

P9185 Project 1: Protocol of a Phase II MATIK Trial

Kate Colvin (kac2301), Chhiring Lama, Emily Carter

I. Introduction

II.

Secondary Objectives

```
baseline <- read.csv("baseline.csv") %>%
  dplyr::select(-X) %>%
  mutate(carryover_period2 = period1,
        carryover_period3 = period2,
        sequence_ind = factor(case_when(
          period1 %in% c("Pill A", "Gel B") & period2 %in% c("Pill A", "Gel B") ~ 0,
          period1 %in% c("Pill A", "Gel C") & period2 %in% c("Pill A", "Gel C") ~ 1,
          .default = 2)),
        seq = factor(paste0(
          case_when(period1 == "Pill A" ~ "A",
                    period1 == "Gel B" ~ "B",
                    period1 == "Gel C" ~ "C"),
          case_when(period2 == "Pill A" ~ "A",
                    period2 == "Gel B" ~ "B",
                    period2 == "Gel C" ~ "C"),
          case_when(period3 == "Pill A" ~ "A",
                    period3 == "Gel B" ~ "B",
                    period3 == "Gel C" ~ "C"))),
        race = factor(case_when(
          race == "black" ~ "Black",
          race == "others" ~ "Other",
          race == "white" ~ "White"), levels = c("Black", "White", "Other")))

baseline_long <- baseline %>%
  pivot_longer(cols = period1:period3,
               names_to = "period",
               values_to = "treatment") %>%
  mutate(bloodVL_before = case_when(
    period == "period1" ~ bviral0,
    period == "period2" ~ bviral2,
    period == "period3" ~ bviral4),
    bloodVL_after = case_when(
      period == "period1" ~ bviral1,
```

```

    period == "period2" ~ bviral3,
    period == "period3" ~ bviral5),
skinVL_before = case_when(
    period == "period1" ~ sviral0,
    period == "period2" ~ sviral2,
    period == "period3" ~ sviral4),
skinVL_after = case_when(
    period == "period1" ~ sviral1,
    period == "period2" ~ sviral3,
    period == "period3" ~ sviral5),
carryover = case_when(
    period == "period1" ~ "0",
    period == "period2" ~ carryover_period2,
    period == "period3" ~ carryover_period3)) %>%
dplyr::select(-c(bviral0:carryover_period3)) %>%
mutate(bloodVL_change = bloodVL_before - bloodVL_after,
       skinVL_change = skinVL_before - skinVL_after) %>%
pivot_longer(cols = c(bloodVL_before:skinVL_after),
              names_to = c(".value", "timepoint"),
              names_sep = "_") %>%
mutate(timepoint = ifelse(timepoint == "before", 0, 4)) %>%
mutate(treatment = factor(treatment, levels = c("Pill A", "Gel B", "Gel C")))

#write.csv(baseline_long, "baseline_longformat.csv")

endpoints <- read.csv("endpoints.csv") %>%
  mutate(carryover_period2 = period1,
         carryover_period3 = period2)

overall_endpoints <- endpoints %>%
  pivot_longer(cols = period1:period3,
               names_to = "period",
               values_to = "treatment") %>%
  mutate(
    AE_pilla_total = rowSums(select(., starts_with("AE_pilla")), na.rm = TRUE),
    AE_gelB_total = rowSums(select(., starts_with("AE_gelB")), na.rm = TRUE),
    AE_gelC_total = rowSums(select(., starts_with("AE_gelC")), na.rm = TRUE),
    Adhere_pilla_total = rowSums(select(., starts_with("Adhere_pilla")), na.rm = TRUE),
    Adhere_gelB_total = rowSums(select(., starts_with("Adhere_gelB")), na.rm = TRUE),
    Adhere_gelC_total = rowSums(select(., starts_with("Adhere_gelC")), na.rm = TRUE),
    overall_safety = case_when(
      treatment == "Pill A" ~ factor(ifelse(AE_pilla_total > 0, 1, 0)),
      treatment == "Gel B" ~ factor(ifelse(AE_gelB_total > 0, 1, 0)),
      treatment == "Gel C" ~ factor(ifelse(AE_gelC_total > 0, 1, 0))),
    overall_adhere = case_when(
      treatment == "Pill A" ~ Adhere_pilla_total,
      treatment == "Gel B" ~ Adhere_gelB_total,
      treatment == "Gel C" ~ Adhere_gelC_total),
    carryover = factor(case_when(
      period == "period1" ~ "0",
      period == "period2" ~ carryover_period2,
      period == "period3" ~ carryover_period3))) %>%
dplyr::select(-c(AE_pilla_week1:carryover_period3), -c(AE_pilla_total:AE_gelC_total), -c(Adhere_pilla_

```

Table 1: **Table 1. Baseline Characteristics**

Characteristic	Overall N = 180 ¹	ABC N = 30 ¹	ACB N = 30 ¹	BAC N = 30 ¹	BCA N = 30 ¹
Age (years)	32 (8) [18, 45]	33 (8) [19, 44]	31 (7) [18, 42]	31 (7) [20, 44]	33 (8) [19, 45]
Race					
Black	64 (36%)	12 (40%)	13 (43%)	12 (40%)	10 (33%)
White	52 (29%)	8 (27%)	9 (30%)	7 (23%)	7 (23%)
Other	64 (36%)	10 (33%)	8 (27%)	11 (37%)	13 (43%)
Female	81 (45%)	13 (43%)	11 (37%)	16 (53%)	14 (47%)

¹Mean (SD) [Min, Max]; n (%)

²Kruskal-Wallis rank sum test; Pearson's Chi-squared test

```

merged <- baseline_long %>%
  filter(timepoint == 0) %>%
  dplyr::select(-timepoint, -bloodVL, -skinVL) %>%
  left_join(overall_endpoints %>% dplyr::select(-treatment, -carryover), by = c("ptid", "period")) %>%
  mutate(overall_safety = factor(ifelse(overall_safety == 1, "Adverse Event", "No Adverse Event")))

baseline_long <- baseline_long %>%
  left_join(overall_endpoints %>% dplyr::select(-treatment, -carryover), by = c("ptid", "period"))

baseline %>%
  dplyr::select(age:gender, seq) %>%
 tbl_summary(
  by = seq,
  statistic = list(
    all_continuous() ~ "{mean} ({sd}) \n [{min}, {max}]",
    all_categorical() ~ "{n} ({p}%)"
  ),
  label = list(
    age ~ "Age (years)",
    race ~ "Race",
    gender ~ "Female"
  )
) %>%
add_p() %>%
add_overall() %>%
bold_labels() %>%
modify_caption("**Table 1. Baseline Characteristics**")

```

preliminary assessment and comparison of systemic and local Pharmacokinetics (PK) of Pill A, Gel B, and Gel C and the correlation of PK with adherence measures and the occurrence of adverse events

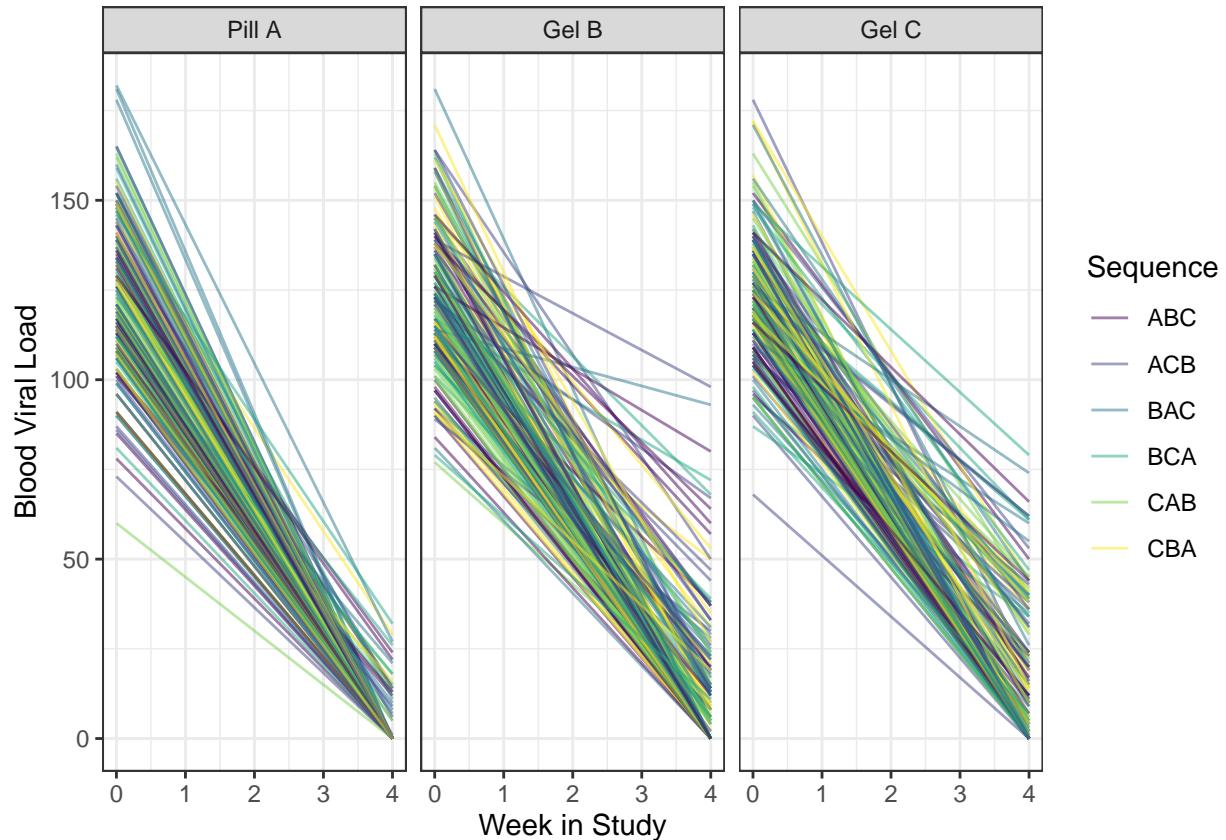
Blood PK

$$Y_{ik} = \mu + b_k + \pi_i + \tau_i + \lambda_i + \alpha_{ik} + \beta_{ik} + \varepsilon_{ik}, \quad b_k \sim N(0, \sigma_b^2), \quad \varepsilon_{ik} \sim N(0, \sigma^2)$$

```

ggplot(baseline_long) +
  geom_line(aes(x = timepoint, y = bloodVL, group = ptid, color = seq), alpha = 0.5) +
  facet_wrap(~ treatment) +
  theme_bw() +
  labs(x = "Week in Study", y = "Blood Viral Load", color = "Sequence")

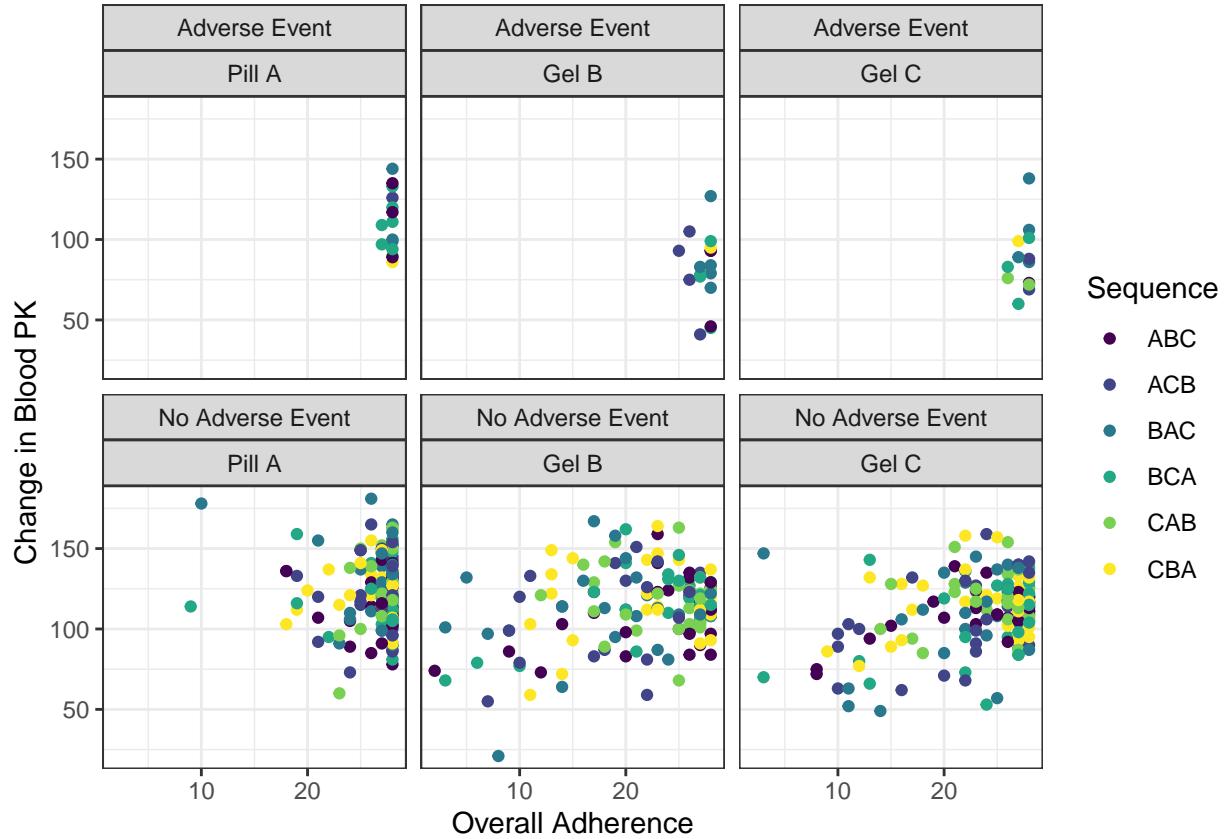
```



```

ggplot(merged) +
  geom_point(aes(y = bloodVL_change, x = overall_adhere, color = seq)) +
  facet_wrap(~ overall_safety + treatment) +
  theme_bw() +
  labs(x = "Overall Adherence", y = "Change in Blood PK", color = "Sequence")

```



```

# 
# ggplot(baseline_long) +
#   geom_line(aes(x = timepoint, y = bloodVL, group = ptid, color = overall_safety), alpha = 0.5) +
#   facet_wrap(~ treatment) +
#   theme_bw()

model_large <- lmer(bloodVL_change ~ treatment + overall_adhere + overall_safety + sequence_ind +
                     period + (1 | ptid),
                     data = merged)

model <- lmer(bloodVL_change ~ treatment + overall_adhere + overall_safety + period + (1 | ptid),
              data = merged)

anova(model, model_large)

## refitting model(s) with ML (instead of REML)

## Data: merged
## Models:
## model: bloodVL_change ~ treatment + overall_adhere + overall_safety + period + (1 | ptid)
## model_large: bloodVL_change ~ treatment + overall_adhere + overall_safety + sequence_ind + period +
##               npar    AIC    BIC  logLik -2*log(L)  Chisq Df Pr(>Chisq)
## model         9 4863.1 4901.7 -2422.6    4845.1
## model_large  11 4866.5 4913.7 -2422.2    4844.5 0.6124  2     0.7362

```

```

summary(model)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: bloodVL_change ~ treatment + overall_adhere + overall_safety +
##      period + (1 | ptid)
## Data: merged
##
## REML criterion at convergence: 4827.4
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -3.5936 -0.5795  0.0153  0.6127  3.0873
##
## Random effects:
##   Groups   Name        Variance Std.Dev.
##   ptid     (Intercept) 23.09    4.805
##   Residual           445.60   21.109
## Number of obs: 540, groups: ptid, 180
##
## Fixed effects:
##                               Estimate Std. Error      df t value
##   (Intercept)                73.4552   6.7058 419.7421 10.954
##   treatmentGel B            -6.7916   2.3555 417.1354 -2.883
##   treatmentGel C            -9.0880   2.2967 391.5595 -3.957
##   overall_adhere             0.9365   0.1882 269.4894  4.975
##   overall_safetyNo Adverse Event 26.4624   3.4675 521.9424  7.632
##   periodperiod2             -2.3295   2.2271 356.0129 -1.046
##   periodperiod3             -1.5433   2.2252 355.4545 -0.694
##                               Pr(>|t|)
##   (Intercept) < 0.0000000000000002 ***
##   treatmentGel B          0.00414 **
##   treatmentGel C          0.000090106833739 ***
##   overall_adhere           0.000001161748206 ***
##   overall_safetyNo Adverse Event 0.000000000000111 ***
##   periodperiod2            0.29629
##   periodperiod3            0.48842
##   ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) trtmGB trtmGC ovrl1_ ov_NAE prdpr2
## treatmentG1B -0.437
## treatmentG1C -0.359  0.538
## overall_dhr -0.832  0.328  0.247
## ovrl1_sfNAE -0.616  0.080  0.032  0.196
## periodperd2 -0.129 -0.009 -0.006 -0.026 -0.038
## periodperd3 -0.173  0.001  0.001  0.004  0.009  0.499

```

```

tbl_regression(
  model,
  intercept = TRUE,
  estimate_fun = function(x) style_number(x, digits = 1)) %>%

```

Comparison	Beta	95% CI	p-value
Intercept (Pill A, Period 1, Sequence 0, No Adherence or AE)	73.5	60.3, 86.6	<0.001
Treatment			<0.001
Gel B vs Pill A	-6.8	-11.4, -2.2	
Gel C vs Pill A	-9.1	-13.6, -4.6	
Additional Day of Adherence	0.9	0.6, 1.3	<0.001
Adverse Event			<0.001
No Adverse Event	26.5	19.7, 33.3	
Period			0.6
Period 2 vs Period 1	-2.3	-6.7, 2.1	
Period 3 vs Period 1	-1.5	-5.9, 2.8	

Abbreviation: CI = Confidence Interval

```

add_global_p() %>%
  modify_table_body(~.x %>%
    mutate(label = case_when(
      label == "(Intercept)" ~ "Intercept (Pill A, Period 1, Sequence 0, No Adherence or AE)",
      label == "Gel B" ~ "Gel B vs Pill A",
      label == "Gel C" ~ "Gel C vs Pill A",
      label == "period2" ~ "Period 2 vs Period 1",
      label == "period3" ~ "Period 3 vs Period 1",
      label == "treatment" ~ "Treatment", # Remove header
      label == "period" ~ "Period",      # Remove header
      label == "Pill A" ~ NA_character_,
      label == "period1" ~ NA_character_,
      label == "sequence_ind" ~ "Sequence",
      label == "Adverse Event" ~ NA_character_,
      label == "0" ~ NA_character_,
      label == "1" ~ "1 vs 0",
      label == "2" ~ "2 vs 0",
      label == "overall_adhere" ~ "Additional Day of Adherence",
      label == "overall_safety" ~ "Adverse Event",
      TRUE ~ label)) %>%
    filter(!is.na(label))) %>%
  modify_header(label ~ "***Comparison***", estimate ~ "***Beta***") %>%
  bold_p(t = 0.05)

```

```
library(emmeans)
```

```

## Welcome to emmeans.
## Caution: You lose important information if you filter this package's results.
## See '? untidy'
```

```
emm_treatment <- emmeans(model, ~ treatment)
summary(emm_treatment)
```

```
##   treatment emmean    SE  df lower.CL upper.CL
```

```

##   Pill A      107.9 2.27 525      103.4      112
##   Gel B      101.1 2.13 523       96.9      105
##   Gel C      98.8 2.18 526       94.5      103
##
## Results are averaged over the levels of: overall_safety, period
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95

# Pairwise comparisons for treatment
pairs(emm_treatment)

##   contrast      estimate    SE   df t.ratio p.value
##   Pill A - Gel B     6.79 2.36 417    2.882  0.0115
##   Pill A - Gel C     9.09 2.30 391    3.956  0.0003
##   Gel B - Gel C     2.30 2.24 359    1.027  0.5602
##
## Results are averaged over the levels of: overall_safety, period
## Degrees-of-freedom method: kenward-roger
## P value adjustment: tukey method for comparing a family of 3 estimates

# Marginal means for overall_adhere
emm_adhere <- emmeans(model, ~ overall_adhere)
summary(emm_adhere)

##   overall_adhere emmean    SE   df lower.CL upper.CL
##                  24     103 1.75 411     99.2     106
##
## Results are averaged over the levels of: treatment, overall_safety, period
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95

# Marginal means for overall_safety
emm_safety <- emmeans(model, ~ overall_safety)
summary(emm_safety)

##   overall_safety emmean    SE   df lower.CL upper.CL
##   Adverse Event     89.4 3.34 500     82.8     95.9
##   No Adverse Event 115.9 1.02 196    113.8    117.9
##
## Results are averaged over the levels of: treatment, period
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95

# Marginal means for period
emm_period <- emmeans(model, ~ period)
summary(emm_period)

##   period emmean    SE   df lower.CL upper.CL
##   period1    104 2.16 527     99.7     108
##   period2    102 2.21 527     97.2     106
##   period3    102 2.14 527     98.2     107

```

```

##  

## Results are averaged over the levels of: treatment, overall_safety  

## Degrees-of-freedom method: kenward-roger  

## Confidence level used: 0.95

pairs(emm_period)

## contrast estimate SE df t.ratio p.value
## period1 - period2 2.330 2.23 355 1.046 0.5484
## period1 - period3 1.543 2.23 355 0.694 0.7674
## period2 - period3 -0.786 2.23 356 -0.353 0.9337
##
## Results are averaged over the levels of: treatment, overall_safety
## Degrees-of-freedom method: kenward-roger
## P value adjustment: tukey method for comparing a family of 3 estimates

resid_scaled <- residuals(model, scaled = TRUE)
fitted_vals <- fitted(model)
random_intercepts <- ranef(model)$ptid[, 1]
model_data <- model@frame

# Plot 1: Q-Q Plot for Scaled Residuals with confidence bands
p1 <- ggplot(data.frame(resid = resid_scaled), aes(sample = resid)) +
  stat_qq(alpha = 0.5) +
  stat_qq_line(color = "red", linewidth = 1) +
  labs(title = "Q-Q Plot: Scaled Residuals",
       x = "Theoretical Quantiles",
       y = "Sample Quantiles") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))

# Plot 2: Histogram of Scaled Residuals with rug
p2 <- ggplot(data.frame(resid = resid_scaled), aes(x = resid)) +
  geom_histogram(aes(y = after_stat(density)), bins = 30,
                fill = "skyblue", color = "black", alpha = 0.7) +
  geom_density(color = "blue", linewidth = 1) +
  stat_function(fun = dnorm, args = list(mean = 0, sd = 1),
                color = "red", linewidth = 1, linetype = "dashed") +
  geom_rug(alpha = 0.3) +
  labs(title = "Histogram: Scaled Residuals",
       x = "Scaled Residuals",
       y = "Density",
       caption = "Red line = Normal(0,1), Blue line = Kernel density") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))

# Plot 3: Q-Q Plot for Random Intercepts
p3 <- ggplot(data.frame(re = random_intercepts), aes(sample = re)) +
  stat_qq(alpha = 0.5) +
  stat_qq_line(color = "red", linewidth = 1) +
  labs(title = "Q-Q Plot: Random Intercepts",
       x = "Theoretical Quantiles",
       y = "Sample Quantiles") +

```

```

theme_bw() +
theme(plot.title = element_text(hjust = 0.5, face = "bold"))

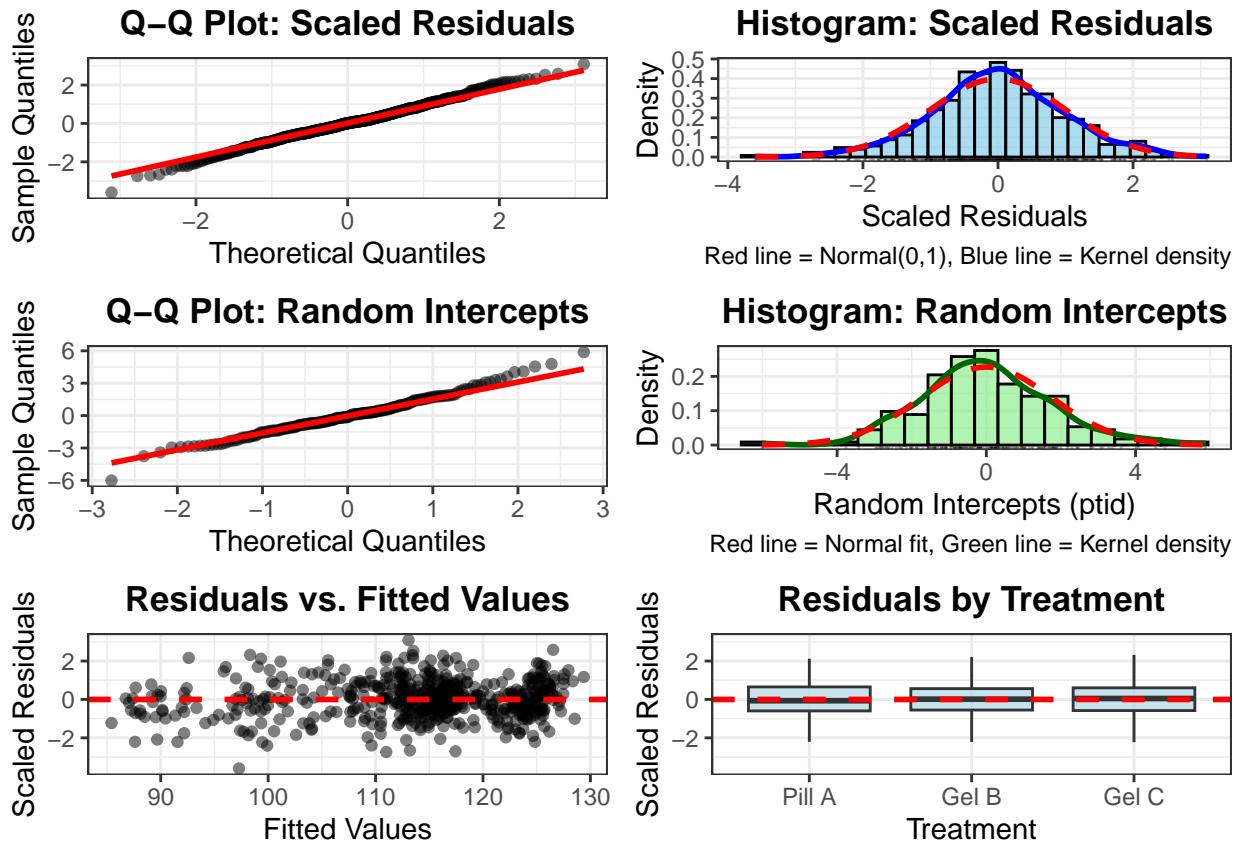
# Plot 4: Histogram of Random Intercepts with rug
p4 <- ggplot(data.frame(re = random_intercepts), aes(x = re)) +
  geom_histogram(aes(y = after_stat(density)), bins = 20,
                 fill = "lightgreen", color = "black", alpha = 0.7) +
  geom_density(color = "darkgreen", linewidth = 1) +
  stat_function(fun = dnorm,
                args = list(mean = 0, sd = sd(random_intercepts)),
                color = "red", linewidth = 1, linetype = "dashed") +
  geom_rug(alpha = 0.3) +
  labs(title = "Histogram: Random Intercepts",
       x = "Random Intercepts (ptid)",
       y = "Density",
       caption = "Red line = Normal fit, Green line = Kernel density") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))

# Plot 5: Residuals vs. Fitted Values with jitter
p5 <- ggplot(data.frame(fitted = fitted_vals, resid = resid_scaled),
              aes(x = fitted, y = resid)) +
  geom_point(alpha = 0.5, position = position_jitter(width = 0.5, height = 0)) +
  geom_hline(yintercept = 0, color = "red", linetype = "dashed", linewidth = 1) +
  #geom_smooth(se = TRUE, color = "blue", method = "loess", linewidth = 1) +
  labs(title = "Residuals vs. Fitted Values",
       x = "Fitted Values",
       y = "Scaled Residuals") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))

# Plot 6: Residuals by Treatment with individual points
p6 <- ggplot(data.frame(treatment = model_data$treatment,
                        resid = resid_scaled),
              aes(x = treatment, y = resid)) +
  geom_boxplot(fill = "lightblue", alpha = 0.7, outlier.shape = NA) +
  #geom_jitter(width = 0.2, alpha = 0.3, color = "darkblue") +
  geom_hline(yintercept = 0, color = "red", linetype = "dashed", linewidth = 1) +
  labs(title = "Residuals by Treatment",
       x = "Treatment",
       y = "Scaled Residuals") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))

# Display all plots
grid.arrange(p1, p2, p3, p4, p5, p6, ncol = 2)

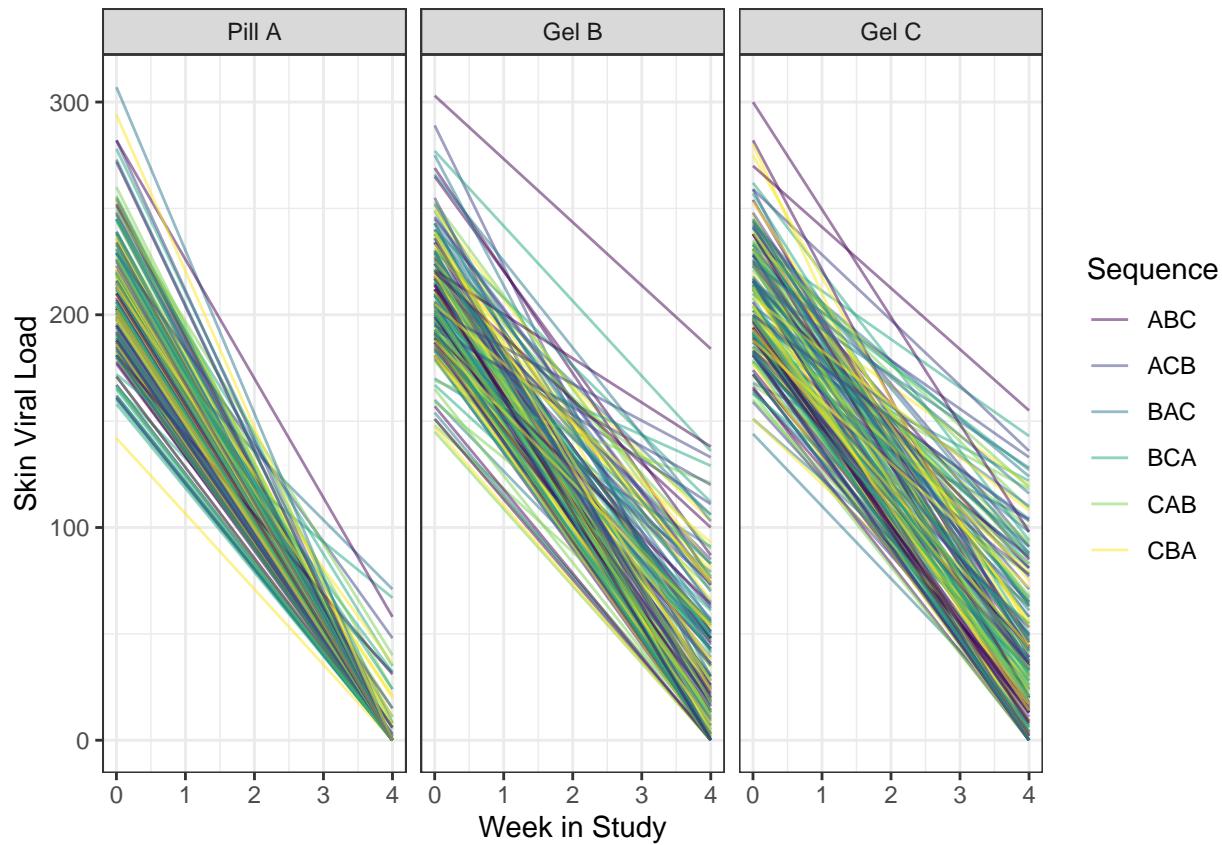
```



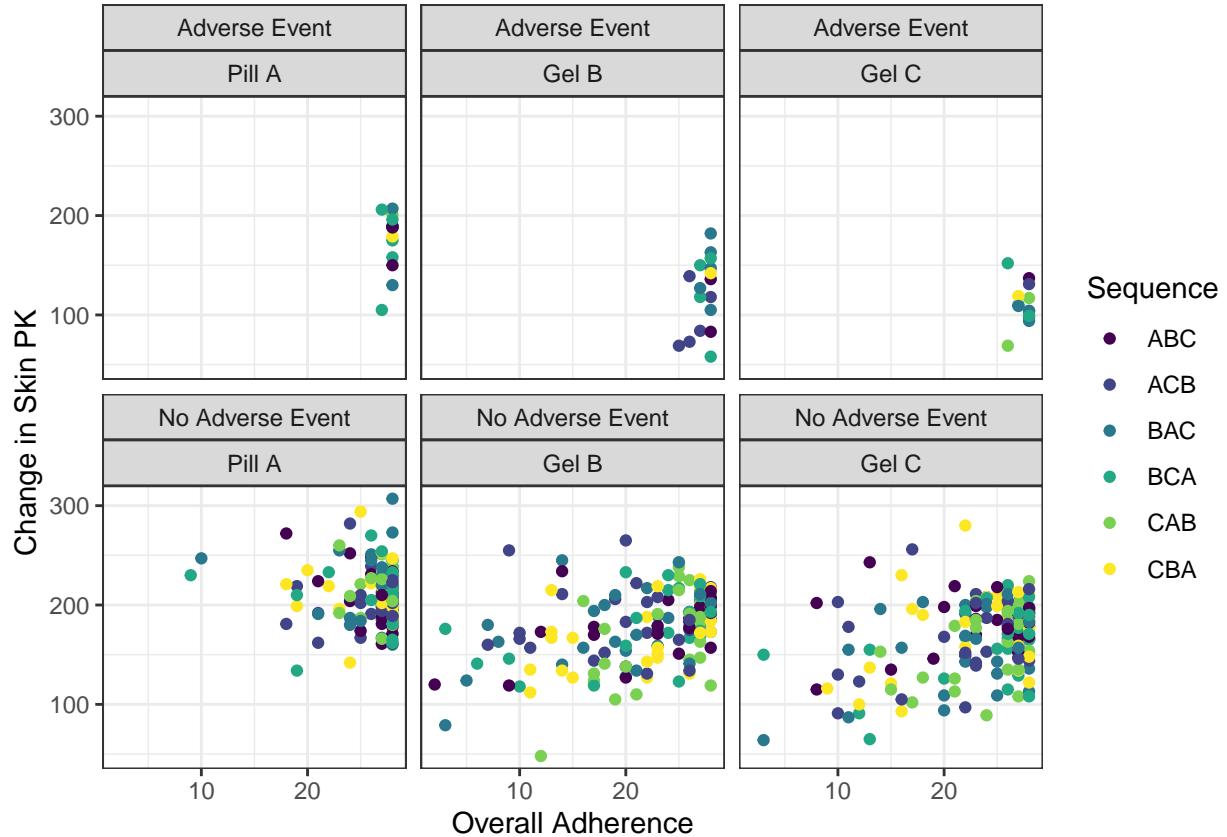
Skin PK

Linear Model

```
ggplot(baseline_long) +
  geom_line(aes(x = timepoint, y = skinVL, group = ptid, color = seq), alpha = 0.5) +
  facet_wrap(~ treatment) +
  theme_bw() +
  labs(x = "Week in Study", y = "Skin Viral Load", color = "Sequence")
```



```
ggplot(merged) +
  geom_point(aes(y = skinVL_change, x = overall_adhere, color = seq)) +
  facet_wrap(~ overall_safety + treatment) +
  theme_bw() +
  labs(x = "Overall Adherence", y = "Change in Skin PK", color = "Sequence")
```



```

model_large <- lm(skinVL_change ~ treatment + overall_adhere + overall_safety + sequence_ind
+ period, data = merged)
model <- lm(skinVL_change ~ treatment + overall_adhere + overall_safety + period,
data = merged)
anova(model, model_large)

```

```

## Analysis of Variance Table
##
## Model 1: skinVL_change ~ treatment + overall_adhere + overall_safety +
##           period
## Model 2: skinVL_change ~ treatment + overall_adhere + overall_safety +
##           sequence_ind + period
##   Res.Df   RSS Df Sum of Sq    F Pr(>F)
## 1     533 571160
## 2     531 568220  2    2940.4 1.3739 0.254

```

```
summary(model)
```

```

##
## Call:
## lm(formula = skinVL_change ~ treatment + overall_adhere + overall_safety +
##       period, data = merged)
##
## Residuals:

```

```

##      Min      1Q   Median      3Q      Max
## -112.388 -22.128    1.035  20.921  115.662
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                105.6003   9.9680 10.594 < 0.0000000000000002
## treatmentGel B             -24.4131   3.6314 -6.723  0.000000000046
## treatmentGel C             -36.9951   3.5493 -10.423 < 0.0000000000000002
## overall_adhere              1.8154   0.2753  6.594  0.000000000103
## overall_safetyNo Adverse Event 55.7940   5.2166 10.696 < 0.0000000000000002
## periodperiod2               0.8567   3.4535  0.248  0.804
## periodperiod3               1.6225   3.4507  0.470  0.638
##
## (Intercept)                 ***
## treatmentGel B              ***
## treatmentGel C              ***
## overall_adhere                ***
## overall_safetyNo Adverse Event ***
## periodperiod2
## periodperiod3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32.74 on 533 degrees of freedom
## Multiple R-squared:  0.3581, Adjusted R-squared:  0.3508
## F-statistic: 49.55 on 6 and 533 DF,  p-value: < 0.0000000000000002

```

```

tbl_regression(
  model,
  intercept = TRUE,
  estimate_fun = function(x) style_number(x, digits = 1)) %>%
  add_global_p() %>%
  modify_table_body(~.x %>%
    mutate(label = case_when(
      label == "(Intercept)" ~ "Intercept (Pill A, Period 1, Sequence 0, No Adherence or AE)",
      label == "Gel B" ~ "Gel B vs Pill A",
      label == "Gel C" ~ "Gel C vs Pill A",
      label == "period2" ~ "Period 2 vs Period 1",
      label == "period3" ~ "Period 3 vs Period 1",
      label == "treatment" ~ "Treatment", # Remove header
      label == "period" ~ "Period",       # Remove header
      label == "Pill A" ~ NA_character_,
      label == "period1" ~ NA_character_,
      label == "sequence_ind" ~ "Sequence",
      label == "Adverse Event" ~ NA_character_,
      label == "0" ~ NA_character_,
      label == "1" ~ "1 vs 0",
      label == "2" ~ "2 vs 0",
      label == "overall_adhere" ~ "Additional Day of Adherence",
      label == "overall_safety" ~ "Adverse Event",
      TRUE ~ label)) %>%
    filter(!is.na(label))) %>%
  modify_header(label ~ "***Comparison***", estimate ~ "***Beta***") %>%
  bold_p(t = 0.05)

```

Comparison	Beta	95% CI	p-value
Intercept (Pill A, Period 1, Sequence 0, No Adherence or AE)	105.6	86.0, 125.2	<0.001
Treatment			<0.001
Gel B vs Pill A	-24.4	-31.5, -17.3	
Gel C vs Pill A	-37.0	-44.0, -30.0	
Additional Day of Adherence	1.8	1.3, 2.4	<0.001
Adverse Event			<0.001
No Adverse Event	55.8	45.5, 66.0	
Period			0.9
Period 2 vs Period 1	0.9	-5.9, 7.6	
Period 3 vs Period 1	1.6	-5.2, 8.4	

Abbreviation: CI = Confidence Interval

```

library(emmeans)
emm_treatment <- emmeans(model, ~ treatment)
summary(emm_treatment)

##  treatment emmean   SE df lower.CL upper.CL
##  Pill A      178 3.42 533     171     185
##  Gel B       154 3.20 533     147     160
##  Gel C       141 3.29 533     135     147
##
## Results are averaged over the levels of: overall_safety, period
## Confidence level used: 0.95

# Pairwise comparisons for treatment
pairs(emm_treatment)

##  contrast      estimate   SE df t.ratio p.value
##  Pill A - Gel B    24.4 3.63 533    6.723 <.0001
##  Pill A - Gel C    37.0 3.55 533   10.423 <.0001
##  Gel B - Gel C    12.6 3.47 533    3.631  0.0009
##
## Results are averaged over the levels of: overall_safety, period
## P value adjustment: tukey method for comparing a family of 3 estimates

# Marginal means for overall_adhere
emm_adhere <- emmeans(model, ~ overall_adhere)
summary(emm_adhere)

##  overall_adhere emmean   SE df lower.CL upper.CL
##                24    158 2.59 533     152     163
##
## Results are averaged over the levels of: treatment, overall_safety, period
## Confidence level used: 0.95

```

```

# Marginal means for overall_safety
emm_safety <- emmeans(model, ~ overall_safety)
summary(emm_safety)

## overall_safety    emmean    SE  df lower.CL upper.CL
## Adverse Event      130 4.99 533      120      139
## No Adverse Event   185 1.47 533      183      188
##
## Results are averaged over the levels of: treatment, period
## Confidence level used: 0.95

# Marginal means for period
emm_period <- emmeans(model, ~ period)
summary(emm_period)

## period    emmean    SE  df lower.CL upper.CL
## period1     157 3.25 533      150      163
## period2     158 3.33 533      151      164
## period3     158 3.23 533      152      165
##
## Results are averaged over the levels of: treatment, overall_safety
## Confidence level used: 0.95

pairs(emm_period)

## contrast      estimate    SE  df t.ratio p.value
## period1 - period2   -0.857 3.45 533   -0.248  0.9666
## period1 - period3   -1.623 3.45 533   -0.470  0.8853
## period2 - period3   -0.766 3.45 533   -0.222  0.9733
##
## Results are averaged over the levels of: treatment, overall_safety
## P value adjustment: tukey method for comparing a family of 3 estimates

library(ggplot2)
library(gridExtra)
library(MASS)

##
## Attaching package: 'MASS'

## The following object is masked from 'package:gtsummary':
## 
##     select

## The following object is masked from 'package:patchwork':
## 
##     area

## The following object is masked from 'package:dplyr':
## 
##     select

```

```

# Fit your model
model <- lm(skinVL_change ~ treatment + period + sequence_ind, data = merged)

# Extract residuals and fitted values
residuals <- residuals(model)
fitted_values <- fitted(model)
standardized_residuals <- rstandard(model)

# 1. Q-Q Plot: Scaled Residuals
p1 <- ggplot(data.frame(residuals = standardized_residuals), aes(sample = residuals)) +
  stat_qq(color = "gray50", size = 2) +
  stat_qq_line(color = "red", linewidth = 1) +
  labs(title = "Q-Q Plot: Scaled Residuals",
       x = "Theoretical Quantiles",
       y = "Sample Quantiles") +
  theme_minimal()

# 2. Histogram: Scaled Residuals
p2 <- ggplot(data.frame(residuals = standardized_residuals), aes(x = residuals)) +
  geom_histogram(aes(y = after_stat(density)), bins = 30, fill = "lightblue", color = "black") +
  stat_function(fun = dnorm, args = list(mean = 0, sd = 1),
                color = "red", linewidth = 1) +
  geom_density(color = "blue", linewidth = 1) +
  labs(title = "Histogram: Scaled Residuals",
       x = "Scaled Residuals",
       y = "Density",
       subtitle = "Red line = Normal(0,1), Blue line = Kernel density") +
  theme_minimal()

# 3. Residuals vs. Fitted Values
p3 <- ggplot(data.frame(fitted = fitted_values, residuals = standardized_residuals),
              aes(x = fitted, y = residuals)) +
  geom_point(color = "gray50", size = 2, alpha = 0.6) +
  geom_hline(yintercept = 0, color = "red", linetype = "dashed", linewidth = 1) +
  geom_smooth(method = "loess", se = FALSE, color = "red", linewidth = 1) +
  labs(title = "Residuals vs. Fitted Values",
       x = "Fitted Values",
       y = "Scaled Residuals") +
  theme_minimal()

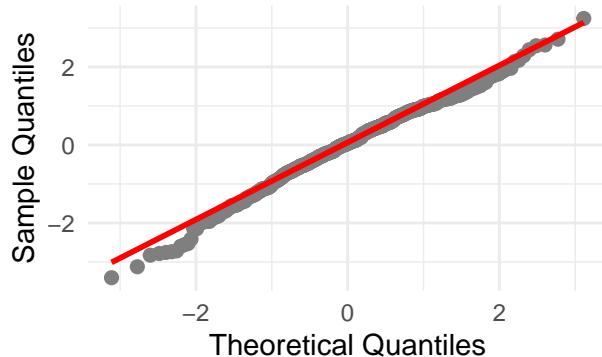
# 4. Residuals by Treatment
p4 <- ggplot(data.frame(treatment = merged$treatment, residuals = standardized_residuals),
              aes(x = treatment, y = residuals)) +
  geom_boxplot(fill = "lightblue", color = "black") +
  geom_hline(yintercept = 0, color = "red", linetype = "dashed", linewidth = 1) +
  labs(title = "Residuals by Treatment",
       x = "Treatment",
       y = "Scaled Residuals") +
  theme_minimal()

# Combine all plots
grid.arrange(p1, p2, p3, p4, ncol = 2)

```

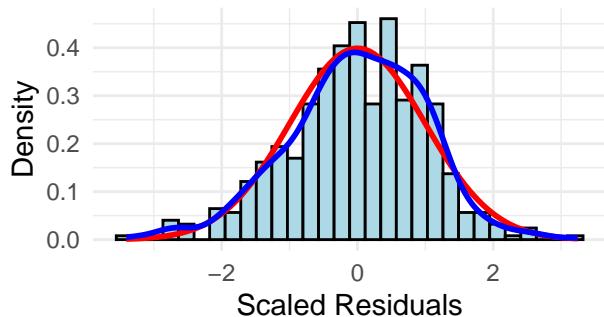
```
## `geom_smooth()` using formula = 'y ~ x'
```

Q-Q Plot: Scaled Residuals

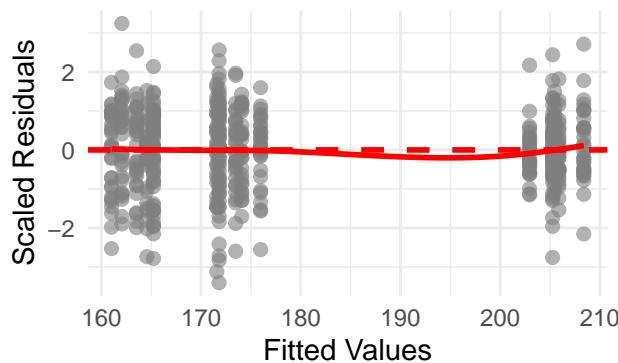


Histogram: Scaled Residuals

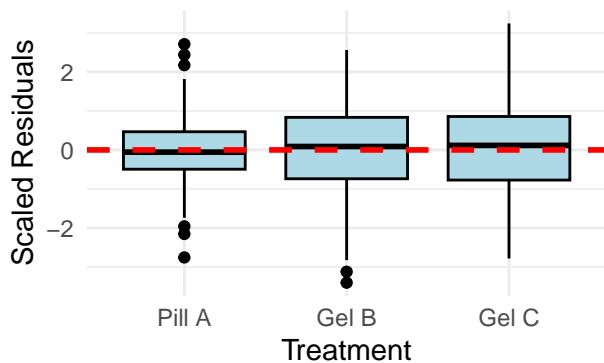
Red line = Normal(0,1), Blue line = Kernel



Residuals vs. Fitted Values



Residuals by Treatment



Linear Mixed Model

```
model_large <- lmer(skinVL_change ~ treatment + overall_adhere + overall_safety + sequence_ind +
                      period + (1 | ptid),
                      data = merged)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
model <- lmer(skinVL_change ~ treatment + overall_adhere + overall_safety + period + (1 | ptid),
               data = merged)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
anova(model, model_large)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: merged
```

```
## Models:
```

```
## model: skinVL_change ~ treatment + overall_adhere + overall_safety + period + (1 | ptid)
## model_large: skinVL_change ~ treatment + overall_adhere + overall_safety + sequence_ind + period + (1 | ptid)
##          npar    AIC    BIC  logLik -2*log(L)   Chisq Df Pr(>Chisq)
## model      9 5310.9 5349.6 -2646.5     5292.9
## model_large 11 5312.1 5359.4 -2645.1     5290.1 2.7872  2     0.2482
```

```

summary(model)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: skinVL_change ~ treatment + overall_adhere + overall_safety +
##     period + (1 | ptid)
## Data: merged
##
## REML criterion at convergence: 5269.4
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -3.4333 -0.6760  0.0316  0.6391  3.5333
##
## Random effects:
##   Groups   Name        Variance Std.Dev.
##   ptid     (Intercept) 0       0.00
##   Residual           1072    32.74
## Number of obs: 540, groups: ptid, 180
##
## Fixed effects:
##                               Estimate Std. Error      df t value
## (Intercept)                105.6003   9.9680 533.0000 10.594
## treatmentGel B             -24.4131   3.6314 533.0000 -6.723
## treatmentGel C             -36.9951   3.5493 533.0000 -10.423
## overall_adhere              1.8154   0.2753 533.0000  6.594
## overall_safetyNo Adverse Event 55.7940   5.2166 533.0000 10.696
## periodperiod2               0.8567   3.4535 533.0000  0.248
## periodperiod3               1.6225   3.4507 533.0000  0.470
##                               Pr(>|t|)
## (Intercept) < 0.0000000000000002 ***
## treatmentGel B          0.000000000046 ***
## treatmentGel C          < 0.00000000000002 ***
## overall_adhere           0.000000000103 ***
## overall_safetyNo Adverse Event < 0.00000000000002 ***
## periodperiod2            0.804
## periodperiod3            0.638
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##   (Intr) trtmGB trtmGC ovrl1_ ov_NAE prdpr2
## treatmntG1B -0.429
## treatmntG1C -0.354  0.534
## overall_dhr -0.826  0.311  0.234
## ovrl1_sfNAE -0.627  0.079  0.033  0.206
## periodperd2 -0.137 -0.008 -0.005 -0.025 -0.037
## periodperd3 -0.180  0.001  0.001  0.004  0.009  0.499
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')

```

Comparison	Beta	95% CI	p-value
Intercept (Pill A, Period 1, Sequence 0, No Adherence or AE)	105.6	86.0, 125.2	<0.001
Treatment			<0.001
Gel B vs Pill A	-24.4	-31.5, -17.3	
Gel C vs Pill A	-37.0	-44.0, -30.0	
Additional Day of Adherence	1.8	1.3, 2.4	<0.001
Adverse Event			<0.001
No Adverse Event	55.8	45.5, 66.0	
Period			0.9
Period 2 vs Period 1	0.9	-5.9, 7.6	
Period 3 vs Period 1	1.6	-5.2, 8.4	

Abbreviation: CI = Confidence Interval

```
tbl_regression(
  model,
  intercept = TRUE,
  estimate_fun = function(x) style_number(x, digits = 1)) %>%
  add_global_p() %>%
  modify_table_body(~.x %>%
    mutate(label = case_when(
      label == "(Intercept)" ~ "Intercept (Pill A, Period 1, Sequence 0, No Adherence or AE)",
      label == "Gel B" ~ "Gel B vs Pill A",
      label == "Gel C" ~ "Gel C vs Pill A",
      label == "period2" ~ "Period 2 vs Period 1",
      label == "period3" ~ "Period 3 vs Period 1",
      label == "treatment" ~ "Treatment", # Remove header
      label == "period" ~ "Period", # Remove header
      label == "Pill A" ~ NA_character_,
      label == "period1" ~ NA_character_,
      label == "sequence_ind" ~ "Sequence",
      label == "Adverse Event" ~ NA_character_,
      label == "0" ~ NA_character_,
      label == "1" ~ "1 vs 0",
      label == "2" ~ "2 vs 0",
      label == "overall_adhere" ~ "Additional Day of Adherence",
      label == "overall_safety" ~ "Adverse Event",
      TRUE ~ label)) %>%
    filter(!is.na(label))) %>%
  modify_header(label ~ "***Comparison***", estimate ~ "***Beta***") %>%
  bold_p(t = 0.05)
```

```
library(emmeans)
emm_treatment <- emmeans(model, ~ treatment)
summary(emm_treatment)
```

```
##   treatment emmean    SE df lower.CL upper.CL
##   Pill A     178 3.43 525      171     185
##   Gel B     154 3.21 525      147     160
```

```

##  Gel C      141 3.29 527      135      147
##
## Results are averaged over the levels of: overall_safety, period
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95

# Pairwise comparisons for treatment
pairs(emm_treatment)

##   contrast      estimate    SE  df t.ratio p.value
## Pill A - Gel B     24.4 3.63 412   6.721 <.0001
## Pill A - Gel C     37.0 3.55 388  10.422 <.0001
## Gel B - Gel C     12.6 3.47 359   3.631  0.0009
##
## Results are averaged over the levels of: overall_safety, period
## Degrees-of-freedom method: kenward-roger
## P value adjustment: tukey method for comparing a family of 3 estimates

# Marginal means for overall_adhere
emm_adhere <- emmeans(model, ~ overall_adhere)
summary(emm_adhere)

##   overall_adhere emmean    SE  df lower.CL upper.CL
##                 24    158 2.6 401      152      163
##
## Results are averaged over the levels of: treatment, overall_safety, period
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95

# Marginal means for overall_safety
emm_safety <- emmeans(model, ~ overall_safety)
summary(emm_safety)

##   overall_safety   emmean    SE  df lower.CL upper.CL
## Adverse Event     130 5.00 485      120      139
## No Adverse Event  185 1.47 196      182      188
##
## Results are averaged over the levels of: treatment, period
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95

# Marginal means for period
emm_period <- emmeans(model, ~ period)
summary(emm_period)

##   period   emmean    SE  df lower.CL upper.CL
## period1     157 3.25 528      150      163
## period2     158 3.34 528      151      164
## period3     158 3.23 528      152      165
##
## Results are averaged over the levels of: treatment, overall_safety
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95

```

```

pairs(emm_period)

## contrast      estimate   SE df t.ratio p.value
## period1 - period2 -0.857 3.45 356 -0.248 0.9666
## period1 - period3 -1.623 3.45 355 -0.470 0.8853
## period2 - period3 -0.766 3.45 356 -0.222 0.9733
##
## Results are averaged over the levels of: treatment, overall_safety
## Degrees-of-freedom method: kenward-roger
## P value adjustment: tukey method for comparing a family of 3 estimates

resid_scaled <- residuals(model, scaled = TRUE)
fitted_vals <- fitted(model)
random_intercepts <- ranef(model)$ptid[, 1]
model_data <- model@frame

# Plot 1: Q-Q Plot for Scaled Residuals with confidence bands
p1 <- ggplot(data.frame(resid = resid_scaled), aes(sample = resid)) +
  stat_qq(alpha = 0.5) +
  stat_qq_line(color = "red", linewidth = 1) +
  labs(title = "Q-Q Plot: Scaled Residuals",
       x = "Theoretical Quantiles",
       y = "Sample Quantiles") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))

# Plot 2: Histogram of Scaled Residuals with rug
p2 <- ggplot(data.frame(resid = resid_scaled), aes(x = resid)) +
  geom_histogram(aes(y = after_stat(density)), bins = 30,
                 fill = "skyblue", color = "black", alpha = 0.7) +
  geom_density(color = "blue", linewidth = 1) +
  stat_function(fun = dnorm, args = list(mean = 0, sd = 1),
                color = "red", linewidth = 1, linetype = "dashed") +
  geom_rug(alpha = 0.3) +
  labs(title = "Histogram: Scaled Residuals",
       x = "Scaled Residuals",
       y = "Density",
       caption = "Red line = Normal(0,1), Blue line = Kernel density") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))

# Plot 3: Q-Q Plot for Random Intercepts
p3 <- ggplot(data.frame(re = random_intercepts), aes(sample = re)) +
  stat_qq(alpha = 0.5) +
  stat_qq_line(color = "red", linewidth = 1) +
  labs(title = "Q-Q Plot: Random Intercepts",
       x = "Theoretical Quantiles",
       y = "Sample Quantiles") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))

# Plot 4: Histogram of Random Intercepts with rug
p4 <- ggplot(data.frame(re = random_intercepts), aes(x = re)) +

```

```

geom_histogram(aes(y = after_stat(density)), bins = 20,
               fill = "lightgreen", color = "black", alpha = 0.7) +
  geom_density(color = "darkgreen", linewidth = 1) +
  stat_function(fun = dnorm,
                args = list(mean = 0, sd = sd(random_intercepts)),
                color = "red", linewidth = 1, linetype = "dashed") +
  geom_rug(alpha = 0.3) +
  labs(title = "Histogram: Random Intercepts",
       x = "Random Intercepts (ptid)",
       y = "Density",
       caption = "Red line = Normal fit, Green line = Kernel density") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))

# Plot 5: Residuals vs. Fitted Values with jitter
p5 <- ggplot(data.frame(fitted = fitted_vals, resid = resid_scaled),
              aes(x = fitted, y = resid)) +
  geom_point(alpha = 0.5, position = position_jitter(width = 0.5, height = 0)) +
  geom_hline(yintercept = 0, color = "red", linetype = "dashed", linewidth = 1) +
  #geom_smooth(se = TRUE, color = "blue", method = "loess", linewidth = 1) +
  labs(title = "Residuals vs. Fitted Values",
       x = "Fitted Values",
       y = "Scaled Residuals") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))

# Plot 6: Residuals by Treatment with individual points
p6 <- ggplot(data.frame(treatment = model_data$treatment,
                        resid = resid_scaled),
              aes(x = treatment, y = resid)) +
  geom_boxplot(fill = "lightblue", alpha = 0.7, outlier.shape = NA) +
  #geom_jitter(width = 0.2, alpha = 0.3, color = "darkblue") +
  geom_hline(yintercept = 0, color = "red", linetype = "dashed", linewidth = 1) +
  labs(title = "Residuals by Treatment",
       x = "Treatment",
       y = "Scaled Residuals") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))

# Display all plots
grid.arrange(p1, p2, p3, p4, p5, p6, ncol = 2)

```

