Interpreting flow cytometry data: a guide for the perplexed

Leonore A Herzenberg, James Tung, Wayne A Moore, Leonard A Herzenberg & David R Parks

Recent advances in flow cytometry technologies are changing how researchers collect, look at and present their data.

Recent advances in fluorescence-activated

subset that is visible in the 'logicle' displays (centered near zero on both axes) seems to be missing from the logarithmic displays. However, as indicated above, this subset is present but is represented mainly by data points and contours that are 'piled up' on the plot axes.

The location of the median fluorescence value in each dimension (**Fig. 1**, dark red crosses) further demonstrates the problem with logarithmic data visualization. By definition, half of the data values in each rectangular region are greater than the median and half are less than that value. However, whereas the locations of the median values in the 'logicle' displays (**Fig. 1**, right) correspond to the visual centers of the subsets, the locations of the median values in the logarithmic displays (**Fig. 1**, left) are substantially offset from the apparent peak of the subset. In fact, each of the subsets in the logarithmic display is broken up into a 'false peak' above the median value, a sparsely populated region between this peak and the baseline, and a 'pileup' of a large fraction of the cells on the baseline.

This artificial subdivision is also visible when data are plotted as a one-dimensional histogram on a logarithmic scale (**Fig. 1**, bottom left). The sharp rise at the lowest value on the scale reflects the 'pileup' of the minimally fluorescent cells at the lowest value on the scale. This 'peak' tends to be overlooked because it is nearly coincident with the axis. However, it represents roughly 40% of the minimally fluorescent population and 20% of the cells in the overall population. The 'valley' and the 'false peak' are also in the histogram, creating a total of three peaks rather than the two that actually exist. This problem does not occur in histograms plotted on 'logicle' axes (**Fig. 1**, bottom right), in which the minimally fluorescent cells form a peak centered at or near zero and extending symmetrically above and below the peak center, thereby representing the cells whose measured fluorescence values are below zero.

In two-dimensional logarithmic displays, cells whose fluorescence measurements are at or below zero in one or both dimensions are 'piled up' on the axes and are represented by contours along both axes in contour displays (**Fig. 1**, middle left) and by colored dots, visible with keen eyes, on the axes in color density

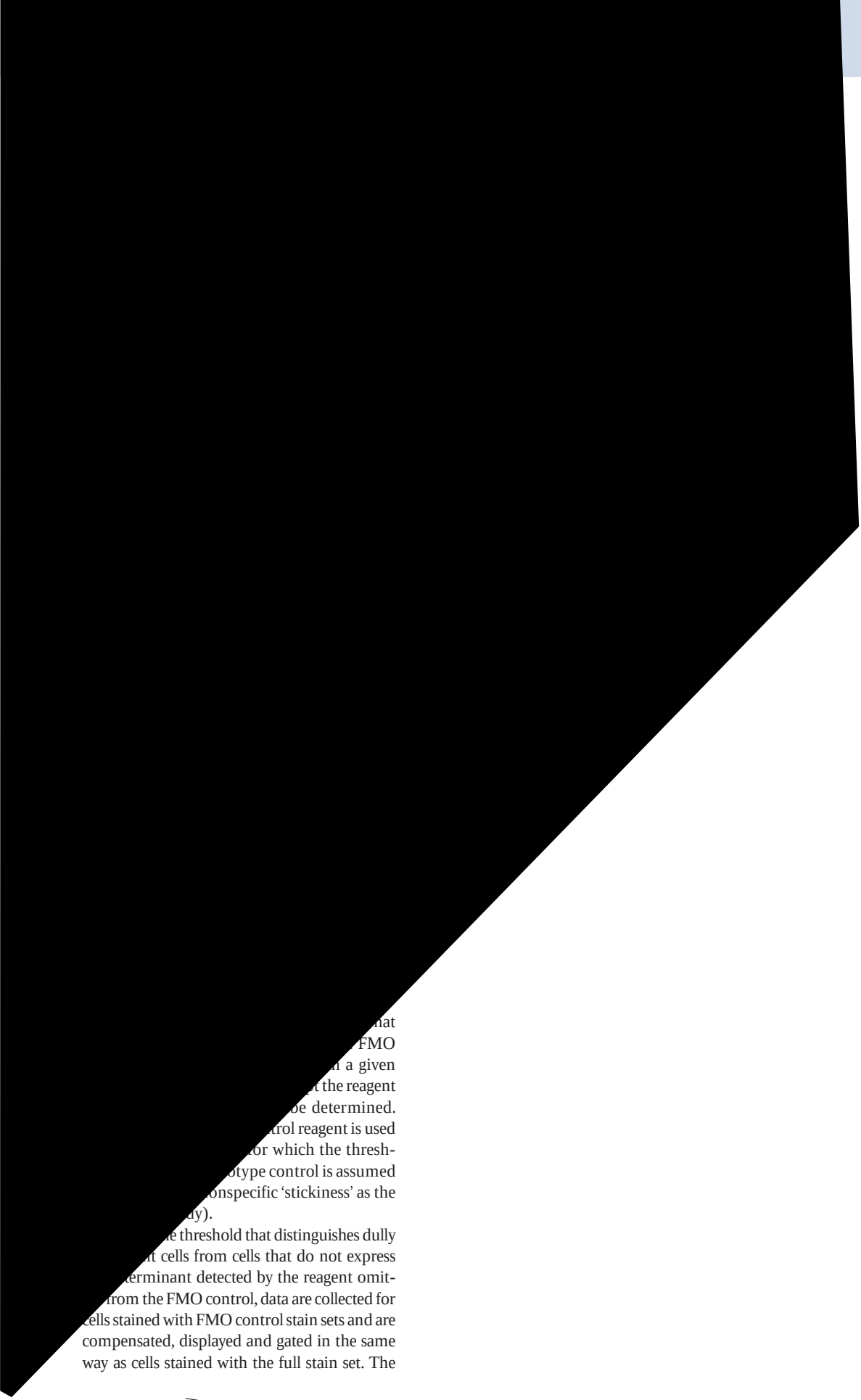
stantially by the effects of spectral overlap between dyes when two or more dyes are used to stain a cell suspension. During FACS data collection, separate detectors are dedicated to each fluorochrome and measure most light from the target fluorochrome. Some of the light from this fluorochrome, however, may be emitted and detected in the wavelength range assigned to a different fluorochrome.

Fluorescence compensation is the process by which the amounts of spectral overlap are estimated and subtracted from the total detected signals to yield an estimate of the actual amount of each dye. Single-laser, two-color FACS experiments in the mid-1970s showed that fluorescence compensation is important[6]. It is now recognized as an obligate step in most FACS analyses.

For measurement of the spectral overlaps, the fluorescence detected on all measurement channels is evaluated for 'compensation control' samples, which have each been labeled with only one of the fluorochromes used in the composite stain set. These measurements are used to make the fluorescence compensation corrections. That is, they are used to estimate and subtract the contribution of the spectral overlap to the overall signal that each detector measures for cells stained with the composite stain set.

When a population has little or no detectable cell-associated fluorescence for a particular dye, the subtraction of overlapping fluorescence by the fluorescence compensation process results in a population distributed symmetrically around zero (or cell autofluorescence) in the corresponding data channel. Statistical uncertainties inherent in primary and overlapping fluorescence measurements will determine the width of this distribution. As these uncertainties become more pronounced as the size of the correction increases, larger corrections will result in broader distributions that nonetheless will still be centered on the average autofluorescence of the cells (**Fig. 2**, bottom right, vertical dimension).

/.0198p.6(r)pe3.-0.0Tc)6(oc)1e.0073 10.1((4 1 no)15mu((4 1 no)-2(cte(4 19By)29.8(t.3r)lw0 tttthe fluteT



nat
FMO
a given
t the reagent
be determined.
trol reagent is used
or which the thresh-
otype control is assumed
onspecific 'stickiness' as the
ty).
e threshold that distinguishes dully
t cells from cells that do not express
terminant detected by the reagent omit-
from the FMO control, data are collected for
cells stained with FMO control stain sets and are
compensated, displayed and gated in the same
way as cells stained with the full stain set. The

Once a cell triggers the placement of a dot on the screen, hundreds or thousands of cells measured at the same fluorescence will effectively be invisible because they will all be represented on the screen by the initial dot. In addition, if enough cells are analyzed, sparse regions will become populated with enough dots at a density of one cell per dot to form a contiguous (black) region indistinguishable from and often merged with highly populated black regions saturated with thousands of dots and potentially thousands of signals per dot (**Fig. 3**).

Curiously, although this problem with 'live' dots plots is fairly well recognized, many investigators have adopted computer-generated dot plots as their preferred method for viewing and publishing data. Some of these investigators believe (erroneously) that dot plots provide the most accurate representation of the subsets that are present. However, quantile contour plots (sometimes called probability plots) and color density plots actually provide a much more accurate representation of the data. By representing the frequencies of cells present at each point in the plot, these plots avoid the dynamic range problems encountered with monochrome dot plots. Thus, they become more rather than less accurate as the number of cells for which data are collected increases (**Fig. 3**). Furthermore, because they facilitate the discrimination of densely and sparsely populated regions, these contour and color density plots enable precise