

excode: Excess Count Detection in Epidemiological Time Series

Benedikt Zacher¹, & Ann Christin Vietor¹

¹ Robert Koch-Institut | Unit 32

Cite

Zacher, B., & Vietor, A. (2025). excode: Excess Count Detection in Epidemiological Time Series. Zenodo. <https://doi.org/10.5281/zenodo.17417083>

Abstract

The repository "excode: Excess Count Detection in Epidemiological Time Series" contains the R package excode with a variety of functions for excess count detection in epidemiological time series. Excess count detection is an important part of public health surveillance.

Table of Content

- Installation
 - Overview
 - Data
 - Administrative and organizational information
 - Funding
 - Collaborate
 - Publication platforms
 - License
-

This repository contains the R package *excode* with a variety of functions for **excess count detection** in epidemiological time series. Excess count detection is an important part of public health surveillance.

Installation

This is an R package. You can use `install_github()` from devtools to install this package.

```
library(devtools)
install_github("robert-koch-institut/excode", subdir = "software")
```

Overview

A variety of algorithms has been developed to identify events such as disease outbreaks or excess mortality. To this end, time series are analysed to detect unusually large (case) counts, that exceed what is normally expected in a certain time period and geographical region. The normal expectancy of cases in a current time period is usually calculated based on historic data. Depending on the time series of interest, the following features need to be taken into account by a model:

- **Seasonal patterns:** Many epidemiological times series that are of public health interest show periodic changes in cases depending on seasons or other calendar periods.
- **Long-term time trends:** The time series may show a long-term increase or decrease in case counts.
- **Historic events:** Events such as disease outbreaks may have caused an excess of case counts in historic data that is used for model estimation. This needs to be considered to avoid overestimation of the normal expectancy for the current time period.

The *excode* package provides a flexible framework that implements well established approaches to control for seasonality, long-term trends and historic events, but also allows the use of customized models. The user can choose between the Poisson and the Negative Binomial distribution, which are the most commonly used probability distributions for modeling count data. By combining hidden Markov models and generalized linear models, *excode* explicitly models normally expected case counts *and* expected excess case counts, i.e. each time point in a time series is labeled either as a normal state or as an excess state. Further descriptions and code examples can be found in the `vignette("excode")`.

Running an excode model

The package's core function, `run_excode()` performs the excess count detection. The output of `run_excode()` is a fitted `excodeModel` object.

The following code example illustrates how to fit a three-state model with sine/cosine functions ('Harmonic') to model seasonal and a natural cubic spline with two knots to capture long-term trends ('Spline2').

```
library(excode)
data(mort_df_germany)
sum_har_nb <- run_excode(surv_ts = mort_df_germany,
                        timepoints = 325,
                        distribution = "NegBinom",
                        states = 3,
                        periodic_model = "Harmonic",
                        time_trend = "Spline2",
                        return_full_model = TRUE)
```

Results can be extracted using the `summary()` and plotted with the `plot_excode_summary()` functions:

```
sum_har_nb <- summary(res_har_nb)
plot_excode_summary(sum_har_nb)
```

Data

The package includes example datasets which can be used to apply the different algorithms.

The following datasets are provided with this package:

mort_df_germany
sarscov2_df
shadar_df.

German all-cause mortality data

The dataset **mort_df_germany** contains the number of weekly deaths in Germany reported to the [German Federal Statistical Office](#).

SARS-CoV-2 infections in Berlin-Neukölln (Germany)

The dataset **sarscov2_df** contains the daily number of reported **SARS-CoV-2** cases from March to July 2020 in **Berlin-Neukölln**. The dataset was downloaded on **2024-10-31** from [this source](#).

Salmonella Hadar cases in Germany (2001-2006)

The dataset **shadar_df** contains the weekly number of reported **Salmonella Hadar** cases from January 2001 to August 2006 in Germany.

Administrative and organizational information

This R package was developed by Benedikt Zacher with contributions from Ann Christin Vietor [Unit 32 |](#)

[Surveillance](#).

The publication of the code as well as the quality management of the metadata is done by department [MF 4 | Domain Specific Data and](#)

[Research Data](#)

[Management](#).

Questions regarding the publication infrastructure

can be directed to the Open Data Team of the Department MF4 at

OpenData@rki.de.

Funding

Benedikt Zacher and Ann Christin Vietor were supported by BMBF (Medical Informatics Initiative: HIGHmed) and the collaborative management platform for detection and analyses of (re-)emerging and foodborne outbreaks in Europe (COMPARE: European Union's Horizon research and innovation programme, grant agreement No. 643476).

Collaborate

If you want to contribute, feel free to fork this repo and send us pull requests.

Publication platforms

This software publication is available on [Zenodo.org](#), [GitHub.com](#) and [OpenCoDE](#):

- <https://zenodo.org/communities/robertkochinstitut>
- <https://github.com/robert-koch-institut>
- <https://gitlab.opencode.de/robert-koch-institut>

License

excode: Excess Count Detection in Epidemiological Time Series is free and open-source software, published under the terms of the [GPL3 license](#).