MethodsX 8 (2021) 101313

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/mex

Method Article

Seasonality, mediation and comparison (SMAC) methods to identify influences on lung function decline

Emrah Gecili[a,1], Anushka Palipana[a,b,1], Cole Brokamp[a,c], Rui Huang[b], Eleni-Rosalina Andrinopoulou[d], Teresa Pestian[a], Erika Rasnick[a], Ruth H. Keogh[e], Yizhao Ni[c,f], John P. Clancy[c,g,h], Patrick Ryan[a,c], Rhonda D. Szczesniak[a,c,h,*]

[a] *Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave, Cincinnati, OH 45229, United States*
[b] *Division of Statistics and Data Science, Department of Mathematics, University of Cincinnati, 155B McMicken Hall, Cincinnati, OH, United States*
[c] *Department of Pediatrics, University of Cincinnati, 3333 Burnet Ave, Cincinnati, OH, United States*
[d] *Department of Biostatistics, Erasmus Medical Center, Dr. Molewaterplein 40, 3015 GD, Rotterdam, Netherlands*
[e] *London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT United Kingdom*
[f] *Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave, Cincinnati, OH, United States*
[g] *Cystic Fibrosis Foundation, 4550 Montgomery Ave, Bethesda, MD, United States*
[h] *Division of Pulmonary Medicine, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave, Cincinnati, OH, United States*

A B S T R A C T

This study develops a comprehensive method to assess seasonal influences on a longitudinal marker and compare estimates between cohorts. The method extends existing approaches by (i) combining a sine-cosine model of seasonality with a specialized covariance function for modeling longitudinal correlation; (ii) performing mediation analysis on a seasonality model. An example dataset and R code are provided. The bundle of methods is referred to as seasonality, mediation and comparison (SMAC). The case study described utilizes lung function as the marker observed on a cystic fibrosis cohort but SMAC can be used to evaluate other markers and in other disease contexts. Key aspects of customization are as follows.

- This study introduces a novel seasonality model to fit trajectories of lung function decline and demonstrates how to compare this model to a conventional model in this context.
- Steps required for mediation analyses in the seasonality model are shown.

* Corresponding author at: Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave, Cincinnati, OH 45229, United States.
E-mail address: Rhonda.Szczesniak@cchmc.org (R.D. Szczesniak).
[1] These authors contributed equally to this work

https://doi.org/10.1016/j.mex.2021.101313

- The necessary calculations to compare seasonality models between cohorts, based on estimation coefficients, are derived in the study.

ARTICLE INFO
Method name: Seasonality, mediation and comparison (SMAC)
Keywords: Climate, Cystic fibrosis, Sine wave model, Mediation analysis, Longitudinal data analysis, Lung disease, Respiratory, Temporal analysis, Time series
Article history: Received 12 January 2021; Accepted 15 March 2021; Available online 21 March 2021

Specifications Table

Subject Area:	Environmental Science
More specific subject area:	*Statistical Science*
Method name:	*Seasonality, mediation and comparison (SMAC)*
Name and reference of original method:	(1) Qvist T, Schluter DK, Rajabzadeh V, Diggle PJ, Pressler T, Carr SB, Taylor-Robinson D. Seasonal fluctuation of lung function in cystic fibrosis: A national register-based study in two northern European populations. J Cyst Fibros. 2019;18(3):390–5. Epub 2018/10/23. doi:10.1016/j.jcf.2018.10.006. (2) Tingley D, Yamamoto, T., Hirose, K., Keele, L., & Imai, K. Mediation: R package for causal mediation analysis. Journal of Statistical Software. 2014;59(5)
Resource availability:	• *NCEP North American Regional Reanalysis: NARR:* https://psl.noaa.gov/data/gridded/data.narr.html • *Jittered clinical data provided as supplemental material for analytic steps from the article* • *Available R software libraries:* • *'base':* https://www.rdocumentation.org/packages/base *(version 4.0.2)* • *'mediation':* https://cran.r-project.org/web/packages/mediation/mediation.pdf *(version 4.5.0)* • *'nlme':* https://cran.r-project.org/web/packages/nlme/nlme.pdf *(version 3.1-140)* • *'dplyr':* https://mran.microsoft.com/web/packages/dplyr/dplyr.pdf *(version 1.0.2)* • *'mvtnorm':*https://mirrors.linux.iu.edu/CRAN/web/packages/mvtnorm/mvtnorm.pdf *(version 1.1-1)* • *'ggpubr':* https://cran.r-project.org/web/packages/ggpubr/ggpubr.pdf *(version 0.4.0)* • *'ggplot2':* https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf *(version 3.4.0)* • *'ellipse':* https://cran.r-project.org/web/packages/ellipse/ellipse.pdf *(version 0.4.1)* • *'lme4':* https://cran.r-project.org/web//packages/lme4/lme4.pdf *(version 1.1-23)* • *'boot':* https://cran.r-project.org/web/packages/boot/boot.pdf • *(version 1.3-22)*

Method details

Data description

The methods described in this research can be used to estimate influences on lung function decline accounting for seasonality, temperature and potential mediating effects of respiratory pathogens. Methods development was motivated by clinical encounter and temperature data acquired from people with cystic fibrosis (CF) based on care received at a Midwest Cystic Fibrosis Care Center and their regional area of residence [1]. The data dictionary is provided in Table 1, including variable names subsequently referenced in implementation code. The merged dataset was structured as one record per person, per clinical encounter.

A four-level categorical variable was created according to when a given clinical encounter occurred. December, January and February were coded as winter; March, April and May corresponded to spring; June, July and August corresponded to summer; September, October and November corresponded to

Table 1
CF Seasonality Data Dictionary.

Feature	Variable name	Description	Frequency recorded
Subject ID	MED.REC	Unique identifier used to index subjects	Repeated for each subject record
Birth cohort	cohort	Birth cohort	Time invariant; repeated for each subject record
Gender	Gender	Gender	Time invariant; repeated for each subject record
Cystic fibrosis - related diabetes	CFRD	Diagnosis of cystic fibrosis related diabetes	Time invariant; repeated for each subject record
Medicaid insurance use	MEDICAID	Medicaid insurance use which corresponds to low socioeconomic status	Time invariant; repeated for each subject record
Pseudomonas aeruginosa	PA	Culturing positive for *Pseudomonas aeruginosa* infection	Time varying; recorded at each encounter
Methicillin-resistant staphylococcus aureus	MRSA	Culturing positive for Methicillin-resistant *Staphylococcus aureus* infection	Time varying; recorded at each encounter
Genotype	F508del	F508del homozygous, heterozygous or neither/unknown	Time invariant; repeated for each subject record
Pancreatic insufficiency	PancreaticEnzymes	Use of pancreatic enzymes	Time invariant; repeated for each subject record
Daily temperature	temp	Daily mean air temperature in Kelvin	
Percent predicted FEV_1	FEV1	Percent predicted forced expiratory volume in 1 s (FEV_1)	Time varying; recorded at each encounter
Visit age	visit_age	Age at the clinic visit	Time varying; recorded at each encounter
Season	season	Season corresponding to each clinic visits	Time varying; recorded at each encounter

autumn. Another variable was created to label the encounter day according day of a given year, with January 1st being day zero and December 31st being day 364/365.

While we are unable to share the clinical data from the CF center, we provide a jittered dataset using the 'jitter' function in 'base' *R* package. This function adds a small amount of noise to observed data, and it is used in this study to mask demographic and clinical data. The temperature data are not restricted and therefore are included with the accompanying dataset. Although running the implementation code below for the included dataset will not exactly reproduce findings from original study data, results are sufficiently close for illustration purposes.

Temperature data acquisition

Temperature data were obtained for the overall geographic study region, which was a catchment area for an academic medical center located in the Midwestern region of the United States. The CF care center from which the cohort's demographic and clinical data were obtained was located within Cincinnati Children's Hospital Medical Center (CCHMC). Daily mean air temperature (Kelvin) was obtained from the North American Regional Reanalysis (NARR). The data were taken as the average values from all 32×32 sq km grids ($n = 9$) that covered the seven county (OH; Hamilton, Clermont, Butler, Warren; KY: Boone, Kenton, Campbell) catchment region for the CF care center to create a daily time series. Thus, temperature was assigned to each patient (who was assumed to live within the CCHMC catchment area) based solely on date. The details of methods used for creating the temperature time series have been described [2].

R *packages*

We utilize the following packages in *R* software (version 3.6.1) (*R* Foundation for Statistical Computing, Vienna, Austria). Each package can be downloaded using the links provided under

resource availability, and we include references below for each *R* package. Specific versions utilized in the article are provided under the Resource availability section.

```
#load required packages
#reason for loading each package provided and reference
library(nlme) # fitting models (3)
library(dplyr) # to fasten data manipulation (4)
library(mvtnorm) # generates data from multivariate normal distribution
(5)
library(ggpubr) # creating panel of figures (6)
library(ggplot2) # creating figures (7)
library(ellipse) # creating joint confidence intervals/ ellipses for
amplitude and horizontal shift (8)
library(mediation) # performing mediation analysis (9)
library(lme4) # fitting lme models for mediation analysis (10)
library(boot) #used for bootstrapping to get confidence intervals for
estimated rate of change by seasons (11)
```

Sourcing the data

The data provided with this article as supplemental material may be sourced using the following commands:

```
#loading the data set
d <- read.csv('Jittered_data_seasonalityMS.csv')
```

We can create the additional variables described in Table 1 as follows:

```
# use winter as the reference level in following output
# F508del use Homozygous as reference level
d$season <- factor(d$season)
d <- within(d, season <- relevel(season, ref = 'winter'))
d$F508del <- factor(d$F508del)
d <- within(d, F508del <- relevel(F508del, ref = 'Homozygous'))
d <- within(d, cohort <- relevel(cohort, ref = "4"))
#temperature in Celsius
d$temp <- d$air.2 m - 273.15
```

Linear mixed effects model with seasonality as a class variable

The first model assumes seasonality impacts lung function, both overall and in terms of rate of decline, in a linear fashion. To account for these impacts, we add seasonality as a class variable. The resulting model equation can be expressed as:

$$y_{ij} = x_i'\beta_1 + x_i' t_{ij}\beta_2 + \alpha_1 \text{Season}_{ij} + \alpha_2 \text{Season}_{ij} \cdot t_{ij} + V_i + W_i(t_{ij}) + Z_{ij} \quad (1)$$

In this model, which we refer to as model (1), y_{ij} is the observed FEV_1 for the ith individual at the jth measurement time t_{ij}; x_i is the vector of covariate values (possibly time-varying) for individual i. β_1 and β_2 are the vector of main and interaction effects for demographic and clinical covariates, respectively. Season_{ij} is a vector of indicators for season and denotes the season in which the jth measurement was made for individual i. The main and interaction effects for the categorical variable

season are denoted by parameter vectors α_1 and α_2. These terms were included in the model through a series of indicator variables to represent the different categories, where the reference category was winter. The interaction between season and encounter time term allows for distinct FEV_1 trajectories over age according to season. The term V_i is a subject-specific random intercept, allowing fluctuation from an individual FEV_1 trajectory relative to the population-level trajectory; $W_i(t_{ij})$ represents a stochastic process accounting for within-subject correlation assuming an exponential covariance function; that is, the covariance matrix for repeated measures within an individual follows an exponentially decaying correlation with increasing time difference; Z_{ij} corresponds to measurement error and residual variation.

Implementation with the 'nlme' package

The terms in Eq. (1) were estimated using the 'nlme' package in *R* [3]. Specific steps are described below with necessary *R* code:

We first create the covariance structure that we will implement, which is based on terms from Eq. (1), and it will be utilized in the subsequent modeling under Eq. (2):

```
#defining the exponential correlation structure
cs1Exp <- corExp(form = ~ visit_age|MED.REC,fixed=F,nugget = T)
cs1Exp <- Initialize(cs1Exp, d)

#fitting the model with seasons as class variable equation (1)
M_season <- lme(FEV1~season*visit_age+
              (Gender+CFRD+MEDICAID+
                 cohort+PA+MRSA+F508del+
                 PancreaticEnzymes)*visit_age, data=d,
              random=~1|MED.REC,method="ML",correlation = cs1Exp)

summary(M_season) #model summary
intervals(M_season) #getting CIs for the parameter estimates
```

Parameter estimates and the corresponding 95% confidence intervals for model (1) can be obtained with above *R* commands. We estimated the rate of change (the first derivative of model (1) with respect to the time variable (visit_age)) for each season and visualized the evolution in FEV_1 over time for the jittered data below.

```
#converting variables to numeric to obtain evaluation in FEV1 and rate of
change for each season and create Fig. 1
age.unique<-sort(unique(d$visit_age))
d$Gender <- ifelse(d$Gender =="F", 0, 1)
d$CFRD <- ifelse(d$CFRD == "Positive", 1, 0)
d$MRSA <- ifelse(d$MRSA == "Yes", 1, 0)
d$PA <- ifelse(d$PA == "Yes", 1, 0)
d$F508Heter <- ifelse(d$F508del=='Heterozygous',1,0)
d$F508Non <- ifelse(d$F508del=='no copies',1,0)
d$pancEnzymes <- ifelse(d$PancreaticEnzymes == "Using", 1, 0)
d$MEDICAID <- ifelse(d$MEDICAID=="Medicaid", 1,0)

#fixed effects part of model (1)
coeff.regressor <- M_season$coefficients$fixed

#getting means for each cohort (cc1 for cohort1; cohort4 is reference)
cc1<-as.numeric(summary(d$cohort)[2])/sum(as.numeric(summary(d$cohort)))
cc2<-as.numeric(summary(d$cohort)[3])/sum(as.numeric(summary(d$cohort)))
cc3<-as.numeric(summary(d$cohort)[4])/sum(as.numeric(summary(d$cohort)))
```

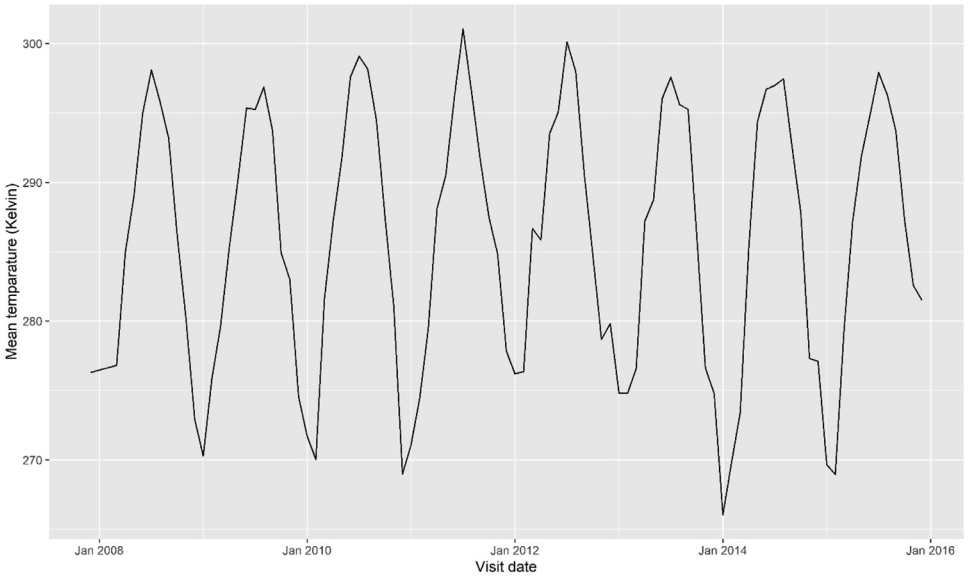

Fig. 1. Daily mean temperature over study period.

```
#obtaining evolution in FEV1 for all seasons
f_winter<-
coeff.regressor[1]+coeff.regressor[5]*age.unique+coeff.regressor[6]*mean
(d$Gender)+coeff.regressor[7]*mean(d$CFRD)+coeff.regressor[8]*mean(d$MED
ICAID)+coeff.regressor[9]*cc1+coeff.regressor[10]*cc2+coeff.regressor[11
]*cc3+coeff.regressor[12]*mean(d$PA)+coeff.regressor[13]*mean(d$MRSA)+
coeff.regressor[14]*mean(d$F508Heter)+coeff.regressor[15]*mean(d$F508Non)
+coeff.regressor[16]*mean(d$pancEnzymes)+coeff.regressor[20]*age.unique*
mean(d$Gender)+coeff.regressor[21]*age.unique*mean(d$CFRD)+coeff.regressor
[22]*age.unique*mean(d$MEDICAID)+coeff.regressor[23]*age.unique*cc1+
coeff.regressor[24]*age.unique*cc2+coeff.regressor[25]*age.unique*cc3+
coeff.regressor[26]*age.unique*mean(d$PA)+coeff.regressor[27]*age.unique*
mean(d$MRSA)+coeff.regressor[28]*age.unique*mean(d$F508Heter)+coeff.
regressor[29]*age.unique*mean(d$F508Non)+coeff.regressor[30]*age.unique*
mean(d$pancEnzymes)

f_autumn<-
coeff.regressor[1]+coeff.regressor[2]+coeff.regressor[5]*age.unique+
coeff.regressor[6]*mean(d$Gender)+coeff.regressor[7]*mean(d$CFRD)+coeff.
regressor[8]*mean(d$MEDICAID)+coeff.regressor[9]*cc1+coeff.regressor[10]*
cc2+coeff.regressor[11]*cc3+coeff.regressor[12]*mean(d$PA)+coeff.
regressor[13]*mean(d$MRSA)+coeff.regressor[14]*mean(d$F508Heter)+coeff.
regressor[15]*mean(d$F508Non)+coeff.regressor[16]*mean(d$pancEnzymes)+
coeff.regressor[17]*age.unique+coeff.regressor[20]*age.unique*mean
(d$Gender)+coeff.regressor[21]*age.unique*mean(d$CFRD)+coeff.regressor
[22]*age.unique*mean(d$MEDICAID)+coeff.regressor[23]*age.unique*cc1+
coeff.regressor[24]*age.unique*cc2+coeff.regressor[25]*age.unique*cc3+
coeff.regressor[26]*age.unique*mean(d$PA)+coeff.regressor[27]*age.unique*
mean(d$MRSA)+coeff.regressor[28]*age.unique*mean(d$F508Heter)+coeff.
```

```
regressor[29]*age.unique*mean(d$F508Non)+coeff.regressor[30]*age.unique*
mean(d$pancEnzymes)

f_spring<-
coeff.regressor[1]+coeff.regressor[3]+coeff.regressor[5]*age.unique+
coeff.regressor[6]*mean(d$Gender)+coeff.regressor[7]*mean(d$CFRD)+coeff.
regressor[8]*mean(d$MEDICAID)+coeff.regressor[9]*cc1+coeff.regressor[10]*
cc2+coeff.regressor[11]*cc3+coeff.regressor[12]*mean(d$PA)+coeff.
regressor[13]*mean(d$MRSA)+coeff.regressor[14]*mean(d$F508Heter)+coeff.
regressor[15]*mean(d$F508Non)+coeff.regressor[16]*mean(d$pancEnzymes)+
coeff.regressor[18]*age.unique+coeff.regressor[20]*age.unique*mean
(d$Gender)+coeff.regressor[21]*age.unique*mean(d$CFRD)+coeff.regressor
[22]*age.unique*mean(d$MEDICAID)+coeff.regressor[23]*age.unique*cc1+
coeff.regressor[24]*age.unique*cc2+coeff.regressor[25]*age.unique*cc3+
coeff.regressor[26]*age.unique*mean(d$PA)+coeff.regressor[27]*age.unique*
mean(d$MRSA)+coeff.regressor[28]*age.unique*mean(d$F508Heter)+coeff.
regressor[29]*age.unique*mean(d$F508Non)+coeff.regressor[30]*age.unique*
mean(d$pancEnzymes)

f_summer<-
coeff.regressor[1]+coeff.regressor[4]+coeff.regressor[5]*age.unique+
coeff.regressor[6]*mean(d$Gender)+coeff.regressor[7]*mean(d$CFRD)+coeff.
regressor[8]*mean(d$MEDICAID)+coeff.regressor[9]*cc1+coeff.regressor[10]*
cc2+coeff.regressor[11]*cc3+coeff.regressor[12]*mean(d$PA)+coeff.
regressor[13]*mean(d$MRSA)+coeff.regressor[14]*mean(d$F508Heter)+coeff.
regressor[15]*mean(d$F508Non)+coeff.regressor[16]*mean(d$pancEnzymes)+
coeff.regressor[19]*age.unique+coeff.regressor[20]*age.unique*mean
(d$Gender)+coeff.regressor[21]*age.unique*mean(d$CFRD)+coeff.regressor
[22]*age.unique*mean(d$MEDICAID)+coeff.regressor[23]*age.unique*cc1+
coeff.regressor[24]*age.unique*cc2+coeff.regressor[25]*age.unique*cc3+
coeff.regressor[26]*age.unique*mean(d$PA)+coeff.regressor[27]*age.unique*
mean(d$MRSA)+coeff.regressor[28]*age.unique*mean(d$F508Heter)+coeff.
regressor[29]*age.unique*mean(d$F508Non)+coeff.regressor[30]*age.unique*
mean(d$pancEnzymes)

#obtaining rate of change (1st derivative) by seasons, the derivative
#of model (1) with respect to time variable (visit_age)
d_winter <-
coeff.regressor[5]+coeff.regressor[20]*mean(d$Gender)+coeff.regressor[21]
*mean(d$CFRD)+coeff.regressor[22]*mean(d$MEDICAID)+coeff.regressor[23]*
cc1+coeff.regressor[24]*cc2+coeff.regressor[25]*cc3+coeff.regressor[26]*
mean(d$PA)+coeff.regressor[27]*mean(d$MRSA)+coeff.regressor[28]*mean
(d$F508Heter)+coeff.regressor[29]*mean(d$F508Non)+coeff.regressor[30]*
mean(d$pancEnzymes)

d_autumn <-
coeff.regressor[5]+coeff.regressor[17]+coeff.regressor[20]*mean(d$Gender)
+coeff.regressor[21]*mean(d$CFRD)+coeff.regressor[22]*mean(d$MEDICAID)+
coeff.regressor[23]*cc1+coeff.regressor[24]*cc2+coeff.regressor[25]*cc3+
coeff.regressor[26]*mean(d$PA)+coeff.regressor[27]*mean(d$MRSA)+coeff.
regressor[28]*mean(d$F508Heter)+coeff.regressor[29]*mean(d$F508Non)+
coeff.regressor[30]*mean(d$pancEnzymes)

d_spring <-
coeff.regressor[5]+coeff.regressor[18]+coeff.regressor[20]*mean(d$Gender)
+coeff.regressor[21]*mean(d$CFRD)+coeff.regressor[22]*mean(d$MEDICAID)+
```

```
coeff.regressor[23]*cc1+coeff.regressor[24]*cc2+coeff.regressor[25]*cc3
+coeff.regressor[26]*mean(d$PA)+coeff.regressor[27]*mean(d$MRSA)+coeff.
regressor[28]*mean(d$F508Heter)+coeff.regressor[29]*mean(d$F508Non)+
coeff.regressor[30]*mean(d$pancEnzymes)

d_summer <-
coeff.regressor[5]+coeff.regressor[19]+coeff.regressor[20]*mean(d$Gender)
+coeff.regressor[21]*mean(d$CFRD)+coeff.regressor[22]*mean(d$MEDICAID)+
coeff.regressor[23]*cc1+coeff.regressor[24]*cc2+coeff.regressor[25]*cc3+
coeff.regressor[26]*mean(d$PA)+coeff.regressor[27]*mean(d$MRSA)+coeff.
regressor[28]*mean(d$F508Heter)+coeff.regressor[29]*mean(d$F508Non)+
coeff.regressor[30]*mean(d$pancEnzymes)

#creating a figure which shows evolution in FEV1 by seasons over time
plot(age.unique,f_winter,lty=1,typ="l", xlab="Age",ylab="FEV1
(%predicted)",main="Evolution in FEV1",col="1",lwd=3)
lines(age.unique,f_spring,lty=2,col="2",lwd=3)
lines(age.unique,f_autumn,lty=3,col="3",lwd=3)
lines(age.unique,f_summer,lty=4,col="4",lwd=3)
text(8101, d_winter,cex = 0.8)
text(7,96, d_spring,cex = 0.8,col="red")
text(16,82, d_autumn,cex = 0.8,col="blue")
text(18.1,84,d_summer,cex = 0.8,col="green")
legend("topright", legend=c("Winter", "Spring","Autumn","Summer"),
       col=1:4,lty=1:4, lwd=3,cex=0.8,bty = "n")
```

The 95% confidence intervals (95% CIs) for estimated rates of change are obtained by a bootstrapping method [12]. The following *R* code obtains the bootstrapped CIs based on 1000 bootstrap replicates for only one of the seasons (winter), it can be obtained similarly for other seasons. For this bootstrapping method we utilized 'boot' package of *R*. We examined increased numbers of replicates, e.g., 5000 replicates, but found estimates were consistent; therefore, we present our approach using 1000 replicates. Fig. 2 shows the fitted lines by season without CIs but these could be added using the above commands.

```
#function to compute rate of change for winter for model (1)
rc_win <- function(formula, data, indices) {
   dd <- subset(data, MED.REC%in% unique(MED.REC)[indices])
   cs1Exp <- corExp(form = ~ visit_age|MED.REC,fixed=F,nugget = T)
   cs1Exp <- Initialize(cs1Exp, dd)# allows boot to select sample
   fit <- lme(formula, data=dd,random=~1|MED.REC,method="ML",
     correlation = cs1Exp)

   coeff.regressor<-fit$coefficients$fixed

   d_win <-coeff.regressor[5]+coeff.regressor[20]*mean(d$Gender2)+
coeff.regressor[21]*mean(d$CFRD2)+coeff.regressor[22]*mean(d$MEDICAID2)+
coeff.regressor[23]*cc1+coeff.regressor[24]*cc2+coeff.regressor[25]*cc3+
coeff.regressor[26]*mean(d$PA2)+coeff.regressor[27]*mean(d$MRSA2)+
coeff.regressor[28]*mean(d$F508Heter2)+coeff.regressor[29]*mean
(d$F508Non2)+coeff.regressor[30]*mean(d$pancEnzymes2)
   return(d_win[1])
}

#bootstrapping to get CIs, R is number of bootstrap replicates
results_win <- boot(data=d, statistic=rc_win,
                    R = 1000, formula=FEV1~season*visit_age+
                      (Gender+CFRD+MEDICAID+cohort+PA+MRSA+F508del+
                         PancreaticEnzymes)*visit_age)
```

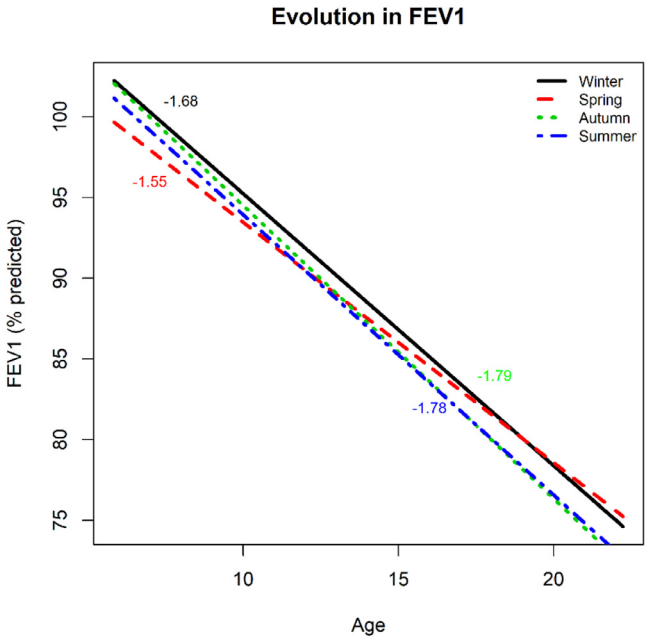

Fig. 2. Estimated population evolution in% predicted FEV_1 (y-axis) over age (x-axis) by season (black, red, green and blue lines are for winter, spring, autumn, and summer, respectively) for the Cincinnati cohort for categorized seasonality (Model 1) for the jittered data. The corresponding estimated rate of change in% predicted FEV_1 are reported in text with color corresponding to a given season. If viewing in black and white, the corresponding patterns are winter (solid line); spring (dashed line); autumn (dotted line); summer (dot-dash line).

```
#view results
results_win
plot(results_win)
CI_win <- boot.ci(results_win, type = "norm") #95% bootstrapped CI
```

Generating fit statistics

The parameter estimates of the model (1) are already obtained in Section 2.1; thus, we can now easily obtain fit statistics, including the Akaike information criterion (AIC), Bayesian information criterion, and −2*log-likelihood (−2LL). Additionally, prediction accuracy metrics root mean square error (RMSE) and mean absolute error (MAE) are provided. For all these statistics, smaller values imply better model fit. Specific steps are described below with necessary R code:

```
#computing metrics in Table S1
AIC(M_season) # getting AIC by using AIC function from 'nlme'
BIC(M_season) # similarly get BIC
LL <- −2* logLik(M_season) #computing −2*log-likelihood
RMSE <- sqrt(mean(residuals(M_season)^2)) #gets RMSE
MAE <- mean(abs(residuals(M_season))) #gets MAE
Table_S1 <- round(data.frame(AIC,BIC,LL,RMSE,MAE),2) #reporting results
colnames(Table_S1) <- c("AIC","BIC","LL","RMSE","MAE")
Table_S_S1 #recalling the table for the results for the jittered data
```

```
> Table_S1
AIC            BIC         LL          RMSE       MAE
53190.44    53421.42    53122.44    12.54       9.35
```

These metrics can be computed for other models by changing the model name "M_season" in the R code provided above.

Sine wave model of seasonality

Following similar notation and the aforementioned published approach [13], the second model is a harmonic seasonal model of lung function, which can be expressed as:

$$y_{ij} = x_i'\beta_1 + x_i't_{ij}\beta_2 + \gamma_0 \sin\left(\frac{2\pi d_{ij}}{T}\right) + \gamma_1 \cos\left(\frac{2\pi d_{ij}}{T}\right) + V_i + W_i(t_{ij}) + Z_{ij} \quad (2)$$

Here, the seasonality variable is included in the model through the sine and cosine terms. In model (2), d_{ij} denotes the day of year on which the measurement was taken. T is the number of time periods described by one sine function over $(0, 2\pi)$ and we let $T = 365.25$ days. The terms γ_0 and γ_1 are the coefficients of the sine and cosine functions that can be used to obtain the amplitude $\alpha = \sqrt{\gamma_0^2 + \gamma_1^2}$ (which represents half the distance between the maximum and minimum values of the sine function) and the horizontal shift $\theta = \frac{T}{2\pi}\arctan(\frac{\gamma_1}{\gamma_0})$ (which represents the days of year on which the function reaches its maximum (peak of the estimated seasonal fluctuation) and minimum (dip of the estimated seasonal fluctuation) values).

Implementation with the 'nlme' package

```
#creating sine and cosine variables for model (2)
day <- d$yrday # defining day variable which shows the day of year

sine <- sin(2*pi*day/365.25)

cosine <- cos(2*pi*day/365.25)

#fitting the model with sine wave model equation (2)
Msin <- lme(FEV1~sine+cosine+

            (Gender+CFRD+MEDICAID+cohort+PA+MRSA+F508del+

               PancreaticEnzymes)*visit_age,
data=d,random=~1|MED.REC,method="ML",correlation = cs1Exp)
```

Model fit statistics and confidence intervals for parameter estimates can be analogously computed using illustrations provided above for model (1) in Section 2.1.1.

Adjustment in model for temperature

The impact of temperature adjustment is assessed by including daily temperature (in Celsius) as covariate in model (2).

$$y_{ij} = \lambda * temp_j + x_i'\beta_1 + x_i't_{ij}\beta_2 + \gamma_0 \sin\left(\frac{2\pi d_{ij}}{T}\right) + \gamma_1 \cos\left(\frac{2\pi d_{ij}}{T}\right) + V_i + W_i(t_{ij}) + Z_{ij} \quad (3)$$

The parameter estimates of this model can be obtained with the following *R* code:

```
Msin_t <-lme(FEV1~sine+cosine+temp+

          (Gender+CFRD+MEDICAID+cohort+PA+MRSA+

             F508del+PancreaticEnzymes)*visit_age,data = d,
           random=~1|MED.REC,method="ML",correlation = cs1Exp)
```

Inclusion of an interaction effect for daily temperature and age worsened the model fit and was therefore excluded from the final model.

Mediation testing steps

By conducting mediation analyses, we can determine the extent to which a binary variable, such as *Pseudomonas aeruginosa* respiratory infection, explains the observed association between seasonality and lung function.

Implementation using the 'mediation' package

We performed mediation analysis using the 'mediation' package in *R*[9]. We implemented the approach with jittered data only for the selected primary model (the sine wave model, Eq. (2)). The following code only shows the mediation analysis for PA (*Pseudomonas aeruginosa)* but it can be similarly performed for other binary variable too.

We first need an outcome model of the direct effect of independent variable (temperature) on our dependent variable (FEV_1), when controlling for our mediator, PA. Below, we used *lmer* function from 'lme4' package instead of *lme* function to estimate parameters of models (1-2) since *lme* function is not supported in mediation package. But, one should not that both *lmer* and *lme* return the same model estimates for the same model. We are not changing the terms of our model except the correlation structure that we ignore now for mediation analysis and this does not have a significant effect on our mediation analysis. The outcome model can be obtained by using the following *R* code:

```
model.y <- lmer(FEV1~sine+cosine+temp+PA+
                  Gender*visit_age+
                  CFRD*visit_age+
                  MEDICAID*visit_age+cohort*visit_age+
                  MRSA*visit_age+
                  F508del*visit_age+
                  PancreaticEnzymes*visit_age+(1|MED.REC),data=d)
```

Now, we implement the mediation model, which models PA, our mediating variable, as a function of temperature. Since PA is a binary variable, we used the *glmer* function with probit link to fit the model. The *glmer* function is available from the lme4 package in *R*.

```
#the model with the mediator predicted by the temperature using probit
link for modeling binary response-PA
model.m <- glmer(PA~sine+cosine+temp+
                  Gender*visit_age+
                  CFRD*visit_age+
                  MEDICAID*visit_age+cohort*visit_age+
                  MRSA*visit_age+
                  F508del*visit_age+
                  PancreaticEnzymes*visit_age+(1|MED.REC),
                  data = d,family = binomial(link = "probit"),
                  control=glmerControl(optimizer="bobyqa",
                                       optCtrl=list(maxfun=2e5)))
```

Then, we combine the fitted models with the *mediate* function, in order to conduct mediation analysis. This analysis provides estimated average casual mediation effect (ACME), average direct effect (ADE), total effect (direct effect + indirect effect), and the proportion of mediated effects by using a three-step procedure [14]. The *mediate* function additionally returns bootstrapped confidence intervals for the estimated effects and the corresponding *p*-values for the significance of the effects. The ACME, which is the indirect effect of the mediator, was used to evaluate statistical significance of the mediating impact of PA on the relationship between seasonality and FEV_1.

```
#combining outcome and mediation models to conduct mediation analysis

med <- mediate(model.m, model.y, treat="temp", mediator = "PA")

summary(med) #presents the results (ACME, ADE, etc.)
```

Below, we provide summary output from our mediation analysis for the jittered data. The ACME, which is the indirect effect of the mediator, was used to evaluate statistical significance of the mediating impact of PA on the relationship between seasonality and FEV_1. Mediating effects of PA were relatively small and not statistically significant for the jittered data.

```
> summary(med)

Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

Output Based on Overall Averages Across Groups

                 Estimate     95% CI Lower     95% CI Upper     p-value
ACME           −0.0000304    −0.0110883           0.01          0.98
ADE            −0.0205774    −0.0959880           0.05          0.57
Total Effect   −0.0206077    −0.0946799           0.05          0.56
Prop. Mediated    0.0144448     −1.0697256          1.47          0.87
```

It is also possible to plot 95% confidence intervals for the ACME, ADE, and total effect with the following R code.

```
plot(med) #visualize the estimated effects
```

Comparison between cohorts

Parameter estimates from model (2) may be directly to compared estimates obtained from prior studies; particularly, we focus on a prior European study of CF cohorts from Denmark and the UK [13] and findings from a Midwest US study [1].

Calculations for comparisons

The following steps enable comparison between cohorts of the seasonality models, which estimate the confidence region for amplitude and horizontal shift. Terms refer to Eq. (2).

1. Assume the coefficient of $\sin\left(\frac{2\pi d_{ij}}{T}\right)$ and $\cos\left(\frac{2\pi d_{ij}}{T}\right)$ follows bivariate normal distribution with mean $= (\widehat{\gamma_0}, \widehat{\gamma_1})$, and cov $= \begin{pmatrix} sd_{r0}^2 & 0 \\ 0 & sd_{r1}^2 \end{pmatrix}$. Following this distribution, we can find the 95% confidence ellipse of the γ_0 and γ_1 . Below, we provided confidence ellipse plots by plotting a sample of 10,000 pairs (γ_0, γ_1) (black dots) with the boundary of confidence ellipse (red circle) for UK, Denmark, and US datasets.

2. Transform each pair of (γ_0^*, γ_1^*) that fall into the 95% confidence ellipse of γ_0 and γ_1, to (horizontal shift, amplitude) using the following formulas:

$$\text{horizontal shift} = \frac{T}{2\pi}\arctan\left(\frac{\gamma_1^*}{\gamma_0^*}\right)$$

$$\text{amplitude} = \sqrt{\gamma_0^{*2} + \gamma_1^{*2}}$$

Implementation

Below, we provide the *R* code to estimate the seasonal variation, amplitude, horizontal shift, peak, and dip date for the Cincinnati, temperature adjusted Cincinnati, UK, and Danish cohorts. For the Cincinnati cohort, we again use the jittered data. Since in previous steps we have already run all the necessary linear mixed effect models, we now directly provide the code to obtain the outcomes of interests and visualize them.

```
#getting coefficient estimates for sine and cosine terms for Msin
coef.si3 <- summary(Msin)$tTable[,1][[2]]
coef.co3 <- summary(Msin)$tTable[,1][[3]]
#compute amplitude and horizontal shift for model Msin
amplitude3 <- sqrt(coef.si3^2 + coef.co3^2) # amplitude for model Msin
horzshft3 <- 365.25/(2*pi)*atan2(coef.co3, coef.si3) # horizontal shift
#estimated seasonal wave for model Msin
wave3 <- coef.si3*sin(2*pi*day/365.25)+coef.co3*cos(2*pi*day/365.25)
data.fit3 <- data.frame(day = day, fitted=wave3)
peak3 <- -horzshft3 + 365.25*0.25 #peak date for Cincinnati cohort
dip3 <- -horzshft3 + 365.25*0.75 #dip date for Cincinnati cohort
ypeak3<- coef.si3*sin(2*pi*peak3/365.25)+coef.co3*cos(2*pi*peak3/365.25)
ydip3 <- coef.si3*sin(2*pi*dip3/365.25)+coef.co3*cos(2*pi*dip3/365.25)
data.hlight3 <- data.frame(hs=horzshft3,y = 0,peak = peak3,
                      dip = dip3,ypeak = ypeak3, ydip = ydip3)
#getting coefficient estimates for sine and cosine terms of model Msin_t
#which is the temperature adjusted model for Cincinnati cohort
coef.si <- summary(Msin_t)$tTable[,1][[2]]
coef.co <- summary(Msin_t)$tTable[,1][[3]]
#compute amplitude and horizontal shift for model Msin_t
amplitude <- sqrt(coef.si^2 + coef.co^2) #amplitude
horzshft <- 365.25/(2*pi)*atan2(coef.co, coef.si) #horizontal shift
#estimated seasonal wave for model Msin_t
wave1 <- coef.si*sin(2*pi*day/365.25)+coef.co*cos(2*pi*day/365.25)
data.fit1 <- data.frame(day = day, fitted=wave1)
```

```
peak <- -horzshft + 365.25*0.25 #peak date for Msin_t

dip <- -horzshft + 365.25*0.75 #dip date for Msin_t

ypeak <- coef.si*sin(2*pi*peak/365.25)+coef.co*cos(2*pi*peak/365.25)

ydip <- coef.si*sin(2*pi*dip/365.25)+coef.co*cos(2*pi*dip/365.25)

data.hlight <- data.frame(hs=horzshft,y = 0,peak = peak,
                          dip = dip,ypeak = ypeak,ydip = ydip)

#estimating seasonal wave for Denmark and UK cohorts by using model
estimates from Qvist et al. (2019)

wave.denmark <- −0.09*sin(2*pi*day/365.25)−0.06*cos(2*pi*day/365.25)

wave.UK <- 0.06*sin(2*pi*day/365.25)−0.13*cos(2*pi*day/365.25)

data.denmark <- data.frame(day=day, denmark =wave.denmark)

data.UK <- data.frame(day=day, UK =wave.UK)

#Creating a Fig. 4 which represent estimated seasonal variations

#Seasons:Spring:3−1~5−31; Summer:6−1~8−31; Fall:9−1~11−30;
Winter:12-#1~2−28

colnames(data.fit3) <- c("x", "y")

colnames(data.denmark) <- c("x", "y")

colnames(data.UK) <- c("x", "y")

colnames(data.fit1) <- c("x", "y")

data.fit3$place <- "cincinnati"

data.UK$place <- "UK"

data.denmark$place <- "denmark"

data.fit1$place <- "cincinnatiAdj"

data.all <- rbind(data.fit3, data.denmark, data.UK, data.fit1)

data.all$place <- factor(data.all$place, levels = c("cincinnati",
"denmark", "UK","cincinnatiAdj"))

p2 <- ggplot() +ylim(−1.5, 1.5)+

  geom_line(data = data.all,aes(x  = x, y = y, group = place,color=place,
linetype = place),size = 1.5)+

  geom_hline(yintercept=0)+geom_vline(xintercept=c(60,152,243,335,−31),
color="darkgrey", linetype = "longdash")+

  geom_point(data=data.hlight,aes(x=-hs,y = y),color="red",size=2) +

  geom_point(data=data.hlight,aes(x=peak,y=ypeak),color="red",size=2)+
```

```
  geom_point(data=data.hlight,aes(x=dip,y=ydip),color="red",size=2) +
  geom_point(data=data.hlight3,aes(x=-hs,y = y),color="red",size=2) +
  geom_point(data=data.hlight3,aes(x=peak3,y=ypeak3),color="red",
size=2) +
  geom_point(data=data.hlight3,aes(x=dip3,y=ydip3),color="red",size=2)+
  geom_text(aes(x = 0, y = 1.5, label="Winter"))+
  geom_text(aes(x = 100, y = 1.5, label="Spring"))+
  geom_text(aes(x = 200, y = 1.5, label="Summer"))+
  geom_text(aes( x = 280, y = 1.5, label="Autumn"))+
  geom_text(aes(x                    =dip3, y = ydip3, label="Aug 9th"),
data=data.hlight3,vjust=1, hjust=-0.2)+
  geom_text(aes(x =peak3, y = ypeak3, label="Feb 7th"),
data=data.hlight3,vjust=-1,hjust=-0.2) +
  geom_text(aes(x =dip, y = ydip, label="Sep 30th"),
data=data.hlight,vjust=1, hjust=-0.2)+
  geom_text(aes(x =peak, y = ypeak, label="Mar 31st"),
data=data.hlight,vjust=-1,hjust=-0.2) +
  theme(panel.grid.minor = element_blank(),
panel.grid.major = element_blank())

p2 + ggtitle("Estimated seasonal fluctuation") +
  scale_color_manual(name="Data",labels=c("Cincinnati",
"Denmark","UK","Cincinnati temp adj"),
  values = c("#000000", "#3399FF","#FF6666","#66ff66"))+
  scale_linetype_manual(name="Data",labels=c("Cincinnati",
"Denmark","UK","Cincinnati temp adj"),values = c("dotted",
"dashed","twodash","solid"))+

  theme(legend.position="top",legend.text=element_text(size=9),
legend.key.width=unit(2,"cm")) + xlab("day") + ylab("fitted")
```

```
#end Fig. 3
```

By implementing the SMAC approach in our published study [1], we were able to assess the potential mediating effects of the PA pathogen on the relationship between seasonality and lung function. In addition, we were able to compare our estimates of seasonal fluctuations in lung function from a Midwest US cohort with those previously reported in cohorts from the UK and Denmark. The SMAC approach provides a guideline and implementation process for future longitudinal data analyses, wherein seasonality and respiratory pathogens may influence lung function patterns.

Now, we present the *R* code for creating the joint 95% confidence region of the amplitude and horizontal shift for a given cohort, and we show how to create a panel of figures for multiple cohorts.

```
#Fig. 5-the joint 95% confidence region of the amplitude and horizontal
shift

#computing 5% confidence region for temp adjusted model(Msin_t)
```

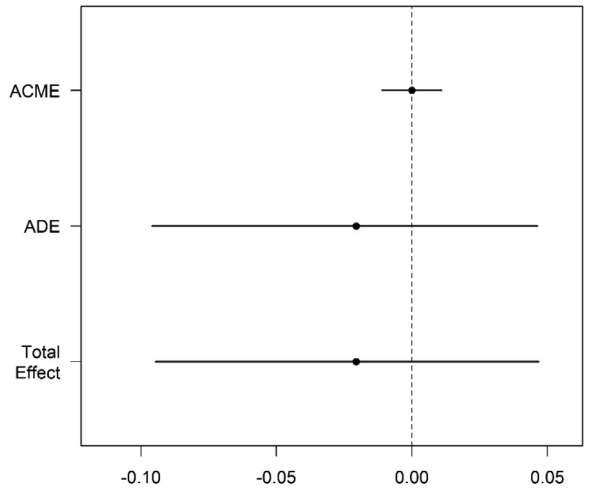

Fig. 3. Estimates (points) and 95% confidence intervals for the average causal mediation effect (ACME), average direct effect (ADE), and total effect. The solid points and lines represent ACME and ADE for the treatment group, and the dotted lines and empty points represent estimates for the control group.

```
mu <- c(coef.si, coef.co) #coefficient estimates from Msin_t
#sigma get standard errors from Msin_t
sigma <- matrix(c((summary(M3_t)$tTable[,2][2])^2,0,0,
                  (summary(M3_t)$tTable[,2][3])^2),2,2)
```

To create confidence ellipse plots, we obtain a sample of 10,000 pairs $(\gamma_0,\ \gamma_1)$ from a multivariate normal distribution with a mean vector that consists of estimated coefficients for sine and cosine terms, and the diagonal elements of the covariance matrix consists of the standard errors of coefficient estimates for sine and cosine terms.

```
#obtaining a sample of 10,000 pairs
data.usa <- data.frame(rmvnorm(10000, mean=mu,sigma))
data.usa <- data.usa%>% mutate(hrztl=−365.25/(2*pi)*atan2(X2,X1),
              amplt <- sqrt(X1^2 + X2^2))

mat.usa <- data.frame(ellipse(sigma,center = mu,level=0.95,
              npoints = 200))
mat.usa <- mutate(mat.usa,hrztl = −365.25/(2*pi)*atan2(y, x),
              amplt = sqrt(x^2 +  y^2),country = "usa")

plot.usa1 <- ggplot()+geom_point(data =
data.usa,aes(x=hrztl,y=amplt),col="#999999")+
geom_line(data=mat.usa, aes(x=hrztl, y=amplt),col="#66ff66",size=2)+
xlim(−180180)+ylim(0,1)+
```

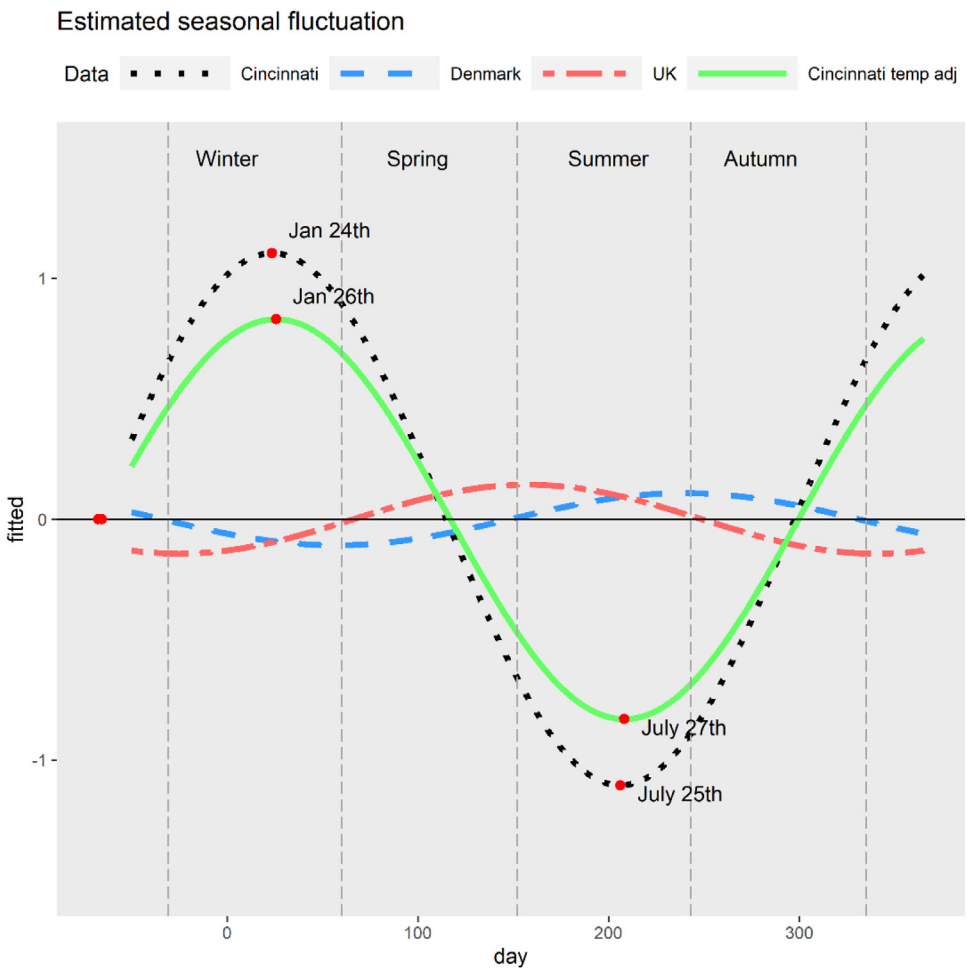

Fig. 4. The estimated seasonal variation in FEV_1 (y-axis) over day of the year, beginning with January 1st (x-axis) for the sine wave (Model (2)) fit to each cohort. Estimated fluctuations shown for the included jittered data are labeled as the Cincinnati cohort (black dashed line) with temperature adjustment (solid green line) and published models (Denmark, shown with red dash-dot line; UK, shown with blue dashed line).

```
  xlab("horizontal shift (days)")+ylab("amplitude (% points of predicted
FEV1)")+ggtitle("(A) Cincinnati temp adjusted")

#to get CI

#sd(data.usa$hrztl)

#summary(data.usa$hrztl)

#getting joint confidence intervals for the UK cohort

mu <- c(0.06,−0.13) #this estimates from Qvist et al. (2019)
```

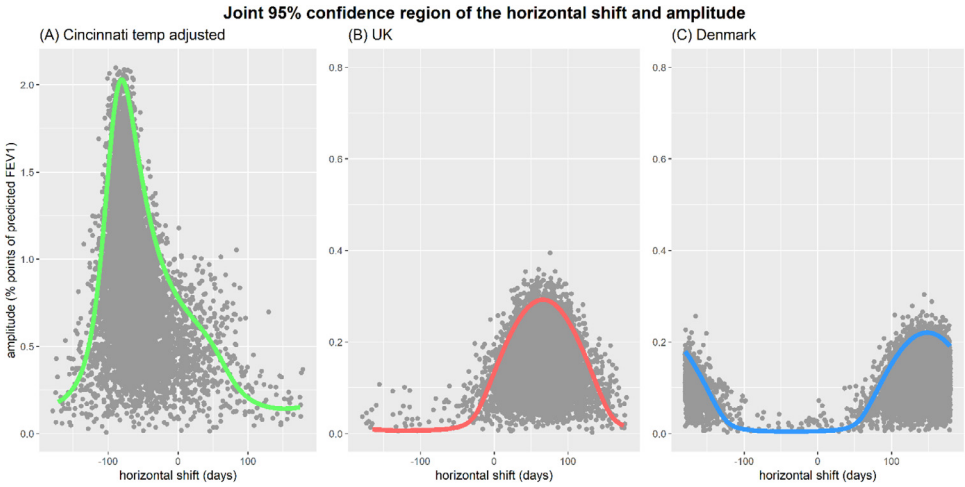

Fig. 5. Joint 95% confidence region of the amplitude (y-axis,% predicted) and horizontal shift in days from January 1st (x-axis) from the (A) Cincinnati sine wave model (adjusted for temperature and based on jittered data); sine wave models from cohorts in the (B) UK and (C) Denmark.

```
sigma <- matrix(c(0.061^2,0,0,0.061^2),2,2) #based on Qvist et al. (2019)

data.usa <- data.frame(rmvnorm(10000, mean=mu,sigma))

data.usa   <-   data.usa%>% mutate(hrztl = −365.25/(2*pi)*atan2(X2, X1),
amplt = sqrt(X1^2 + X2^2))

mat.usa = data.frame(ellipse(sigma,center = mu,level=0.95,
                      npoints = 200))

mat.usa = mutate(mat.usa,hrztl = −365.25/(2*pi)*atan2(y, x),
                 amplt = sqrt(x^2 + y^2),country = "usa")

plot.uk = ggplot()+geom_point(data = data.usa,
aes(x=hrztl,y=amplt),col="#999999")+

  geom_line(data=mat.usa, aes(x=hrztl, y=amplt),col="#FF6666",size=2)
+xlim(−180180)+ylim(00.8)+

  xlab("horizontal shift (days)") +theme(axis.title.y =
element_blank())+ggtitle("(B) UK")

#getting joint confidence intervals for Denmark cohort

mu<- c(−0.09,−0.06) #this for mu and sigma are from Qvist et al. (2019)

sigma=matrix(c(0.04591837^2,0,0,0.04591837^2),2,2)

data.usa = data.frame(rmvnorm(10000, mean=mu,sigma))

data.usa = data.usa%>% mutate(hrztl = −365.25/(2*pi)*atan2(X2, X1),
amplt = sqrt(X1^2 + X2^2))
```

```
mat.usa = data.frame(ellipse(sigma,center = mu,level=0.95,
                        npoints = 200))
mat.usa = mutate(mat.usa,hrztl = −365.25/(2*pi)*atan2(y, x),
                amplt = sqrt(x^2 + y^2),country = "usa")

plot.denmark <- ggplot()+geom_point(data=data.usa,aes(x=hrztl,y=amplt),
col="#999,999")+ geom_line(data=mat.usa, aes(x=hrztl, y=amplt),
col="#3399FF",size=2)+xlim(−180,180)+ylim(0,0.8)+ xlab("horizontal
shift (days)")+theme(axis.title.y = element_blank())+ggtitle("(C)
Denmark")

#creatingFig. 5 as panel of above three figures
figure=ggarrange(plot.usa1,plot.uk,plot.denmark, ncol=3, nrow=1)
annotate_figure(figure,
                top = text_grob("Joint 95% confidence region of the
horizontal shift and amplitude", face = "bold", size = 15))
```

Conclusion

In this paper, we propose a comprehensive approach to SMAC testing and provide the requisite implementation code in freely available statistical software and dataset for application. A practical example is given through use of jittered data based on seasonality and CF lung function decline. Although the case study here was motivated by the CF context, care in other lung conditions relies on FEV_1 and research in these areas face similar challenges with covariance, modeling, seasonal fluctuations and other influential factors, for example, chronic obstructive pulmonary disease. Furthermore, these methods could be utilized to assess lung function changes in healthy populations.

Declaration of Competing Interest

The authors have no competing interests to declare.

Acknowledgments

This work was supported by Grant R01 HL141286 from the National Institutes of Health and Grant GECILI20F0 from the Cystic Fibrosis Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Cystic Fibrosis Foundation. We thank the people with cystic fibrosis and their families who contributed the data which motivated these methodologic developments.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.mex.2021.101313.

References

[1] E. Gecili, C. Brokamp, A. Palipana, R. Huang, E.-R. Andrinopoulou, T. Pestian, E. Rasnick, R.H. Keogh, Y. Ni, J.P. Clancy, P. Ryan, R.D. Szczesniak, Seasonal variation of lung function in cystic fibrosis: longitudinal modeling to compare a Midwest US cohort to international populations, STOTEN 776 (145945) (2021) Epub March 2021.

[2] C. Brokamp, R. Jandarov, M. Hossain, P. Ryan, Predicting daily urban fine particulate matter concentrations using a random forest model, Environ. Sci. Technol. 52 (7) (2018) 4173–4179 Epub 2018/03/15, doi:10.1021/acs.est.7b05381. PubMed PMID: 29537833.

[3] J.B.D. Pinheiro, S. DebRoy, D. Sarkar and R. Core Team. linear and nonlinear mixed effects models. R package version 3.1-137 ed2018.
[4] H.F.R. Wickham, L. Henry, K. Müller Dplyr: a Grammar of Data Manipulation. R package version 0.7.4 ed2017.
[5] A.B.F. Genz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, T. Hothorn mvtnorm: multivariate normal and t distributions. R package version 1.0-7 ed2019.
[6] Kassambara A. ggpubr: 'ggplot2' based publication ready plots. R package version 0.2 ed2018.
[7] H. Wickham, Ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York, 2009.
[8] Murdoch D C.E. Ellipse: functions for drawing ellipses and ellipse-like confidence regions. R package version 0.4.2 ed2020.
[9] D. Tingley, T. Yamamoto, K. Hirose, L. Keele, K. Imai, Mediation: R package for causal mediation analysis, J. Stat. Softw. 59 (5) (2014).
[10] D. Bates, M. Mächler, B. Bolker, S Walker, Fitting linear mixed-effects models using lme4, J. Stat. Softw. 67 (1) (2015) 1–48, doi:10.18637/jss.v067.i01.
[11] Canty A. RB. boot: bootstrap R (S-Plus) functions. R package version 1.3-20 ed2017.
[12] B. Efron, R. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall/CRC, Boca Raton, FL, 1994.
[13] T. Qvist, D.K. Schluter, V. Rajabzadeh, P.J. Diggle, T. Pressler, S.B. Carr, D. Taylor-Robinson, Seasonal fluctuation of lung function in cystic fibrosis: a national register-based study in two northern European populations, J. Cyst. Fibros. 18 (3) (2019) 390–395 Epub 2018/10/23PubMed PMID: 30343891; PMCID: PMC6559396, doi:10.1016/j.jcf.2018.10.006.
[14] K. Imai, L. Keele, D. Tingley, T. Yamamoto, Unpacking the black box of causality: learning about causal mechanisms from experimental and observational studies, Am. Political Sci. Rev. 105 (2011) 765–789.