copies of any questionnaire or standardized tests you plan to use. If your study involves interviews, submit an outline of the types of questions you will include. Your research outline should include a description of:*

A. Experimental procedures:

This study consists of interviews with practitioners from the cell and gene therapy manufacturing industry and supportive testing services who perform analytical measurements related to determination of characteristics important for the manufacture and measurement of cell and gene therapy products.

Participants will be asked to attend a 1 hour interview. The interview questions are aimed at determining the participant's responsibilities for performing the analytical measurements under discussion, their background and training in the field, and their use of a particular measurement or manufacturing platform. They will be asked to complete an exercise within this to label the various sources of variation they identify within this platform contributed by human factors. The interview will follow a 15 minute briefing, where the participant will be made aware of the aims of the study and their rights as participants, and will provide informed consent. Example of the types of questions which will be asked during the interview are submitted to COUHES in a separate document.

B. Study population, maximum number of subjects and age range:

- Participants will be relevant experts/users of specific measurement equipment used within cell and gene therapy manufacturing environments or at supporting testing facilities.

- A maximum of 20 subjects will be interviewed.
- Interview participants will be adults (≥ 18 years)

C. Subject Compensation: (describe all plans to pay subjects in cash or other forms of payment i.e. gift certificate).

N/A

D. Method of recruitment *(attach recruitment materials flyer, poster, email message, Internet posting, etc.)*

Participants in the study will be recruited through personal contacts that individuals at the Center for Biomedical Innovation have with members of the cell and gene therapy manufacturing community and the contract laboratories which support this field.

E. Length of subject involvement:

Approximately 1.25 hours. The initial ~15 minutes will be a briefing that outlines the aims of the study, provides information about the interview, and obtains the participant's informed consent. The remaining ~1 hour will consist of the subject matter interview.

F. Location of the research:

The interviews will take place at a location of the participant's choosing. This could be on-site, at their facility, at the MIT Center for Biomedical Innovation, or at an independent location.

G. Procedures for obtaining informed consent *(if you are requesting a waiver or alteration of informed consent, complete a Waiver or Alteration of Informed Consent Request form):*

A copy of the Informed Consent form is attached. This form will be emailed to the participant and ahead of time so that they have the opportunity to review and consider it prior to the scheduled interview. Immediately prior to the interview, the participant will receive a briefing which outlines the aims of the study and the information needed for them to provide informed consent.

H. Describe procedures to ensure confidentiality and explain in detail how research data will be secured:

Interview information will be maintained confidential within the Study Team. Participants will receive a unique Identification Number which will render them anonymous for data analysis and possible publication. All data will be anonymized when the interviews are transcribed from the audiotapes or from written notes. The anonymization codes will be stored separately from the study materials. All of the original data files will be stored securely in locked locations on MIT campus. Electronic data will be stored in secure workspaces on the MIT Secure server and the Loughborough University server.

3. HIPAA Privacy Rule. *If you are in any way working with individually identifiable health information for a research study that is sponsored by MIT Medical, an MIT Health Plan or another healthcare provider, then the Health Insurance Portability and Accountability Act ("HIPAA") likely applies to your study and you must comply with HIPAA in the conduct of your study. However, we expect that if you are applying for exempt status, you will only receive de-identified health information from participants in connection with your study. If you expect to receive identifiable health information from or about research participants in your study, you should complete the standard COUHES application form rather than this application form. You may consult with COUHES staff if you have questions about the exempt/non-exempt status of your proposed research study.*

N/A

Signature of Investigator Anthony J Sinskey **Date** 6.28.18

Signature of Faculty Sponsor Anthony J Sinskey **Date** 6.28.18

Signature of Department Head Alan Grossman **Date** 6.28.18

Print Full Name and Title Alan D. Grossman, Dept Head, Biology

The electronic file should be sent as an attachment to an e-mail: couhes@mit.edu. In addition, two single sided hard copies (one with original signatures) should be sent to the COUHES office: Building E25-Room 1438.

COUHES – EXEMPT APPLICATION– *ver.* 3/31/2017 –

- 5 -

ORGANISATION
Introductory stage • Basic information about the purpose of the interview and the research project of which it is a part • Clear idea of why they have been asked • Idea of probably length of the interview; and that I would like to record it and why • A clear idea of where and when the interview will take place *You are participating in a research project which aims to identify sources of variation within cell and gene therapy manufacturing & measurement platforms, attributed by human factors involved. This will have a specific focus on operator subjectivity within data analysis and processing, whilst also encompassing differences between operator experience and training.* *Your input helps us identify new areas for research and focus, to reduce variation within cell and gene therapy manufacturing processes by understanding where human factors play a key role in determination of final measures and decisions. This will ultimately improve the consistency and quality of the final product manufactured, and therefore a better quality of treatment for the recipient.* *You were selected as a possible participating this study because of your expertise and experience in carrying out analytical evaluations for the purpose of manufacturing cell or gene therapy products or in performing analytical testing in support of those manufacturers. This interview is voluntary, you have the right not to answer any question, and to stop the interview at any time or for any reason. We expect that the interview will take one hour.*
INTERVIEWEE PROFILE 1. What are your roles/responsibilities and what techniques do you work with most commonly? • Create a profile for interviewees career/technique history to aid this question.

2. How much experience have you had with cellular imaging for viability/confluency (in years & months)?

TECHNIQUE USE

3. How frequently do you use the cell imaging platform?

4. What do you use these techniques to measure?
- (Cell measurement, organic compounds, metabolites etc)

EXERCISE

5. Could you outline the process of obtaining a measurement with the respective technique, and identify the relevant tasks within the sample preparation and optimisation, instrument setup and data acquisition and analysis stages of the measurement process? (*Example provided, interviewee to talk through process to confirm what they do, along with the next question).*

6. From the process stages you have identified, could you highlight which of these are affected by the operator's judgement and their respective variability, and how much human contact time they take?

DATA ANALYSIS

7. Focusing on the data analysis elements, do they require human or machine learning interpretation to provide a quantitative result?

8. Are you familiar with how the software computes to obtain a final answer? What does it look for to determine differences between cells? Do different analysis packages use different methods/algorithms?

9. Could you go through this data analysis process in detail and explain variables that can be optimised/changed/altered to obtain the image (or analyse it) and how significant the variation could be? *(Yes/No/detailed answer)*

TRAINING

10. Are you responsible for training new users on this instrumentation?

11. Do you see differences between new and experienced users when they interpret data from this technique?

12. How are users trained in your facility (process)? Are there recognised training courses/material that are encouraged to use as reference material/information to help with quantification/diagnosis?

STANDARD OPERATION 13. Do you use Standard Operating Procedures for analysis of data? Is this internal/external? (Yes/No; Internal/External) 14. Do you use proprietary software for data processing (i.e. comes with the instrument), or do you use 3rd party software? If the latter, why?
CLOSURE • Give interviewee a verbal summary of what you feel you have gained from the interview and thank them for their time and contribution. • State time of interview end and interview duration.

Short Description	Detailed Description	Inclusion Criteria	Exclusion Criteria
Sample Preparation First Cycle Code	defined as liquid handling and sample handling required for an analytical measurement. This does not include any instrumental setup for measurement platforms or data processing. It also does not include ant cell culture required for expansion.		
Product interations	Refers to matrix effects of additional materials used within culture that can affect the final measurement, or effects of reagents used in the measurement process.	Matrix effects, mAb conjugation, buffers and reagents	Instrumental setup or protocol specifications.
Protocol Optimisation	Refers to optimised steps of the protocol, which contribute variability due to time, volume, concentration or other environmental factors.	Longitudinal effects, volume, concentration environment	Instrumental setup, product interactions and biological variation
Process Familiarity	Refers to operator human factors when repeating process steps. This includes factors such as experience, use frequency, competence, confidence and speed of work that impact how an operator completes a process and impacts variability	Experience, troubleshooting, use frequency, competence, confidence, speed of work	Protocol steps and detail, training
Copy Errors	Processing errors that are introduced when an operator is required to intervene and transfer information manually from one platform to another.	Copying values from instruments to excel docs, further manual calculation, file naming and overwriting	Does not include creating macros or automated file name generation/ information transfer

Haptic Variables	Variables affected by the operator completing a physical task or manual manipulation of the product.	Pipetting error, removal of superanatant	Does not include software based variables that require users to understand workings of parameters and adjustment thereof
Non-Haptic Variables	Variables affected by the operator which are not seen as physical in completion. These often involve software-based tasks where the operator is using a computer to select and test a variety of parameters.	Manual size gating of images	Pipetting error, removal of superanatant
In-Process First Cycle Code	In-process is defined as an instrumental setup required to facilitate the measurement. It does not include any prior sample handling, or post-processing of data if required. It related to setup parameters required to be tuned in order to take the measurement.		
Protocol Optimisation	These are parameters that have been set in a protocol, defined by prior testing and validation. It does not include parameters that need adaptive tuning for each measurand.	Cell line selection, concentrations, volumes, doubling times, cycle times, temperatures.	Variables tuned for each measurement, e.g. focus/ autofocus on cells.
Digital Image Parameters	These are variables which can be adjusted to improve the visual image of the product, often relating to camera variables. It does not include the spatial positioning of the camera or dynamic ranges thereof.	Brightness, contrast, masking, image stitching	Dynamic range of image, cell focusing, size thresholds
Spatial Image Parameters	These are area parameters reating to the chosen area for imaging in the x, y and z planes of focus and measurement. It also relates to the frequency of image capture and size thresholds if captured in real time.	Centering on well, size thresholds, z-scan height.	Brightness, Gain, Contrast.
Process Setup	These are human factor decisions and actions that are required to be made during the process of instrumental setup. They require the user to know the correct decisions have been made and troubleshooting if not. It does not include any digital or spatial parameters that can be adjusted.	Choosing correct optical channel, focusing, correct objective.	Brightness, gain, choosing area for image.

Copy Errors	Processing errors that are introduced when an operator is required to intervene and transfer information manually from one platform to another.	Copying values from instruments to excel docs, further manual calculation, file naming and overwriting	Does not include creating macros or automated file name generation/ information transfer
Post-Process Cycle Code	Post-processing is defined as any processing completed on the image/data once the measurement/ image has been taken. This includes any thresholding for background and cell/pixel size. It does not include any image setup parameters used to take the image.		
Copy Errors	Processing errors that are introduced when an operator is required to intervene and transfer information manually from one platform to another.	Copying values from instruments to excel docs, further manual calculation, file naming and overwriting	Does not include creating macros or automated file name generation/ information transfer
Spatial Image Parameters	These are area parameters reating to the chosen area for imaging in the x, y and z planes of focus and measurement. It also relates to the frequency of image capture and size thresholds if captured in real time.	Centering on well, size thresholds, z-scan height.	Brightness, Gain, Contrast.
Digital Image Parameters	These are variables which can be adjusted to improve the visual image of the product, often relating to camera variables. It does not include the spatial positioning of the camera or dynamic ranges thereof.	Brightness, contrast, masking, image stitching	Dynamic range of image, cell focusing, size thresholds
Non-Haptic Variables	Variables affected by the operator which are not seen as physical in completion. These often involve software-based tasks where the operator is using a computer to select and test a variety of parameters.	Manual size gating of images	Pipetting error, removal of superanatant
Protocol Optimisation	These are parameters that have been set in a protocol, defined by prior testing and validation. It does not include parameters that need adaptive tuning for each measurand.	Cell line selection, concentrations, volumes, doubling times, cycle times, temperatures.	Variables tuned for each measurement, e.g. focus/ autofocus on cells.