УТВЕРЖДЕНА

приказом Национального института

общественного здоровья

от «___» _____ 20___ г. № _____

ИНСТРУКЦИЯ

по краткосрочному прогнозированию зарегистрированных случаев COVID-19 в Кыргызской Республике

Оглавление

Глава 1. Общие положения

Глава 2. Основные понятия

Глава 3. Организация проведения прогнозирования

Глава 4. Подготовка программной среды

Глава 5. Входные данные

Глава 6. Подготовка данных

Глава 7. Метод прогнозирования и параметры расчета

Глава 8. Выполнение расчетов

Глава 9. Выходные данные прогнозирования

Глава 10. Подготовка аналитического отчета и интерпретация результатов

Глава 11. Ограничения модели

Глава 12. Оценка качества прогноза

Глава 13. Контроль корректности расчетов и устранение ошибок

Глава 14. Пересмотр методики прогнозирования

Приложение 1. Последовательность выполнения расчетов

Приложение 2. Текст скрипта прогнозирования COVID-19 на 15 дней

Приложение 3. Структура аналитического отчета по результатам краткосрочного прогнозирования COVID-19

Приложение 4. Журнал выполнения и проверки расчетов

Глава 1. Общие положения

1.1. Настоящая Инструкция устанавливает единый порядок подготовки данных, проведения расчетов и представления результатов краткосрочного прогнозирования зарегистрированных случаев COVID-19 на 15 календарных дней в Кыргызской Республике и по административным территориям, представленным во входном наборе данных.

1.2. Прогноз используется для оперативной оценки возможной динамики зарегистрированной заболеваемости и планирования противоэпидемических мероприятий. Прогноз не является точным предсказанием количества всех заражений и не заменяет эпидемиологическую оценку.

1.3. Расчет выполняется на основе оценок эффективного репродуктивного числа R и имитационного прогнозирования.

Глава 2. Основные понятия

Термин	Значение
Зарегистрированные случаи	Случаи COVID-19, включенные в официальную систему регистрации за соответствующую дату.

R (эффективное репродуктивное число)	Оценка среднего числа вторичных случаев, связанных с одним случаем, в текущих условиях; зависит от качества и полноты данных.
Серийный интервал	Промежуток времени между появлением симптомов у последовательных случаев в цепочке передачи; в модели задан распределением.
Сценарий	Расчет при выбранном значении R: минимальном, среднем или максимальном среди пяти последних оценок.
Горизонт прогноза	15 календарных дней после последней даты входных данных.
Симуляция	Одна случайная траектория прогноза, рассчитанная моделью.

Глава 3. Организация проведения прогнозирования

3.1. Прогнозирование включает проверку исходных данных, оценку эффективного репродуктивного числа R, расчет трех сценариев на 15 календарных дней, проверку результатов и подготовку аналитического отчета. Для каждого расчета фиксируют дату последнего наблюдения, период прогноза и используемую версию программного кода.

3.2. Ввод ежедневных данных по территориям выполняет [должность / ФИО].

3.3. Оценку R, расчет сценариев и подготовку аналитического отчета выполняет [должность / ФИО].

3.4. Отчет утверждает и направляет в Министерство здравоохранения Кыргызской Республики [должность / ФИО].

3.5. Прогноз пересчитывается [указать периодичность: ежедневно / еженедельно / по решению НИОЗ]. Срок направления отчета: [указать].

Глава 4. Подготовка программной среды

- Установить R и RStudio; обеспечить доступ к рабочей папке и входной таблице.
- Проверить наличие пакетов R, вызываемых в приложении 2: EpiEstim, incidence, distcrete, ggplot2, projections, epitrix и magrittr.
- Сохранить копию используемого скрипта и входных данных для воспроизводимости расчета. Результаты каждого запуска хранить отдельно с указанием даты.
- При использовании команд чтения данных из буфера обмена Windows (clipboard) и открытия графического устройства x11() необходимо учитывать особенности операционной системы.

Глава 5. Входные данные

5.1. Входные данные — ежедневное число вновь зарегистрированных случаев COVID-19 по территориям в фиксированном порядке. Приведенный в приложении 2 скрипт предусматривает 10 числовых столбцов.

5.2. Порядок территорий по столбцам входной таблицы (Приложение 2) фиксирован и должен соблюдаться:

№ столбца	Территория
1	г. Бишкек
2	г. Ош
3	Чуйская область
4	Ошская область
5	Жалал-Абадская область
6	Баткенская область
7	Таласская область
8	Нарынская область
9	Иссык-Кульская область
10	Кыргызская Республика (национальный итог)

Скрипт не проверяет соответствие порядка столбцов этому перечню: при несовпадении порядка результаты по территориям будут ошибочно сопоставлены без предупреждения.

5.3. Каждая строка соответствует одной календарной дате, без пропусков календарных дней. Все ряды должны охватывать одинаковый период. Отрицательные значения и текстовые обозначения отсутствующих данных не допускаются без предварительного разрешения и документированного исправления.

5.4. Сумма территориальных рядов сопоставляется с национальным итогом с учетом особенностей официальной регистрации; обнаруженные расхождения фиксируются, а не исправляются автоматически.

5.5. В журнале указываются источник, дата выгрузки, начальная и конечная даты, число дней и территорий, сведения о пересмотрах прошлых данных.

Глава 6. Подготовка данных

6.1. Получить официальную выгрузку и сохранить исходный файл без изменений. Создать рабочую копию для расчета.

6.2. Проверить даты, порядок территорий, непрерывность календарного ряда, отсутствие дублей и соответствие числа строк числу дат.

6.3. В скрипте приложения 2 заданы начальная дата 01.09.2020 и n=170. Перед расчетом необходимо привести дату начала и число наблюдений в соответствие с используемыми данными. Длина вектора дат должна совпадать с числом строк входной таблицы.

6.4. Скрипт дважды читает таблицу ежедневных случаев из буфера обмена Excel. Перед вторым чтением таблицу необходимо повторно скопировать: первый этап записывает оценки R в буфер обмена и тем самым заменяет его содержимое.

Глава 7. Метод прогнозирования и параметры расчета

Параметр	Значение в исходном 15-дневном алгоритме
Периодичность входа	Ежедневные случаи

Оценка R	EpiEstim, method="parametric_si"
Серийный интервал при оценке R	Среднее 4,8 дня; стандартное отклонение 2,3 дня
Используемые оценки R	Пять последних доступных оценок по каждой территории
Три сценария	min(R), mean(R), max(R) за пять оценок
Серийный интервал при прогнозировании	Дискретизированное гамма-распределение с параметрами 4,8 и 2,3
Горизонт	15 дней
Число симуляций	1000 на сценарий и территорию
Итоговое значение	Среднее по симуляциям для каждого прогнозного дня

7.1. Для каждой территории оценивают R методом parametric_si (EpiEstim; среднее серийного интервала 4,8 дня, стандартное отклонение 2,3 дня). Из пяти последних оценок R выбирают минимальное, среднее и максимальное значения для трех сценариев. Для каждого сценария рассчитывают 1000 имитационных траекторий на 15 дней; итоговое значение за день — среднее по траекториям. Сценарные значения R не являются границами статистического прогнозного интервала.

7.2. При изменениях тестирования, регистрации случаев или эпидемиологической ситуации необходимо учитывать возможное влияние этих изменений на точность прогноза.

Глава 8. Выполнение расчетов

- Открыть RStudio и загрузить скрипт из приложения 2 либо его технически проверенную рабочую версию.
- Проверить дату начала, число дней n, количество столбцов и соответствие фактической выгрузке.
- Скопировать подготовленную таблицу из Excel в буфер обмена и выполнить чтение данных.
- Запустить расчет R и проверить наличие пяти конечных числовых оценок для каждой территории.
- Выполнить расчет трех сценариев и проверить наличие результатов на все 15 дней для каждой территории.
- Сохранить результаты в файлы с однозначными названиями, датой расчета и обозначением сценария.
- Выполнить графическую часть, если она включена в утвержденный скрипт; проверить шкалы, подписи и единицы измерения.

Скрипт использует буфер обмена Windows. Перед каждым чтением данных следует повторно копировать соответствующую таблицу из Excel. Команды write.table(..., "clipboard") заменяют содержимое буфера обмена; при отсутствии повторного копирования следующий этап может получить неверные входные данные.

Глава 9. Выходные данные прогнозирования

9.1. Основные выходные данные — три матрицы размером 15 × 10: средний, минимальный и максимальный сценарии, рассчитанные для девяти территорий и Кыргызской Республики. Каждая строка соответствует дню прогноза, столбец — территории. Названия территорий и даты в матрицы автоматически не записываются.

9.2. При переносе результатов в аналитический отчет к матрицам добавляют даты прогнозных дней, наименования территорий и обозначения сценариев. В скрипте результаты выводятся через буфер обмена, поэтому каждый из трех массивов необходимо сохранить отдельно.

9.3. Графический блок приложения 2 использует дополнительно подготовленные ряды для построения графиков и не формирует автоматически показатели заболеваемости на 100 000 населения. Эти показатели не следует указывать как результат основного 15-дневного расчета.

Глава 10. Подготовка аналитического отчета и интерпретация результатов

10.1. По итогам расчета составляют аналитический отчет о 15-дневном прогнозе зарегистрированных случаев COVID-19 по Кыргызской Республике и территориям, представленным во входной таблице. Состав отчета приведен в приложении 3.

10.2. В отчете указывают дату последнего фактического наблюдения, 15-дневный период прогноза, значения трех сценариев, таблицы ежедневных прогнозных случаев и пояснения к результатам. Графики приводят только после проверки их соответствия исходным данным и расчетному периоду.

Глава 11. Ограничения модели

- Прогноз основан на зарегистрированных случаях и чувствителен к неполноте регистрации.
- Фиксированное распределение серийного интервала может не соответствовать новым эпидемиологическим условиям.
- Пять последних оценок R могут быть нестабильны при малом числе случаев и резких изменениях регистрации.
- Сценарии по минимуму, среднему и максимуму R не описывают всю неопределенность модели.
- При существенном изменении ситуации 15-дневный прогноз требует внеочередного пересчета и экспертной интерпретации.

Глава 12. Оценка качества прогноза

12.1. После появления фактических данных за прогнозный период сопоставлять прогноз и зарегистрированные случаи по одинаковым датам и территориям.

12.2. Для центрального сценария рассчитывают абсолютную ошибку по каждой дате и среднюю абсолютную ошибку (MAE) за доступный период. Дополнительно проверяют

направление изменения и долю фактических значений между минимальным и максимальным сценариями; эта доля не является оценкой покрытия доверительного интервала.

12.3. Хранить результаты проверок в журнале качества. При систематических отклонениях проводить анализ причин и документировать предложения по пересмотру модели без несанкционированного изменения рабочего скрипта.

Глава 13. Контроль корректности расчетов и устранение ошибок

Ситуация	Действие
Число строк не совпадает с n	Уточнить даты, пропуски и длину ряда; не запускать прогноз до исправления.
Вместо чисел считаны строки или NA	Проверить формат Excel, разделители, пустые ячейки и порядок столбцов.
Нет пяти конечных оценок R	Проверить исходный ряд, ошибки EpiEstim и достаточность данных; не подменять оценки вручную.
Ошибка clipboard или x11	Проверить среду Windows и доступность графического устройства; согласовать техническую адаптацию.
Отрицательные, бесконечные или отсутствующие прогнозы	Прекратить выпуск отчета, проверить расчеты и исходные данные.
Несоответствие суммарных и национальных значений	Проверить порядок столбцов и методологию учета; зафиксировать различия.

Глава 14. Пересмотр методики прогнозирования

Пересмотр методики необходим при изменении правил регистрации случаев, структуры входных данных, состава территорий, расчетных параметров или алгоритма прогнозирования. Изменения фиксируются с указанием даты и версии скрипта.

Приложение 1. Последовательность выполнения расчетов

1. Загрузка R-пакетов и подготовка вектора ежедневных дат.

2. Чтение десяти числовых столбцов ежедневной заболеваемости из Excel через буфер обмена.

3. Оценка R для каждой территории методом parametric_si.

4. Извлечение пяти последних оценок R и расчет минимума, среднего и максимума.

5. Формирование дискретизированного серийного интервала.

6. Построение трех прогнозных сценариев на 15 дней с 1000 симуляций.

7. Расчет средних ежедневных прогнозных значений по симуляциям.

8. Выгрузка результатов и дополнительное построение графиков недельной заболеваемости.

Приложение 2. Текст скрипта прогнозирования COVID-19 на 15 дней

Перед расчетом проверяют соответствие дат и числа строк исходной таблицы, а также порядок десяти столбцов. Скрипт содержит фиксированные значения дат и длины рядов. Графический блок требует отдельной подготовки данных и не является автоматическим продолжением расчета трех сценариев.

```
library(EpiEstim)
library(incidence)
n=170
dates1=rep(as.Date('2020-09-01'), n)
dates1=dates1+0:(n-1)
dates1
inc=read.table('clipboard', header = F, sep='\t')# вставит с Excel
rt=matrix(NA, 10, 5)
for (i in 1:10) {
  covid=inc[,i]
  covid=as.incidence(covid, dates=dates1)
  g=estimate_R(covid,method="parametric_si",
               config = make_config(list(mean_si = 4.8, std_si = 2.3)))
  rt[i,]=g$R[3][(dim(g$R[3])[1]-4):dim(g$R[3])[1],1]
  write.table(rt, 'clipboard', row.names = F, sep='\t')
}
rt
mean_r=apply(rt, 1, mean)#sredni
min_r=apply(rt, 1, min)#min
max_r=apply(rt, 1, max)#max
mean_r
min_r
max_r
library(distcrete)
library(ggplot2)
library(projections)
library(epitrix)
library(magrittr)
mu <- 4.8
sigma <- 2.3
cv <- sigma / mu
params <- gamma_mucv2shapescale(mu, cv)
params
si <- distcrete("gamma", shape = params$shape,
                scale = params$scale,
                interval = 1, w = 0.5)
si# ot mu do si все вместе запускать
covid=read.table('clipboard', header = F, sep='\t') #заново вставим с Excel
df1=matrix(NA, 15, 10) #средняя
df2=matrix(NA, 15, 10)#минимальная
df3=matrix(NA, 15, 10)#максимальная
for (j in 1:10){
  inc=covid[,j]
  inc=as.incidence(inc, dates1)
  pred_1 <- project(inc, R  = mean_r[j], si = si, n_days = 15, n_sim = 1000) #средняя
  df1[,j]=rowMeans(pred_1)
  pred_2 <- project(inc, R  = min_r[j], si = si, n_days = 15, n_sim = 1000) #минимальная
  df2[,j]=rowMeans(pred_2)
  pred_3 <- project(inc, R  = max_r[j], si = si, n_days = 15, n_sim = 1000) #максимальная
  df3[,j]=rowMeans(pred_3)
}
write.table(df1, 'clipboard', row.names = F, col.names=F, sep='\t') #средняя
#вставляем данные в Excel для среднего прогноза желтый
write.table(df2, 'clipboard', row.names = F, col.names=F, sep='\t') #минимальная
#вставляем данные в Excel для минимального прогноза зел
write.table(df3, 'clipboard', row.names = F, col.names=F, sep='\t') #максимальная
#вставляем данные в Excel для максимального прогноза красный
wi=read.table('clipboard', sep='\t', header=F) #фактические данные желт
dates2=rep(as.Date('2020-09-07'), 179)
dates2=dates2+0:178
```

```
dates2
w_min=read.table('clipboard', sep='\t', header=F) #фактические данные зел
w_max=read.table('clipboard', sep='\t', header=F) #фактические данные красный
obl=c('Бишкек', 'г.Ош', "Чуй", "Ош обл.",  "Жалал-Абад",
      "Баткен", "Талас", "Нарын", "Иссык-Куль", "КР")
par(mfrow=c(3,3))
x11(h=9, w=16)  #запускать все вместе
par(mfrow=c(3,3))
for (i in 1:9) {
  plot.new()
  plot.window(xlim=c(dates2[1], dates2[179]), ylim=c(0, ceiling(max(wi[, i])/10)*11.5))
  axis(2, at=seq(0, ceiling(max(wi[, i])/10)*10, 11.5), las=2)
  axis(1, dates2, format(dates2, "%b %d"), cex.axis = .7)
  lines(x=dates2[1:164], y=wi[,i], lwd=2)
  lines(x=dates2[164:179], y=w_min[,i], col='green', lwd=2)
  lines(x=dates2[164:179], y=w_max[,i], col='red', lwd=2)
  abline(h=20, lty='dashed', col='orange', lwd=1)
  abline(h=40, lty='dashed', col='red', lwd=1)
  abline(h=80, lty='dashed', col='red', lwd=2)
  title(main=obl[i], xlab='Дата', ylab='Недельная заболеваемость')
}
x11(h=9, w=16) #KR
for (i in 10:10) {
  plot.new()
  plot.window(xlim=c(dates2[1], dates2[179]), ylim=c(0, ceiling(max(wi[, i])/10)*10))
  axis(2, at=seq(0, ceiling(max(wi[, i])/10)*10, 10), las=2)
  axis(1, dates2, format(dates2, "%b %d"), cex.axis = .7)
  lines(x=dates2[1:164], y=wi[,i], lwd=2)
  lines(x=dates2[164:179], y=w_min[,i], col='green', lwd=2)
  lines(x=dates2[164:179], y=w_max[,i], col='red', lwd=2)
  abline(h=20, lty='dashed', col='orange', lwd=1)
  abline(h=40, lty='dashed', col='red', lwd=1)
  abline(h=80, lty='dashed', col='red', lwd=2)
  title(main=obl[i], xlab='Дата', ylab='Недельная заболеваемость')
}
```

Приложение 3. Структура аналитического отчета по результатам краткосрочного прогнозирования COVID-19

1. Дата выполнения расчета, дата последнего фактического наблюдения и период 15-дневного прогноза.

2. Источник данных, охват территорий и отметка о проверке исходных рядов.

3. Результаты 15-дневного прогноза: три сценария, таблицы ежедневных прогнозных случаев по Кыргызской Республике и территориям.

4. Сопоставление фактической динамики с прогнозными значениями по тем же датам и территориям. Графики включают при наличии подготовленных данных для их построения.

5. Оценки R, использованные для трех сценариев, и описание их интерпретации.

6. Интерпретация расчетных результатов и ограничения прогноза. Минимальный и максимальный сценарии не являются статистическими границами прогнозного интервала.

7. Дата подготовки отчета и исполнитель.

Приложение 4. Журнал выполнения и проверки расчетов

Поле	Заполняемое значение
Дата и время расчета	_____
Источник и дата выгрузки	_____
Последняя фактическая дата	_____

Количество дней / территорий	_____
Версия скрипта	_____
Проверка входных данных	Пройдена / не пройдена
Проверка результатов	Пройдена / не пройдена
Файл отчета и место хранения	_____
Исполнитель / согласовавший	_____