

ABSTRACT

Preclinical research in Oncology requires using a wide variety of animal models. Robust, clinically relevant distribution methods including randomized and non-randomized methods appropriate to the model are essential to ensure unbiased group assignments and reliable experimental outcomes. Our thorough analysis of over 10,000 publications published in the last 5 years revealed that while ~80% publications randomize animals into treatment groups, <1% use block randomization or stratified randomization procedures where two parameters (such as tumor volume-body weight) were involved. While these randomization strategies have been well described in Biostatistics literature for decades, their adaptations to research in oncology has been extremely slow, at least in part because of lack of tools to perform these procedures. Significant advancements in computational biology are yet to percolate into the scientific methods used in Vivarium. These inconveniences are easily overcome using an animal study workflow software. Strategies for randomizing individual animals into groups and randomizing mouse cages into groups can be implemented with exclusion criteria, multi-parameter randomization, re-randomization etc. and additional features of workflow software such as improved process integrity, research quality, robust documentation and reproducibility come with it. These strategies are highly relevant in the context of much needed improvements in dismally low predictability of clinical outcomes.

BACKGROUND and RATIONALE

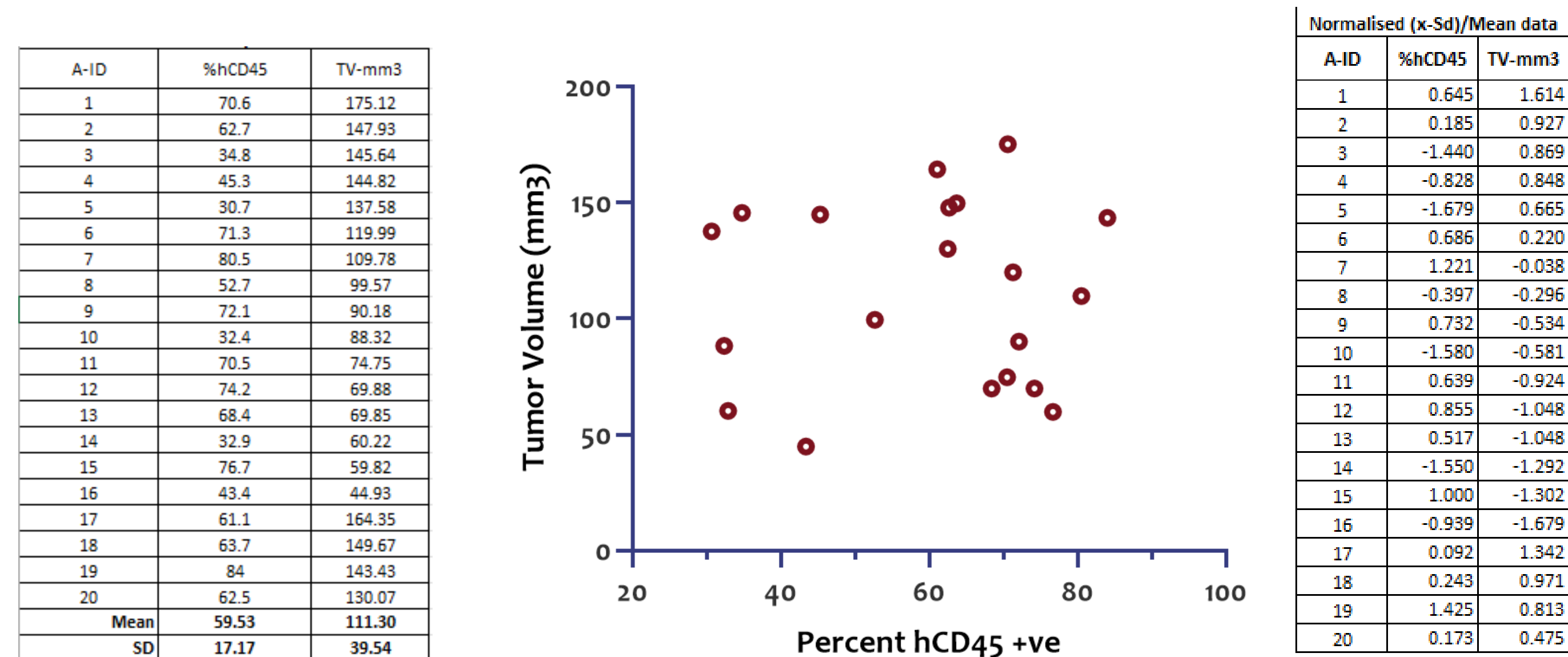
In animal studies where two or more characteristics are being investigated (for example, Tumor volume and % human CD45 cells in the circulation of humanized mice, circulating TGF-beta & tumor volume, bw & tv etc.), simple randomization (or worse, lack of it) is not ideal. Block randomization, or stratified randomization and covariate adaptive randomization may need to be used.

Our recent analysis of 5,809 scientific publications (Published between 2019 and 2024 [total of 66,937], full text freely available on PubMed [47,386], and associated animal data available [5,830] and not retracted[5,809]), showed that less than one percent of such studies used block/stratified randomization strategies

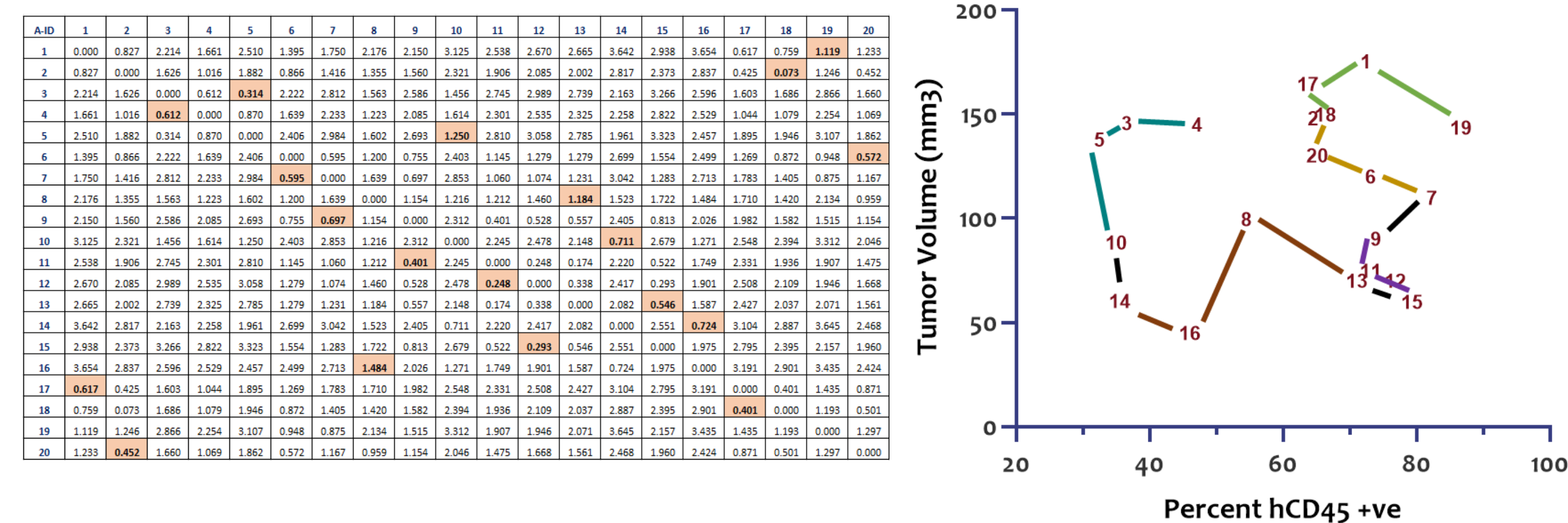
Hence, we sought to develop a simple-to-use block randomization method to enhance usability of relevant statistical methods in animal research

METHODS AND RESULTS: Approach 1

Visualization of Tumor volume Vs % hCD45 and Normalization



Calculation of path length, Euclidian distance



Euclidean distance $D = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]}$, where (x_1, y_1) and (x_2, y_2) are the Cartesian coordinates of the two points

Hierarchical clustering of the data

Similarity matrix between all data points and combining the two nearest sampling points using a "bottom-up" or "Agglomerative" approach. This is a "bottom-up" strategy in which each observation begins in its own cluster, and pairs of clusters are merged as one progresses up the hierarchy. ("hclust" in R)

Optimize the order (leaf order structure)

The optimal leaf ordering maximizes the similarity of adjacent objects in ordering. This principle is well-expressed in the traveling salesman problem (TSP, the salesman visits all cities without returning and travels for the shortest distance possible). Note the sequence, and the points fall on the order to be utilized for allocating the points (animals) for the block.

Animal allocation to blocks

The block is constructed based on the number of animals per group. To build five groups of three animals, we will need three blocks with five animals each. Based on the sequence obtained in Step 3, assign the animals (bottom-up) to each block. So that each block would be assigned similar animals with both variables.

Block#	Animal-ID			
Block 1	4	3	5	10
Block 2	14	16	8	13
Block 3	15	12	11	9
Block 4	7	6	20	2
Block 5	18	17	1	19

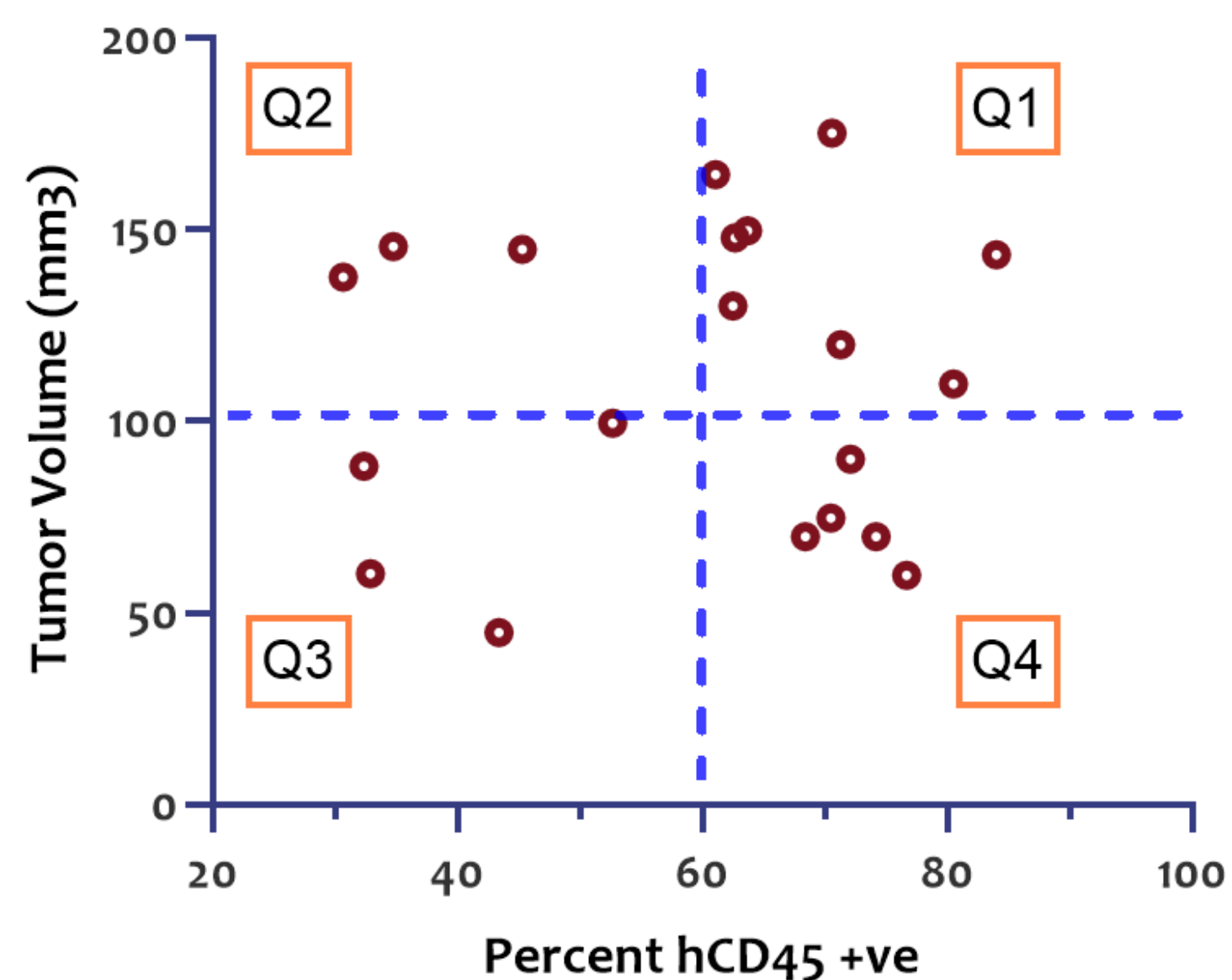
Block Randomization

Intended to randomize animals into groups with equivalent sample sizes and maintain a balanced sample size across groups. Among 13824 possible combinations, one example is shown in the table below

Group 1	Group 2	Group 3	Group 4
A	B	C	D
3	10	1	2
7	11	5	4
12	16	6	8
13	18	14	9
17	20	15	19

Relevance of block randomization

The rationale is for each block to consist animals of comparable characteristics and each treatment group should have representation from each block. In simple terms, each group should have animals with high AND low tumor volumes & high AND low CD45 +ve cells. This is required to get a readout of efficacy of the treatment on animals with diverse characteristics



METHODS AND RESULTS: Approach 2

```
python
from sklearn.cluster import KMeans
import numpy as np

# Given data: Mouse, Tumor volume (TV), and hCD45
mouse_data = {
    ('1', 175.118944, 70.6),
    ('2', 147.833792, 62.7),
    ('3', 145.642436, 34.8),
    ('4', 144.826224, 45.3),
    ('5', 137.57652, 30.7),
    ('6', 119.985173, 71.3),
    ('7', 109.7624, 80.5),
    ('8', 99.574272, 52.7),
    ('9', 90.17568, 72.1),
    ('10', 88.319925, 32.4),
    ('11', 74.75, 70.5),
    ('12', 69.876587, 76.2),
    ('13', 69.8544, 66.4),
    ('14', 60.2176, 32.9),
    ('15', 59.82343, 76.7),
    ('16', 41.925912, 43.4),
    ('17', 154.346597, 61.1),
    ('18', 149.673244, 63.7),
    ('19', 143.434368, 84),
    ('20', 130.870816, 62.5)
}

# Extract tumor volume and hCD45 data
data = np.array([(tv, hcd45) for _, tv, hcd45 in mouse_data])

# Perform k-means clustering with 4 clusters
kmeans = KMeans(n_clusters=4, random_state=0)
kmeans.fit(data)

# Assign each mouse to a cluster
clusters = kmeans.labels_

# Initialize groups
groups = [[] for _ in range(4)]

# Divide mice into groups based on cluster assignment
for i, mouse_info in enumerate(mouse_data):
    cluster_index = clusters[i]
    groups[cluster_index].append(mouse_info)

# Display the grouped mice
for group_index, group in enumerate(groups):
    print(f"Group {group_index + 1}:")
    for mouse in group:
        print(f"Mouse {mouse[0]} - Tumor volume: {mouse[1]}, hCD45: {mouse[2]}")
    print()
```

Group 1:
Mouse 6 - Tumor volume: 119.985173, hCD45: 71.3
Mouse 7 - Tumor volume: 109.7624, hCD45: 80.5
Mouse 8 - Tumor volume: 99.574272, hCD45: 52.7
Mouse 9 - Tumor volume: 90.17568, hCD45: 72.1
Mouse 10 - Tumor volume: 88.319925, hCD45: 32.4

Group 2:
Mouse 1 - Tumor volume: 175.118944, hCD45: 70.6
Mouse 2 - Tumor volume: 147.833792, hCD45: 62.7
Mouse 17 - Tumor volume: 164.346597, hCD45: 61.1
Mouse 18 - Tumor volume: 149.673244, hCD45: 63.7
Mouse 19 - Tumor volume: 143.434368, hCD45: 84

Group 3:
Mouse 3 - Tumor volume: 145.642436, hCD45: 34.8
Mouse 4 - Tumor volume: 144.826224, hCD45: 45.3
Mouse 5 - Tumor volume: 137.57652, hCD45: 30.7
Mouse 20 - Tumor volume: 130.870816, hCD45: 62.5

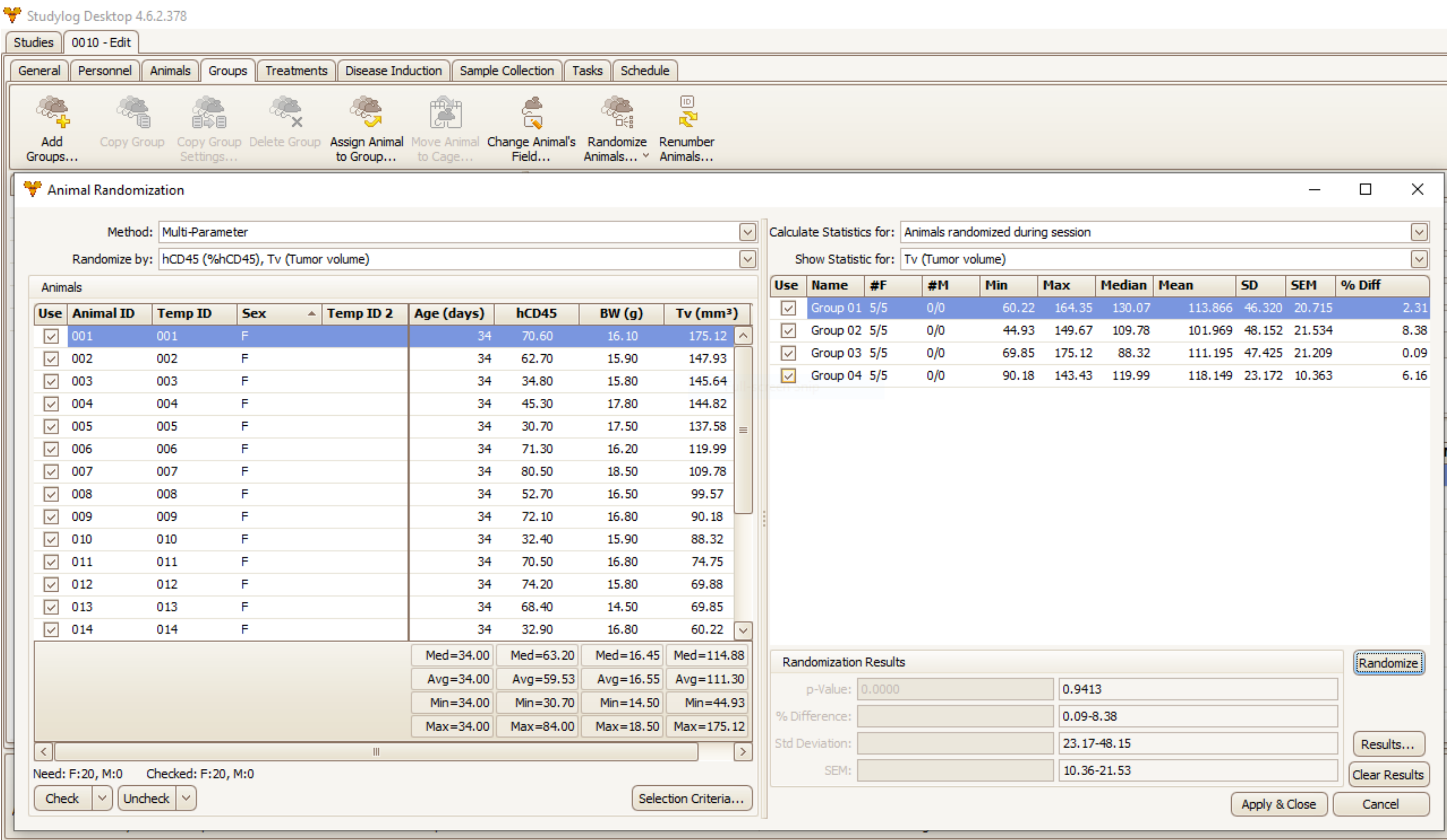
Group 4:
Mouse 11 - Tumor volume: 74.75, hCD45: 70.5
Mouse 12 - Tumor volume: 69.876587, hCD45: 76.2
Mouse 13 - Tumor volume: 69.8544, hCD45: 66.4
Mouse 14 - Tumor volume: 60.2176, hCD45: 32.9
Mouse 15 - Tumor volume: 59.82343, hCD45: 76.7
Mouse 16 - Tumor volume: 41.925912, hCD45: 43.4

Block randomization via Python or R project

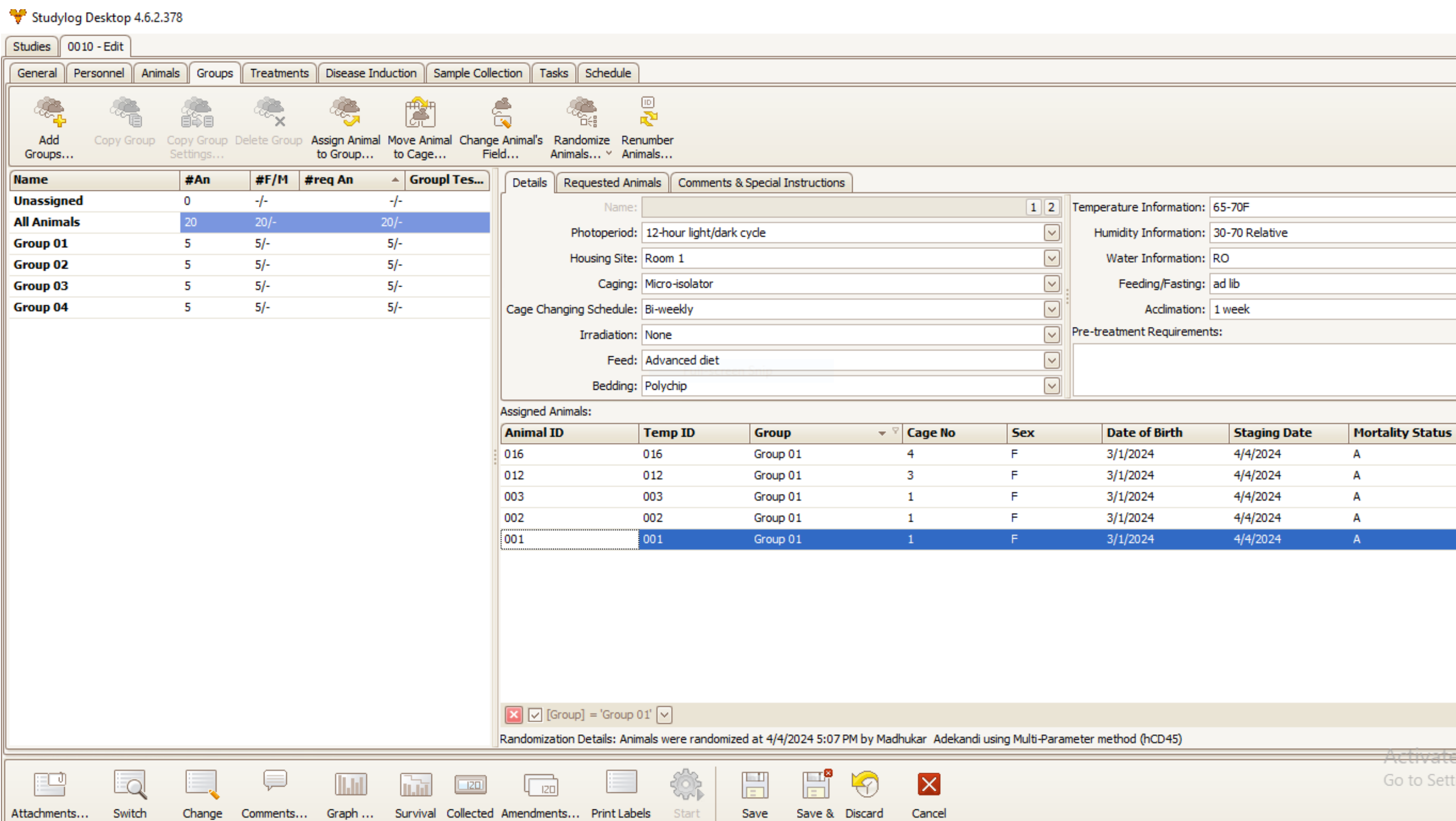
A series of computer programs can be utilized to perform randomization/grouping based on the requirements. The level of knowledge required is dependent on the complexity of experiment, such as number of variables, animals in each group, prerequisites / dependencies for the software etc.

METHODS AND RESULTS: Approach 3

Block randomization with Studylog ® Desktop



Animal allocation with Studylog ® Desktop



Studylog is designed with focus on ease-of-use and a plethora of features are available for laboratory animal experimentation

Conclusions

It is essential to adopt more relevant statistical methods in laboratory animal studies than what is currently being practiced. We found animal workflow software such as Studylog® to be of great relevance in implementation of randomization and grouping strategies. Animal workflow software also come with other advantages such as improved process integrity, robust documentation, reproducibility and archival capabilities

References and Acknowledgements:

1. Studylog ® Desktop Version 4.6.2
2. Stephen J. Iturria, Pharmaceutical Statistics, 2011, 10 169
3. Rios-Doria J, et al. J Immunother Cancer 2020;8:e000416
4. R-project Version 4.4.0



Scan for
Soft copy: