A model for harmonizing flow cytometry in clinical trials

Holden T Maecker, J Philip McCoy Jr & the FOCIS Human Immunophenotyping Consortium

Complexities in sample handling, instrument setup and data analysis are barriers to the effective use of flow cytometry to monitor immunological parameters in clinical trials. The novel use of a central laboratory may help mitigate these issues.

Flow cytometry has historically not constituted a large segment of the monitoring assays in clinical trials. The technology tends to be complex and expensive and often uses idiosyncratic methods for instrument setup and analysis. However, new technologies have evolved that can substantially enable the use of flow cytometry in clinical trial settings. These include lyophilized, preformatted, multiwell staining plates that decrease operator time and error[1]; multicolor analysis of many cell subsets simultaneously from a single stained cell sample[2]; automated instrument setup and compensation routines[3]; and batched analysis from templates that can include dynamic gates to allow for run-to-run staining differences[4].

Together with those technological advancements, there is growing interest in collecting more information from preclinical and early phase clinical trials for better prediction of the performance of candidate drugs and identification of risks before an expensive, late-stage clinical trial[5]. As a platform for monitoring immune function and immunotoxicity, flow cytometry is extremely powerful. In fact, it is arguably the most powerful single-cell analysis technology available at present. It is therefore not surprising that there is growing interest in the use of flow cytometry as a tool for monitoring clinical trials.

Despite the technological advancements listed above, there are still many factors that impede the widespread use of flow cytometry in clinical trials. These can be roughly catego-

A full list of authors and affiliations appears at the end.
e-mail: maecker@stanford.edu

rized as issues of sample handling, instrument setup and data analysis (**Fig. 1**). Here we will briefly review each of these areas and then propose a model that could in many cases minimize the effect of these variables in multicenter clinical trials.

Under the heading of 'sample handling' are many questions about how to process and ship blood samples so their viability and function are not unduly compromised while still following a workflow that can be used by all laboratories involved in the study. This is perhaps the most vexing area, as invariably compromises must be made between the preservation of sample integrity and maintenance of a practical workflow.

Sample handling issues

Sample handling begins with blood collection, the timing of which may introduce circadian variation, as has been demonstrated for the frequency[6] and functions[7,8] of various lymphocyte subsets and for serum cytokine production[9]. The type of anticoagulant used for blood collection can also influence the phenotype and function of peripheral blood mononuclear cells[10–12]. The length of time from blood draw to sample processing can be crucial for the counting of certain cell types[13], staining of labile markers[10] and preservation of function[11]. Density-gradient centrifugation, frequently used for the isolation of peripheral blood mononuclear cells, can also result in differences in staining patterns and subset distributions and function relative to those obtained by the analysis of whole blood[14–16]. One approach that allows sample batching and deferred analysis is cry-

opreservation, but this can introduce additional changes in labile phenotypic markers and subset distributions[17–19].

Functional assays, such as intracellular cytokine staining, analysis of proliferation or flow cytometry to assess phosphorylated epitopes, incur additional variation related to the method of *in vitro* stimulation of cells. The stimulation media, source and lot of stimulation reagent, titer, stimulation time and type of stimulation vessel can all influence the degree of activation or proliferation seen in such assays[1].

The actual staining of samples can introduce variability; for example, whether whole blood samples are stained before or after erythrocyte lysis, the time and temperature of staining, the fluorochrome conjugate, titer and even the lot of staining reagent used can all affect the readout. The increasing use of intracellular staining protocols for the examination of cytokines or phosphorylated signaling molecules introduces additional variables; for example, the fixation and permeabilization system used and whether cell-surface markers are stained before or after fixation and permeabilization can have an effect. Many cell surface epitopes are mostly destroyed by fixation and permeabilization, at least with the harsher fixation-permeabilization schemes needed to detect certain intracellular epitopes. Finally, the use of tandem dyes (conjugates of two fluorochromes that create a greater shift in emission wavelength than do single dyes) must be considered. Some of these tandem dyes (such as phycoerythrin-indotricarbocyanine and allophycocyanin-indotricarbocyanine) are particularly labile in the presence of light, fixation and higher
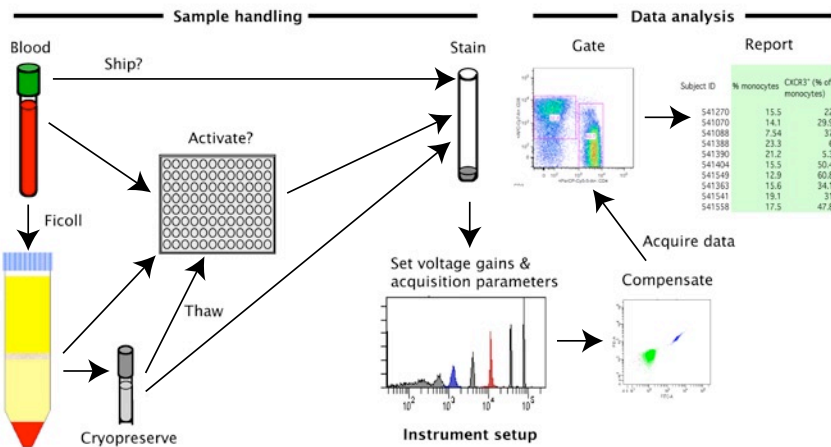
Figure 1 Flow chart of decisions that can affect the quality of data produced in clinical trials in which immunological parameters are monitored by flow cytometry.

temperatures[20]. They can also have greater lot-to-lot variability in optical spillover properties than do single dyes. Thus, reproducible and well-controlled sample handling, as well as consistency of reagent lots, becomes critical.

Instrument setup and data analysis issues

In the area of instrument setup, much standardization can be achieved with the software packages available with newer digital cytometers. However, this instrumentation and software is not yet widely used in clinical research organizations or other clinical trial–associated laboratories. In addition, these systems do not continuously track performance but instead assume that the cytometer does not change over the course of a day. If sufficient warm-up time has been given (up to 2 hours for some types of lasers), this assumption might hold true. However, analyzing a control bead population before and after each experiment is also advisable for the detection of any performance changes that might occur over the course of a run. Finally, instrument setup software does not necessarily address standardization across different cytometers, particularly if those cytometers vary in their configuration. For example, cytometers equipped with green lasers (usually 532-nm emission) have better sensitivity for phycoerythrin itself and tandem phycoerythrin dyes[21] than do those that use blue (usually 488-nm) lasers. In such cases, there will invariably be performance differences that cannot be overcome.

Analysis of multicolor flow cytometry necessarily involves compensation for optical spillover between detectors[22]. Fortunately, automated algorithms are now available with most acquisition and analysis software that calculate compensation from a set of single-color controls. Additionally, the use of software-based compensation on newer digital instruments allows adjustment of compensation, if necessary, even after sample acquisition. However, variability and potential inaccuracy can still be introduced into the process via the following parameters: the type of single-color controls chosen (such as beads or cells), the antibody used to stain each control, the handling of those controls relative to the handling of experimental samples and the choice of a negative population associated with each compensated parameter. Normally compensation should not have to be adjusted after it has been computed by the software, but depending on the variables outlined above, some corrections may occasionally be necessary and this then becomes a subjective process and a source of variability.

Of course, the degree of optical spillover in a particular experiment is dependent on the choice of fluorochromes and antibodies used[20]. Suboptimal panel design will negatively affect the quality of data because of the use of fluorochromes that are too dim for particular markers and/or that have excessive optical spillover. In general, efforts should be made to standardize reagent panels for particular purposes so data are comparable and development time is minimized. However, because research questions constantly change, there is always pressure to redesign existing panels, and the addition of even one more reagent often requires extensive rearrangement of fluorochrome-antibody combinations so acceptable performance is maintained. This is especially true as the number of fluorochromes in the experiment increases.

In addition to variability in compensation controls, there can be variability in the choice of gating controls used to determine positive-negative boundaries in the data[23]. Fluorescence-minus-one controls[22] include all the experimental staining reagents except one and can be useful for setting gates when staining is dim or smeared. However, these controls do not take into account background staining of the reagent that has been left out. This can be estimated by substitution of a non-staining antibody of the same isotype as the experimental reagent (isotype-matched control antibody), but the amount of background may still not be accurately assessed because of differences in concentration, the fluorochrome/protein ratio and inherent nonspecific binding. Also, it is still necessary to use isotype-matched control antibodies in the context of the other staining reagents to account for optical spillover between reagents.

Another useful type of control, the so-called 'process control', can be added to verify the performance of certain steps in the assay. For example, prestained lyophilized cells can be used to verify instrument setup and gating independently of sample handling and staining. Alternately or additionally, serial aliquots of a single cryopreserved sample may be thawed for each assay and stained to simultaneously verify the performance of that day's staining, instrument setup and gating.

In the area of data analysis, there have been advances in gating tools and batch-analysis options. However, the analysis software now available does not allow efficient archiving and retrieval of large amounts of data or analysis across multiple experiments. The tools available are still highly focused on experiment-specific analysis and are generally insufficient to achieve the ultimate goal of reliable, single-step transformation of raw data into quantified results for large numbers of files.

Perhaps the largest single contributor to variability in flow cytometry is differences in gating. In one example of this, as part of a multisite standardization study[1], prestained lyophilized cells were distributed to 15 experienced laboratories and researchers were asked to acquire the samples and then analyze the data, and to also send the raw data files to a single laboratory for central analysis. The data from individual laboratory analyses showed a mean coefficient of variation of 20.5% across four samples, whereas the data from central analysis showed a mean coefficient of variation of 4%. This means that instrument setup and statistical counting errors accounted for only a very minor proportion of the variability, whereas individualized gating methods accounted for the vast majority of the inter-laboratory variation.

In the study described above, the inclusion of 'dim' populations for key markers such as

CD4 and CD8 accounted for most of the gating variability noted. When populations are tightly clustered and easily discriminated from each other, such variability will of course be less. This means that a certain amount of gating variability can be avoided by optimal design of reagent panels. However, the remaining variability needs to be handled through the use of either a shared gating template or central analysis by a single operator. The shared template can still suffer from problems, as some adjustment of gates may be required between donors and between experiments, so there will still be a degree of operator bias. This can be minimized in some cases by the use of dynamic gates (available in some analysis software) that adjust to shifting data[4]. However, such gates need to be rigorously tested and their settings must be optimized to ensure the desired results, and they might not be feasible for use in some situations.

Most flow cytometry data are reported as the percentage of cells positive for a particular marker or set of markers, with the denominator of the percentage being a chief subset of interest, such as $CD4^+$ or $CD8^+$ T cells, B cells and so on. Because the numbers of these subsets can vary, particularly in certain conditions such as infection with human immunodeficiency virus, it is sometimes desirable to convert percentages to absolute counts per microliter of blood (or per milliliter, for rarer subsets). This is straightforward if an absolute counting test for the subset of interest is done concomitantly with the blood draw for which immunophenotyping is done. However, such conversion is not routinely done in the vast majority of clinical immunomonitoring studies.

In cases in which a cell population displays a continuous distribution of staining intensity, rather than discrete positive and negative populations, it can be more appropriate to report the median fluorescence intensity of the entire cell population. However, differences in staining and instrument setup from experiment to experiment warrant the use of some type of standard to ensure the reproducibility of this approach. This could involve simply calculating the ratio of the median fluorescence for the experimental sample to that of a sample stained with isotype-matched control antibody. Alternately, so-called 'quantitation beads', which contain a known number of fluorochrome molecules per bead, can be used as a reference for converting raw fluorescence units to fluorochrome molecules per cell[24]. If a 1:1 conjugate of antibody/fluorochrome is used for staining, these numbers are identical to the antibodies bound per cell.

Beyond simply reporting the results of a flow cytometry experiment, there are efforts
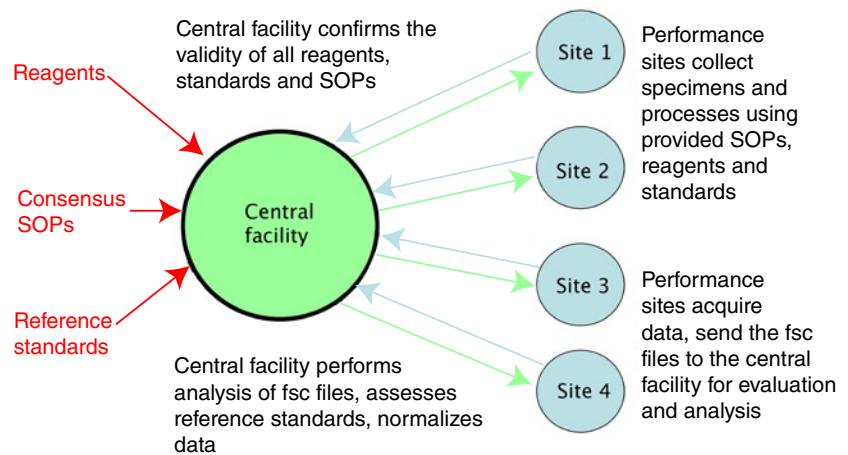
Figure 2 Proposed mixed model of the operation of a multisite clinical trial.

under way to encourage more complete and consistent reporting of the methodology used to achieve that result. For example, the MIFlowCyt (minimal information about a flow cytometry experiment) standard[25] has been approved by the International Society for Advancement of Cytometry for the reporting of any flow cytometry results. Basically, the standard specifies information that should be supplied with any experiment under the following headings: Experiment Overview, Flow Sample/Specimen Details, Instrument Details and Analysis Details.

For particular classes of experiments, other standards are being developed. For example, the MIATA (minimal information about T cell assays) approach[26] aims to set standards about the reporting of tetramer and intracellular cytokine staining and other related T cell assays. Obviously, consistent adherence to such standards would increase the transparency of published data, making the data easier to interpret and reproduce.

Strategies to minimize these issues

For inter-institutional or cooperative studies, or even separate studies attempting to produce comparable data, the issues described above present barriers to the generation of accurate and precise data with the least variation among different sites and studies. This represents a huge obstacle in clinical research, as data from one study or institution may have little meaning in the context of data gathered separately without control of many variables. Several options can be considered to address the issue of standardization of data collection, each having associated drawbacks and benefits. In a broad context these can be grouped into three models: the remote model, the central model and the mixed model.

In the remote model, inter-institutional studies are undertaken in every institution, each operating under (hopefully) standard or 'harmonized' protocols determined before the start of the study. The key advantages of doing cytometry at remote sites is that these sites are more proximal to the patients; thus, issues of sample handling before staining are minimized and the potential of obtaining data on labile cells or markers is maintained. Disadvantages of this model include variations in protocols (and protocol adherence) among the participating institutions. Subjective nuances can be introduced even if there is an attempt to follow an identical protocol at different sites. Clearly, this model has little chance of minimizing variability among sites unless strict standardization procedures are implemented.

In the central model, all remote samples are sent to a central facility for processing and analysis. Although this can more easily ensure standardization of process, instruments and analysis, it does introduce the vagaries associated with the necessity of shipping samples. There is also the introduction of time as a factor in the evaluation of which markers can be analyzed and the subsequent interpretation of the results obtained. Essentially, this model acts as a clinical reference laboratory would.

The mixed model (**Fig. 2**) would blend the desired aspects of the remote and central models into one to minimize variation as much as possible while still allowing each institution separate 'ownership' of the respective laboratories. In this model, samples would be obtained, processed and acquired at local sites through the use of strict standard operating procedures (SOPs) to expedite sample handling and processing. The central laboratory would 'harmonize' the remote sites by confirming SOPs were used, training researchers and so forth.

In this scenario, care must be taken to ensure that instruments are standardized at the various sites so each is able to detect the anticipated staining in a universally consistent way. Major clinical reference laboratories with multiple locations have pioneered this approach, and similar procedures could be implemented in academic institutions. Furthermore, technical staff should be centrally trained to ensure that all procedures are done in a uniform manner and SOPs are strictly followed. Ideally, in this scenario, all sites would use identical reagents, including the same lots of reagents tested and distributed to each site. If possible, lyophilization of reagents, which prevents any alteration in reagents during shipping, would be used. All reagents and SOPs would be tested by the central laboratory and then distributed to remote laboratories. Additionally, quality-assurance samples could be shipped to all laboratories, similar to the proficiency testing among laboratories certified by the Clinical Laboratory Improvement Amendments, to ensure consistency among the sites.

One issue discussed before is the variability that arises when flow cytometry data are analyzed by different researchers not working under strict guidelines[1]. Variability due to subjective gating and how positive versus negative events are delineated can result in considerable deviations. At a minimum, templates for acquisition and data analysis would be distributed among the remote sites. Perhaps centralization of data analysis would be even more desirable. This could be accomplished easily if acquisition of data at remote sites were followed by secure, electronic transmission of the data files to the central laboratory for analysis.

Thus, in this mixed model, all sample procurement, staining, and acquisition of flow cytometry data would be conducted at local or institutional laboratories. This would be done under the aegis of the central facility through the use of strict SOPs and training procedures. The central laboratory's role would include training, SOP development, titration, validation and distribution of reagent lots to remote sites, proficiency testing and instrument standardization at remote sites, and centralized analysis of the flow cytometry standard files generated remotely.

Clearly, the operation of a central laboratory in the context of the mixed model would add up-front costs to inter-institutional group studies. It is possible that some of these costs could be offset by centralized reagent procurement and validation, as well as centralized data analysis. Additionally, cost savings (relative to the costs of the central model) might be gained from the lack of a need to ship specimens by express courier. Furthermore, once the initial investment is made, the cost per study may not be much different for the mixed model versus the central model.

Concluding remarks

Regardless of the model used for the monitoring of multicenter clinical trials, the application of complex (and powerful) flow cytometry assays in this setting needs to be carefully planned. Investments in training and personnel, as well as use of the appropriate hardware and software tools, are necessary to ensure the production of consistent and accurate data in a study and ideally to create data that can be compared by meta-analyses of the same assays across studies. Building an infrastructure that can support the generation of such data is a tremendous challenge for the immunology community. However, this is a challenge that must be met if immunologists are to realize the potential of human translational immunology.

1. Maecker, H.T. *et al. BMC Immunol.* **6**, 13 (2005).
2. Perfetto, S.P., Chattopadhyay, P.K. & Roederer, M. *Nat. Rev. Immunol.* **4**, 648–655 (2004).
3. Maecker, H.T. *Methods Mol. Biol.* **485**, 375–391 (2009).
4. Suni, M.A. *et al. BMC Immunol.* **4**, 9 (2003).
5. Davis, M.M. *Immunity* **29**, 835–838 (2008).
6. Levi, F.A., Canon, C., Touitou, Y., Reinberg, A. & Mathe, G. *J. Clin. Invest.* **81**, 407–413 (1988).
7. Arjona, A. & Sarkar, D.K. *Brain Behav. Immun.* **20**, 469–476 (2006).
8. Bollinger, T. *et al. Clin. Exp. Immunol.* **155**, 231–238 (2009).
9. Dimitrov, S., Lange, T., Tieken, S., Fehm, H.L. & Born, J. *Brain Behav. Immun.* **18**, 341–348 (2004).
10. Shalekoff, S., Page-Shipp, L. & Tiemessen, C.T. *Clin. Diagn. Lab. Immunol.* **5**, 695–702 (1998).
11. Weinberg, A., Betensky, R.A., Zhang, L. & Ray, G. *Clin. Diagn. Lab. Immunol.* **5**, 804–807 (1998).
12. Kumar, P. & Satchidanandam, V. *Clin. Diagn. Lab. Immunol.* **7**, 578–583 (2000).
13. Hodge, G.L., Flower, R. & Han, P. *J. Immunol. Methods* **225**, 27–38 (1999).
14. Renzi, P. & Ginns, L.C. *J. Immunol. Methods* **98**, 53–56 (1987).
15. Appay, V. *et al. J. Immunol. Methods* **309**, 192–199 (2006).
16. Hoffmeister, B., Bunde, T., Rudawsky, I.M., Volk, H.D. & Kern, F. *Eur. J. Immunol.* **33**, 3484–3492 (2003).
17. Romeu, M.A. *et al. J. Immunol. Methods* **154**, 7–10 (1992).
18. Seale, A.C. *et al. J. Clin. Lab. Anal.* **22**, 153–158 (2008).
19. Weinberg, A. *et al. Clin Vaccine Immunol* **16**, 1176–1186 (2009).
20. Maecker, H.T., Frey, T., Nomura, L.E. & Trotter, J. *Cytometry A* **62**, 169–173 (2004).
21. Perfetto, S.P. & Roederer, M. *Cytometry A* **71**, 73–79 (2007).
22. Roederer, M. *Cytometry* **45**, 194–205 (2001).
23. Maecker, H.T. & Trotter, J. *Cytometry A* **69**, 1037–1042 (2006).
24. Wang, L., Gaigalas, A.K., Marti, G., Abbasi, F. & Hoffman, R.A. *Cytometry A* **73**, 279–288 (2008).
25. Lee, J.A. *et al. Cytometry A* **73**, 926–930 (2008).
26. Janetzki, S. *et al. Immunity* **31**, 527–528 (2009).

Holden T. Maecker is with the Institute for Immunity, Transplantation and Infection, Stanford University, Stanford, California, USA. J. Philip McCoy Jr. is with the National Heart, Lung, and Blood Institute, US National Institutes of Health, Bethesda, Maryland, USA. The FOCIS (Federation of Clinical Immunology Societies) Human Immunophenotyping Consortium includes the following: Michael Amos, John Elliott, Adolfas Gaigalas and Lili Wang are with the National Institute of Standards and Technology, Gaithersburg, Maryland, USA. Richard Aranda is with Bristol Myers-Squibb, Princeton, New Jersey, USA. Jacques Banchereau is with Baylor Institute for Immunology Research, Dallas, Texas, USA. Chris Boshoff is with the Foundation for the National Institutes of Health, Bethesda, Maryland, USA. Jonathan Braun, Yael Korin and Elaine Reed are at the University of California, Los Angeles, California, USA. Judy Cho and David Hafler are at Yale University, New Haven, Connecticut, USA. Mark Davis, C. Garrison Fathman and William Robinson are at Stanford University, Stanford, California, USA. Thomas Denny and Kent Weinhold are at Duke University, Durham, North Carolina, USA. Bela Desai is with Schering Plough, Palo Alto, California, USA. Betty Diamond and Peter Gregersen are with the Feinstein Institute for Medical Research, Manhasset, New York, USA. Paola DiMeglio, Frank Nestle, Mark Peakman and Federica Villnova are at King's College, London, UK. John Ferbas is with Amgen, Thousand Oaks, California, USA. Elizabeth Field is at the University of Iowa, Iowa City, Iowa, USA. Aaron Kantor is with the Immune Tolerance Institute, Menlo Park, California, USA. Thomas Kawabata is with Pfizer, Groton, Connecticut, USA. Wendy Komocsar is with Eli Lilly, Indianapolis, Indiana, USA. Michael Lotze is at the University of Pittsburgh, Pittsburgh, Pennsylvania, USA. Jerry Nepom is with the Benaroya Research Institute, Seattle, Washington, USA. Hans Ochs is at the University of Washington, Seattle, Washington, USA. Raegan O'Lone is with the International Life Sciences Health and Environmental Sciences Institute, Washington, DC, USA. Deborah Phippard is with the Immune Tolerance Network, Bethesda, Maryland, USA. Scott Plevy is at the University of North Carolina, Chapel Hill, North Carolina, USA. Stephen Rich is at the University of Virginia, Charlottesville, Virginia, USA. Mario Roederer and Dan Rotrosen are with the National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, USA. Jung-Hua Yeh is with Genentech, South San Francisco, California, USA.