

**SUPPLEMENT zu**

Morfeld P, Timmermann B, Groß JV, Lewis P, Erren TC  
 COVID-19: Wie änderte sich die Sterblichkeit? - Mortalität von Frauen und Männern in  
 Deutschland und seinen Bundesländern bis Oktober 2020  
 COVID-19: How did mortality change? - Mortality of women and men in Germany and its  
 federal states until October 2020  
 DMW 2020

**Das Supplement besteht aus 4 Teilen:**

SUPPL1 - sonderauswertung-sterbefaelle\_2020-11-27.xlsx (Input Statistisches Bundesamt)  
 SUPPL2 - smr\_CI\_WOCHE\_LNV.txt (Stata-Auswerteprogramm)  
 SUPPL3 - SMR\_MONAT\_LNV\_27Nov2020\_all.xls (SMR-Ergebnisse pro Monat)  
 SUPPL4 - SMR\_WOCHE\_LNV\_27Nov2020\_all.xls (SMR-Ergebnisse pro Woche)

**SUPPL1 - sonderauswertung-sterbefaelle\_2020-11-27.xlsx**

Originaldatei des Statistischen Bundesamts zur Zahl der Todesfälle in Deutschland und in den 16 Bundesländern pro Jahr 2016 bis 2020, bereitgestellt vom Statistischen Bundesamt am 27. November 2020:

Destatis. Sterbefälle - Fallzahlen nach Tagen, Wochen, Monaten, Altersgruppen und Bundesländern für Deutschland 2016 - 2020 2020.

URL: <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Sterbefaelle-Lebenserwartung/Tabellen/sonderauswertung-sterbefaelle.html?nn=375478>

Aus dieser Datei wurden die 6 Datenblätter BL\_2016\_2020\_Monate\_AG\_Weiblich; BL\_2016\_2020\_Monate\_AG\_Männlich, BL\_2016\_2020\_Monate\_AG\_Ins, BL\_2016\_2020\_KW\_AG\_Weiblich, BL\_2016\_2020\_KW\_AG\_Männlich, BL\_2016\_2020\_KW\_AG\_Ins ohne Tabellenkopf extrahiert. Diese 6 Datenblätter enthalten die monatlich bzw. wöchentlich aufsummierten Zahlen der Todesfälle an allen Todesursachen, aufgeteilt nach Geschlecht (weiblich, männlich, insgesamt) und Alter bei Tod (< 65 Jahre, ≥ 65 Jahre, insgesamt), für 17 Regionen (Deutschland und die 16 Bundesländer) von Januar bis Dezember 2016-2019 und Januar bis Oktober 2020 bzw. Kalenderwoche 1 bis 52 in 2016-2019 und Kalenderwoche 1 bis 44 in 2020.

Die Datenblätter werden umformatiert, so dass Stata-Dateien pro Bundesland, Altersklasse, Monat- bzw. Wochenbasis, und Geschlecht zum Import in die Auswerteprogramme (siehe SUPPL2) vorliegen: z.B. Inputdatei BL1\_D\_0-65\_m\_male.dta, d.h. Bundesland 1 = D = Deutschland, 0-65: Altersklasse < 65, m = Monatsbasis, mal = Männer mit folgender Satzstruktur (Beispieldaten für Deutschland im Januar für Männer < 65 Jahre):

date	BL	BL_nr	Alt_Kat	Monat	J_16	J_17	J_18	J_19	J_20
27Nov2020	Deutschland	1	0-65	1	7991	8239	8056	7988	7881

Zur Bedeutung der Variablen “date” bis “Monat” siehe die Erläuterungen zu SUPPL3 und SUPPL4, “J\_16” bis “J\_20” enthalten die Todesfallzahlen pro Monat in den Jahren 2016 bis 2020.

**SUPPL2 - smr\_CI\_WOCHE\_LNV.txt**

Stata-Programm [1] zur Auswertung der extrahierten Daten (vgl. SUPPL1). Die Datei enthält den Klartext des Programms und kann mit einem beliebigen Texteditor geöffnet werden. Um als Stata-Programm lauffähig zu sein, muss lediglich die Endung .txt in die Endung .do umbenannt werden.

Das Beispielprogramm berechnet Standardisierte Mortalitätsratios SMR mit Konfidenzintervallen für jede Woche und insgesamt in 2020 unter Annahme einer Lognormalverteilung der Todesfallzahlen zwischen den Referenzjahren 2016 bis 2019. Zum Aufbau des Dateinamens und zu den Namen der Variablen im Programm siehe die Erläuterungen zu SUPPL3 und SUPPL4.

Wir berechnen vereinfachte Standardized Mortality Ratios (SMRs) als beobachtete Zahl an Todesfällen (Zahl der Toten pro Monat bzw. Woche im Indexjahr 2020) geteilt durch die erwartete Zahl der Todesfälle (entsprechende Durchschnittswerte der Referenzjahre 2016 – 2019) und 95% Konfidenzintervalle zu den SMRs auf Basis einer Poissonverteilung der Todesfallzahlen innerhalb von 2020 [2, 3]. Wir bestimmen die erwarteten Zahlen als arithmetische Mittelwerte und als geometrische Mittelwerte, was der Annahme einer Normal- bzw. Lognormalverteilung der Todesfallzahlen zwischen den Jahren 2016 – 2019 entspricht. Um die Variation der Todesfallzahlen zwischen den Jahren 2016 – 2019 zu berücksichtigen, addieren wir die Varianzen der Zufallsfehler in 2020 und der Fehler zwischen 2016 – 2019 auf einer logarithmischen Skala, um so erweiterte 95% Konfidenzintervalle für die SMRs zu bestimmen, also so, wie für random-effects Analysen empfohlen [4]. Bei Annahme einer Normalverteilung schätzen wir entsprechend 95% Konfidenzintervalle der erwarteten Todeszahlen als Erwartet  $\pm 1,96 \cdot (\text{Standardfehler SE der Todesfallzahlen in 2016-2019})$ , im Falle einer Lognormalverteilung als  $\exp(\log(\text{Erwartet}) \pm 1,96 \cdot (\text{SE des Logarithmus der Todesfallzahlen 2016-2019}))$ .

Diese SMRs sind vereinfacht berechnet, da die übliche indirekte Standardisierung nach Geschlechtern und 5-Jahres-Altersklassen [2, 3] aufgrund fehlender Populationsdaten für das Jahr 2020 nicht durchführbar und die vom Statistischen Bundesamt vorgegebene Aufteilung der Todesfälle in nur zwei Altersklassen (< 65 Jahre,  $\geq 65$  Jahre) sehr grob ist.

Das Programm schätzt zudem adaptierte (reduzierte) Zahlen beobachteter und erwarteter Todesfälle, die mit den erweiterten Konfidenzintervallschätzungen bei Annahme einer Poissonverteilung korrespondieren, so dass nachgeschaltete Modellierungen der SMRs die erweiterten Konfidenzintervalle im Sinne einer random-effects-Analyse unmittelbar berücksichtigen können (zweischrittige Approximation im Programm über: beobachtet =  $1/\text{Varianz}(\log \text{ SMR})$  and erwartet = beobachtet/SMR) [2].

Wir berechneten nachgeschaltet zum SMR-Programm Poissonmodelle mit robusten Varianzschätzern, z.B. um Trends der SMRs über Monaten bzw. Wochen in 2020 zu bestimmen, wobei gleichzeitig nach Alter, Geschlecht und Bundesländern adjustiert wurde [2, 3, 5], und wir berechneten entsprechende, erweiterte Modelle mit Überdispersionsparametern (Annahme einer negativen Binomialverteilung) [6].

Anmerkung: Die vereinfachte Berechnung der SMRs kann zu verzerrten Schätzungen führen, wenn sich Umfang und Alters-Geschlechter-Verteilungen der Populationen zwischen 2016 und 2020 ändern (die Zahl der in Deutschland lebenden Menschen und der Anteil älterer Personen wachsen, was eine tendenziell steigende Zahl an Todesfällen bedingt [7], also potentiell überschätzte SMRs). Allerdings beeinflusst diese Verzerrung kaum Vergleiche der SMRs in 2020. Unsere Auswertung des Indexjahrs 2020 bezieht sich auf die vom Statistischen Bundesamt gewählten Referenzjahre 2016 – 2019. Die Ergebnisse stellen nicht unbedingt die SMRs dar, die sich im kontrafaktischen Vergleich mit dem Jahr 2020 ohne COVID-19 ergeben würden, denn

die Referenzjahre unterscheiden sich vom Indexjahr durch z.B. unterschiedlich stark und zu unterschiedlichen Zeiten verlaufende Grippe- und Hitzewellen ([https://www.rki.de/DE/Content/InfAZ/I/Influenza/IPV/IPV\\_Node.html](https://www.rki.de/DE/Content/InfAZ/I/Influenza/IPV/IPV_Node.html), <https://www.umweltbundesamt.de/indikator-heisse-tage#die-wichtigsten-fakten>). Für den Zeitraum April/Mai ist dieser Unterschied aber kaum von Bedeutung.

Haftungsausschluss: die Autoren uebernehmen keine Garantie für die Richtigkeit aller Ergebnisse. Insbesondere bei Anwendung des Programms auf einen anderen Datenbestand können bei ungünstiger Datenlage Probleme auftreten, z.B. beim Vorkommen von Nullen als Fallzahlen in einigen Zellen.

Literatur zu diesem Abschnitt:

1. StataCorp. Stata Statistical Software: Release 14. Texas, USA: StataCorp LP; 2015
2. Breslow NE, Day, NE. Statistical Methods in Cancer Research. Volume II – The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer; 1987
3. Rothman KJ, Greenland, S, Lash, TL. Modern Epidemiology. 3. Aufl. Philadelphia: Lippincott Williams & Wilkins; 2008
4. Sutton AJ, Abrams KR, Jones DR et al. Methods for Meta-analysis in Medical Research. Chichester, U.K: Wiley; 2000
5. Checkoway H, Pearce N, Kriebel D. Research Methods in Occupational Epidemiology. 2. Aufl. New York: Oxford University Press; 2004
6. Cameron AC, Trivedi PK. Microeconometrics. Methods and Applications. New York: Cambridge University Press; 2005
7. zur Nieden F, Sommer B, Lüken S. Sonderauswertung der Sterbefallzahlen 2020 - Daten zur Einordnung einer zeitweisen Übersterblichkeit im Zusammenhang mit der Corona-Pandemie. Statistisches Bundesamt. WISTA:4, 2020

**SUPPL3 - SMR\_MONAT\_LNV\_27Nov2020\_all.xls (SMR-Ergebnisse pro Monat)**

**SUPPL4 - SMR\_WOCHE\_LNV\_27Nov2020\_all.xls (SMR-Ergebnisse pro Woche)**

Diese beiden Beispielergebnisdateien enthalten alle Daten, die in der Publikation dargestellt sind sowie alle Aufgliederungen nach Bundesländern, Altersklassen und Geschlechtern.

Dateinamen der Beispieldateien:

WOCHE bedeutet, dass die Daten auf Wochenbasis dargestellt sind, MONAT auf Monatsbasis. LNV bedeutet, dass eine Lognormalverteilung der Todesfallzahlen zwischen den Jahren 2016 – 2019 zur Berechnung der erwarteten Fallzahlen angesetzt wurde (vgl. SUPPL2). Das Datum 27Nov2020 nennt das Veröffentlichungsdatum der zugrunde liegenden Datei des Statistischen Bundesamtes (vgl. SUPPL1).

Variablen (Spalten) der Beispieldateien:

Die Variable "date" enthält das Veröffentlichungsdatum der Basisdaten des Statistischen Bundesamtes (siehe SUPPL1), "BL" die Namen und "BL\_nr" Nummern für Deutschland (BL\_nr = 1) und die 16 Bundesländer (BL\_nr = 2 bis 17). "Alt\_Kat" kennzeichnet die drei Altersklassen: 0-65 (<65 Jahre), 65plus (≥65 years) und total (insgesamt); „Sex“ die drei Geschlechtergruppen: female (weiblich), male (männlich) und total (insgesamt). "Monat" gibt den Monat von 1 (Januar) bis 10 (Oktober) in 2020 bzw. „Woche“ die Kalenderwoche von 1 bis 44 in 2020 an; der Eintrag 9999 kennzeichnet die Zeile mit dem Gesamtwert über alle Monate bzw. Wochen. LNV-Dateien enthalten in den Variablen "LJ\_16" to "LJ\_20" die natürlichen Logarithmen der Anzahlen der Todesfälle der Referenzjahre 2016 bis 2019; "Deaths\_lo\_1619" und "Deaths\_hi\_1619" sind die Grenzen der 95% Konfidenzintervalle dieser Todesfallzahlen, geschätzt auf Basis einer

Lognormalverteilung. "Observed" ist die Zahl der beobachteten Todesfälle in 2020, "Expected" enthält die erwartete Zahl an Todesfällen, geschätzt als geometrischer Mittelwert der Todesfallzahlen 2016-2019 in LNV-Dateien. Die Spalte "SMR" enthält die Standardisierte Mortalitätsratio = Quotient aus "Observed" und "Expected". "SMR\_95\_lo" und "SMR\_95\_up" sind 95%-Konfidenzintervallgrenzen auf Basis einer Poissonverteilung; "SMR\_lo" und "SMR\_up" sind die erweiterten 95%-Konfidenzintervalle; "Obs\_corr" und "Exp\_corr" die adaptierten (reduzierten) Fallzahlen, die mit erweiterten Konfidenzintervallen korrespondieren.