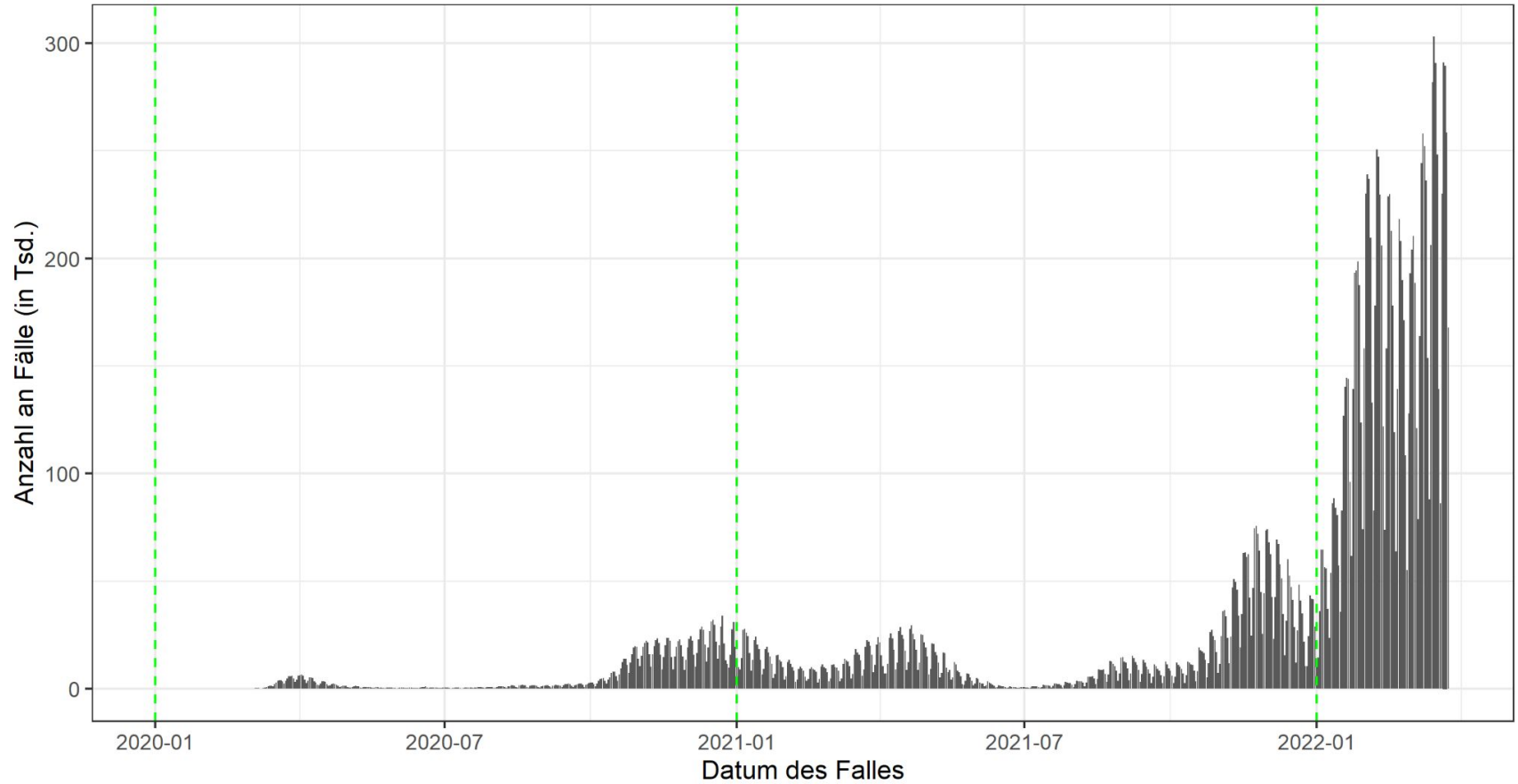


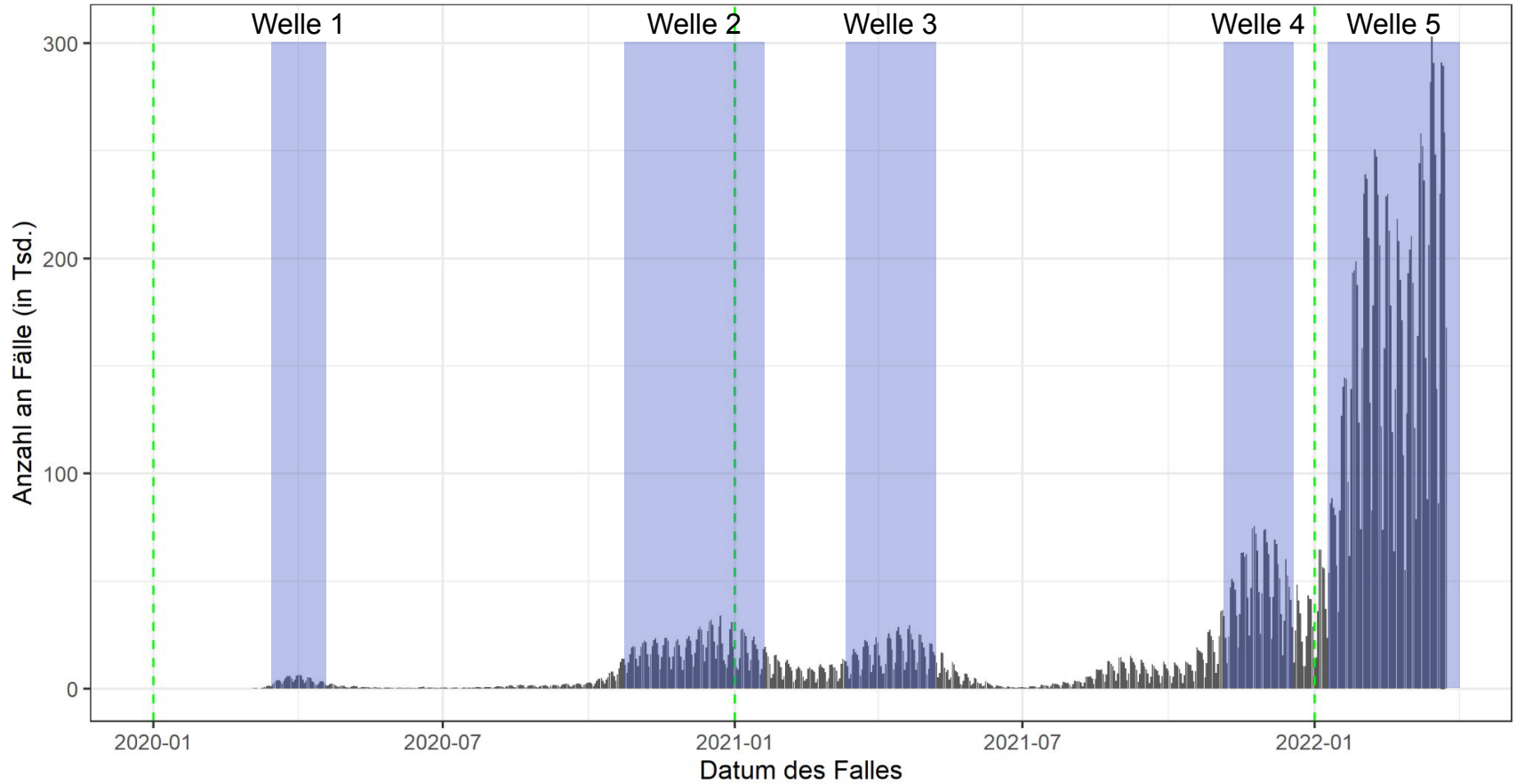
Anfängerpraktikum: Meldeverzug der Neuinfektionen in den RKI Corona-Daten 28.03.2022

Vortragsgruppe:	Quang Le, Qingyan Lu, Wolfgang Stefani, Joshua Weinert
Betreuer:	Patrick Kaiser
Projektpartner:	StaBLab (Maximilian Weigert)

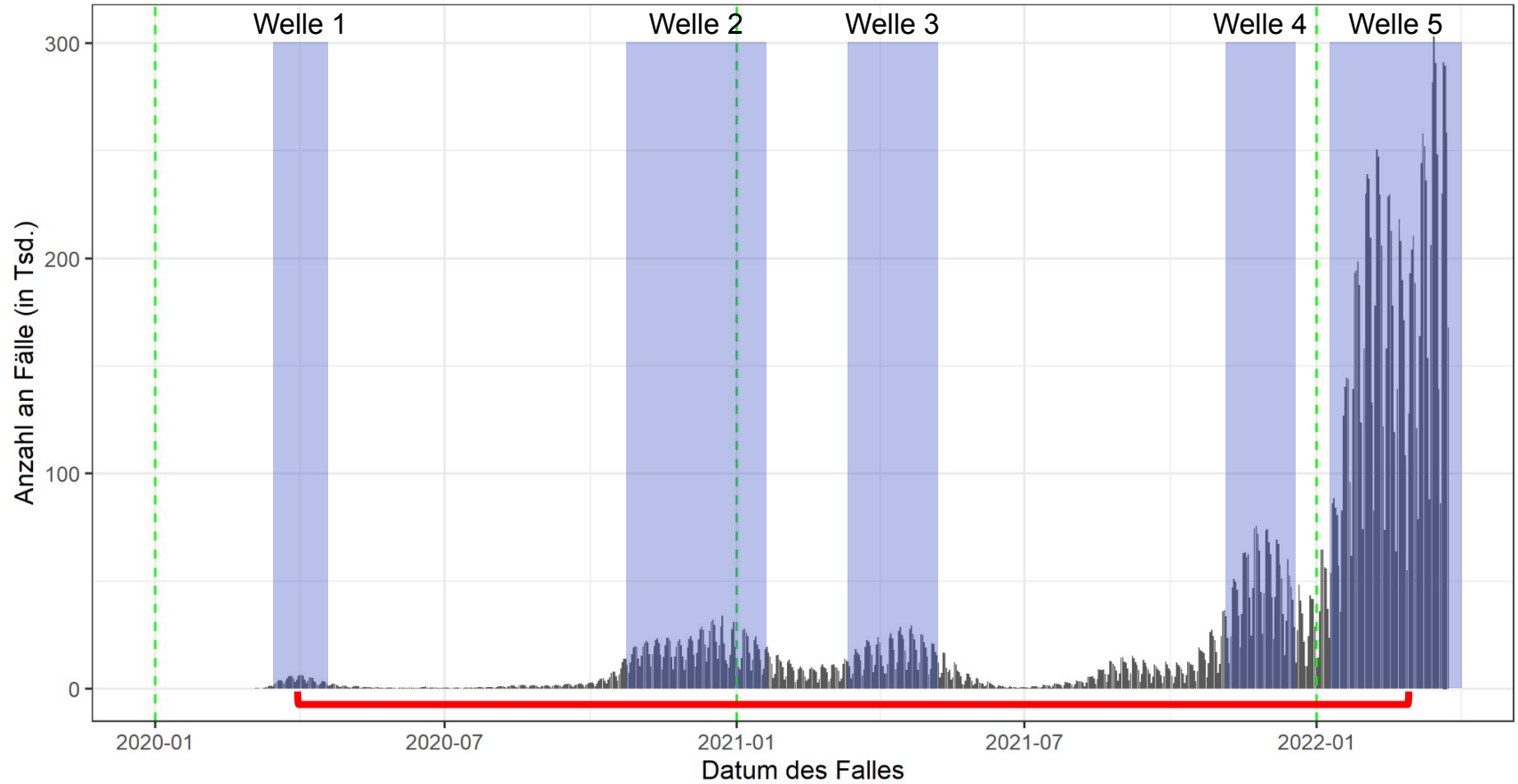
Der Verlauf der Pandemie: Neue Coronafälle pro Tag



Der Verlauf der Pandemie: Neue Coronafälle pro Tag

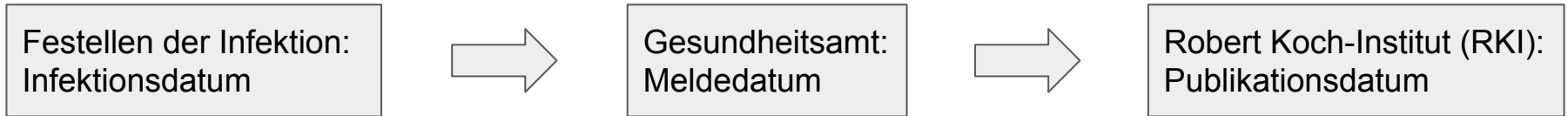


Der Verlauf der Pandemie: Neue Coronafälle pro Tag



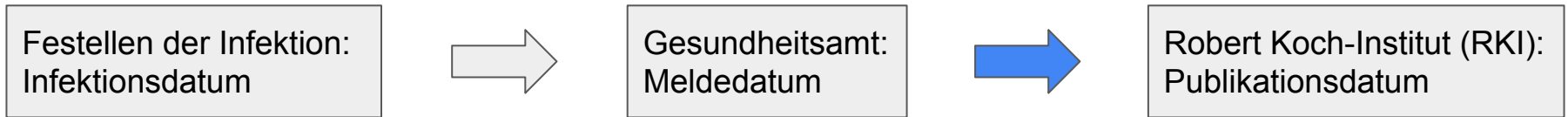
Definition Meldeverzug

Meldeverlauf eines Corona-Falls



Definition Meldeverzug

Meldeverlauf eines Corona-Falls

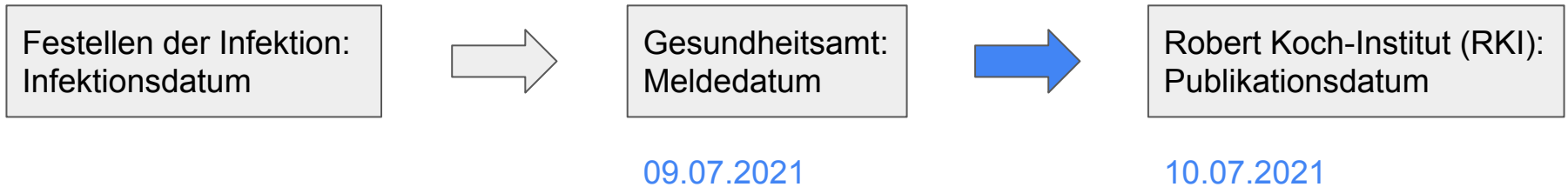


Definition:

$$\text{Meldeverzug} = \text{Publikationsdatum} - \text{Meldedatum} - 1$$

Definition Meldeverzug

Meldeverlauf eines Corona-Falls



Definition:

Meldeverzug = Publikationsdatum - Meldedatum - 1

$$(10.07.2021 - 09.07.2021) - 1 = 1 - 1 = 0$$

Der Auftrag

Forschungsfragen

unseres Projektpartners, des StaBLab:

- Wie verändert sich der Meldeverzug im Zeitverlauf?

Forschungsfragen

unseres Projektpartners, des StaBLab:

- Wie verändert sich der Meldeverzug im Zeitverlauf?
- Bestehen Unterschiede im Meldeverzug zwischen den Altersgruppen und Bundesländern?

Forschungsfragen

unseres Projektpartners, des StaBLab:

- Wie verändert sich der Meldeverzug im Zeitverlauf?
- Bestehen Unterschiede im Meldeverzug zwischen den Altersgruppen und Bundesländern?
- Welchen zeitlichen Verlauf haben die erfassten Todesfälle?

Erstellung des Gesamtdatensatz

Einheitlicher Gesamtdatensatz

Ausgangssituation

- Zeitraum: 27.03.2020 - 22.02.2022, d.h. 698 Tage

Einheitlicher Gesamtdatensatz

Ausgangssituation

- Zeitraum: 27.03.2020 - 22.02.2022, d.h. 698 Tage
- Hierzu 698 Tagesdatensätze

Einheitlicher Gesamtdatensatz

Ausgangssituation

- Zeitraum: 27.03.2020 - 22.02.2022, d.h. 698 Tage
- Hierzu 698 Tagesdatensätze, jedoch doppelter Datensatz für

[cases_GermanTemporal_2021-08-10.rds](#)

[cases_GermanTemporal_2021-08-10\(1\).rds](#)

Einheitlicher Gesamtdatensatz

Ausgangssituation

- Zeitraum: 27.03.2020 - 22.02.2022, d.h. 698 Tage
- Hierzu 697 Tagesdatensätze

Einheitlicher Gesamtdatensatz

Ausgangssituation

- Zeitraum: 27.03.2020 - 22.02.2022, d.h. 698 Tage
- Hierzu 697 Tagesdatensätze, jedoch Duplikate für

Einheitlicher Gesamtdatensatz

Ausgangssituation

- Zeitraum: 27.03.2020 - 22.02.2022, d.h. 698 Tage
- Hierzu 697 Tagesdatensätze, jedoch Duplikate für 6 Tage-Paare

Einheitlicher Gesamtdatensatz

Ausgangssituation

- Zeitraum: 27.03.2020 - 22.02.2022, d.h. 698 Tage
- Hierzu 691 Tagesdatensätze

Einheitlicher Gesamtdatensatz

Ausgangssituation

- Zeitraum: 27.03.2020 - 22.02.2022, d.h. 698 Tage
- Hierzu 691 Tagesdatensätze
- Mit 6 Varianten an Spalten

Einheitlicher Gesamtdatensatz

Ausgangssituation

- Zeitraum: 27.03.2020 - 22.02.2022, d.h. 698 Tage
- Hierzu 691 Tagesdatensätze
- Mit 6 Varianten an Spalten, dabei 2 Codierungen

Einheitlicher Gesamtdatensatz

Ausgangssituation

- Zeitraum: 27.03.2020 - 22.02.2022, d.h. 698 Tage
- Hierzu 691 Tagesdatensätze
- Mit 6 Varianten an Spalten, dabei 2 Codierungen

...	cases	deaths	date	new_case	new_fatality	...

...	cases	deaths	date	...	newcase	newdeath	...

Einheitlicher Gesamtdatensatz

Ausgangssituation

- Zeitraum: 27.03.2020 - 22.02.2022, d.h. 698 Tage
- Hierzu 691 Tagesdatensätze
- Mit 6 Varianten an Spalten, dabei 2 Codierungen

Einheitlicher Gesamtdatensatz

Zusammenführung

- 2 Codierungen

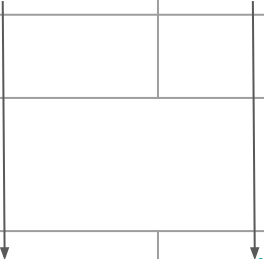
Einheitlicher Gesamtdatensatz

Zusammenführung

- 2 Codierungen

...	cases	deaths	date	new_case	new_fatality	...

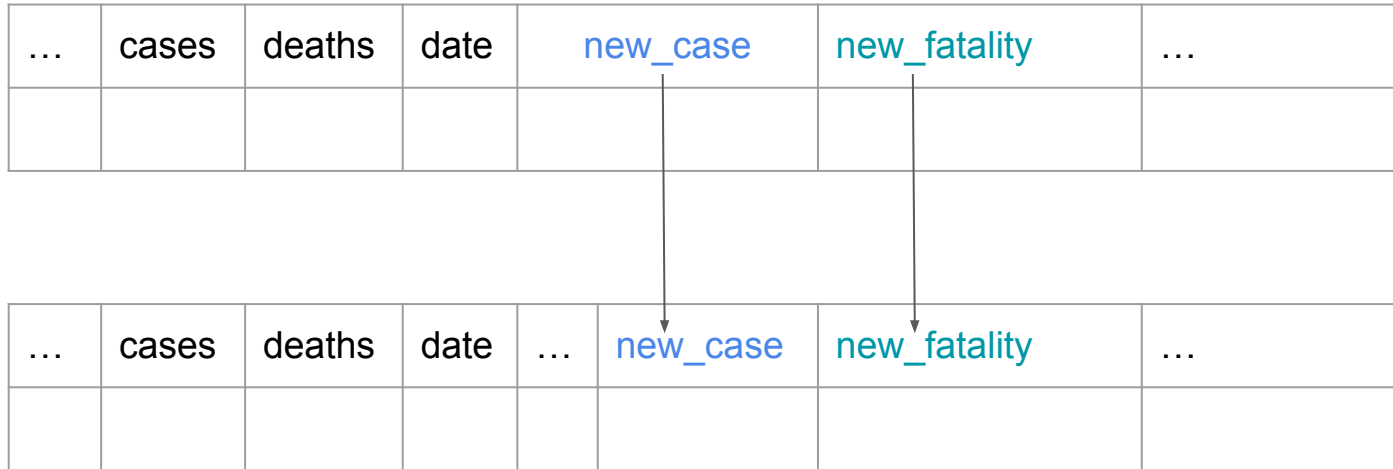
...	cases	deaths	date	...	newcase	newdeath	...



Einheitlicher Gesamtdatensatz

Zusammenführung

- 2 Codierungen harmonisiert



Einheitlicher Gesamtdatensatz

Zusammenführung

- 2 Codierungen harmonisiert
- 7 Datumsformate, z.B. 10.08.2021, 00:00 Uhr

Einheitlicher Gesamtdatensatz

Zusammenführung

- 2 Codierungen harmonisiert
- 7 Datumsformate harmonisiert

Einheitlicher Gesamtdatensatz

Zusammenführung

- 2 Codierungen harmonisiert
- 7 Datumsformate harmonisiert
- Selektieren der neuen Fälle je Datensatz

Einheitlicher Gesamtdatensatz

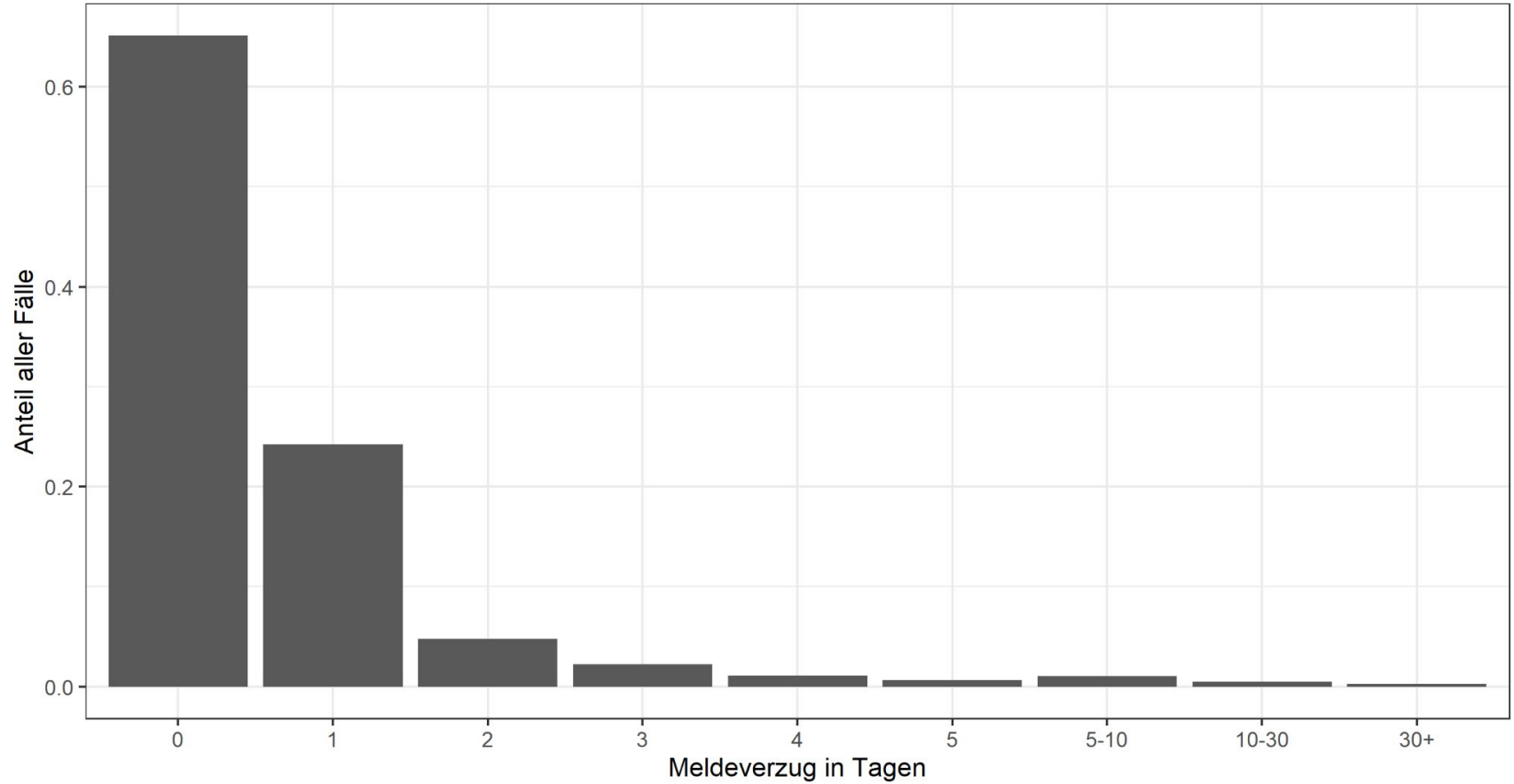
Zusammenführung

- 2 Codierungen harmonisiert
- 7 Datumsformate harmonisiert
- Selektieren der neuen Fälle je Datensatz
- Zusammenfügen mit neuer Spalte “publication_date”

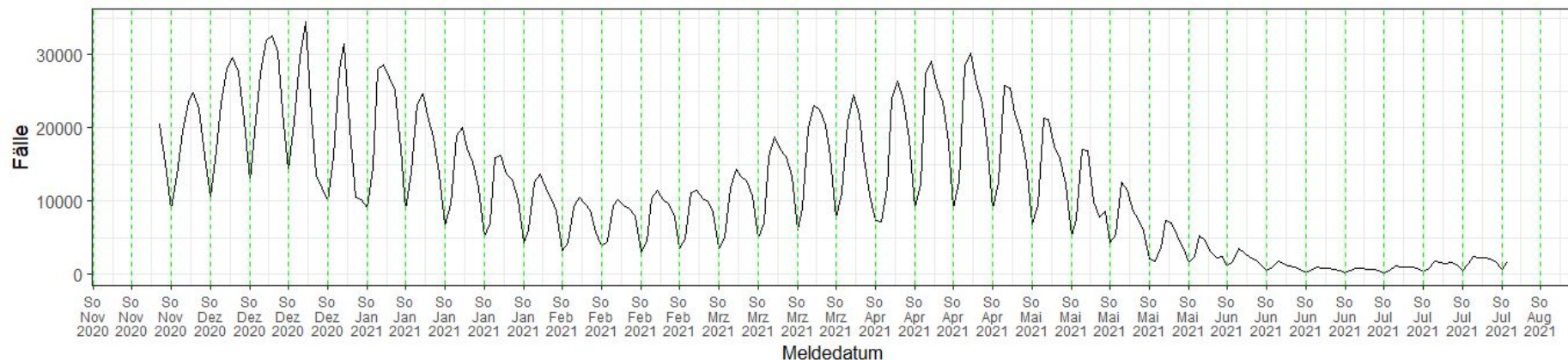
...	cases	deaths	date	new_case	new_fatality	publication_date	...

Analyse des Meldeverzugs

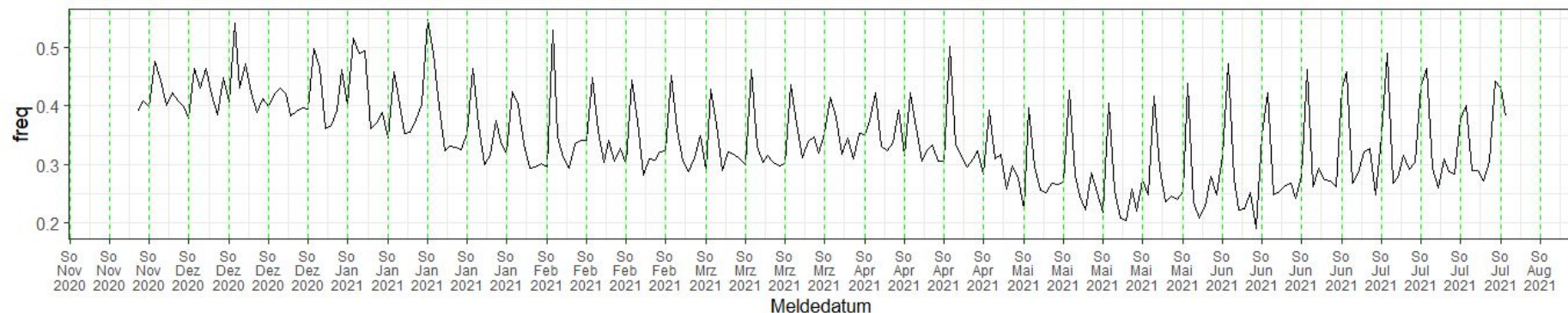
Anteil an Fällen mit Meldeverzug



Wöchentlicher Zyklus der Anzahl der Neuinfektionsmeldungen im Zeitraum 27.11.2020 - 26.07.2021

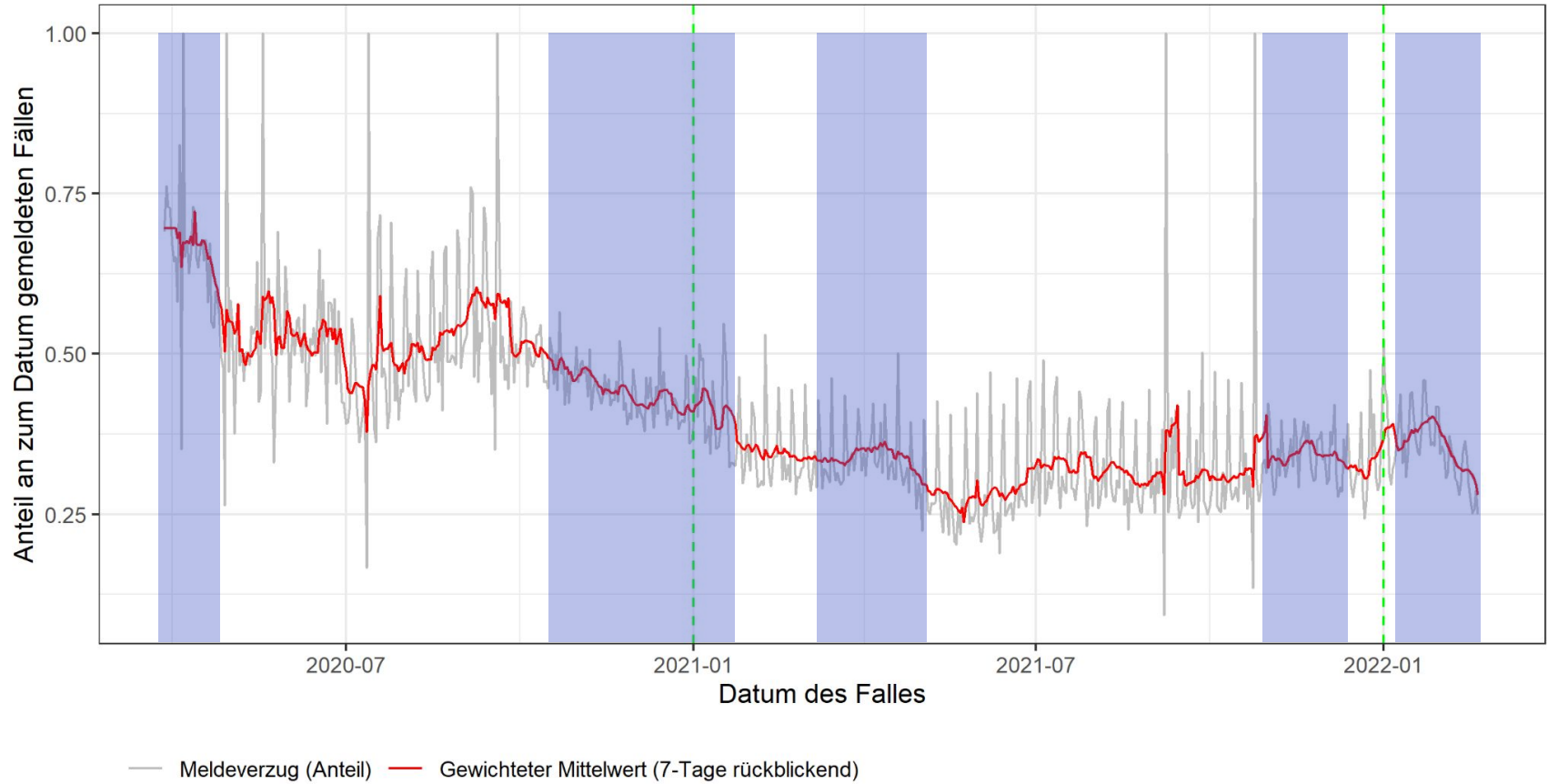


Wöchentlicher Zyklus des Anteils der Spätmeldungen* im Zeitraum 27.11.2020 - 26.07.2021

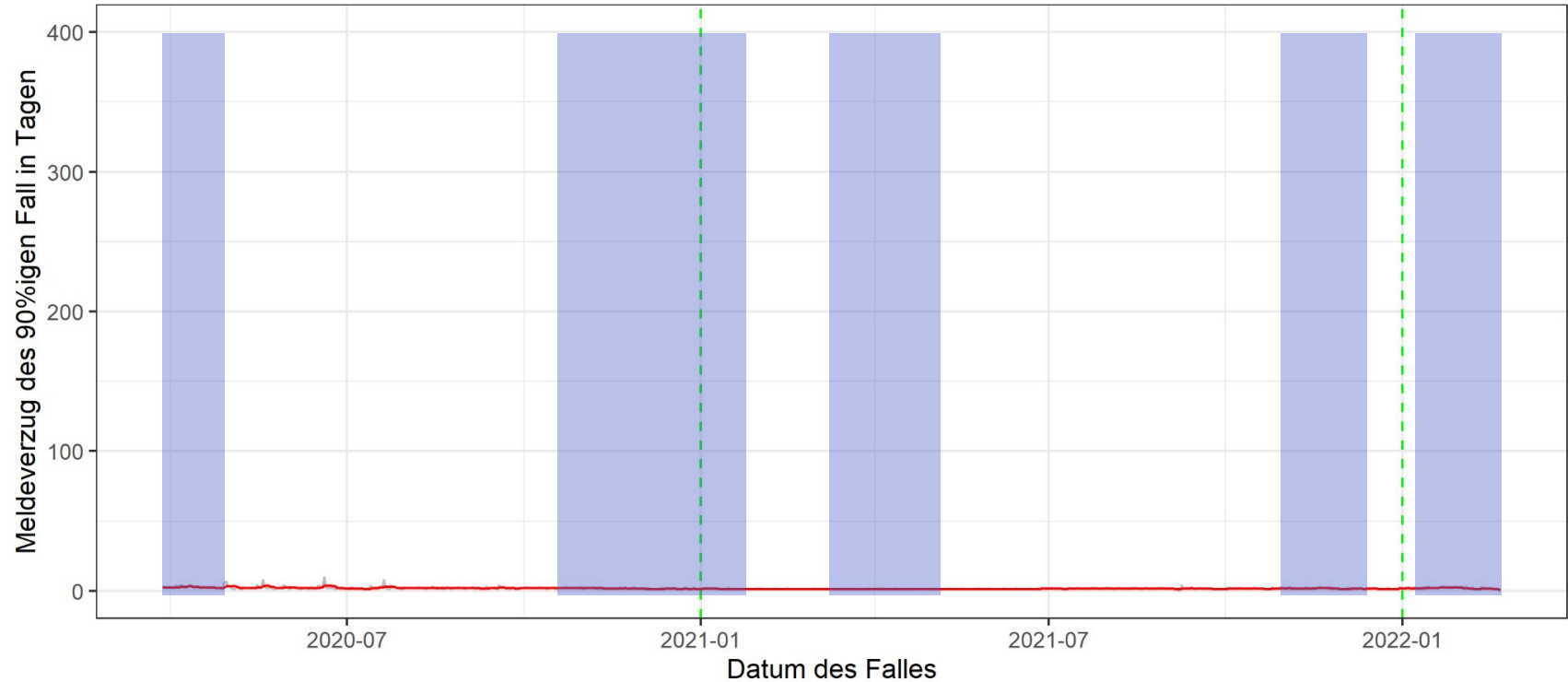


*Meldungen mit Verzug mind. 1 Tag.
 $\text{freq} = n(\text{Spätmeldungen am Meldedatum}) / n(\text{alle Meldungen am Meldedatum})$

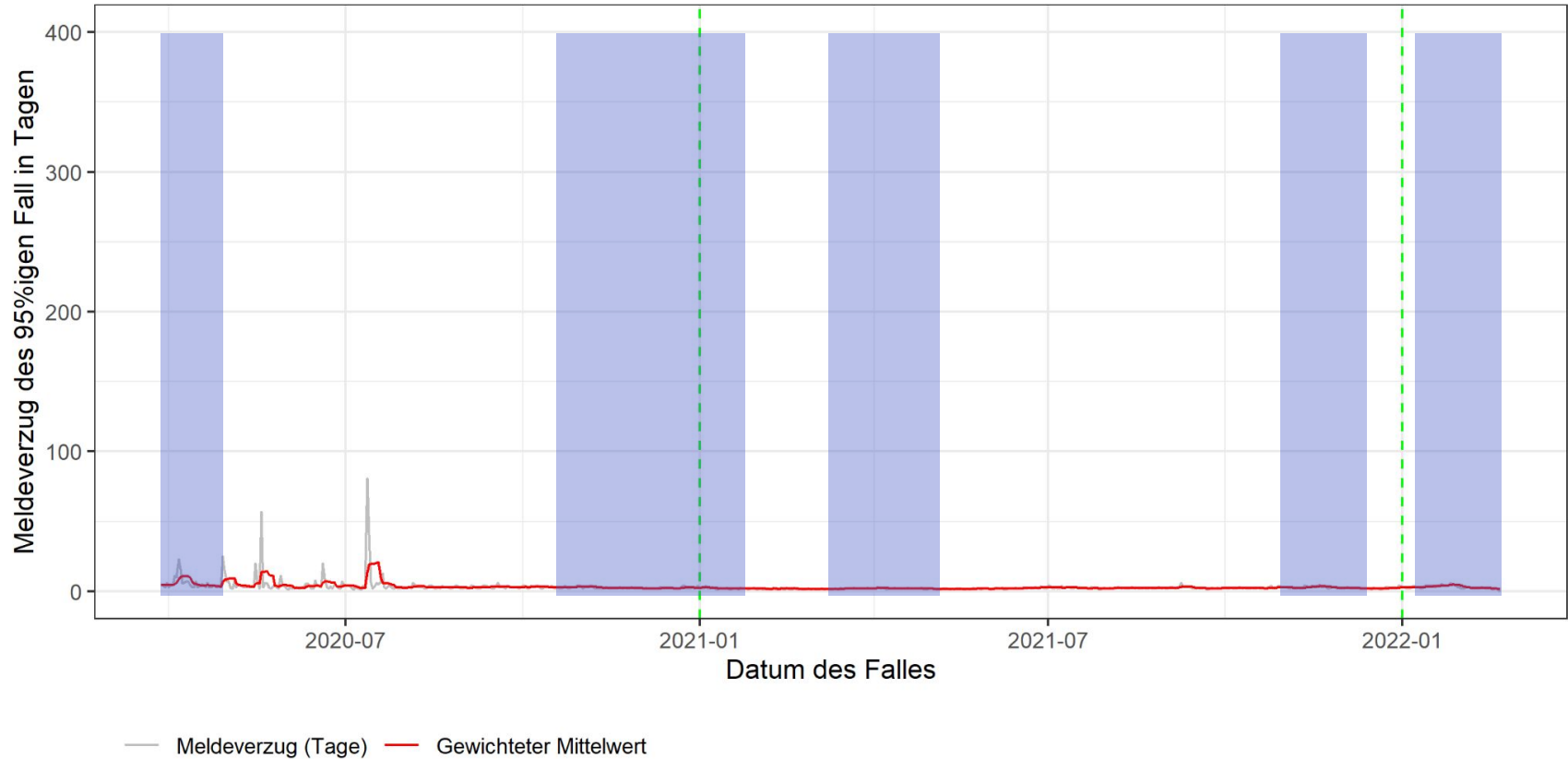
Täglicher Anteil an Fällen mit Meldeverzug



Zeit nach der 90% der Fälle berichtet wurden

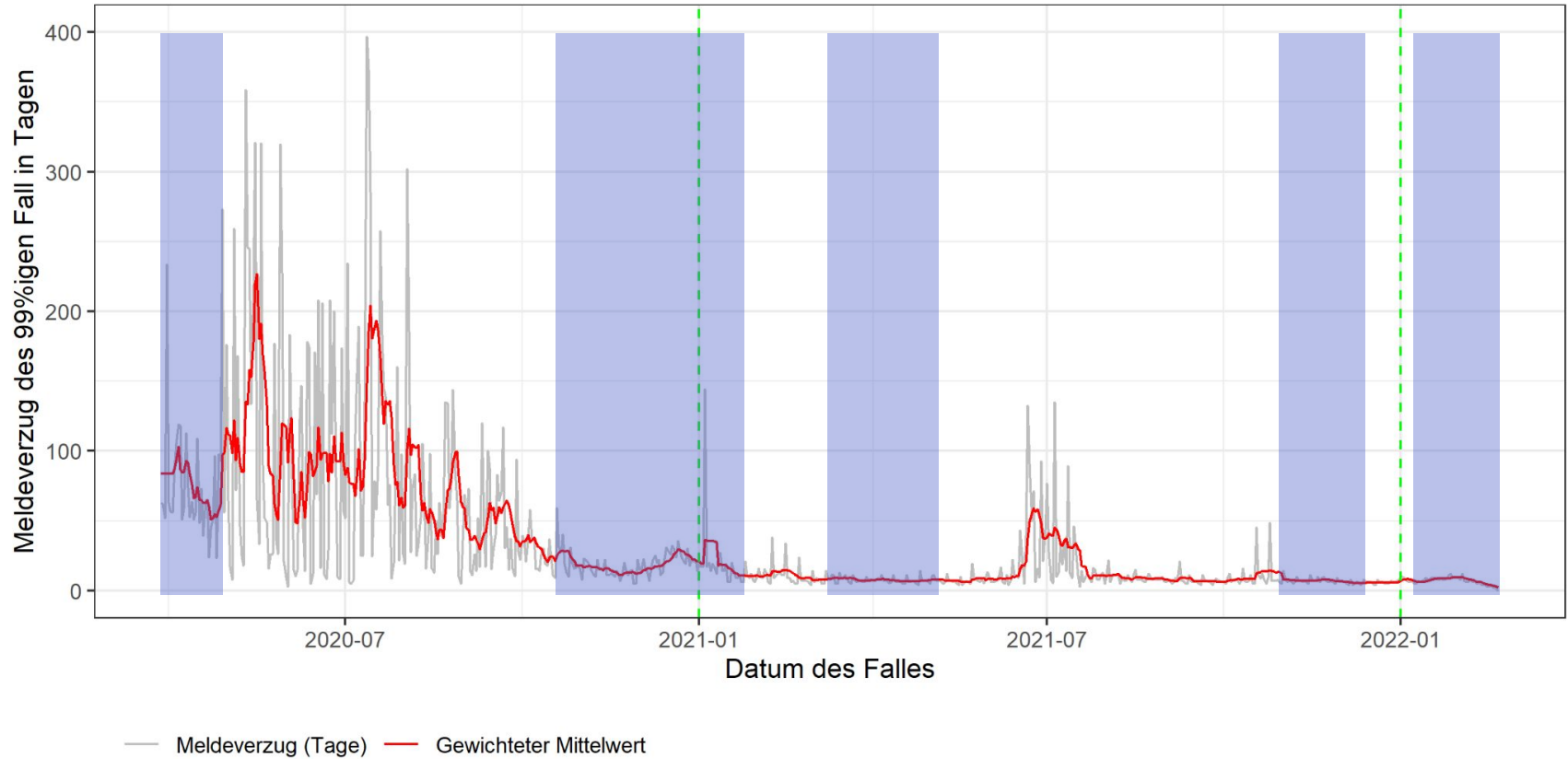


Zeit nach der 95% der Fälle berichtet wurden



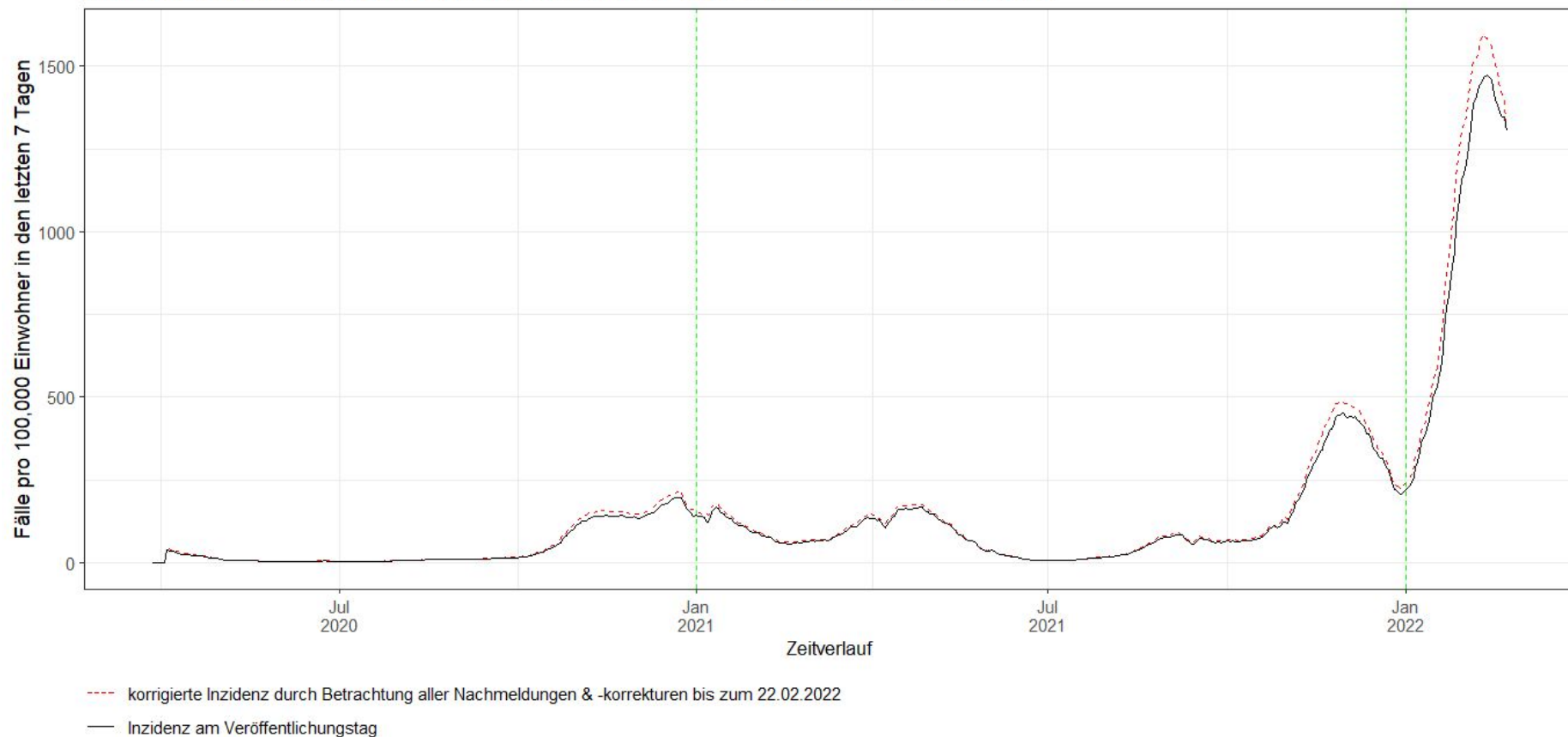
Meldeverzögerung auf individuelle Fälle gewichtet. Gleitender Mittelwert (7-Tage rückblickend) bei dem NAs ersetzt wurden.

Zeit nach der 99% der Fälle berichtet wurden



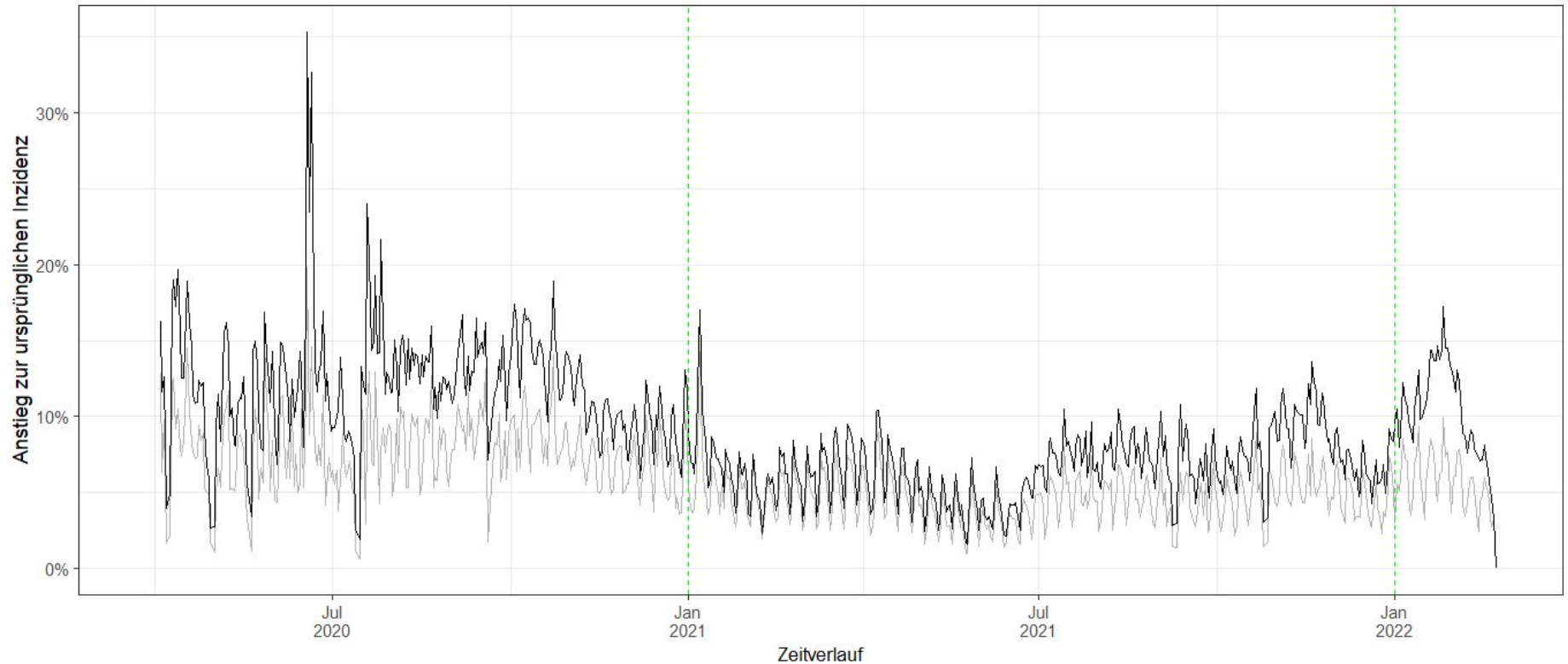
Meldeverzögerung auf individuelle Fälle gewichtet. Gleitender Mittelwert (7-Tage rückblickend) bei dem NAs ersetzt wurden.

Einfluss von Nachmeldungen & -korrekturen auf die 7-Tage-Inzidenz* der Neuinfektionen



*basiert auf wann der Fall beim Gesundheitsamt gemeldet ist, nicht wann der RKI ihn erfasst

Um wie viel Prozent erhöht sich die ursprüngliche 7-Tage-Inzidenz wegen Nachmeldungen



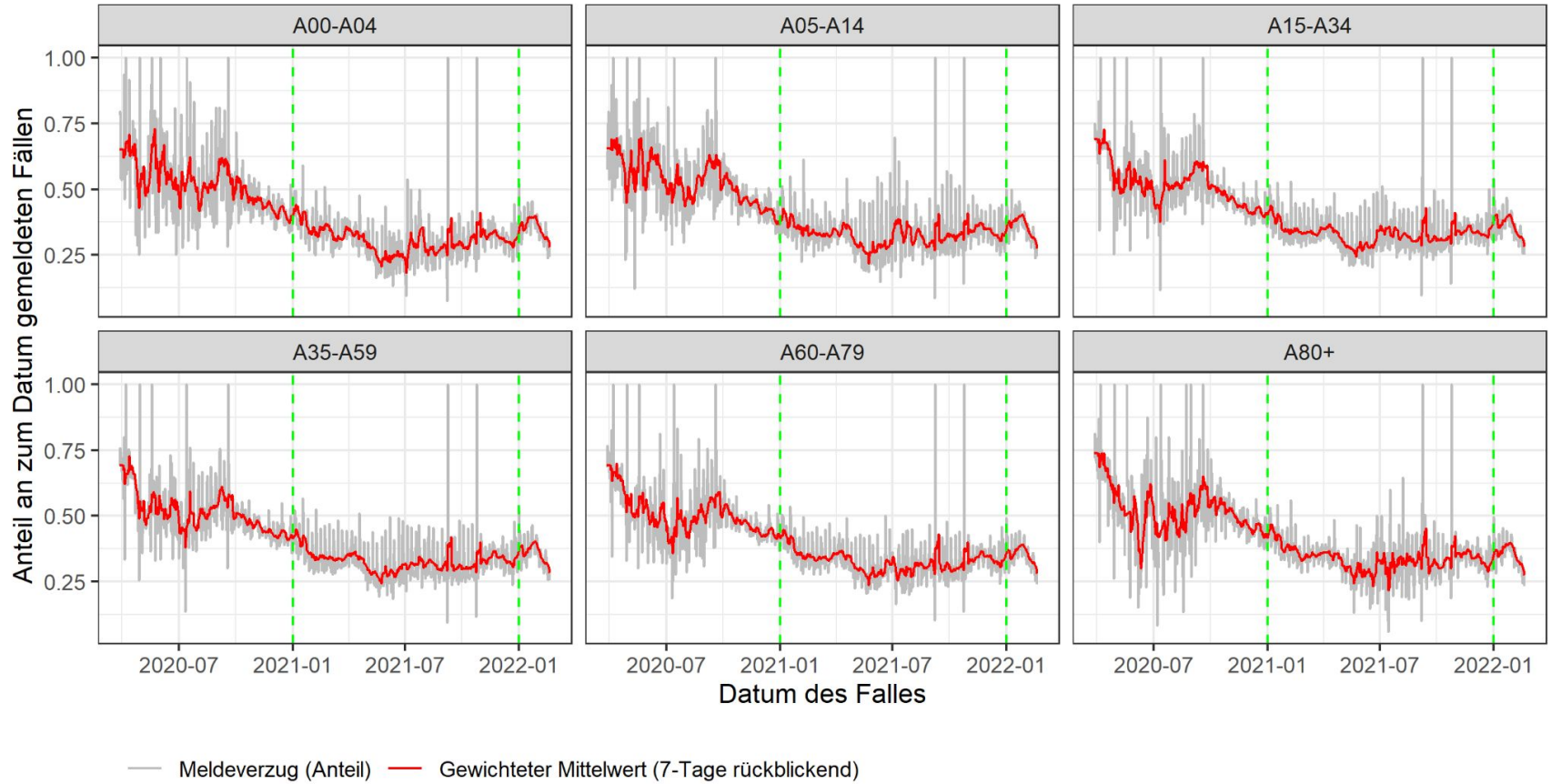
— nach Betrachtung aller Nachmeldungen & -korrekturen bis zum 22.02.2022

— nach Betrachtung aller Nachmeldungen & -korrekturen bis zum 1 Tag nach dem Veröffentlichungsdatum der ursprünglichen Inzidenz

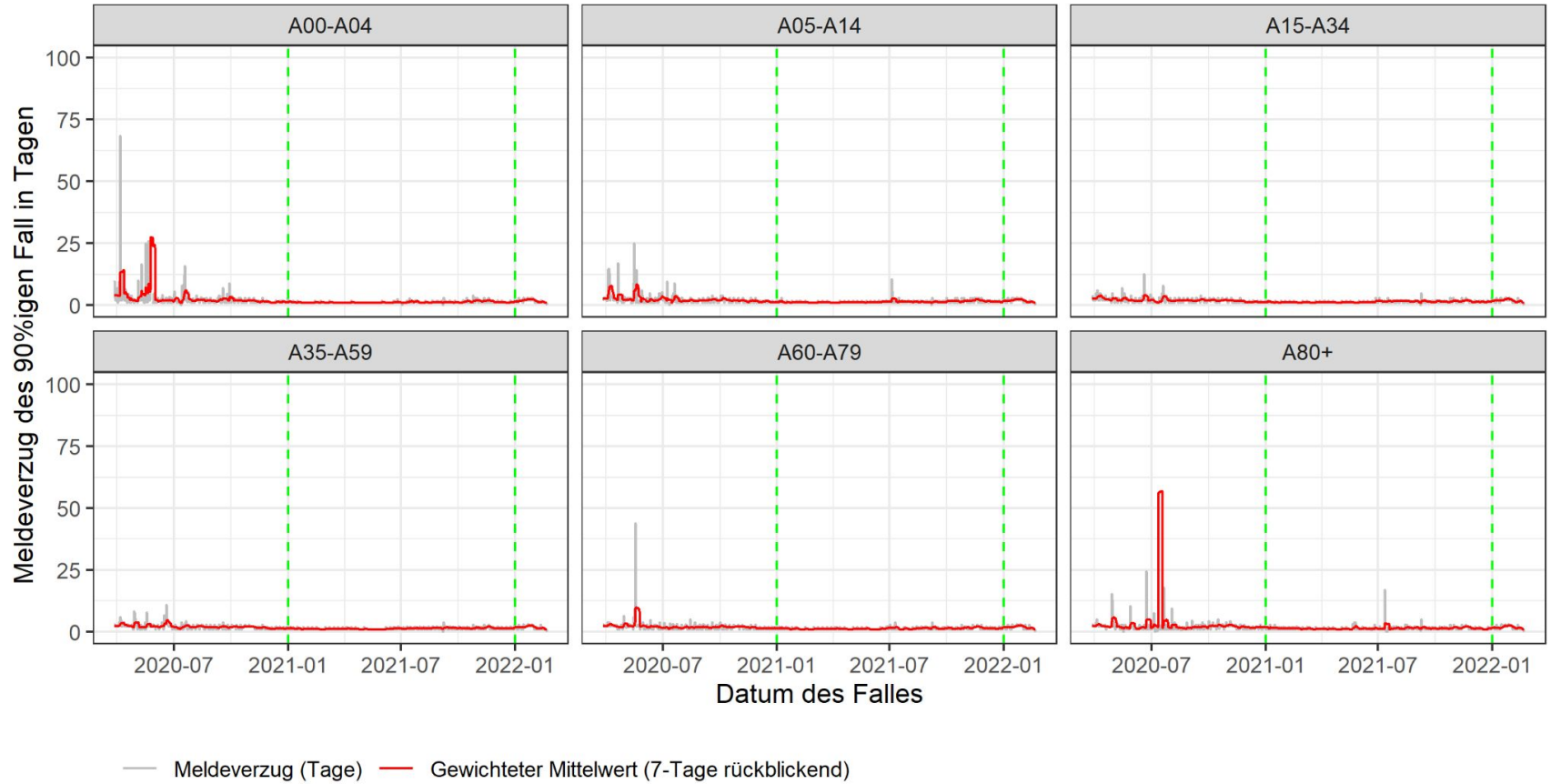
*basiert auf wann der Fall beim Gesundheitsamt gemeldet ist, nicht wann der RKI ihn erfasst

Meldeverzug pro Altersgruppe

Täglicher Anteil an Fällen mit Meldeverzug (pro Altersgruppe)

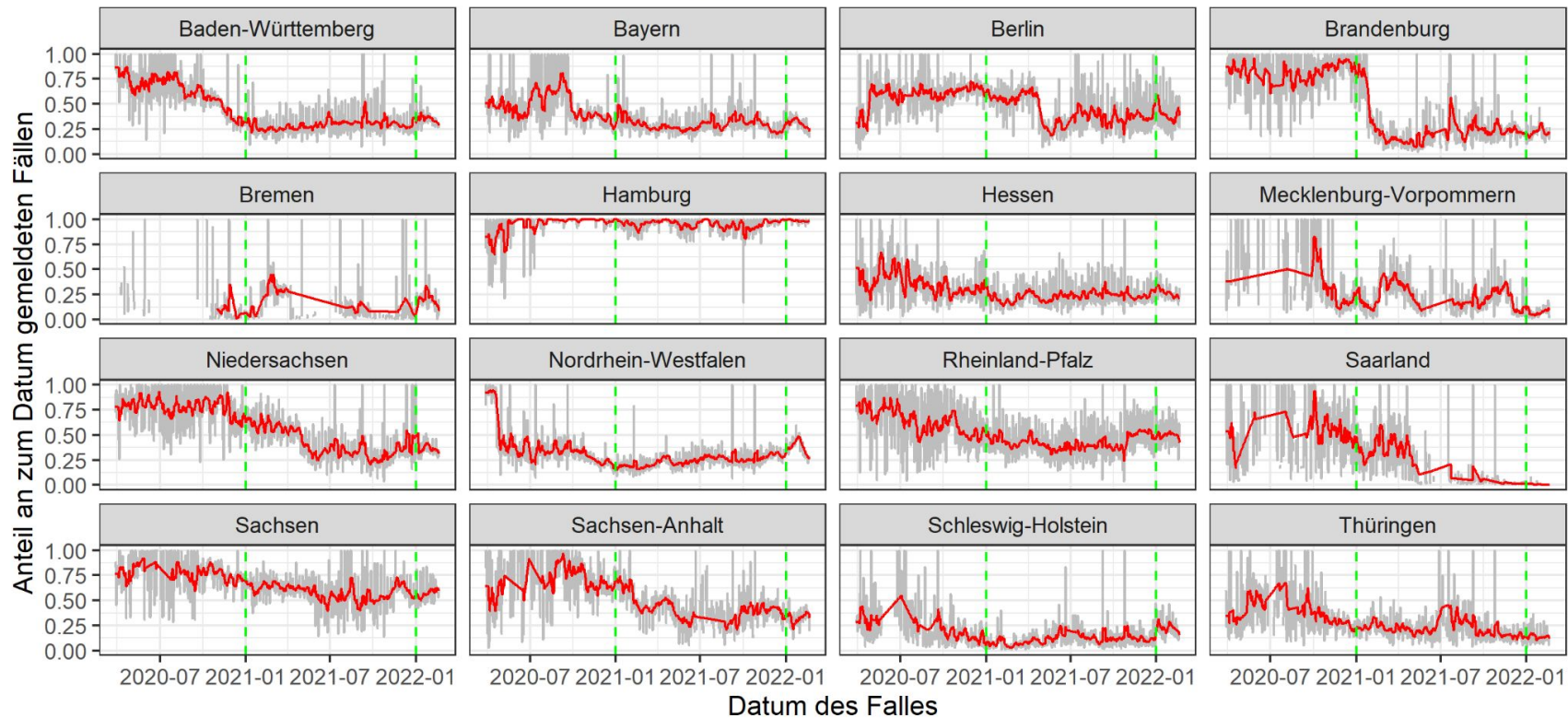


Zeit nach der 90% der Fälle berichtet wurden (pro Altersgruppe)



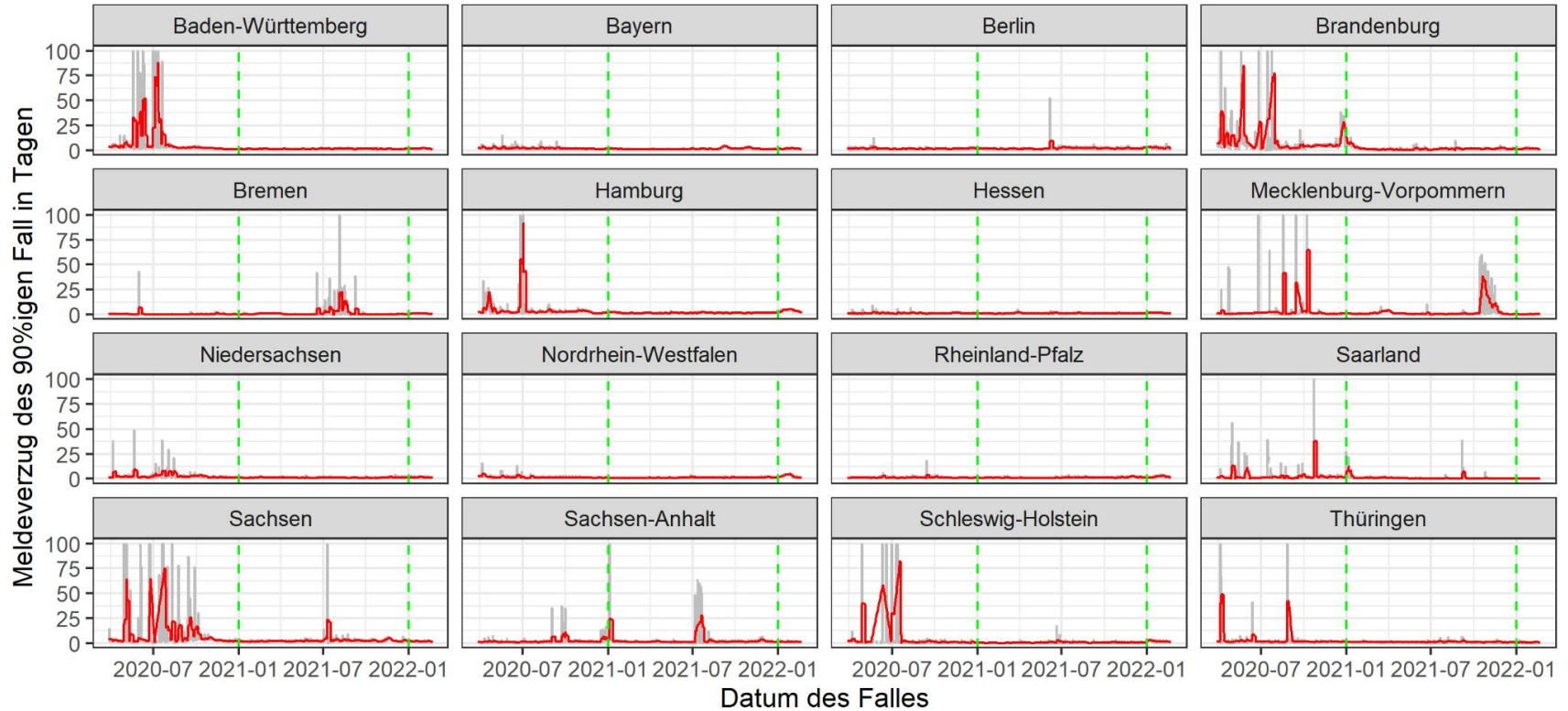
Meldeverzug (pro Bundesland)

Täglicher Anteil an Fällen mit Meldeverzug (pro Bundesland)



— Meldeverzug (Anteil) — Gewichteter Mittelwert (7-Tage rückblickend)

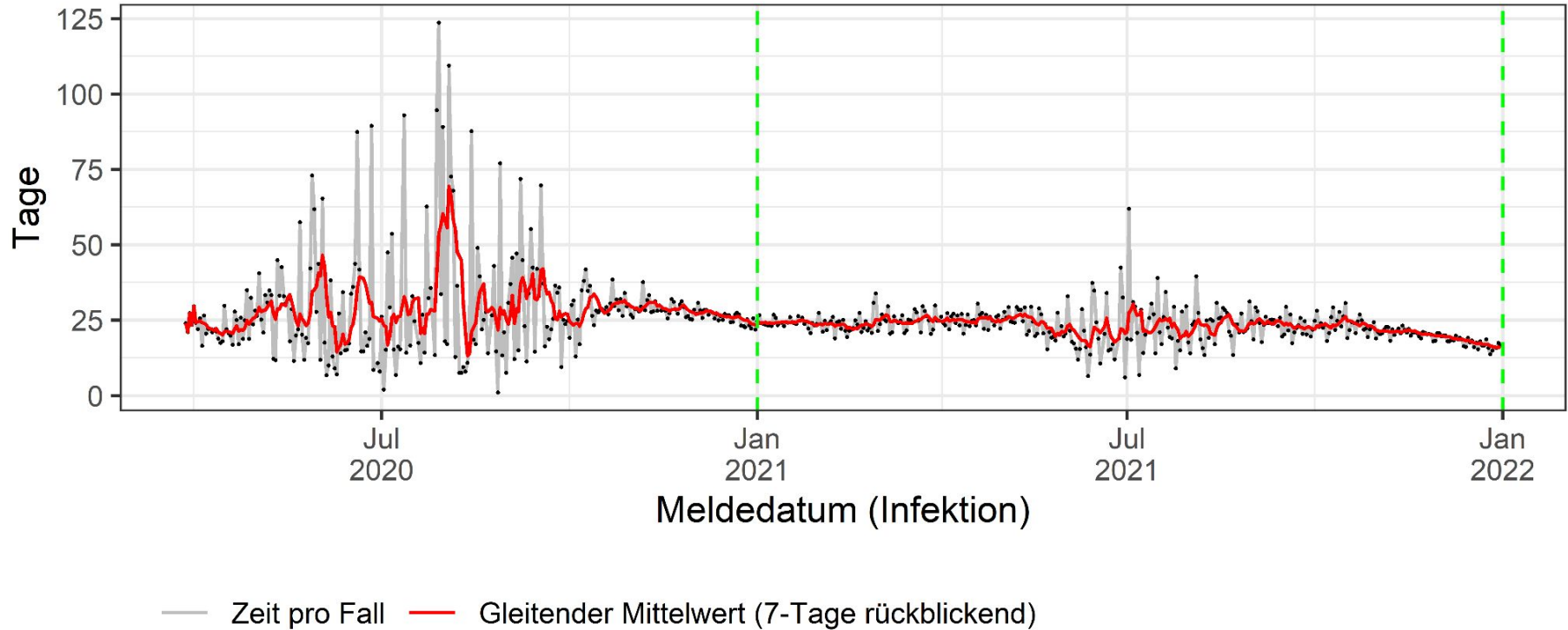
Zeit nach der 90% der Fälle berichtet wurden (pro Bundesland)



— Meldeverzögerung (Tage) — Gewichteter Mittelwert (7-Tage rückblickend)

Todesfälle

Durchschnittliche Zeit zwischen Meldedatum und Todesmeldung



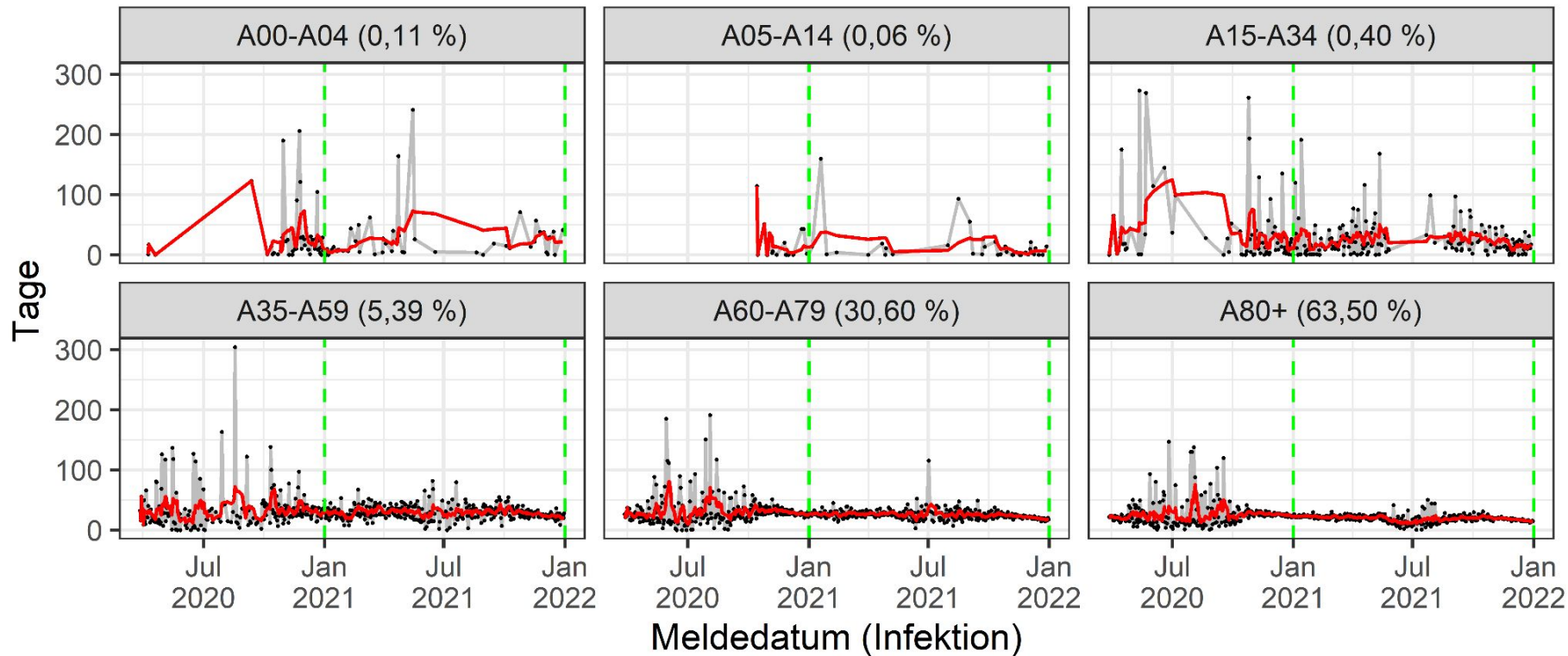
Zeitraum 27.03.2020 bis 31.12.2021

Highlights

- Der 7-Tage Zyklus gilt auch für den Meldeverzug
- Der Meldeverzug verringert sich über den Verlauf der Pandemie
- Wegen des Meldeverzugs liegt die Inzidenz meist 5-15% unter dem tatsächlichen Tageswert
- Hamburg meldet immer mit Meldeverzug
- Todesfälle sind mit den RKI Daten nur begrenzt analysierbar

Appendix

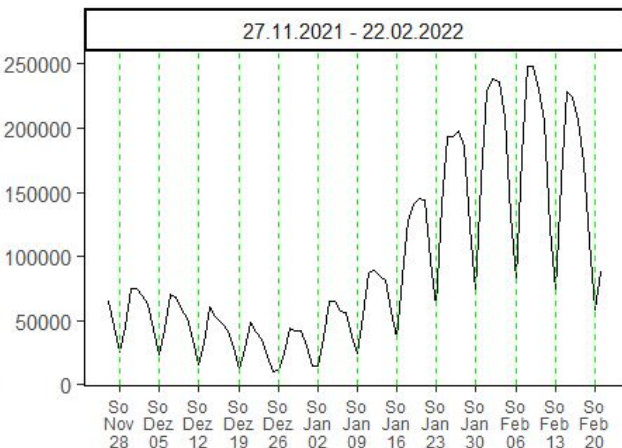
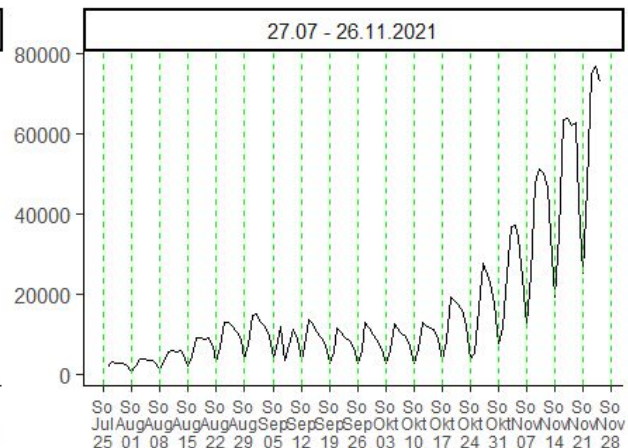
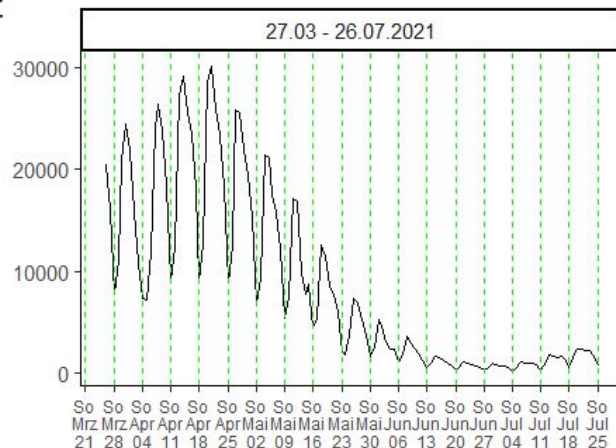
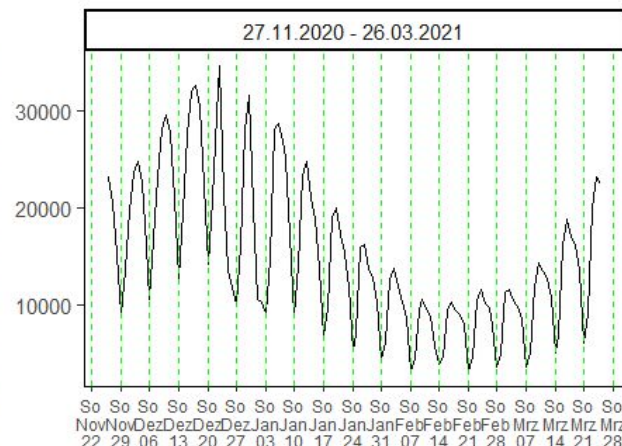
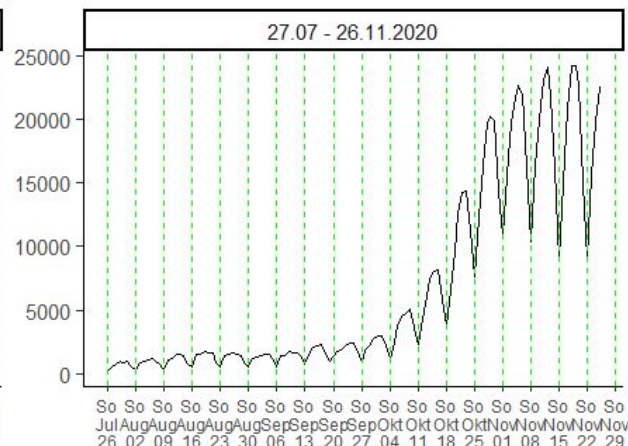
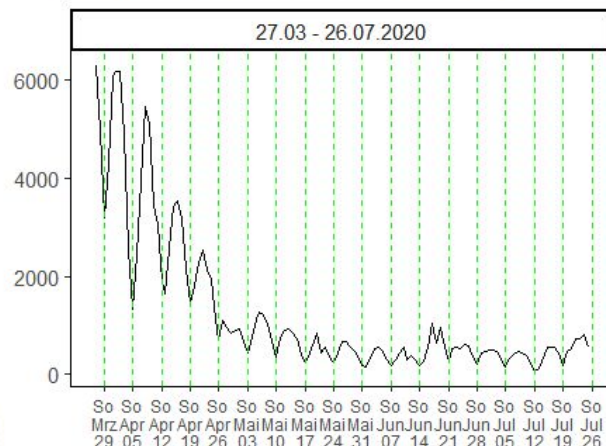
Durchschnittliche Zeit zwischen Meldedatum und Todesmeldung



7 Tage SMA, ohne "unbekannt", n(Todesmeldungen) = 117471, Zeitraum 27.03.2020 bis 31.12.2021

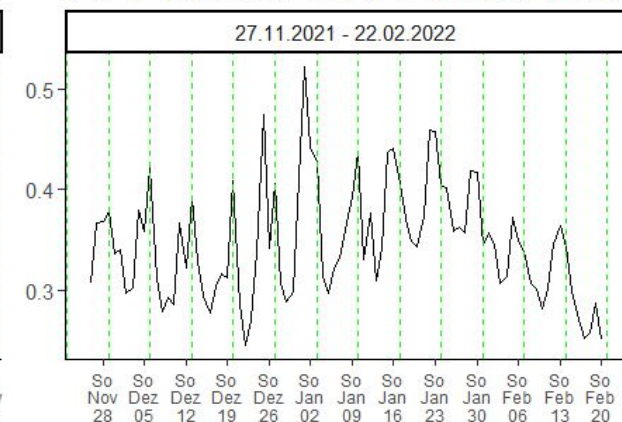
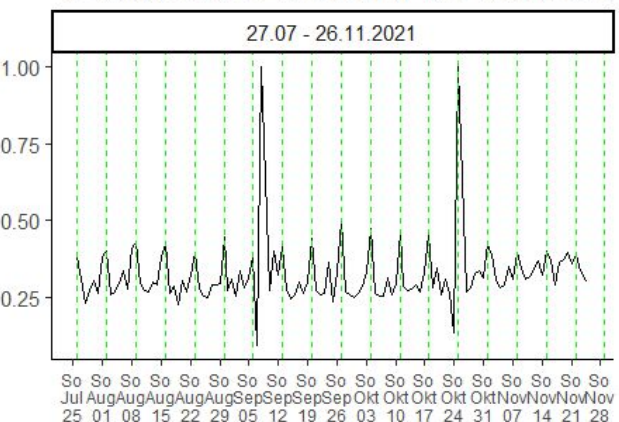
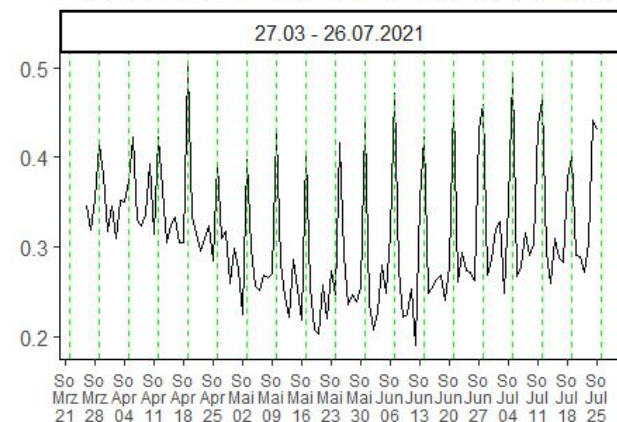
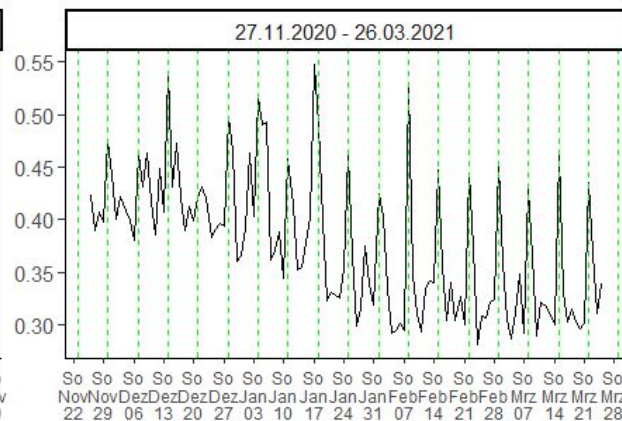
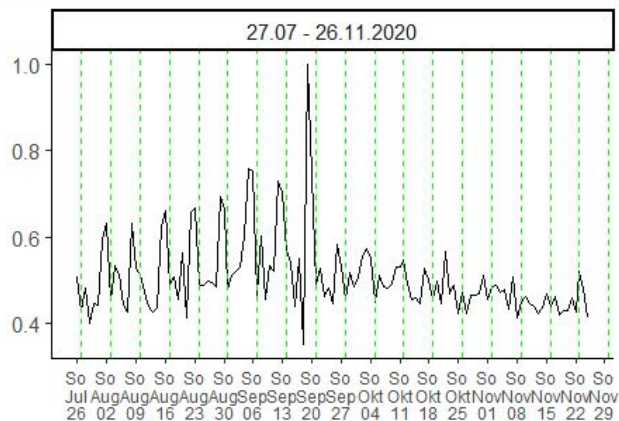
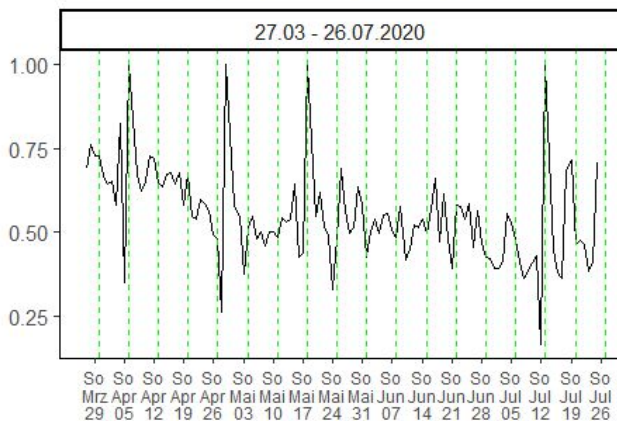
Anzahl der Neuinfektionsmeldungen hat einen wöchentlichen Zyklus

Fälle



Meldedatum

Anteil der Spätmeldungen* hat auch einen wöchentlichen Zyklus



Meldedatum

*Meldungen mit Verzug mind. 1 Tag.
 $y = n(\text{Spätmeldungen am Meldedatum}) / n(\text{alle Meldungen am Meldedatum})$