

# Mehrfacherfassungen von Hepatitis-B- und Hepatitis-C-Meldungen in den Gesundheitsämtern – Identifikation und statistische Schätzung

## Hintergrund

Die Weltgesundheitsorganisation (WHO) schätzt, dass weltweit ca. 296 Millionen Menschen mit dem Hepatitis-B-Virus (HBV) und ca. 58 Millionen Menschen mit dem Hepatitis-C-Virus (HCV) infiziert sind und jährlich mehr als 1,1 Millionen Menschen an einer Hepatitis B oder Hepatitis C versterben.<sup>1,2</sup> In Deutschland werden HBV- und HCV-Infektionen gemäß Infektionsschutzgesetz (IfSG) seit 2001 epidemiologisch überwacht. Eine dauerhafte und systematische Surveillance der Hepatitis-B- und C-Fälle ist für die Planung, Durchführung und Evaluation gezielter Präventions- und Eindämmungsmaßnahmen essenziell. Den Gesundheitsämtern obliegt dabei eine zentrale Rolle, da sie für die Datenerhebung und Ermittlung zu den gemeldeten Fällen zuständig sind.

Dabei liegen den Gesundheitsämtern zu den Meldungen alle Informationen vor, die für die Nachverfolgung und die Einschätzung des Übertragungsriskos von Bedeutung sind. An die Landesbehörden und das Robert Koch-Institut (RKI) werden nur die Informationen übermittelt, die zur Einschätzung der epidemiologischen Situation eine wichtige Rolle spielen; diese lassen keine Identifizierung infizierter Personen zu. Zu den übermittelten Informationen gehören Geburtsjahr und Geburtsmonat, Geschlecht, Landkreis, Stadtkreis oder Bezirk des Wohnorts, klinische Angaben und Labornachweise.

Am 25.7.2017 erfolgte eine Änderung des IfSG in Bezug auf die Meldepflicht von HBV- und HCV-Infektionen.<sup>3,4</sup> Seitdem wurde sie gemäß §7 IfSG auf alle Nachweise ausgedehnt, unabhängig vom klinischen Bild (symptomatisch oder asymptomatisch) und Stadium (akut oder chronisch) (s. INFOBOX 1). Zusätzlich erfüllen seit 2019 nicht nur akute, sondern auch chronische HBV-Infektionen die Falldefinition und werden in der Statistik veröffentlicht. Für beide Infektionen wurde in den vergangenen Jahren ein Anstieg der Meldungen registriert.<sup>3,4</sup>

Da HBV und HCV bei chronischen Infektionen oft jahre- oder jahrzehntelang nachweisbar sind, kann es durch die neue Meldepflicht für chronische Infektionen vorkommen, dass Gesundheitsämter im Verlauf multiple Labormeldungen für eine einzelne Person erhalten. Bei jeder Meldung einer HBV- oder HCV-Infektion gleicht das zuständige Gesundheitsamt zum einen anhand der personenbezogenen Daten ab, ob der Fall im eigenen Gesundheitsamt be-

### INFOBOX 1

#### Daten aus dem Meldewesen

Laut Infektionsschutzgesetz (IfSG) werden dem Gesundheitsamt gemäß §6 Abs.1 Nr.1 Buchst. e der Krankheitsverdacht, die Erkrankung sowie der Tod an akuter Virushepatitis sowie gemäß §7 Abs.1 Nr.21 bzw. Nr.22 IfSG der Nachweis von Hepatitis-B-Virus (HBV) bzw. Hepatitis-C-Virus (HCV) (Meldepflicht für alle Nachweise) namentlich gemeldet. Der gemeldete Fall wird gemäß §11 Abs.1 IfSG an die zuständige Landesbehörde übermittelt und erfüllt die sogenannte Referenzdefinition, wenn mindestens einer der folgenden Labornachweise vorliegt:

#### Hepatitis B

- ▶ Nukleinsäurenachweis (z. B. Polymerasekettenreaktion [PCR]) oder
- ▶ HBs-Antigennachweis (z. B. Enzyme-linked Immunosorbent Assay [ELISA]), bestätigt durch mindestens eine der drei folgenden Methoden:
  - ▶ Zusatztest (z. B. HBsAg-Neutralisationstest)
  - ▶ Anti-HBc-Gesamt-Antikörpernachweis (indirekter [serologischer] Nachweis)
  - ▶ HBe-Antigennachweis

#### Hepatitis C

- ▶ HCV-Core-Antigennachweis (z. B. Immunoassay) oder
- ▶ Nukleinsäurenachweis (z. B. PCR)

reits bekannt ist. Ist der Fall noch nicht bekannt, wird zum anderen beim meldenden Labor, dem meldenden Arzt/der meldenden Ärztin und/oder der betroffenen Person ermittelt, ob die Infektion bereits vorher bekannt war. Ergibt sich dabei, dass der Fall bereits einem anderen Gesundheitsamt gemeldet war, wird der Fall in der Datenbank nicht neu erfasst.

Bei dem beschriebenen Vorgang kann es zu mehreren Szenarien kommen, die eine Mehrfacherfassung innerhalb des Gesundheitsamts oder zwischen mehreren Gesundheitsämtern verursachen:

1. Die personenbezogenen Daten zu einer früher gemeldeten Infektion werden im Gesundheitsamt nicht oder nicht detailgenau\* abgeglichen, da z. B. die Kapazitäten nicht vorhanden sind oder die personenbezogenen Daten zur Erfüllung von Datenschutzvorgaben bereits gelöscht wurden;
2. die Infektion wurde vorher durch eine andere Ärztin/einen anderen Arzt oder ein anderes Labor diagnostiziert und dem jetzigen Meldenden liegt diese Information nicht vor,
3. Informationen zu einer bereits diagnostizierten Infektion sind beim Patienten/bei der Patientin nicht einzuholen (z. B. durch fehlende Erreichbarkeit oder Sprachbarrieren) oder
4. ein Abgleich der personenbezogenen Daten ist nicht möglich, da die Probe anonymisiert aus einem Testzentrum ins meldende Labor eingesandt wurde.

Ziel dieser Studie war es herauszufinden, welcher Anteil der Hepatitis-B- und Hepatitis-C-Meldungen seit der IfSG-Novellierung im Jahr 2017 auf Mehrfacherfassungen zurückzuführen ist. Das erste Unterziel bestand darin, mehrfach erfasste Personen (Personen-Dubletten) in den Meldekategorien Hepatitis B und Hepatitis C innerhalb der Gesundheitsämter zu identifizieren und zu quantifizieren (Teil 1). Ein weiteres Unterziel war es, durch die Simulation statistisch erwarteter Zufalls-Dubletten eine Methode zu entwickeln, um die Anzahl der Personen-Dubletten auf Basis der Anzahl von Meldedaten insgesamt vorherzusagen (Teil 2).

## Methoden

An den Landesstellen Hamburg und Rheinland-Pfalz (Arbeitsorte der Autorinnen Bühler und Schoeps) wurde nach Daten-Dubletten unter den Hepatitis-B- und Hepatitis-C-Meldefällen (vgl. Begriffsbestimmungen in INFOBOX 2) innerhalb der einzelnen Gesundheitsämter\*\* gesucht, die sich in Geburtsjahr, Geburtsmonat und Geschlecht gleichen. In die Dublettensuche wurden alle Fälle einbezogen, die keine fehlenden Werte im Geburtsjahr enthielten, die die Referenzdefinition erfüllten und bei denen das Meldedatum ab Januar 2017 lag (s. Abb. 1).

Nach einer Pilotierungsphase wurden die Daten im Oktober 2022 aus der Datenbank der Meldesoftware SurvNet abgefragt und Listen mit Daten-Dubletten an diejenigen Gesundheitsämter in Hamburg und Rheinland-Pfalz verschickt, in denen mindestens eine Daten-Dublette identifiziert worden war. Die Gesundheitsämter wurden um den Abgleich mittels identifizierender Informationen und um Rückmeldung an die Landesstellen gebeten.

## Statistische Auswertung (Teil 1)

Nach der Rückmeldung der Gesundheitsämter wurde für jedes Gesundheitsamt einzeln und jeweils für alle Gesundheitsämter in Rheinland-Pfalz und Hamburg der Anteil der Personen-Dubletten unter den Daten-Dubletten ( $\text{Anteil}_{\text{PD}}$ ) sowie der Anteil der Personen-Dubletten unter allen Meldefällen ( $\text{Anteil}_{\text{PM}}$ ) berechnet. Die Anteile berechnen sich wie folgt:

$$\text{Anteil}_{\text{PD}} = \frac{\text{Anzahl Personen-Dubletten}}{\text{Anzahl Daten-Dubletten}} \quad (\text{Formel 1})$$

und

$$\text{Anteil}_{\text{PM}} = \frac{\text{Anzahl Personen-Dubletten}}{\text{Anzahl Meldefälle}} \quad (\text{Formel 2})$$

\* Es kann beispielweise vorkommen, dass sich die Schreibweise des Namens auf der Meldung unterscheidet oder die Zahlen des Geburtsdatums verdreht sind, so dass ein automatisierter Abgleich nach Name und Geburtsdatum nicht erfolgreich ist.

\*\* Aufgrund der großen Anzahl von Zufalls-Dubletten in Rheinland-Pfalz bzw. Hamburg wurden in der aktuellen Arbeit nur Daten-Dubletten **innerhalb** der einzelnen Gesundheitsämter gesucht. Bei den 2.590 gemeldeten HBV-Fällen ab 2017 in ganz Rheinland-Pfalz wären rein statistisch 700 Zufalls-Zweifach-Dubletten, 378 Zufalls-Dreifach-Dubletten, 186 Zufalls-Vierfach-Dubletten, 83 Zufalls-Fünffach-Dubletten, 34 Zufalls-Sechsfach-Dubletten usw. von Fällen mit gleichen Eigenschaften **über alle** Gesundheitsämter zu erwarten. Auch bei den 1.601 gemeldeten HCV-Fällen ab 2017 wäre die Anzahl der Zufalls-Dubletten zu groß für eine Recherche zwischen den Gesundheitsämtern.

## INFOBOX 2

**Begriffsbestimmungen**

**Daten-Dubletten:** In einer Datenbank bezeichnet der Begriff Daten-Dublette einen mehrfach vorhandenen Fall. In dieser Auswertung wird der Begriff Daten-Dublette verwendet, wenn bei der Abfrage für den Zuständigkeitsbereich eines Gesundheitsamtes Geburtsjahr, Geburtsmonat und Geschlecht mehrerer Fälle übereinstimmen. Ob es sich dabei um eine *Personen-Dublette* oder eine *Zufalls-Dublette* handelt, wurde durch das zuständige Gesundheitsamt auf Basis identifizierender Informationen ermittelt. Im Folgenden wird der Begriff Dublette sowohl verwendet, wenn der gleiche Fall genau zweimal vorkommt (Zweifach-Dublette) oder auch, wenn ein Fall mehrfach vorkommt (z. B. als Dreifach- oder Vierfach-Dublette).

**Personen-Dubletten:** Hat ein Gesundheitsamt eine Person tatsächlich mehrfach in der Datenbank angelegt, obwohl es sich um mehrere Meldungen zur gleichen Person handelt, entspricht dies einer Personen-Dublette. Personen-Dubletten sind eine Teilmenge der Daten-Dubletten und können entstehen, wenn identifizierende Informationen bei der Meldung nicht zugänglich, unvollständig oder fehlerhaft sind.

**Zufalls-Dubletten:** Bei Zufalls-Dubletten handelt es sich um diejenige Teilmenge der Daten-Dubletten, die nach Prüfung durch das Gesundheitsamt keine Personen-Dubletten sind. Dabei handelt es sich also um zwei oder mehr Personen unterschiedlicher Identität, die nur zufällig das gleiche Geburtsjahr, den gleichen Geburtsmonat und das gleiche Geschlecht haben.

**Erwartete Zufalls-Dubletten:** Es gibt nur eine begrenzte Anzahl von Kombinationsmöglichkeiten der drei Variablen Geburtsjahr (ca. 90 Ausprägungen), Geburtsmonat (12 Ausprägungen) und Geschlecht (zwei Ausprägungen\*). Daraus ergeben sich  $90 \times 12 \times 2 = 2.160$  verschiedene Kombinationen mit unterschiedlicher Wahrscheinlichkeit des Auftretens abhängig von der Alters- und Geschlechterverteilung unter den gemeldeten Fällen. Eine Simulation auf Basis der Verteilung in den Meldedaten zeigt, dass bei 100 gemeldeten HBV-Fällen statistisch 3,7 Zufalls-Zweifach-Dubletten und 0,1 Zufalls-Dreifach-Dubletten erwartet werden. Bei 300 gemeldeten HBV-Fällen werden statistisch 29,6 Zufalls-Zweifach-Dubletten, 2,7 Zufalls-Dreifach-Dubletten und 0,2 Zufalls-Vierfach-Dubletten erwartet. Aufgrund der Alters- und Geschlechterverteilung liegt die erwartete Anzahl bei HCV-Fällen leicht höher.

**Erwartete Personen-Dubletten:** Aus der Differenz der Daten-Dubletten und der erwarteten Zufalls-Dubletten ergibt sich die Anzahl der erwarteten Personen-Dubletten. Nähere Erläuterungen zur Berechnung und den zugrundeliegenden Annahmen finden sich unter Statistische Auswertung (Teil 2).

\* Auf die Kategorien divers und fehlend wurde aufgrund der sehr kleinen Fallzahlen in dieser Berechnung verzichtet, da dies die Berechnung verfälscht hätte.

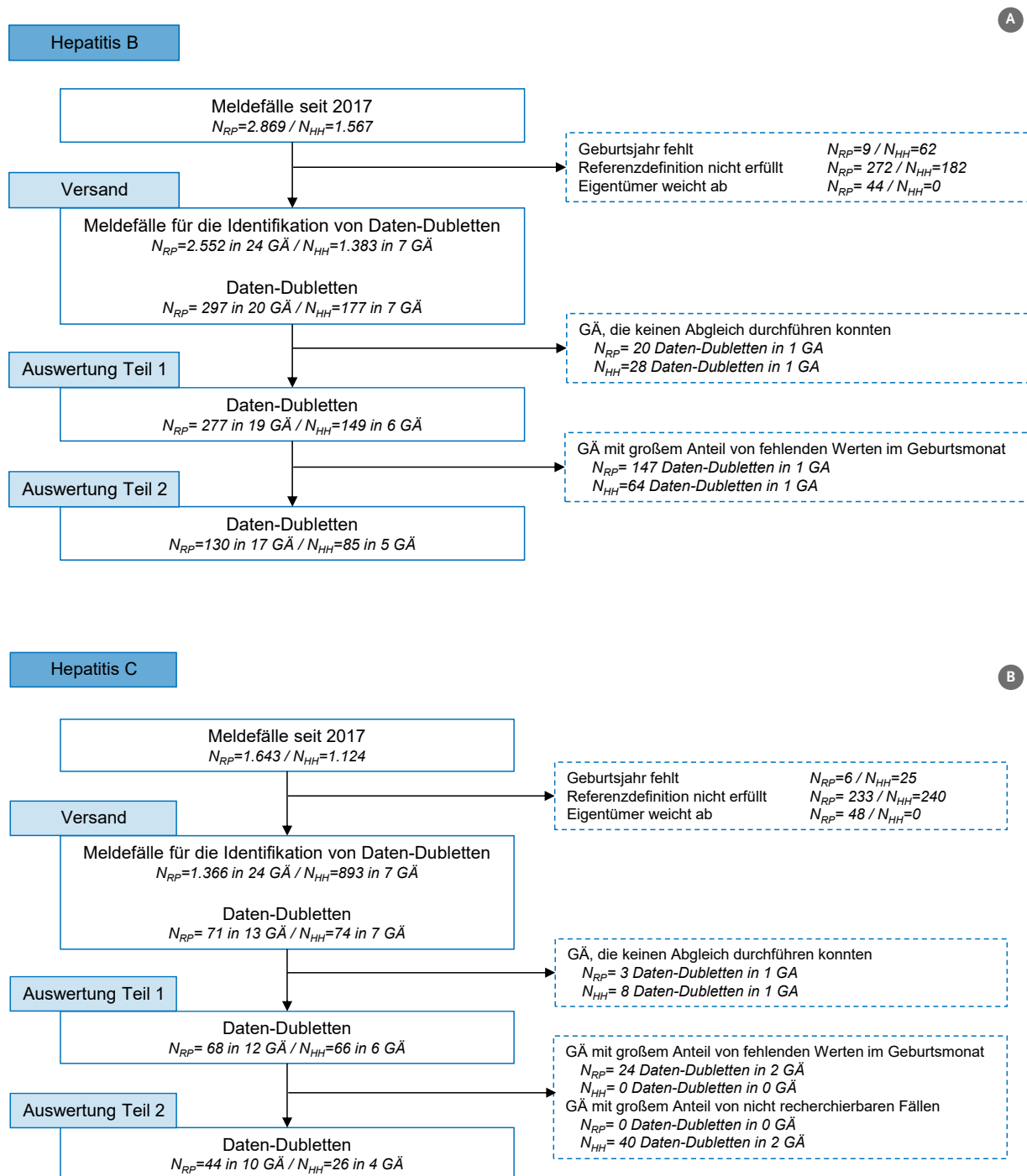
Die Berechnung der Anzahl von Daten-Dubletten wurde unter Berücksichtigung von Mehrfachdubletten wie Dreifach- oder Vierfach-Dubletten durchgeführt. Es gilt, dass bei jeder Daten-Dublette mindestens ein Datensatz originär sein muss. Somit kann es sich bei einer Zweifach-Dublette bei maximal einem Datensatz um eine Personen-Dublette handeln. Analog gilt für Dreifach-Dubletten, dass hier maximal zwei Datensätze Personen-Dubletten sein können, für Vierfach-Dubletten maximal drei Datensätze usw. Die Gesamtzahl der Daten-Dubletten berechnete sich daher wie folgt:

$$\text{Anzahl Daten-Dubletten} = \sum_{k=2}^n (k-1)x_k, \quad (\text{Formel 3})$$

wobei  $n$  die maximale Größe der Dubletten und  $x_k$  die jeweilige Anzahl der Daten-Dubletten der Dublettengröße  $k$  bezeichnet.

**Statistische Auswertung (Teil 2)**

Mittels einer Simulation wurde die Anzahl der erwarteten Zufalls-Dubletten bezogen auf die Anzahl der Hepatitis-B-Meldefälle bestimmt. Hierfür wurde ein Datensatz aus 1.000 Sets zu je 1.000 Fällen



**Abb. 1** | Flussdiagramm für den Einschluss von Gesundheitsämtern und Meldefällen für Hepatitis B (Feld A) und Hepatitis C (Feld B); (GÄ = Gesundheitsämter, GA = Gesundheitsamt, RP = Rheinland-Pfalz, HH = Hamburg); „Eigentümer weicht ab“ bedeutet, dass das Feld für den Eigentümer in der Datenbank nicht der für die Gesundheitsämter der jeweiligen Bundesländer festgelegten Kodierung entsprach.

generiert. Die Merkmale Geburtsjahr, Geburtsmonat und Geschlecht wurden den so entstandenen 1.000.000 Fällen entsprechend der Verteilung unter den rheinland-pfälzischen Meldefällen zufällig zugewiesen (s. INFOBOX 3).

Aus diesem Datensatz wurden zur Bestimmung der erwarteten Zufalls-Dubletten bei  $n$  Meldefällen die ersten  $n$  Fälle der jeweils 1.000 Sets gezogen und auf Zufalls-Dubletten untersucht. Auch bei der Berechnung der Anzahl der Zufalls-Dubletten wurden

## INFOBOX 3

**Simulation**

Für die Simulation wurden den Fällen im ersten Schritt **Geburtsjahre** aus der Häufigkeitsverteilung des Geburtsjahres in den Meldedaten in Rheinland-Pfalz zugeordnet. Für die darauffolgende Zuordnung von Geschlecht und Geburtsmonat wurden die simulierten Fälle in je 10 Geburtsjahrgängen zusammengefasst.

Da das Geschlecht in den Meldedaten nicht gleichmäßig über die Jahrgänge verteilt war, wurde den so generierten Fällen in einem zweiten Schritt auf Basis dieser Jahrgangsguppen ein **Geschlecht** zugewiesen (männlich oder weiblich). Auf die Zuordnung der Kategorien divers und fehlend wurde aufgrund der sehr kleinen Fallzahlen verzichtet.

Bei der Zuordnung nach **Geburtsmonat** musste die Besonderheit berücksichtigt werden, dass die Ausprägung „1“ in den Meldedaten sowohl für Januar als auch für einen fehlenden Wert stehen kann. Eine gesonderte Ausprägung für fehlende Werte existiert in der Meldedatenbank somit nicht. Da sich die Anzahl der fehlenden Werte zwischen den Jahrgängen unterschied, wurde der Geburtsmonat 1 den Jahrgangsguppen gemäß der Verteilung in den Meldedaten zugewiesen. Den übrigen Fällen wurden die Geburtsmonate 2 bis 12 gemäß der Anzahl ihrer Tage im Kalenderjahr zugewiesen.

Die Simulation wurde getrennt für Hepatitis B und Hepatitis C durchgeführt, da sich die Verteilung über Geburtsjahr, Geburtsmonat und Geschlecht unterschied. Für die Gesundheitsämter in Hamburg wurde die Simulation auf Basis der rheinland-pfälzischen Daten zugrunde gelegt, da die Simulation von der Landesstelle Rheinland-Pfalz durchgeführt wurde.

Mehrfachdubletten berücksichtigt. Die Anzahl der erwarteten Zufalls-Dubletten ergab sich aus dem Mittelwert über die 1.000 Sets. Das 95 % Konfidenzintervall entspricht der 5 %- und der 95 %-Perzentile der 1.000 Sets.

Die Anzahl der erwarteten Personen-Dubletten konnte dann aus der Differenz zwischen den Daten-

Dubletten und den erwarteten Zufalls-Dubletten vorhergesagt werden. Die zugrundeliegende Annahme ist, dass die Anzahl der Daten-Dubletten im Mittel identisch mit der Anzahl der erwarteten Zufalls-Dubletten sein sollte, falls sich unter den Daten-Dubletten keine Personen-Dubletten befinden:

$$\text{erwartete Personen-Dubletten} = \text{Anzahl Daten-Dubletten} - \text{erwartete Zufalls-Dubletten} \quad (\text{Formel 4})$$

In Gesundheitsämtern mit einer großen Anzahl von Personen-Dubletten ist davon auszugehen, dass die Anzahl der Zufalls-Dubletten überschätzt wurde, da die Simulation auf der Anzahl der Meldefälle basiert, welche wiederum durch das Vorhandensein der Personen-Dubletten künstlich erhöht ist. Um dies zu korrigieren, wurde die Anzahl der erwarteten Zufalls-Dubletten iterativ auf Basis der Meldefälle abzüglich der erwarteten Personen-Dubletten erneut zugewiesen.

Auf Grundlage der Rückmeldungen der Gesundheitsämter zu den Personen-Dubletten wurde die Schätzung von erwarteten Personen-Dubletten validiert. Dabei wurden zwei Gesundheitsämter von der Analyse ausgeschlossen, bei denen der Anteil der Meldefälle mit Geburtsmonat Januar signifikant über dem erwarteten Anteil lag, was auf einen hohen Anteil an fehlenden Werten hindeutet (vgl. [Abb. 1](#)).

**Ergebnisse**

Insgesamt wurden an 22 der 24 rheinland-pfälzischen Gesundheitsämter und an alle sieben Hamburger Gesundheitsämter Listen mit Daten-Dubletten verschickt. Jeweils ein Gesundheitsamt in Rheinland-Pfalz und Hamburg konnte die Daten-Dubletten nicht abgleichen, in einigen weiteren Hamburger Gesundheitsämtern war nur ein partieller Abgleich möglich, da dort selbst keine identifizierenden Informationen zu den Fällen mehr vorlagen.\*

\* Ein rheinland-pfälzisches Gesundheitsamt war Opfer eines Cyberangriffs geworden. In Hamburg waren die Daten gemäß des bestehenden Löschkonzepts zum Datenschutz nicht mehr verfügbar.

### Identifikation von Personen-Dubletten (Teil 1)

Aus den 3.935 Meldefällen von **Hepatitis B** wurden an den Landesstellen 474 Daten-Dubletten identifiziert, von denen 426 abgeglichen werden konnten: 277 in Rheinland-Pfalz und 149 in Hamburg (s. Tab.1). In Rheinland-Pfalz wurden darunter 52 Personen-Dubletten identifiziert, was einem Anteil von

etwa 20 % unter den Daten-Dubletten entspricht und einem Anteil von etwa 2 % am gesamten Meldeaufkommen (Spannweite 0 %–10 %). In Hamburg waren diese Anteile etwa doppelt so hoch: Es handelte sich bei 34 % der 149 Daten-Dubletten um Personen-Dubletten, was einem Anteil von etwa 4,5 % am gesamten Meldeaufkommen entsprach

Gesundheitsamt (pseudonymisiert)	Anzahl Meldungen seit 1.1.2017	Daten-Dubletten	Personen-Dubletten		Anteil Personen-Dubletten	
				fehlende Werte <sup>A</sup>	unter allen Daten-Dubletten	unter allen Meldefällen ab 2017
<b>Rheinland-Pfalz</b>						
A	14	1	0	0	0 %	0,0 %
B	22	1	0	0	0 %	0,0 %
C	30	0	–	–	–	0,0 %
D	29	0	–	0	–	0,0 %
E	40	3	2	0	67 %	5,0 %
F	41	3	3	0	100 %	9,8 %
G	63	4	2	0	50 %	3,2 %
H	29	1	1	0	100 %	3,4 %
I	69	3	2	0	67 %	2,9 %
J	62	2	0	0	0 %	0,0 %
K	40	0	–	–	–	0,0 %
L	27	1	0	0	0 %	0,0 %
M	91	5	2	0	40 %	3,3 %
N	110	5	1	0	20 %	1,8 %
O	21	0	–	–	–	0,0 %
P	132	9	3	0	33 %	3,0 %
Q	104	5	1	0	20 %	1,0 %
R	45	0	–	–	–	0,0 %
S	82	3	1	0	33 %	1,2 %
T	256	27	4	1	15 % (19%) <sup>B</sup>	1,6 % (2,0%) <sup>B</sup>
U	484	147	20	5	14 % (17%) <sup>B</sup>	4,1 % (5,2%) <sup>B</sup>
V	246	20	5	0	25 %	2,0 %
W	237	(20) <sup>C</sup>	–	–	–	–
X	278	37	5	0	14 %	2,2 %
<b>Gesamt RP</b>	<b>2.552</b>	<b>277</b>	<b>52</b>	<b>6</b>	<b>19 % (21%)<sup>B</sup></b>	<b>2,1 % (2,2%)<sup>B</sup></b>
<b>Hamburg</b>						
A	139	5	3	0	60 %	2,2 %
B	214	(28) <sup>C</sup>	–	–	–	–
C	55	1	0	0	0 %	0,0 %
D	304	57	41	0	72 %	13,5 %
E	153	15	3	0	20 %	2,0 %
F	131	7	1	1	14 % (29%) <sup>B</sup>	0,8 % (1,5%) <sup>B</sup>
G	387	64	2	4	3 % (9%) <sup>B</sup>	0,5 % (1,6%) <sup>B</sup>
<b>Gesamt HH</b>	<b>1.383</b>	<b>149</b>	<b>50</b>	<b>5</b>	<b>34 % (37%)<sup>B</sup></b>	<b>4,3 % (4,7%)<sup>B</sup></b>

**Tab. 1 |** Dubletten unter Meldefällen von Hepatitis B nach Gesundheitsamt

A Anzahl der Daten-Dubletten, bei denen unklar ist, ob es sich um Personen-Dubletten handelt.

B Maximalwert unter Annahme, dass alle unbekanntenen Daten-Dubletten Personen-Dubletten sind.

C Zahlen in Klammern wurden in der Berechnung der Gesamtzahl für die Bundesländer nicht berücksichtigt.

(Spannweite 0%–13%). Es zeigten sich relativ große Unterschiede, wobei der Anteil der Personen-Dubletten in den Gesundheitsämtern *RP-F* und *HH-D* (pseudonymisiert) besonders hoch waren. Aufgrund der relativ kleinen Zahlen pro Gesundheitsamt sollten diese Unterschiede allerdings nicht überinterpretiert werden.

Bei **Hepatitis C** waren sowohl die Anzahl der Meldefälle im gleichen Zeitraum als auch die Anteile der Personen-Dubletten unter den Daten-Dubletten und unter den Gesamtfällen geringer (s. Tab. 2). In Rheinland-Pfalz wurden insgesamt 11 Personen-Dubletten identifiziert, was einem Anteil von weniger als einem Prozent unter den Meldefällen ent-

Gesundheitsamt (pseudonymisiert)	Anzahl Meldungen seit 1.1.2017	Daten-Dubletten	Personen-Dubletten		Anteil Personen-Dubletten	
				fehlende Werte <sup>A</sup>	unter allen Daten- Dubletten	unter allen Meldefällen ab 2017
<b>Rheinland-Pfalz</b>						
A	21	0	0	0	–	0,0%
B	15	0	0	0	–	0,0%
C	13	0	0	0	–	0,0%
D	22	2	0	0	0%	0,0%
E	35	1	0	0	0%	0,0%
F	26	0	0	0	–	0,0%
G	34	2	1	0	50%	2,9%
H	7	0	0	0	–	0,0%
I	37	0	0	0	–	0,0%
J	20	0	0	0	–	0,0%
K	33	2	1	0	50%	3,0%
L	27	0	0	0	–	0,0%
M	127	5	3	0	60%	2,4%
N	82	5	1	0	20%	1,2%
O	21	0	0	0	–	0,0%
P	100	8	3	0	38%	3,0%
Q	80	0	0	0	–	0,0%
R	37	2	0	0	0%	0,0%
S	71	3	0	0	0%	0,0%
T	143	10	1	0	10%	0,7%
U	32	0	0	0	–	0,0%
V	191	21	0	0	0%	0,0%
W	77	(3) <sup>C</sup>	–	–	–	–
X	115	7	1	0	14%	0,9%
<b>Gesamt RP</b>	<b>1.366</b>	<b>68</b>	<b>11</b>	<b>3</b>	<b>16%</b>	<b>0,8%</b>
<b>Hamburg</b>						
A	130	11	1	6	9% (64%) <sup>B</sup>	0,8% (5,4%) <sup>B</sup>
B	87	(8) <sup>C</sup>	–	–	–	–
C	18	0	0	0	0%	0,0%
D	170	15	11	0	73%	6,5%
E	147	10	1	2	10% (30%) <sup>B</sup>	0,7% (2,0%) <sup>B</sup>
F	96	1	0	0	0%	0,00%
G	245	29	0	9	0% (31%) <sup>B</sup>	0,0% (3,7%) <sup>B</sup>
<b>Gesamt HH</b>	<b>893</b>	<b>66</b>	<b>13</b>	<b>17</b>	<b>20% (45%)<sup>B</sup></b>	<b>1,5% (3,4%)<sup>B</sup></b>

**Tab. 2 |** Dubletten unter Meldefällen von Hepatitis C nach Gesundheitsamt

A Anzahl der Daten-Dubletten, bei denen unklar ist, ob es sich um Personen-Dubletten handelt.

B Maximalwert unter Annahme, dass alle unbekanntenen Daten-Dubletten Personen-Dubletten sind.

C Zahlen in Klammern wurden in der Berechnung der Gesamtzahl für die Bundesländer nicht berücksichtigt.

sprach (Spannweite 0 %–3 %). In Hamburg wurden 13 Personen-Dubletten gefunden. 17 Daten-Dubletten konnten nicht abgeglichen werden, woraus sich ein Anteil zwischen 1 % und 4 % der Gesamtfälle ergab, die mehrfach angelegt worden waren.

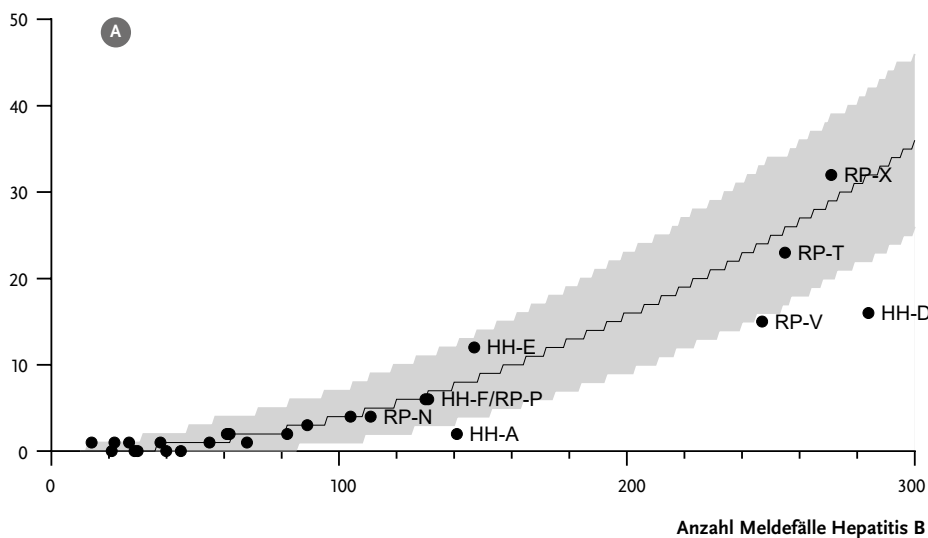
### Vorhersage von Personen-Dubletten (Teil 2)

Die Anzahl erwarteter Zufalls-Dubletten steigt exponentiell mit der Anzahl der Meldedefälle (s. Abb. 2). Bei

der Mehrzahl der Gesundheitsämter lag die Anzahl der gefundenen Zufalls-Dubletten im 95% Konfidenzintervall der erwarteten Zufalls-Dubletten.

Bei jeweils drei Gesundheitsämtern wurde die Anzahl der Personen-Dubletten auf Basis der erwarteten Zufalls-Dubletten signifikant zu niedrig eingeschätzt. So wurden beispielsweise im rheinland-pfälzischen Gesundheitsamt *RP-V* (pseudonymi-

#### Zufalls-Dubletten



#### Zufalls-Dubletten

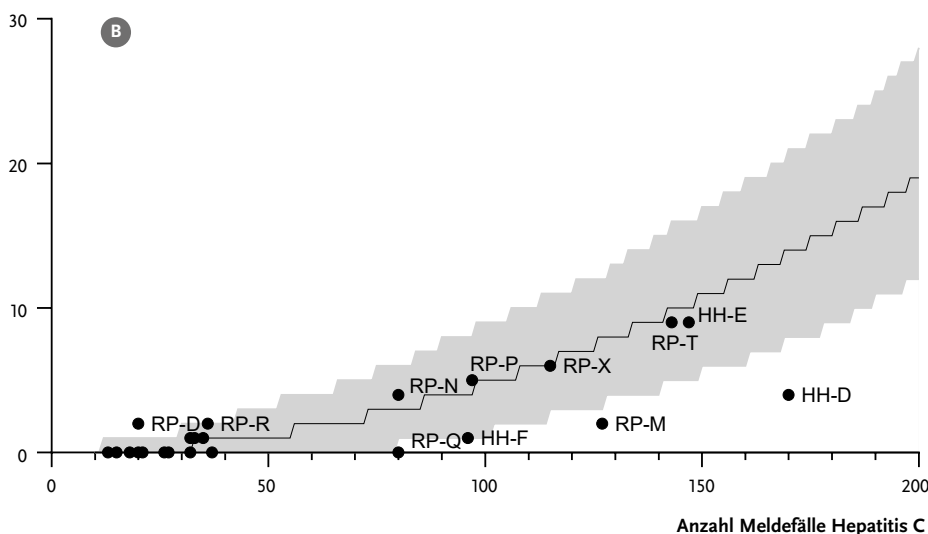


Abb. 2 | Erwartete Zufalls-Dubletten und gefundene Zufalls-Dubletten im Verhältnis zur Anzahl der Meldedefälle für Hepatitis B (Feld A) und Hepatitis C (Feld B)

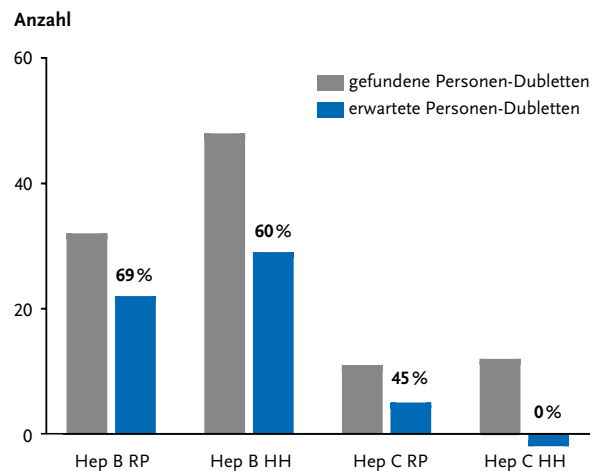


siert) auf Basis der 247 Hepatitis-B-Meldefälle 24 Zufalls-Dubletten erwartet. Die Suche identifizierte allerdings nur 20 Daten-Dubletten, also mindestens vier Zufalls-Dubletten weniger als statistisch erwartet. Unter diesen 20 Daten-Dubletten befanden sich nach Abgleich durch das Gesundheitsamt außerdem fünf Personen-Dubletten, so dass es am Gesundheitsamt 15 Zufalls-Dubletten gab und somit neun weniger als statistisch erwartet. Auch für die Gesundheitsämter (pseudonymisiert) *HH-A*, *HH-D*, *RP-M* und *RP-Q* wurde die Anzahl der Personen-Dubletten unter den Hepatitis-B- und/oder den Hepatitis-C-Fällen auf Basis der Simulation signifikant unterschätzt. Für das Gesundheitsamt mit den meisten Hepatitis-B-Meldefällen (*HH-D*) wären statistisch 32 Zufalls-Dubletten zu erwarten gewesen. Es handelte sich allerdings sogar bei 41 der 57 Daten-Dubletten um Personen-Dubletten und somit nur bei 16 Daten-Dubletten um Zufalls-Dubletten. Im Gegensatz dazu wurde die Anzahl der Personen-Dubletten für das Gesundheitsamt *RP-D* überschätzt, wobei es sich allerdings aufgrund der kleinen Anzahl von 22 Hepatitis-C-Fällen wahrscheinlich um eine zufällige Abweichung handelt.

Im Mittel unterschätzte die Vorhersage also die Personen-Dubletten, so dass es auf Basis der Daten aus der Simulation möglich war, einen Teil der Personen-Dubletten vorherzusagen (s. Abb. 3). Die Unterschätzung der Personen-Dubletten für die Hepatitis-B-Fälle in Rheinland-Pfalz wurde wie oben beschrieben durch das Gesundheitsamt *RP-V* verursacht. Auch in Hamburg hatten die Ausreißer einen starken Einfluss auf die Abweichung der vorhergesagten von den gefundenen Werten.

## Diskussion

Die Auswertung aus der Recherche der Gesundheitsämter ergab, dass es sich nur bei einem Anteil von weniger als 5 % der Meldefälle um Personen-Dubletten innerhalb der Gesundheitsämter handelte. Daher sind der Anstieg von Hepatitis-B- und Hepatitis-C-Fällen in den letzten Jahren nicht oder zu einem nicht bedeutenden Anteil auf Mehrfach-erfassungen innerhalb der Gesundheitsämter zurückzuführen. Während es durchaus Unterschiede zwischen den Bundesländern und zwischen den einzelnen Gesundheitsämtern gab, zeigt diese Aus-



**Abb. 3** | Anzahl gefundener Personen-Dubletten (grau) und erwarteter Personen-Dubletten (blau) und Anteil erwarteter Personen-Dubletten an gefundenen Personen-Dubletten (Prozentwerte) für Hepatitis B (Hep B) und Hepatitis C (Hep C) in den Bundesländern Rheinland-Pfalz (RP) und Hamburg (HH)

wertung, dass die Gesundheitsämter bei der Verarbeitung der Meldungen von Hepatitis-B- und Hepatitis-C-Fällen sehr gute Arbeit leisten.

Unterschiedliche Rahmenbedingungen begünstigen die Mehrfacherfassung von Hepatitis-Meldefällen in der Surveillance. Seit der Änderung der Falldefinition für Hepatitis-B-Fälle im Jahr 2019 erfüllen nicht nur akute Fälle und Fälle mit unbekanntem Infektionsstadium die Falldefinition, sondern auch chronische Fälle, so dass diese auch in die veröffentlichten Statistiken eingehen. Zusammen mit der Änderung der Meldepflicht im Jahr 2017 kann dies potenziell dazu geführt haben, dass alte, bereits bekannte chronische Hepatitis-B-Fälle bei erneuter Testung nun gemeldet und somit in die Fallzahlen mit eingerechnet wurden. Eine weitere Schwierigkeit ergibt sich aus der Vorschrift, alle personenbezogenen Daten von Hepatitis-C-Meldefällen nach Ablauf von drei Jahren zu löschen. Diese Vorschrift gilt bundesweit seit 2017 nicht mehr, allerdings wurde das Vorgehen in einigen Regionen noch nicht aktualisiert, so dass bei Eingang einer neuen Hepatitis-C-Meldung nicht überprüft werden kann, ob diese Meldung zu einem Fall gehört, der dem Gesundheitsamt mehr als drei Jahre zuvor schon gemeldet und dort in der Datenbank angelegt worden war. Ein weiteres Hindernis für den Abgleich stellen

potenzielle Umzüge infizierter Personen zwischen Stadtbezirken, Landkreisen oder Verwaltungsbezirken dar. Zieht eine chronisch mit Hepatitis B oder Hepatitis C infizierte Person um, werden Meldungen ab diesem Zeitpunkt an das zuständige Gesundheitsamt des neuen Wohnortes geschickt. Liegen der Meldung keine Informationen über eine frühere Diagnose bzw. Meldung bei, legt das Gesundheitsamt den Fall neu an, so dass der Fall somit in zwei unterschiedlichen Gesundheitsämtern angelegt ist und gezählt wird. Ein Abgleich von Meldungen aus verschiedenen Gesundheitsämtern auf Länder- oder Bundesebene ist aufgrund fehlender identifizierender Daten nicht möglich.

Die hier verwendete Simulation konnte aus der Differenz von statistisch erwarteten Zufalls-Dubletten und Daten-Dubletten etwa zwei Drittel der Personen-Dubletten von Hepatitis B vorhersagen (64%). Das bedeutet, dass es möglich ist, die Meldedaten für die Veröffentlichung von Statistiken über den zeitlichen Verlauf der Hepatitis-B-Inzidenzen auf Basis dieser Simulation zu korrigieren ohne die Gesundheitsämter mit dem Abgleich der einzelnen Meldedaten belasten zu müssen. Obwohl Mehrfach-erfassungen von Meldedaten in dem von uns untersuchten Zeitraum aufgrund des sehr kleinen Anteils an Personen-Dubletten am gesamten Meldeaufkommen keine entscheidende Rolle gespielt haben, könnte sich das mit Einführung der Meldesoftware des deutschen elektronischen Melde- und Informationssystems für den Infektionsschutz (DEMIS) verändert haben, da Erregernachweise aus den Laboren seitdem automatisch an die Gesundheitsämter gemeldet werden. Der extrem starke Anstieg von Hepatitis-B- und Hepatitis-C-Meldedaten seit 2022 lässt einen solchen Zusammenhang zumindest vermuten. Die Vorhersage von Personen-Dubletten bei Hepatitis-C-Meldedaten war durch die kleinen Fallzahlen und die fehlende Möglichkeit des Abgleichs durch gelöschte personenbezogene Daten schwieriger.

Die Überprüfung von Mehrfach-erfassungen konnte nur innerhalb der einzelnen Gesundheitsämter durchgeführt werden, da eine Überprüfung aller Daten-Dubletten zwischen allen Gesundheitsämtern innerhalb eines Bundeslandes eine zu hohe Arbeitsbelastung für die Gesundheitsämter darstellt

hätte. Zudem wären datenschutzrechtliche Probleme beim Teilen einzelner identifizierender Merkmale zum Abgleich zwischen den Gesundheitsämtern aufgetreten. Es ist anzunehmen, dass es durch Umzüge innerhalb von Deutschland vorkommt, dass verschiedene Gesundheitsämter den gleichen Fall erfassen. Das ist besonders wahrscheinlich in den Großstädten mit mehreren Verwaltungsbezirken wie Hamburg und Berlin. Auf Grundlage dieser Studie ist es nicht möglich, den Anteil der Mehrfach-erfassungen zwischen den Gesundheitsämtern zu quantifizieren. Durch eine Erweiterung des hier verwendeten Vorhersagealgorithmus mit Korrektur für multiple Vergleiche könnte eine Quantifizierung allerdings angestrebt werden.

Die Vorhersage verwendete simulierte Zufalls-Dubletten auf Grundlage der Verteilung der Merkmale Geburtsjahr, Geburtsmonat und Geschlecht unter den Meldedaten von Hepatitis B und Hepatitis C in der rheinland-pfälzischen Bevölkerung. Sollten sich diese Eigenschaften in der Hamburger Bevölkerung stark unterscheiden, könnte das zu Ungenauigkeiten in der Vorhersage geführt haben. Die Anzahl der erwarteten Zufalls-Dubletten hat sich allerdings in verschiedenen Simulationen relativ robust gezeigt, so dass die Verwendung der rheinland-pfälzischen Daten für die Gesundheitsämter in Hamburg keine größere Unter- oder Überschätzung verursacht haben sollte.

Die größte Herausforderung für die Vorhersage von Mehrfach-erfassungen waren fehlende Werte im Geburtsmonat. Bei Daten-Dubletten mit fehlendem Geburtsmonat ist die Wahrscheinlichkeit, dass es sich um Personen-Dubletten handelt geringer, da der Geburtsmonat mit 92% Wahrscheinlichkeit in Wirklichkeit nicht identisch ist. Das Problem für diese Auswertung bestand darin, dass fehlende Werte in der Datenbank nicht als fehlend gekennzeichnet werden können, da nur Werte zwischen eins und 12 in die Variable eingetragen werden können. Somit bestand die einzige Möglichkeit, fehlende Werte im Geburtsmonat zu identifizieren darin, zu untersuchen, ob der Geburtsmonat Januar überproportional häufig vorkam. In der Auswertung für Hepatitis B und Hepatitis C mussten jeweils zwei Gesundheitsämter von der Analyse ausgeschlossen

werden, da der Anteil der Fälle mit Geburtsmonat Januar signifikant über den erwarteten 8 % lag.

In Zukunft könnten zwei Veränderungen in der Datenbank dafür sorgen, dass eine Überprüfung auf Mehrfacherfassungen von Landes- und Bundeseite erleichtert wird. Zum einen sollte eine Kategorie für fehlende Werte für die Variable Geburtsmonat eingeführt werden, so dass Fälle mit Geburtsmonat Januar von Fällen mit fehlendem Geburtsmonat unterschieden werden können. Zum anderen könnte mittels einer *Hash*-Verschlüsselung auf Basis von Name und Geburtstag eine einzigartige Identifikation (ID) erstellt werden, die mit den nicht personenbezogenen Daten an die Landesstellen und das RKI übermittelt werden könnte. Auf diesen Ebenen könnten

mehrfach erfasste Fälle identifiziert und zwischen den Gesundheitsämtern abgeglichen werden.

Die Frage, ob Mehrfacherfassungen zum Anstieg von Hepatitis-B- und Hepatitis-C-Meldefällen zwischen 2017 und 2022 beigetragen haben, kann nur bedingt beantwortet werden. Unsere Auswertung zeigt, dass Mehrfacherfassungen innerhalb der einzelnen Gesundheitsämter keine bedeutende Rolle gespielt haben. Allerdings sollten Mehrfacherfassungen zwischen den Gesundheitsämtern und potenzielle Mehrfacherfassungen seit der Einführung von DEMIS untersucht werden, um aussagekräftige Auswertungen der Inzidenzen von Hepatitis B und Hepatitis C über die Zeit gewährleisten zu können.

## Literatur

- 1 World Health Organization, Hepatitis B. <https://www.who.int/news-room/fact-sheets/detail/hepatitis-b> (abgerufen am 27. Juni 2023)
- 2 World Health Organization, Hepatitis C. <https://www.who.int/news-room/fact-sheets/detail/hepatitis-c> (abgerufen am 27. Juni 2023)
- 3 Zimmermann R: Hepatitis-C-Melddaten nach IfSG, 2016 – 2018: Auswirkungen der Änderungen von Falldefinition und Meldepflicht. *Epid Bull* 2019;30:275-285 | DOI 10.25646/6206.2
- 4 Dudareva A, Kremer K, Harder T, Zimmermann R: Virushepatitis B und D im Jahr 2018. *Epid Bull* 2019; 29:261-270 | DOI 10.25646/6203.1

## Weiterführende Informationen

### Hepatitis B

<https://www.rki.de/hepatitisb>

### Hepatitis C

<https://www.rki.de/hepatitisc>

## Autorinnen und Autoren

<sup>a,b,c,f)</sup> Anja Schoeps | <sup>b,c,d)</sup> Silja Bühler | <sup>e)</sup> Ruth Zimmermann | <sup>e)</sup> Sandra Dudareva | <sup>a,f)</sup> Philipp Zanger | <sup>e)</sup> Matthias an der Heiden

<sup>a)</sup> Landesuntersuchungsamt Rheinland-Pfalz

<sup>b)</sup> Robert Koch-Institut, Abt. 3 Infektionsepidemiologie, Postgraduiertenausbildung für angewandte Epidemiologie

<sup>c)</sup> European Programme for Intervention Epidemiology Training (EPIET), European Centre for Disease Prevention and Control (ECDC)

<sup>d)</sup> Infektionsepidemiologisches Landeszentrum, Institut für Hygiene und Umwelt, Freie und Hansestadt Hamburg

<sup>e)</sup> Robert Koch-Institut, Abt. 3 Infektionsepidemiologie

<sup>f)</sup> Heidelberger Institut für Global Health, Universitätsmedizin Heidelberg

**Korrespondenz:** [anja.schoeps@lua.rlp.de](mailto:anja.schoeps@lua.rlp.de)

---

### Vorgeschlagene Zitierweise

Schoeps A, Bühler S, Zimmermann R, Dudareva S, Zanger P, an der Heiden M: Mehrfacherfassungen von Hepatitis-B- und Hepatitis-C-Meldungen in den Gesundheitsämtern – Identifikation und statistische Schätzung

Epid Bull 2023;30:3-14 | DOI 10.25646/11657

---

### Interessenkonflikt

Alle Autorinnen und Autoren geben an, dass kein Interessenskonflikt besteht.

---

### Danksagung

Wir bedanken uns bei den rheinland-pfälzischen und Hamburger Gesundheitsämtern, die diese Auswertung durch die detaillierte Überprüfung der Daten-Dubletten ermöglicht haben: Gesundheitsamt Altenkirchen, Gesundheitsamt Altona, Gesundheitsamt Alzey, Gesundheitsamt Bad Ems, Gesundheitsamt Bad Kreuznach, Gesundheitsamt Bad Neuenahr, Gesundheitsamt Bergedorf, Gesundheitsamt Bitburg, Gesundheitsamt Cochem, Gesundheitsamt Daun, Gesundheitsamt Eimsbüttel, Gesundheitsamt Germersheim, Gesundheitsamt Hamburg-Mitte, Gesundheitsamt Hamburg-Nord, Gesundheitsamt Harburg, Gesundheitsamt Idar-Oberstein, Gesundheitsamt Kaiserslautern, Gesundheitsamt Kirchheimbolanden, Gesundheitsamt Koblenz, Gesundheitsamt Landau, Gesundheitsamt Mainz, Gesundheitsamt Montabaur, Gesundheitsamt Neustadt an der Weinstraße, Gesundheitsamt Neuwied, Gesundheitsamt Simmern, Gesundheitsamt Trier, Gesundheitsamt Wandsbek und Gesundheitsamt Wittlich.  
Herzlichen Dank für die gute Zusammenarbeit!