Effects of e-cig vapor on stem cell function

Pilot data analysis and experimental design

Joshua Chang, Shen Huang, Jing Liao, Arnold Seong, Tong Zou

UCI Dept. of Statistics

Table of contents

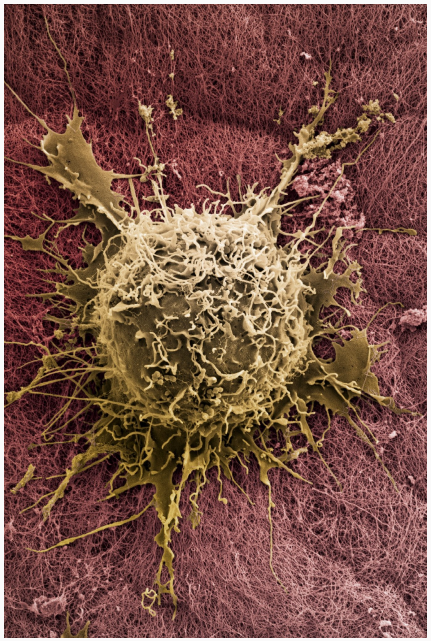
Introduction

Motivation

- Vaping is increasing in prevalence
- Effects of vaping are poorly understood
- How does vaping affect stem cells?

Motivation

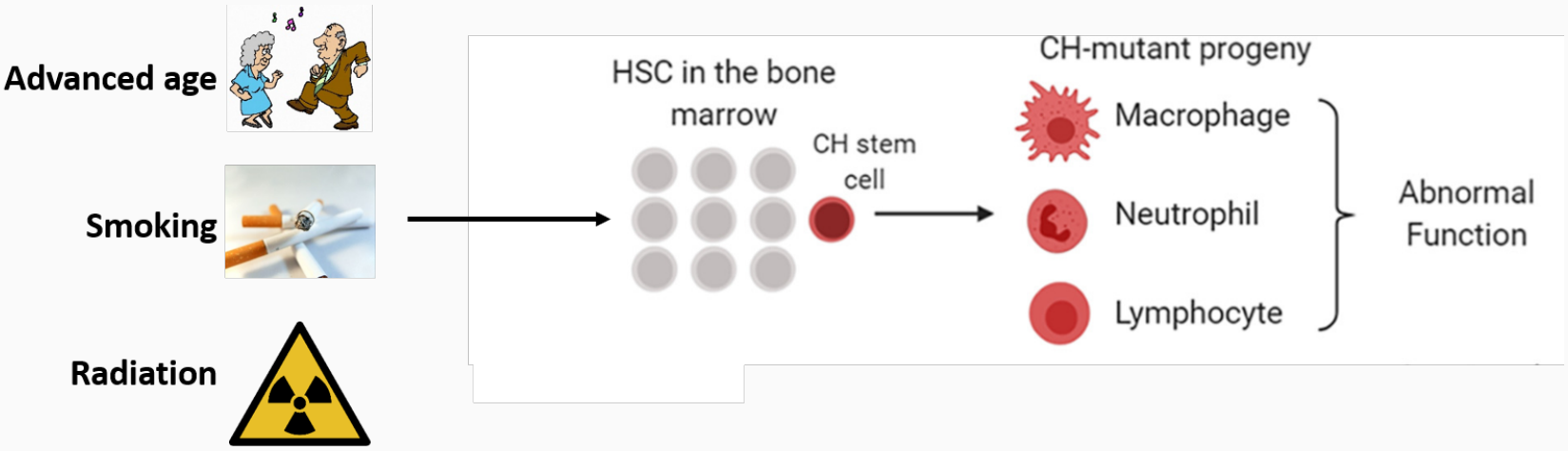
Hematopoietic Stem Cells (HSCs) $\Longrightarrow$ all blood cells



- Cigarette smoke exposure results in ”abnormal hematopoiesis” and ”dysfunctional” niches (Siggins et al, 2014)
- Smoking associated with Clonal Hematopoiesis of Indeterminate Potential (CHIP)
- CHIP is an established risk factor for hematologic cancers, heart attacks, stroke, ...
- Changes in number and function of HSCs can increase risk of CHIP

CHIP Risk Factors

CHIP: Clonal Hematopoiesis of Indeterminate Potential



Hematopoietic Stem Cells / Progenitor Cells

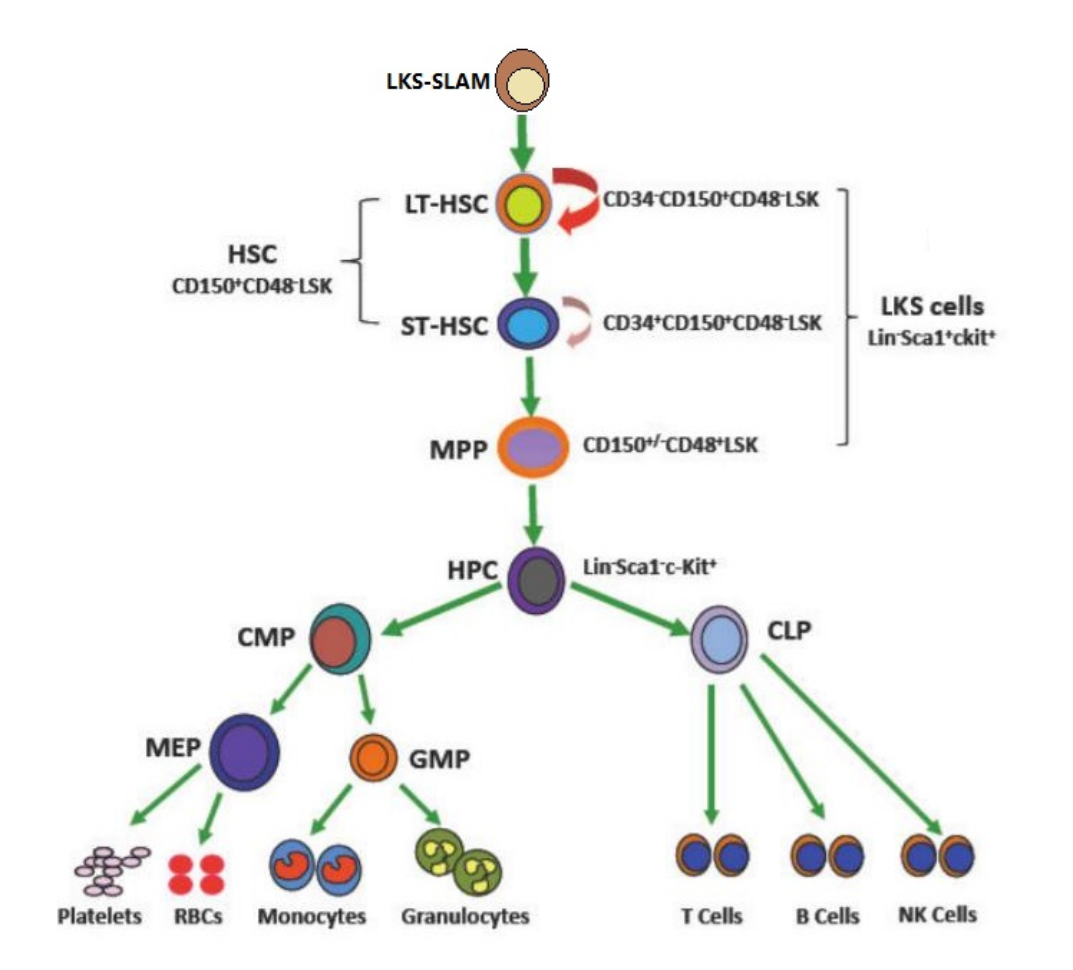


Figure 1: Murine Hematopoietic System

Purpose of Analysis

Hypothesis:

- Similarly to smoking, vaping is likely to induce chronic oxidative stress and inflammation that can affect HSC population size and function (promoting CHIP).

Goals:

- establish effect of E-cig vapor via pilot study
- hypothesis generation for larger trial
- inform experimental design of larger trial

Experimental Methods

Objectives I

Note: Experiment conducted in 2 phases

Primary Objectives

- Phase I: Test whether *ongoing* exposure to E-cigarette vapor impacts LKS-SLAM cell frequencies in mouse animal model;
- Phase II: Investigate the impact of *past* exposure to E-cigarette vapor on HSC function via functional assay. In particular, assess competitive advantage/disadvantage in repopulation ability during early post-transplant period.

Objectives II

Secondary Objectives

- Phase I: Establish and quantify effect of ongoing exposure on other progenitor cell types;
- Phase I & II: Perform power analyses based on pilot study results;
- Phase II: Assess current study design for impact of past exposure.

Experimental Methods: Overview

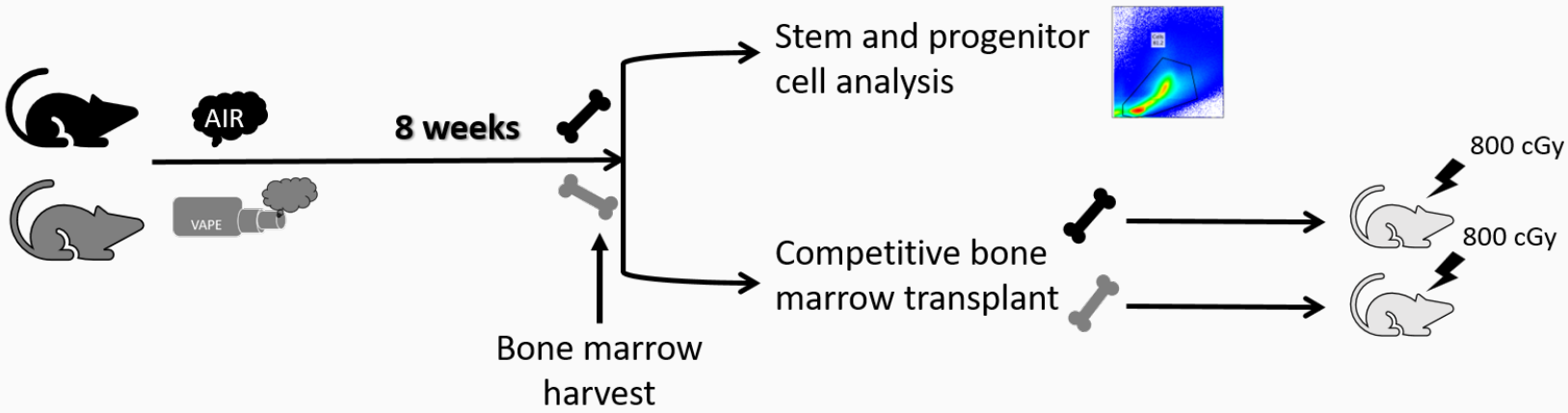


Figure 2: Experimental Design: overview

Experimental Methods: Detailed

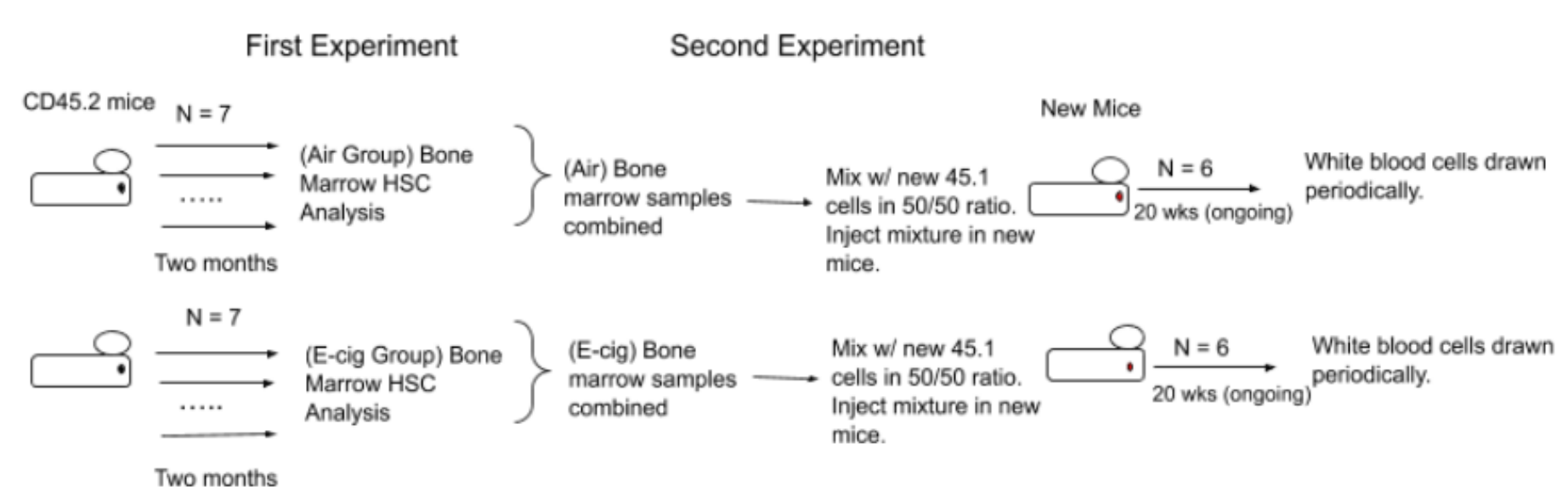


Figure 3: Experimental Design

Experimental hiccups

Some experimental difficulties

- 1 mouse died during Phase 1. **RIP.**
- positive controls excluded in Phase II
- different staining method in wk 16
- lack of standardized techniques for drawing/handling samples
- frequency of blood draws changed at wk 16

Statistical Methods/Analysis

Phase 1: EDA

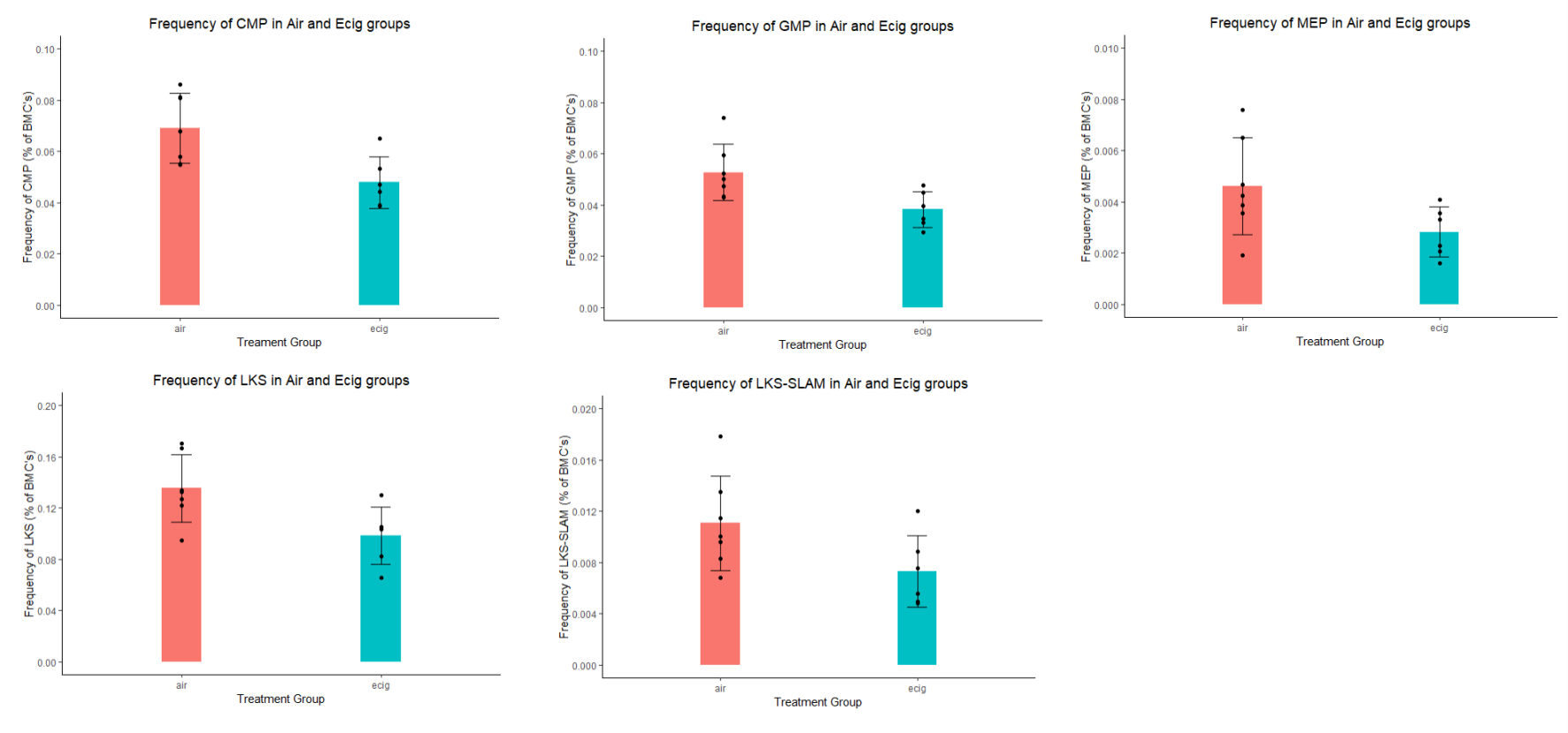


Figure 4: Bar Plot of Mean Frequency of Cell Type

Decrease in cell frequency across bone marrow HSPC in e-cig group.

Phase 1: Hypotheses

H_0: E-cigarette has no effect on the in hematopoietic stem cell and progenitor populations, which indicates that the percentage of LKS-SLAM, LKS, CMP, MEP and GMP are the same in both groups.

H_1: E-cigarette truly has effect on the in hematopoietic stem cell and progenitor populations.

Phase 1: Multiple Testing Correction

The Holm–Bonferroni method is used to conduct our multiple comparisons.

- Let $H_{01}, ...H_{05}$ be a family of $m = 5$ null hypotheses and $P_1, ..., P_5$ the corresponding p-values.
- Start by ordering the p-values (from lowest to highest) $P_{(1)}, ..., P_{(5)}$ and let the associated hypotheses be $H_{(1)}, ..., H_{(5)}$.
- For a given significance level $\alpha = 0.05$, let k be the minimal index such that

$$P_{(k)} > \frac{\alpha}{m + 1 - k}$$

.

- Reject the null hypotheses $H_{(1)}, ..., H_{(k-1)}$ and do not reject $H_{(k)}, ..., H_{(5)}$.

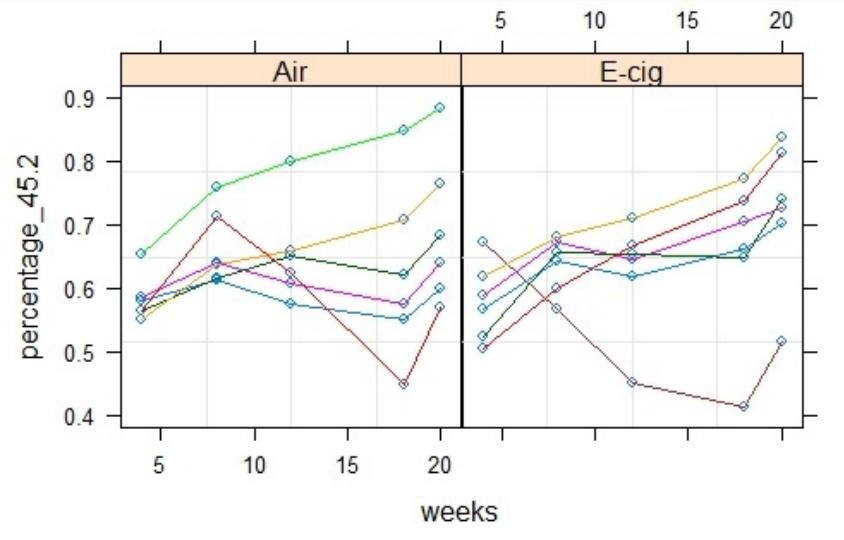
Phase 1: Results

Applied the *t.test()* in R to conduct the two sample t test and used Holm-Bonferroni Method to deal with familywise error rates for multiple hypothesis tests.

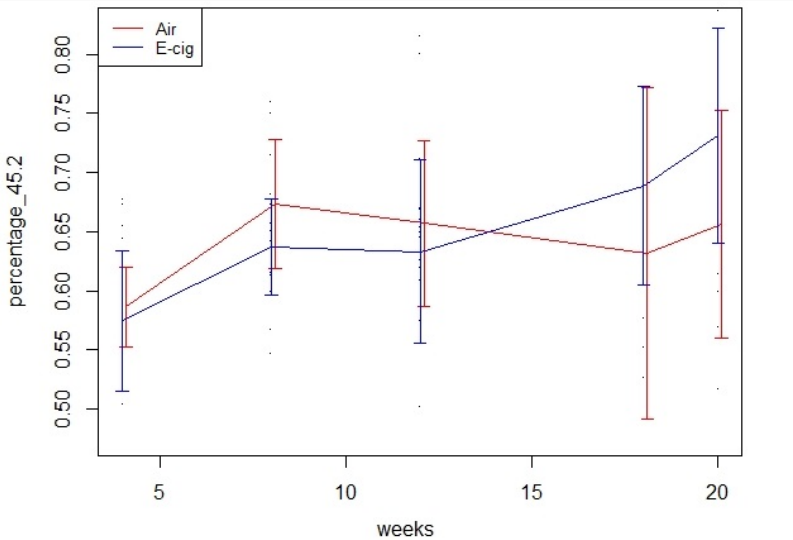
			a=0.05	Holm-Bonferroni	
Primary Outcome	test statistics	P value	Reject H0	alpha level	Reject H0
CMP	3.24	0.008	Yes	0.01	Yes
GMP	2.86	0.016	Yes	0.0125	No
LKS	2.75	0.018	Yes	0.0167	No
MEP	2.21	0.054	Yes	0.025	No
LKS_SLAM	2.08	0.061	No	0.05	No

We could not reject the most of the Ho under Holm-Bonferroni Method. There is no enough statistical evidence to support that there is a different between E-cigarette and air on GMP, LKS, MEP and LKS_SLAM

Phase 2: EDA I



(a) Spaghetti plots of CD45.2 proportions



(b) Mean plots of CD45.2 proportions

- Subject 2042 (Air) and 2054 (E-cig) are potential outliers that may require further examination.
- Primary objective is to infer if there is significant difference between two groups in early period. (will be modeled at week 12)
- The E-cig group grows approximately linearly over time while Air group shows fluctuation.
- Mean plot is weighted by the number of total CD45.1 and CD45.2 cells to improve accuracy.

Phase 2: EDA II

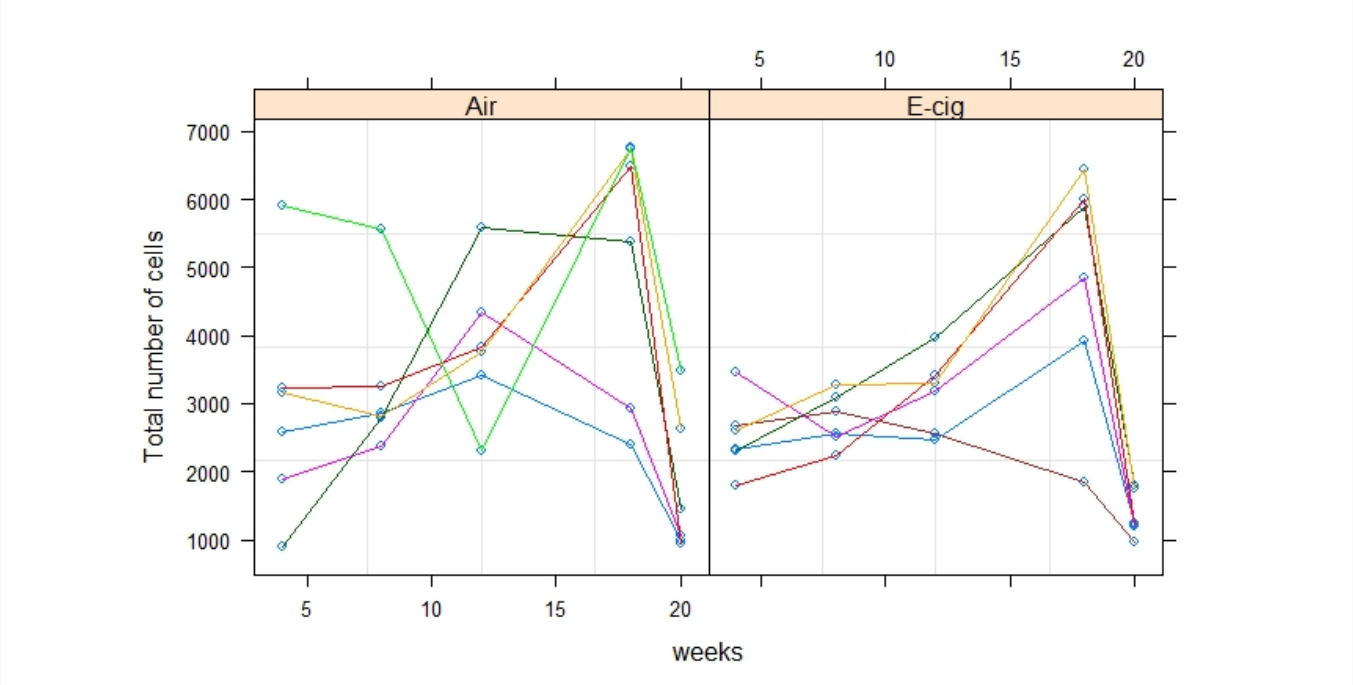
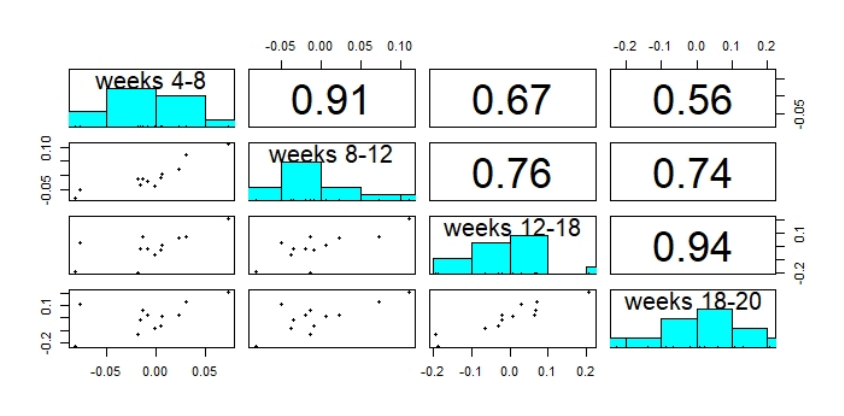


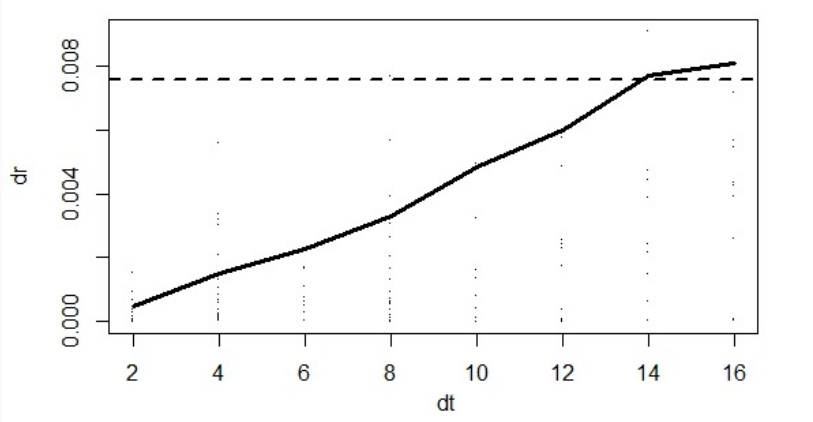
Figure 6: Total number of CD45.1 and CD45.2 cells

The sample sizes of cells obtained at each observation varies a lot was not controlled during experiment (they should be), therefore the proportion/percentage should be weighted by the number of cells. This gives us the motive to use a binomial model.

Phase 2: Variance assumption EDA



(a) Pairwise plot and correlations of residuals between periods



(b) Variogram of residuals

- Residuals of the CD45.2 proportion are obtained by fitting natural cubic splines with respect to time.
- The pairwise plot, their correlations and the variogram all suggest the relation between observations decays over time.

Phase 2: model

We use the following generalized linear mixed-effect model (GLMM) to account for binomial data and longitudinal structure :

$$Y_{ij} \sim Bin(\mu_{ij}, m_{ij})$$

$$log(\frac{\mu_{ij}}{1-\mu_{ij}}) = b_{0j} + b_{1j}t_{ij} + \beta_2 Air_i \times I_{(t_{ij} \leq 12)} + \beta_3 Air_i \times I_{(t_{ij} > 12)}$$

$$\begin{pmatrix} b_{0j} \\ b_{1j} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix} \right)$$

- E-cig group is selected as the reference group based on the the mean plot from EDA.
- Since the major interest is the effect of E-cigarette vapor exposure in earlier period, the model split the timeline at week 12. ($H_0 : \beta_2 = 0$)
- Random intercept and random slope are used to model the correlation decay.
- Package *lme4* is used for implementation.
- Generalized estimating equation (GEE) is not considered due to the limited size of data. (a robust estimate may not be valid)

Phase 2: Results

We use the following generalized linear mixed-effect model (GLMM) to account for binomial data and longitudinal structure :

$$Y_{ij} \sim Bin(\mu_{ij}, m_{ij})$$

$$log(\frac{\mu_{ij}}{1-\mu_{ij}}) = b_{0j} + b_{1j}t_{ij} + \beta_2 Air_i \times I_{(t_{ij} \leq 12)} + \beta_3 Air_i \times I_{(t_{ij} > 12)}$$

$$\begin{pmatrix} b_{0j} \\ b_{1j} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix} \right)$$

The coefficient estimates are listed in the table below :

	Est	exp(Est)	ci95.lo	ci95.hi	z value	Pr(>|z|)
β_0	0.20	1.22	1.01	1.46	2.12	0.03
β_1	0.04	1.04	1.01	1.06	3.16	0.00
$\boldsymbol{\beta_2}$	**0.02**	**1.02**	**0.85**	**1.23**	**0.23**	**0.82**
β_3	-0.40	0.67	0.55	0.82	-3.97	0.00

Phase 2: Results

Nota Bene: odds

- Binomial model treats each cell as a coin flip, with 'heads' (success) here defined as the cell being a CD45.2-type cell.
- The *odds* is the probability of success (μ_{ij}) over the probability of failure (1-μ_{ij}).
- The odds of getting heads on a fair coin is 1 (1-to-1 chance)

Bene Nota Bene: odds ratios

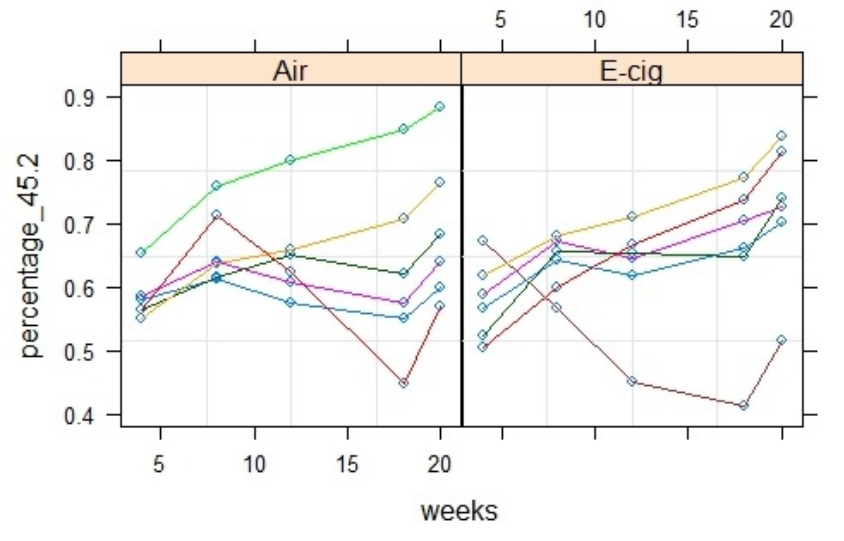
- The odds ratio between X and Y is the odds of X over the odds of Y.

Phase 2: Results

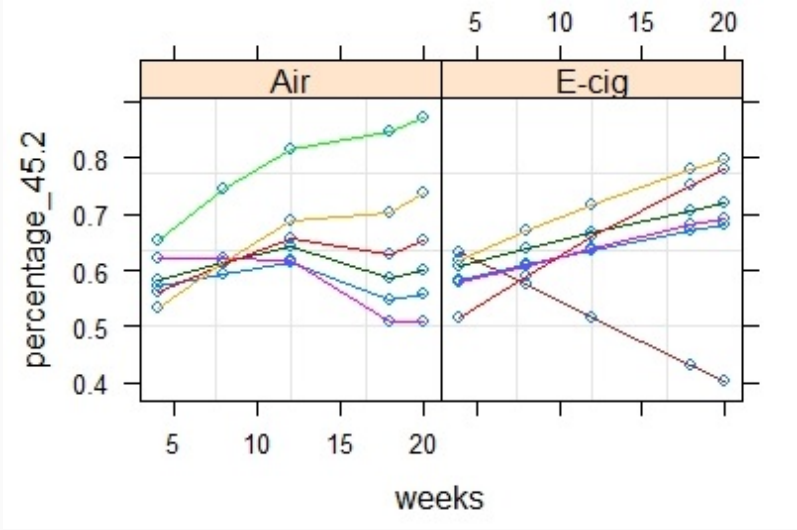
Interpretation of coefficient of interest

- From the results above, we estimate that the odds that a cell will be CD45.2 will be 0.02 times higher in the air group than in the treatment group. At a 95% confidence level, we believe it will be between .15 times lower to .23 times higher.
- As you can tell from that statement, it is fairly inconclusive (not statistically significant)!

Phase 2: Results



(a) Original spaghetti plots of CD45.2 proportions



(b) Fitted spaghetti plots of CD45.2 proportions

Power Analyses

Phase I: methods

• We applied ***pwr.t.test*** function in ***PWR*** package to estimated the sample sizes for Phase I since pwr.t.test is specific for 2 sample t test.

• We assume the sample sizes in e-cigarettes and air groups are equal, then we input

- the α level ($\alpha = 0.05$ and $\alpha = 0.1$)
- compute effect size of the Phase I from below formulas

$$d = \frac{|\mu_1 - \mu_2|}{\sigma}$$

  σ is a pooled standard error.
- power: set Power=0.7, 0.8, and 0.9

to estimate the effect size .

Phase I: results

We fixed α=0.05 and 0.1 to estimate the sample sizes that we need under differential statistical power.

	Significant level					
	a=0.05			a=0.1		
	Power					
	0.7	**0.8**	0.9	0.7	**0.8**	0.9
LKS-SLAM	11	**14**	18	9	**11**	14
LKS	7	**8**	11	5	**7**	9
CMP	6	**7**	8	4	**5**	7
MEP	11	**13**	17	8	**10**	14
GMP	7	**8**	10	5	**7**	9

Since LKS-SLAM and LKS are primary variables of our study interests. If we could apply 14 mice in e-cigarettes and air group in Phase I, the statistical power of LKS-SLAM could achieved 80%.

Phase II: methods

For power analysis, we use the following procedures to simulate data:

$$m_{ij} \sim 100 \times Poisson(30)$$

$$\begin{pmatrix} b_{0j} \\ b_{1j} \end{pmatrix} \sim N\left(\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}, \begin{pmatrix} \hat{\sigma}_0{}^2 & \hat{\rho}\hat{\sigma}_0\hat{\sigma}_1 \\ \hat{\rho}\hat{\sigma}_0\hat{\sigma}_1 & \hat{\sigma}_1{}^2 \end{pmatrix} \right)$$

$$\eta_{ij} = b_{0j} + b_{1j}t_{ij} + \beta_2 Air_i \times I_{(t_{ij} \leq 12)} + \hat{\beta}_3 Air_i \times I_{(t_{ij} > 12)}$$

$$\mu_{ij} = \frac{exp(\eta_{ij})}{1 + exp(\eta_{ij})}$$

$$Y_{ij} \sim Bin(\mu_{ij}, m_{ij})$$

- Choices of the value of $exp(\beta_2)$ are 1.01, **1.02**, 1.03, 1.05, 1.1, 1.2, 1.5.
- Choices of the sample size are **12**, 16, 20, 30, 50, 100.
- For each setting 500 groups of data are generated and each fitted by GLMM to estimate the power

Phase II: results

We fixed α=0.05 and 0.1 to estimate the power.

Size/OR	1.01	1.02	1.03	1.05	1.10	1.20	1.50
N = 12	0.11	**0.08**	0.10	0.10	0.22	0.44	0.73
N = 16	0.07	0.08	0.06	0.14	0.22	0.47	0.74
N = 20	0.07	0.08	0.09	0.10	0.24	0.60	0.74
N = 30	0.05	0.06	0.07	0.08	0.29	0.66	0.74
N = 50	0.05	0.05	0.09	0.17	0.45	0.72	0.74
N = 100	0.05	0.08	0.11	0.26	0.63	0.72	0.73

Size/OR	1.01	1.02	1.03	1.05	1.10	1.20	1.50
N = 12	0.15	**0.13**	0.17	0.17	0.31	0.51	0.73
N = 16	0.12	0.13	0.11	0.17	0.30	0.57	0.74
N = 20	0.12	0.12	0.13	0.15	0.33	0.66	0.74
N = 30	0.11	0.12	0.11	0.13	0.37	0.68	0.74
N = 50	0.11	0.11	0.16	0.23	0.55	0.74	0.74
N = 100	0.09	0.14	0.16	0.35	0.68	0.72	0.73

Discussion

Phase 1 Discussion

Experiment Phase 1

- Phase I need multiple testing correction since analysis involves multiple simultaneous statistical tests.
- There is no significant statistical evidence to support that e-cigarettes and air have different impacts on critical Stem cells(LKS,LKS_SlAM) once multiple testing correction applied.
- The Statistical Power of Phase I was pretty low. There is only 60% statistical power with 6 mice in each group.
- Pilot study data indicates (fairly clearly) that just need higher n to reach desired significance level.

Phase 2 Discussion

Phase 2: future investigations & modeling

Trend for 'Air' (control) group does not appear linear

- may need to consider which group's trajectory is more linear when conducting analysis with knots (make it the 'default' group).
- semi-parametric model (e.g. spline-based general additive model) may be helpful, but inference can be more difficult.

Phase 2: Other design considerations

Phase 2: More standardization needed

Consistent blood draw/sample procedure would likely lead to less noise

- draw same number / volume of blood cells
- same staining technique
- keep the same frequency

Phase 2 Discussion: Crossing Beams

The mice do not start in the same place

- Competitive transplant is 50-50 by *volume*.
- Phase I results indicate that donor air and ecig bone marrow contains different number of stem cells by volume.

The 45.2 E-cig stem cells appear to have competitive advantage

- May start at lower numbers but replicate faster.
- More appropriate way to assess competitive advantage may be to measure the relationship between percentages and the interaction between treatment and time over the whole time period.
- Include Phase 1 results as covariate (proxy for wk0 baseline) and take measurements every two weeks.