



**UNIVERSITÄTS  
KLINIKUM**  
Jena

**Methodenlehre  
Med. Informatik  
WS 2024/2025**

Institut für Medizinische Statistik,  
Informatik und Datenwissenschaften



# Lernzielkatalog Methodenlehre

## **Seminar 1 (Literaturrecherche) – Klinisch orientierte Literaturrecherche mit PubMed:**

**Klinisches Beispiel:** *Basistherapie bei Multipler Sklerose (NePs: POL-Fall „Erna Diestel“)*. **Methoden:** Literaturrecherche auf Grundlage einer realen klinischen Fragestellung, Nutzung des PICO-Modells, Formulierung von Suchstrategien und Begriffen (MeSH-Suche), Vergleich der Suchstrategien mit Google und Google-Scholar, Nutzung von Evidence-Based-Medicine Datenbanken (Cochrane) und ärztlichen Leitlinien (AWMF,G-I-N), Studien des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) sowie Empfehlungen des Krankheitsbezogenen Kompetenznetzwerkes Multiple Sklerose (KKNMS), Grundlagen und praktische Übungen zur Dokumentation von Leistungsabrechnungen im Fallpauschalensystem.

## **Seminar 2 (Literaturdatenbanken) – Literaturverwaltung und -zitation mit EndNote:**

**Klinisches Beispiel:** *Staphylococci-Infektion - computerised clinical decision support system (HELP-CDSS), Beispieltext einer konkreten Publikation*. **Methoden:** Erstellen eigener Literaturdatenbanken im EndNote, Abfragen in externen Datenbanken im Internet (PubMed- und Web-of-Science-Abfragen), Zitieren von Literaturstellen und Erstellen eines Literaturverzeichnisses (Cite-While-You-Write) in Publikationen, Berichten und der Doktorarbeit mit Hilfe einer EndNote-Datenbank, Einbinden des Literaturstyles (jnamed.ens) für die Promotion an der Medizinischen Fakultät der FSU Jena, EndNote-Style-Datenbank im Internet.

## **Seminar 3 (Datensicherheit) – Praktischer Datenschutz und Datensicherheit:**

**Klinisches Beispiel:** *Entlassungsbrief „Erna Diestel“ aus POL-Fall*. **Methoden:** Datenschutz-/Datensicherheitsrisiken erläutern; Sichere Passwörter erstellen und nutzen; Passwörter in einem Passwort-Manager verwalten; Daten auf Datenträgern containerbasiert verschlüsseln (Symmetrische Verschlüsselungsverfahren); Konzept der asymmetrischen Verschlüsselung erläutern; Nachrichten-/Datenaustausch - insbesondere E-Mails - mit Public-Private-Key-Verfahren absichern; Konzept der digitalen Fingerabdrücke erläutern (Hash-Funktionen); Authentizität von Daten durch digitale Signaturen absichern und überprüfen.

## **Seminar 4 (Tabellenkalkulation) – Einführung in Tabellenkalkulation mit Excel:**

**Klinisches Beispiel:** *Daten aus Studie „Myokardinfarkt“*. **Methoden:** Vermittlung vertiefter Kenntnisse in der Anwendung der Tabellenkalkulation, Grundstrukturen bei Dateneingabe, -formatierung und -import, Programmieren von Formeln und Zellbezügen sowie Nutzung von Tabellenfunktionen für die Datenauswertung, Erstellen von Diagrammen, Nutzung von Logik- und Datumsfunktionen sowie Erstellen von bedingten Formatierungen, Regeln zum Sortieren von Datensätzen.

## **Seminar 5 (R-Seminar) – Datenmanagement mit R:**

**Klinisches Beispiel:** *Öffentlich verfügbare Genexpressions- und (anonymisierte) Patientendaten zu wechselnden klinischen Indikationen*. **Methoden:** Variablenzuweisungen, Basisoperationen und Funktionsdeklarationen in der Programmiersprache R formulieren; Vektoren deklarieren und als Datenstrukturen nutzen; Bedingte Anweisungen und Wiederholschleifen erläutern und programmieren; R-Pakete einbinden; Datentabellen einlesen, verbinden, sortieren; Daten in Datentabellen selektieren und filtern; Integrierte Entwicklungsumgebung (IDE) für R nutzen (Grundfunktionen) und mit R-Notebooks arbeiten; Beispiele für die Datenvisualisierung mittels R beschreiben.

**Seminar 6 (Datenbanken) – Einführung in Datenbanken mit Access:**

**Klinisches Beispiel:** *Integration von Daten einer Beobachtungsstudie.* **Methoden:** Einführung in die Struktur relationaler Datenbanken, Strukturierung von Tabellen, Definition von Gültigkeitsregeln, Nutzung von Nachschlageprüfungen zum Aufbau von Beziehungen zwischen Tabellen, Import von Daten aus externen Quellen, Einrichtung von Zugriffsbeschränkungen, Entwurf von Auswahlabfragen, Nutzung von Abfragen zur Berechnung neuer Daten, Aggregieren von Daten mittels Abfragen, Eingabe von und Navigieren durch Daten mit Formularen.

**Laut Scheinvergabeordnung QB1 ist die Teilnahme an 4 der 6 angebotenen Methodenlehre-Seminare nötig!**



## 1 Klinisch orientierte Literaturrecherche

### *Motivation zur klinisch orientierten Literaturrecherche*

Literaturrecherchen zu klinisch relevanten Fragestellungen sind ein wichtiger Bestandteil der ärztlichen Entscheidungsfindung. Die Nutzung von Recherchesystemen, wie z. B. **PubMed** (Service der U.S. National Library of Medicine), **Google** bzw. **Google Scholar**, **Cochrane** (unabhängige Datenbank der Cochrane Collaboration, die 1993 geschaffen und nach dem Epidemiologen Archie Cochrane (1909 - 1988) benannt wurde - listet insbesondere systematische Reviews und Evidence Based Medicine) und **AWMF** (Webseite der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften, die z. B. Leitlinien für das ärztliche Handeln veröffentlicht) bzw. **G-I-N** (Guidelines International Network) oder aber auch die Nutzung von Studien des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (**IQWiG**) sollte deshalb beherrscht werden und Bestandteil des klinischen Alltags sein. Nach der Vermittlung genereller Suchstrukturen und -befehle sollen anhand eines Fallbeispiels aus der Neurologie Suchstrategien entwickelt, konkrete Aspekte der Suche bzw. Fehlerquellen vermittelt sowie unterschiedliche Suchmöglichkeiten aufgezeigt und verglichen werden.

### *Verwendung des PICO-Modells*

Allgemein lässt sich eine Literaturrecherche mittels des PICO-Modells vorbereiten:

**P**opulation oder Patient, Problem:

Welche Gruppe von Patienten bzw. welches Problem soll untersucht werden?

**I**ntervention = Technologien, diagnostisches/therapeutisches Verfahren:

Welche Intervention ist Gegenstand der gegenwärtigen Untersuchung?

**C**omparison oder Control = Vergleichsintervention:

Was ist die Hauptalternative, mit der die Intervention verglichen werden kann?

**O**utcome = Zielgröße:

Was soll erreicht werden?

### *Vorstellung des Fallbeispiels*

Das Fallbeispiel wird anhand eines Anamnese-Videos demonstriert, das in Zusammenarbeit mit der Klinik für Neurologie erarbeitet wurde.

**Fallbeispiel:** *Welches Medikament in Tablettenform ist für die Basistherapie von Multipler Sklerose geeignet?*

### 1.1 Das MeSH-Konzept in PubMed

PubMed ist ein Service der U.S. National Library of Medicine, der über 33 Millionen Zitierungen aus MEDLINE und anderen Life-Science-Journalen umfasst. MeSH (Medical Subject Headings) ist ein von der U.S. National Library of Medicine kontrolliertes Vokabular, das zur Indexierung von Artikeln in MEDLINE/PubMed genutzt wird. MeSH Terme bieten eine konsistente Art und Weise, Informationen zu sammeln, die verschiedene Terminologien für dasselbe Konzept nutzen. Die MeSH Datenbank kann verwendet werden, um medizinische Fachausdrücke zu suchen und eine Suchstrategie aufzubauen. Derzeit sind mehr als 23000 Deskriptoren in MeSH enthalten. Grundsätzlich ist zu beachten, dass PubMed (wie fast alle relevanten Suchsysteme) die englische Sprache verwendet.

**Aufgabe 1.1:** Anhand des Begriffes „multiple sclerosis“ sollen Grundfertigkeiten der Suche in der MeSH Datenbank erläutert werden. Die Verbindung von verschiedenen MeSH Termen soll mittels einer Kombination der Begriffe „multiple sclerosis“ und „prevention“ vermittelt werden.

**Lösungsschritte:**

**Schritt 1 - Einfache Suche mit MeSH:**

- ⇒ PubMed wird geöffnet, indem der Begriff unter Google gesucht und der entsprechende Link genutzt wird (<https://www.ncbi.nlm.nih.gov/pubmed>). Auf der Startseite wird rechts unten unter **Explore** „MeSH Database“ angeklickt. Anschließend wird in dem Suchfeld oben rechts der Begriff „multiple sclerosis“ eingegeben und durch **Search** bestätigt. Daraufhin erscheint eine Liste verschiedener MeSH Terme, die im Zusammenhang mit dem Suchbegriff „multiple sclerosis“ stehen.
- ⇒ Unter diesen Termen wird jeweils eine kurze Erklärung angeboten. Durch Anklicken der einzelnen Terme erscheint die vollständige Anzeige, die zusätzliche Informationen und Suchoptionen enthält, z. B. Subheadings oder die Position des jeweiligen Terms in der MeSH Hierarchie. Rechts neben den Termen kann unter **Related Information** die Suche weiter eingeschränkt werden.

The screenshot displays the MeSH database search results for the term "multiple sclerosis". The search bar at the top shows "MeSH" selected and "multiple sclerosis" entered. Below the search bar, there are several panels:

- Search results:** Shows 6 items. The first item, "Multiple Sclerosis", is selected and circled in red. A red arrow points from this item to the detailed view below.
- Multiple Sclerosis (Detailed View):** Contains a definition: "An autoimmune disorder mainly affecting young adults and characterized by destruction of myelin in the central nervous system. Pathologic findings include multiple sharply demarcated areas of demyelination throughout the white matter of the central nervous system. Clinical manifestations include visual loss, extra-ocular movement disorders, paresthesias, loss of sensation, weakness, dysarthria, spasticity, ataxia, and bladder dysfunction. The usual pattern is one of recurrent attacks followed by partial recovery (see MULTIPLE SCLEROSIS, RELAPSING-REMITTING), but acute fulminating and chronic progressive forms (see MULTIPLE SCLEROSIS, CHRONIC PROGRESSIVE) also occur. (Adams et al., Principles of Neurology, 6th ed, p903)". Below this is a "PubMed search builder options" section with "Subheadings:" circled in red. It lists various subheadings such as analysis, anatomy and histology, blood, cerebrospinal fluid, etc., each with a checkbox.
- Related information:** A panel on the right side, also circled in red, showing options like PubMed, PubMed - Major Topic, Clinical Queries, NLM MeSH Browser, dbGaP Links, and MedGen.
- Search details:** Shows the search query: "multiple sclerosis"[MeSH Terms].
- Recent Activity:** Shows a list of recent searches, including "Multiple Sclerosis" and "multiple sclerosis (6)".

- ⇒ In unserer Suche ist „Multiple Sclerosis“ der dem Begriff „multiple sclerosis“ entsprechende MeSH Term. Beim Anklicken des Terms „Multiple Sclerosis“ (Betrachtung und Einschränkung hinsichtlich Subheadings oder die Position des jeweiligen Terms in der MeSH Hierarchie) sind insbesondere die Optionen **Restrict to MeSH Major Topic** und **Do not include MeSH terms found below this term in the MeSH hierarchy** zu beachten.
- ⇒ Unter **Related information** kann nun entweder mittels **PubMed** auf die vollständige Datenbank aller zu diesem Thema vorhandenen Publikationen zugegriffen werden oder durch Auswahl von **PubMed Major Topic** nur die Betrachtung von Artikeln, in denen „multiple sclerosis“ das Hauptthema der Untersuchungen ist, genutzt werden. Unter **Clinical Queries** gibt es weiterhin die Möglichkeit, z. B. nach bestimmten Kategorien klinischer Studien zu suchen (**Search by Clinical Study Category**).
- ⇒ Wurde eine dieser Möglichkeiten genutzt (hier: PubMed Major Topic), erscheint als Resultat eine Liste aller vorhandenen Publikationen. Nun können rechts neben der Ergebnisliste noch Kategorien wie z. B. **Article types - Review** (Reviews sind Übersichtsartikel, die grundlegende Ergebnisse zu einer bestimmten Fragestellung zusammenfassen) oder **Text availability - Free Full Text** (der Artikel soll frei verfügbar sein – entweder als PDF-File oder URL) ausgewählt werden.

The screenshot shows the PubMed search interface. At the top, the search bar contains the query "Multiple Sclerosis"[MeSH Major Topic], which is circled in red. Below the search bar, there are buttons for "Advanced", "Create alert", and "Create RSS". The search results are displayed on page 1 of 6,465, with the total number of results (64,646) circled in red. On the left side, there are several filter sections: "MY NCBI FILTERS", "RESULTS BY YEAR" (with a bar chart showing an increase in results over time from 1945 to 2025), "TEXT AVAILABILITY" (with options for Abstract, Free full text, and Full text, where "Free full text" is circled in red), "ARTICLE ATTRIBUTE" (with an option for Associated data), and "ARTICLE TYPE" (with options for Books and Documents, Clinical Trial, Meta-Analysis, Randomized Controlled Trial, Review, and Systematic Review, where "Review" is circled in red). The main results area shows a list of articles, with the first three articles circled in red:

- Anti-CD20 Therapies in Drug-Naive Patients With Primary Progressive Multiple Sclerosis: A Multicenter Real-Life Study.**  
Cite: Hay M, Rollot F, Casey R, Kerbrat A, Edan G, Mathey G, Labauge P, De Sèze J, Vukusic S, Laplaud DA, Papeix C, Moreau T, Thouvenot E, Defer G, Lebrun-Frénay C, Ciron J, Berger E, Stankoff B, Clavelou P, Maillart E, Heinzlief O, Zéphir H, Ruet A, Cazez O, Moulin S, Al-Khedr A, Bourre B, Pelletier J, Magy L, Neau JP, Camdessanché JP, Doghri I, Wahab A, Tchikviladzé M, Labeyrie C, Hankiewicz K, Le Page E, Michel L; as the OFSEP Investigators.  
Neurology. 2024 Oct 22;103(8):e209886. doi: 10.1212/WNL.0000000000209886. Epub 2024 Sep 25. PMID: 39321406
- Non-inferiority design for trials in multiple sclerosis.**  
Cite: Sormani MP, Wiendl H.  
Lancet Neurol. 2024 Oct;23(10):970. doi: 10.1016/S1474-4422(24)00336-3. PMID: 39304262 No abstract available.
- Diagnosis of multiple sclerosis: challenges and opportunities.**  
Cite: Wallin M.  
Lancet Neurol. 2024 Oct;23(10):958-960. doi: 10.1016/S1474-4422(24)00368-5. PMID: 39304250 No abstract available.

## Schritt 2 - Verbindung verschiedener MeSH Terme:

⇒ Es wird noch einmal „multiple sclerosis“ in das obige MeSH-Suchfeld eingegeben. Nach Erscheinen der Termliste wird der 1. Term (hier: „Multiple Sclerosis“) durch ein Kreuz im entsprechenden Feld links daneben aktiviert. Unterhalb des **PubMed search builders** in der rechten Menüleiste wird **Add to search builder** ausgewählt. Der Term erscheint nun in der Search Box.

⇒ Analog zum Schritt 1 wird die MeSH Suche für den Begriff „prevention“ durch Eingabe in das Suchfeld durchgeführt (Löschen von „multiple sclerosis“, Eingabe von „prevention“, **Search**). Hier kann z. B. der gefundene MeSH Term „secondary prevention“ (3. Term in der Termliste) ausgewählt und ebenfalls zur Search Box gesendet werden. Sind beide Terme in der Search Box vorhanden und damit (standardmäßig mit logisch „AND“) miteinander verknüpft, kann über den Link **Search PubMed** eine Suche durchgeführt werden. Bei der logischen Verknüpfung ist je nach inhaltlicher Fragestellung und Auswahl der MeSH-Terme auch die Auswahl von „OR“ oder „NOT“ möglich bzw. kann nötig sein.

MeSH search results for "prevention". The search results list includes:

- [prevention and control \[Subheading\]](#)
  - Used with disease headings for increasing human or animal resistance against disease (e.g., immunization), for control of transmission agents, for **prevention and control** of environmental hazards, or for **prevention and control** of social factors leading to disease. It includes preventive measures in individual cases. Year introduced: 1966
- [Tertiary Prevention](#)
  - Measures aimed at providing appropriate supportive and rehabilitative services to minimize morbidity and maximize quality of life after a long-term disease or injury is present. Year introduced: 2009
- [Secondary Prevention](#)
  - The **prevention** of recurrences or exacerbations of a disease or complications of its therapy. Year introduced: 2009

The search builder shows the query: `("Multiple Sclerosis"[Mesh]) AND "Secondary Prevention"[Mesh]`. The **Add to search builder** button is highlighted with a red circle, and a red arrow points from it to the PubMed search interface below.

PubMed search results for the query: `("Multiple Sclerosis"[Mesh]) AND "Secondary Prevention"[Mesh]`. The search results show 339 results. The first result is:

- [Differential diagnosis of multiple sclerosis.](#)
  - Cavanagh JJ, Levy M. Presse Med. 2021 Jun;50(2):104092. doi: 10.1016/j.jlpm.2021.104092. Epub 2021 Oct 27. PMID: 34715293 Review.
- [Natalizumab Versus Fingolimod in Patients with Relapsing-Remitting Multiple Sclerosis: A Subgroup Analysis From Three International Cohorts.](#)
  -

The number 339 results is circled in red.

## 1.2 Entwicklung einer Suchstrategie für das Fallbeispiel: Welches Medikament in Tablettenform ist für die Basistherapie von Multipler Sklerose geeignet?

Grundsätzlich existieren verschiedene Recherche-Systeme, die zur Lösung der Problemstellung aus dem oben genannten Fallbeispiel eingesetzt werden können. Ein wichtiger Aspekt bei der Benutzung von verschiedenen Systemen ist, dass die Verarbeitung derselben Suchbegriffe zu verschiedenen Ergebnissen führen kann und in der Regel auch wird. Darüber hinaus unterscheiden sich verschiedene Recherche-/Suchsysteme noch hinsichtlich der Gesicherheit und Aktualität der Ergebnisse. Die entsprechenden Effekte sollen einleitend durch Beispielsuchen unter Google, Google Scholar und PubMed untersucht werden. Weiterhin sind in der Problemstellung verschiedene Teilfragestellungen enthalten, die schrittweise bearbeitet werden müssen. Dazu werden neben PubMed auch Cochrane, AWMF und G-I-N eingesetzt. Die einzelnen Teilaufgaben werden separat, aber als Teil der gesamten Suchstrategie behandelt.

**Aufgabe 1.2:** *Um die Frage der Medikation aus dem Fallbeispiel zu lösen, soll mit Hilfe von verschiedenen Recherchesystemen gearbeitet werden. Es soll untersucht werden:*

- welche Resultate eine einfache Suche mit Google bzw. Google Scholar ergibt,
- wie die Problemstellung bei einer MESH-Suche in PubMed behandelt wird,
- wo man unabhängig von PubMed große Reviews zu diesem Thema finden kann,
- welche Leitlinien auf den Seiten deutscher Fachgesellschaften oder international existieren,
- welche Zusatzinformationen man ggfs. z. B. aus Qualitätssicherheitsberichten, Kompetenznetzwerken, Presse oder auch Medien sichten kann.

### Lösungsschritte:

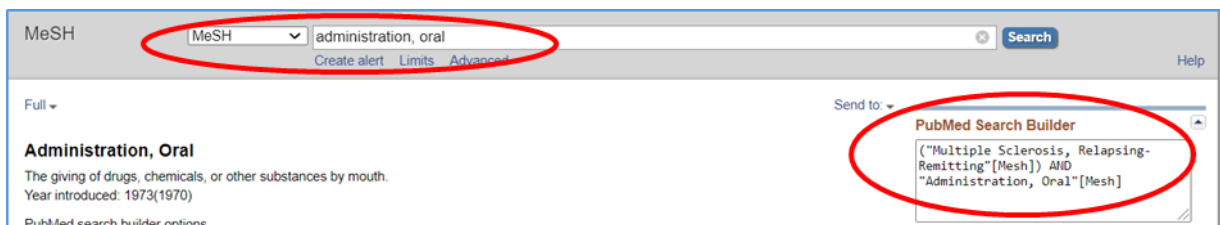
Schritt 1 (Google bzw. Google Scholar):

- ⇒ Google (<http://www.google.de>) wird geöffnet und z. B. die Begriffe „ms therapie tabletten“ (deutsch) oder „multiple sclerosis therapy“ (englisch) eingegeben. Dasselbe wird unter Google Scholar (<http://scholar.google.de>) wiederholt und z.B. auch noch mit „oral“ kombiniert. Bei der Suche unter Google/Google Scholar sind die verschiedenen Optionen der Erweiterten Suche zu beachten, insbesondere die logischen Verknüpfungen sowie Datums- bzw. Zeit- und Spracheinschränkungen.
- ⇒ Bewertet man die Ergebnisse hinsichtlich ihrer Gesicherheit und Aktualität, kann man feststellen, dass unter Google eine Vielzahl unspezifischer Ergebnisse erscheinen, die sich schwer hinsichtlich ihrer Abgesicherheit einordnen lassen. Unter Google Scholar ist dies schon besser möglich, es erscheinen konkrete Hinweise auf die Quellen der gefundenen Ergebnisse.
- ⇒ Unabhängig davon eignen sich beide Suchstrategien aber dazu, ganz allgemeine Informationen zum Suchgegenstand zu sammeln (z. B. bei deutscher Suche: Webseite der forschenden Pharmaunternehmen - Basiswissen: schubförmige und primär-progrediente Form der MS, Basistherapie schubförmige MS: Betainferone, Antikörper, Tablettenform seit 2011, Wirkstoffe Tablettenform: Fingolimod, Teriflunomid, Dimethylfumarat, Cladribin; z. B. bei englischer Suche: Webseite der Mayo Clinic - Basiswissen englische Begriffe: relapsing-remitting multiple sclerosis, primary-progressive multiple sclerosis).

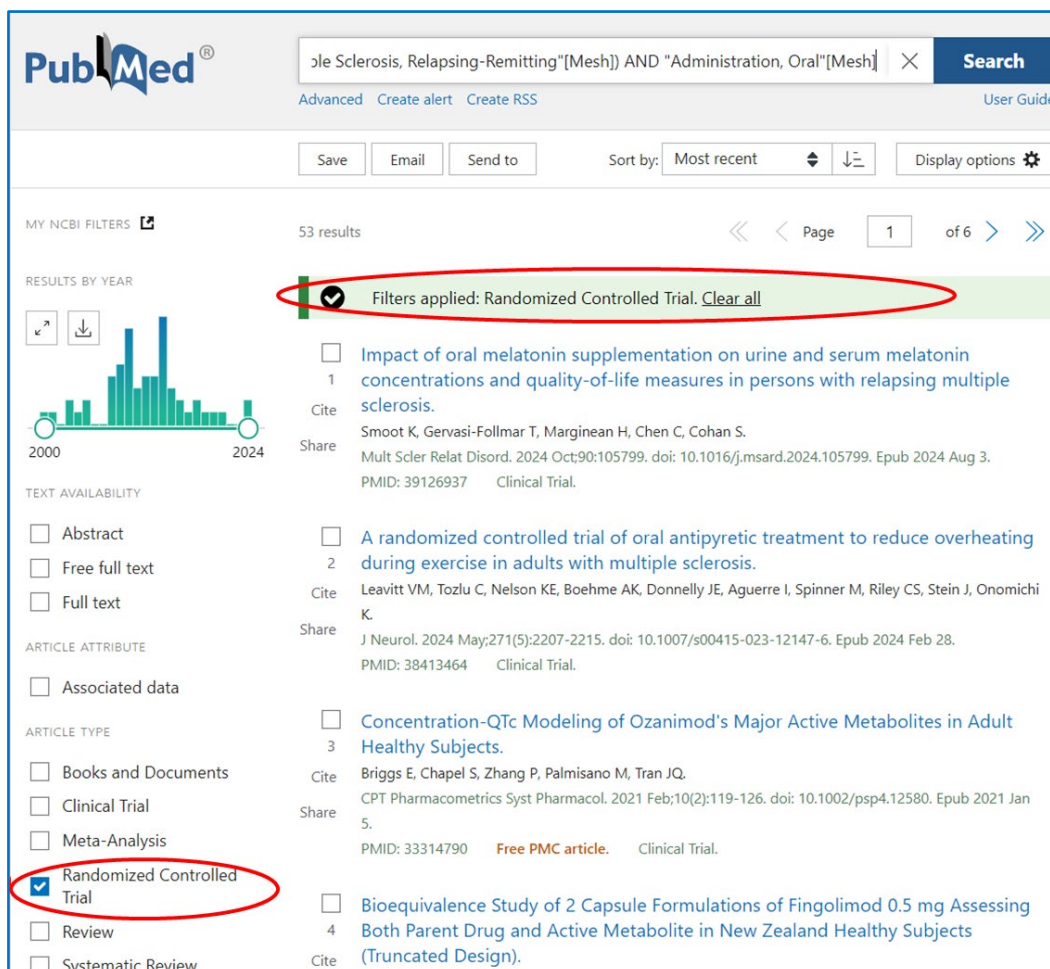


Schritt 2 (PubMed):

- ⇒ PubMed wird geöffnet und **MeSH Database** aktiviert. Im MeSH-Suchfeld wird „Multiple Sclerosis“ eingegeben und in der resultierenden MeSH-Termliste jetzt „Multiple Sclerosis, Relapsing Remitting“ ausgewählt und in den **PubMed search builder** gebracht. Unter Nutzung der in Aufgabe 1.1, Lösungsschritt 2 erlernten Möglichkeiten der Verbindung mehrerer MeSH Terme wird nun nach „Administration, oral“ gesucht und beide Begriffe („Multiple Sclerosis, Relapsing Remitting“ AND „Administration, oral“) im **PubMed search builder** verbunden.



- ⇒ Zur schnelleren Übersicht hinsichtlich der klinischen Relevanz bietet sich eine Limitierung auf hochwertige klinische Studien an. Dazu wird in der Ansicht, in der alle Artikel sichtbar sind, links neben der Ergebnisliste als Artikeltyp **Randomized Controlled Trial** gewählt. Nicht sichtbare Artikeltypen können unter **more...** angewählt werden. In der resultierenden Ergebnisliste wird diese Limitierung mittels eines blau hinterlegten Ausrufezeichens **Filters activated** angezeigt und bleibt so lange gültig, bis sie vor einer neuen Suche wieder aktiv beendet wird (**clear all**).



- ⇒ Eine alternative Suchmöglichkeit wäre z. B. auch die Verbindung von „Multiple Sclerosis, Relapsing Remitting“ AND „secondary prevention“ und die Auswahl von **Review** als Artikeltyp. Verschiedene Kombinationen von MeSH-Termen und Artikeltypen sollen gesichtet werden.
- ⇒ Die gefundenen Artikel sollen gesichtet werden; auf eventuell vorhandene Volltextversionen (**Full Text**) ist zu achten. Dabei sollen Ergebnisse durchaus kritisch und unter Berücksichtigung der Aktualität beleuchtet werden

## Schritt 3 (Übersichtsartikel in der **Cochrane-Datenbank**):

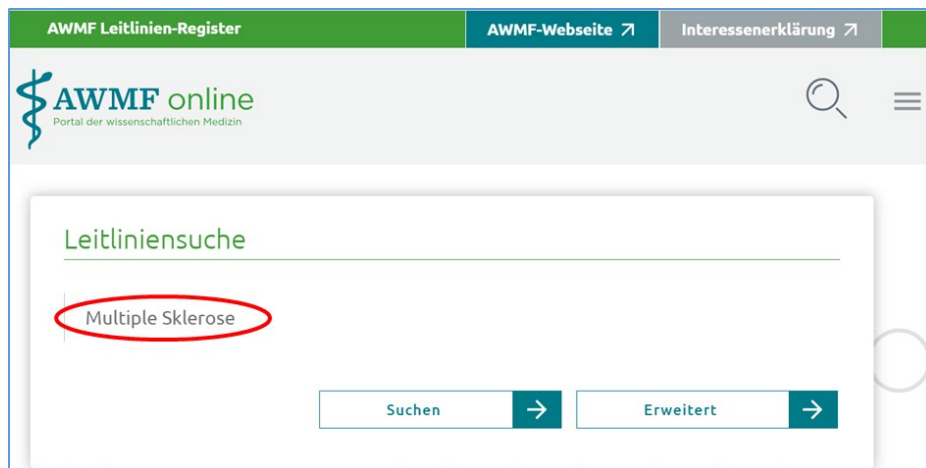
- ⇒ Alternativ zur Auswahl des Reiters **Reviews** bei der PubMed-Suche können auch andere Systeme zur Suche von Übersichtsartikeln verwendet werden. Die Webseite von Cochrane Library (unabhängige Datenbank für Systematische Reviews/Evidence Based Medicine) soll geöffnet werden (<http://www.thecochranelibrary.com/>).
- ⇒ Das rechts oben erscheinende Menu wird geöffnet, im Suchbereich **Advanced Search** angeklickt und dann z. B. „multiple sclerosis oral treatment“ eingegeben. Es erscheinen verschiedene Übersichtsartikel. Einzelne Übersichtsartikel können ausgewählt und angesehen werden. Volltext-Versionen bzw. Zusammenfassungen können als PDF heruntergeladen werden. Einen guten ersten Überblick bietet jeweils die „Plain Language Summary“ am Ende des Abstracts.

The screenshot displays the Cochrane Library website's search interface. At the top, the Cochrane Library logo and tagline "Trusted evidence. Informed decisions. Better health." are visible. The navigation bar includes "Cochrane Reviews", "Searching for trials", "Clinical Answers", "About", "Help", and "About Cochrane". The "Advanced Search" section is highlighted with a red circle. Below it, the search input field contains "multiple sclerosis oral treatment" and is also circled in red. The search results section shows 18 Cochrane Reviews matching the search criteria. The results are listed in a table with columns for "Cochrane Reviews", "Cochrane Protocols", "Trials", "Editorials", "Special Collections", "Clinical Answers", and "More". The first three results are:

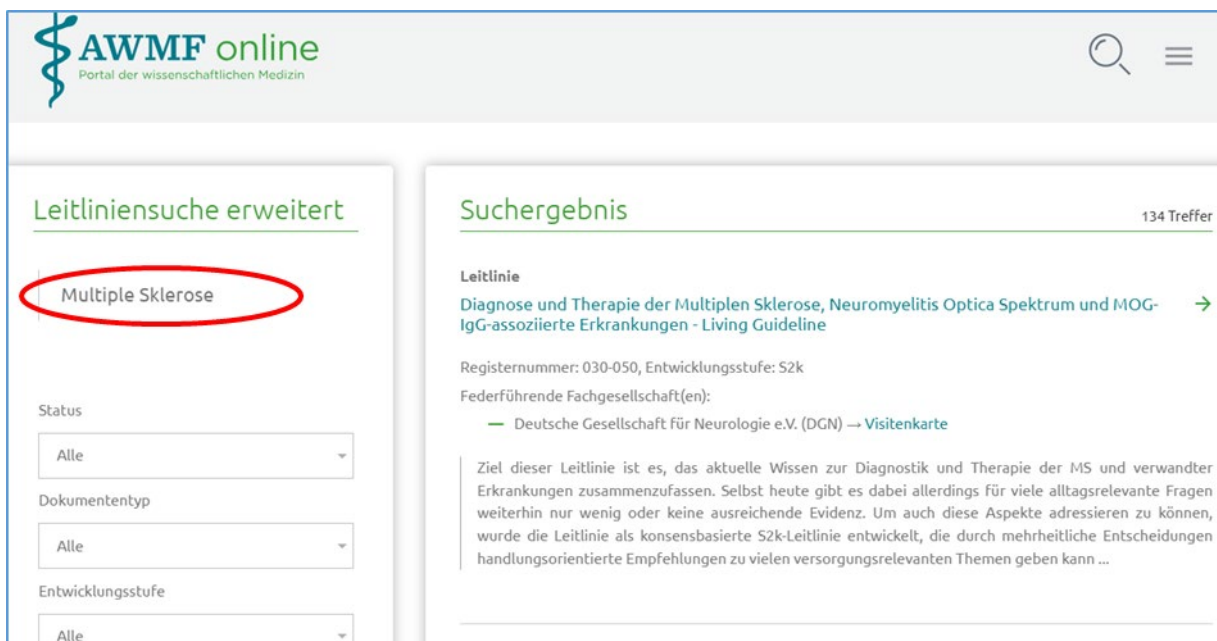
Rank	Title	Authors	Intervention	Review Date	Access
1	Laquinimod for multiple sclerosis	Dian He, Kai Han, Xiangdong Gao, Shuai Dong, Lan Chu, Zhanhui Feng, Shan Wu	Intervention	6 August 2013	Free access
2	Statins for multiple sclerosis	Jin Wang, Yousheng Xiao, Man Luo, Hongye Luo	Intervention	7 December 2011	New search
3	Siponimod for multiple sclerosis	Liujiu Cao, Meixuan Li, Liang Yao, Peijing Yan, Xiaoqin Wang, Zhen Yang, Yongfeng Lao, Huijuan Li, Kehu Yang <sup>a</sup> , Ka Li <sup>a</sup>	Intervention	16 November 2021	Free access

Schritt 4 (Leitlinien in AWMF):

- ⇒ Die Webseite der AWMF (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften) soll geöffnet werden (<http://www.awmf-online.de/>). AWMF ist das Portal der Wissenschaftlichen Medizin in Deutschland. Dort sind u. a. Leitlinien aller medizinischen Fachgesellschaften verfügbar. Unter dem Menubottom oben rechts wird **Leitlinien** ausgewählt, die Leitliniensuche öffnet sich unter **AWMF-Leitlinienregister**.
- ⇒ Unter der Stichwortsuche wird z. B. „Multiple Sklerose“ eingegeben und vorerst keine weitere Einschränkung vorgenommen. Alternativ kann auch nach „Multiple Sklerose Therapie“ gesucht werden. Die Ergebnisse sollen verglichen werden.



- ⇒ Bei Einschränkungen kann insbesondere der **Status** (aktuelle Leitlinie, angemeldete Leitlinie), die **Entwicklungsstufe** (siehe Anlage 1.1) und die federführende **Fachgesellschaft** (z. B. Deutsche Gesellschaft für Neurologie-DGN) berücksichtigt werden. Weiterhin kann beim **Dokumententyp** ausgewählt werden, ob Leitlinien als Kurz- oder Langfassung oder z. B. Patientenleitlinien angezeigt werden sollen.



- ⇒ Die Suche ergibt verschiedene Treffer. Bei den oben aufgeführten Suchen existiert z. B. eine Leitlinie „Diagnose und Therapie der Multiplen Sklerose, Neuromyelitis Optica Spektrum und MOG-IgG-

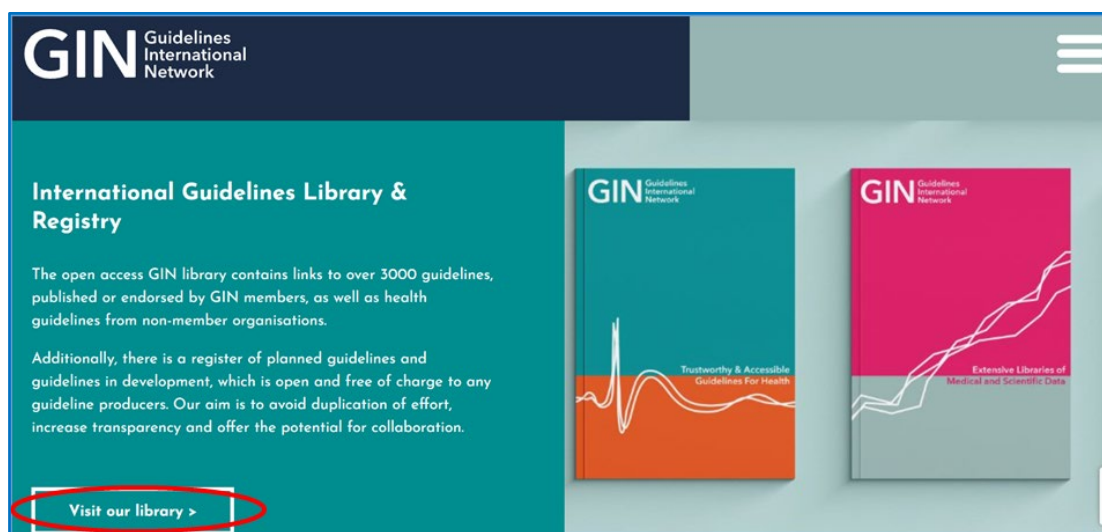


assoziierte Erkrankungen – Living Guideline“ der Deutschen Gesellschaft für Neurologie (DGN) der Stufe S2k. Diese soll gesichtet und nach Vorgaben zur Basistherapie und insbesondere hinsichtlich Empfehlungen zur Gabe in Tablettenform durchsucht werden (z. B. in der Langfassung der Leitlinie, die heruntergeladen werden kann).

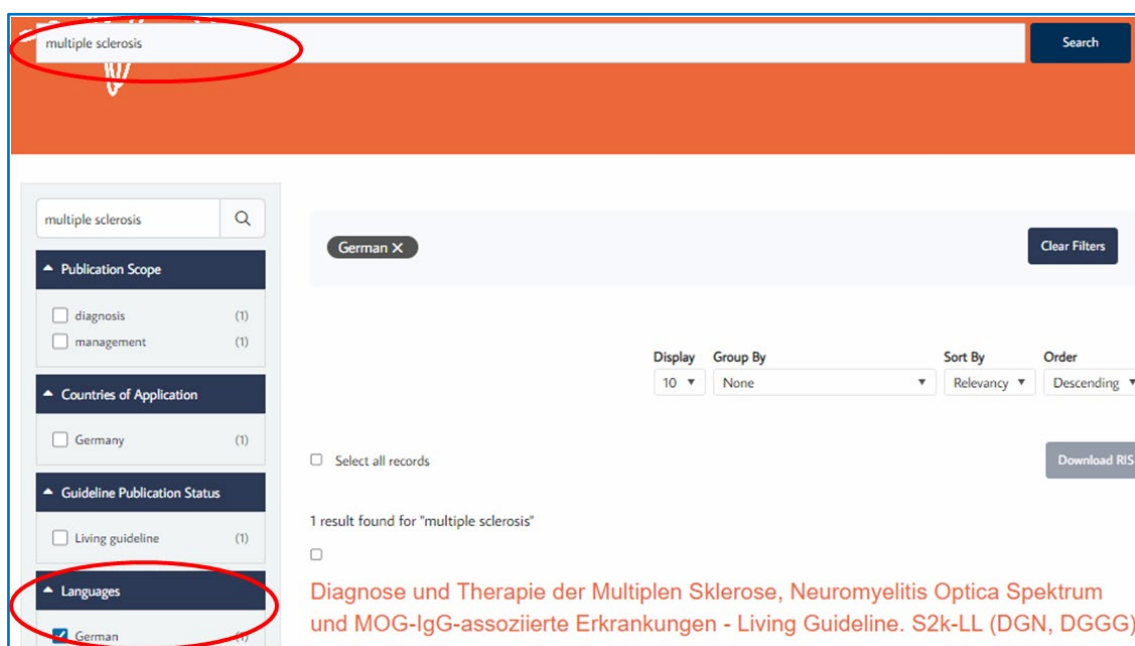
⇒ Alternativ ist auch eine Suche nach Leitlinien auf der Webseite der Deutschen Gesellschaft für Neurologie (DGN) möglich.

### Schritt 5 (Internationale Leitlinien in G-I-N):

⇒ G-I-N (Guidelines International Network) ist ein globales Netzwerk, das evidenzbasierte Gesundheitsfürsorge und eine verbesserte Gesundheitspflege durch die Einführung von internationalen Standards und Leitlinien fördert. Die Webseite von G-I-N soll geöffnet und dort **visit our library** ausgewählt werden.



⇒ In der Suchleiste wird „multiple sclerosis“ eingegeben“. Bei der Suche kann nun weiterhin nur auf bestimmte Sprachen (also z. B. „german“) eingeschränkt werden. Bei der Sichtung der Ergebnisse obiger Suche findet man u.a. die unter AWMF gefundenen Leitlinie der DGN.



### Schritt 6 (Zusatzinformationen: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen - IQWiG - Nutzenbewertungen und Pressemitteilungen):

- ⇒ Als letzter Schritt soll sich auf der Webseite des IQWiG über aktuelle und abgeschlossene Projekte, Pressemitteilungen usw. informiert werden. Mit Hilfe der Suchfunktion (Eingabe von „Multiple Sklerose oral“) können entsprechende Projekte bzw. relevante Pressemitteilungen gefunden und ausgewählt werden. Sowohl eine Kurzfassung des jeweiligen Abschlussberichts als auch weitere Informationen können abgerufen werden.

The screenshot shows the IQWiG website interface. At the top, there is a navigation bar with the IQWiG logo and menu items: ÜBER UNS, PROJEKTE, SICH EINBRINGEN, VERANSTALTUNGEN, PRESSECENTER, and KONTAKT. A search bar is located below the navigation bar, containing the text "Multiple Sklerose oral". To the right of the search bar, there is a "Zeitraum:" dropdown menu set to "Gesamter Zeitraum". Below the search bar, there are filters for "Nach Kategorie filtern:" with buttons for "Alle Ergebnisse (68)", "Projekte (0)", "Pressemitteilungen (1)", "Seiten (0)", and "Gesundheitsinformation.de (0)". The search results are listed below, with the third result, "Teriflunomid (multiple Sklerose) - Nutzenbewertung gemäß § 35a SGB V", circled in red. The other results are "Zusatznutzen von Cannabis Sativa bei Spastik infolge Multipler Sklerose ist nicht belegt" and "Cladribin (multiple Sklerose) - Nutzenbewertung gemäß § 35a SGB V".

Leichte Sprache | Gebärdensprache | Mein IQWiG | DE | EN

**IQWiG** ÜBER UNS PROJEKTE SICH EINBRINGEN VERANSTALTUNGEN PRESSECENTER KONTAKT

Suchbegriff:  Zeitraum:

Ergebnislink kopieren > Alles zurücksetzen

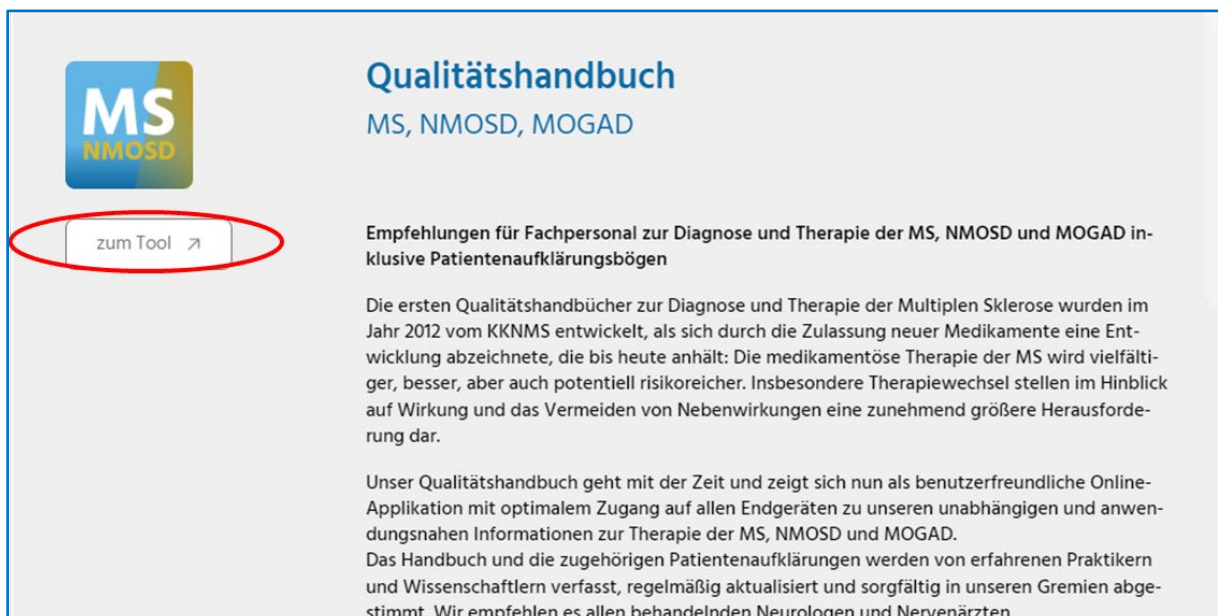
Nach Kategorie filtern:

Alle Ergebnisse (68) Projekte (0) Pressemitteilungen (1) Seiten (0) Gesundheitsinformation.de (0)

**Teriflunomid (multiple Sklerose) - Nutzenbewertung gemäß § 35a SGB V**  
02.11.2021 | Dossierbewertung

### Schritt 7 (Zusatzinformationen: Kompetenznetz Multiple Sklerose – Qualitätshandbuch und Patientenaufklärung):

- ⇒ Das Krankheitsbezogene Kompetenznetzwerk Multiple Sklerosis (KKNMS) ist eines von 21 Kompetenznetzen in der Medizin, die vom Bundesministerium für Bildung und Forschung (BMBF) initiiert wurden. Sie alle verfolgen das Ziel, Forscher zusammenzubringen, um die Patientenversorgung zu verbessern. Das KKNMS wurde seit 2009 vom BMBF finanziell unterstützt.
- ⇒ Das KKNMS entwickelt Empfehlungen zu Diagnose und Therapie der Multiplen Sklerose für Fachpersonal. Die Webseite der KKNMS wird geöffnet und entweder unter dem Reiter **TOOLS** das **Qualitätshandbuch MS, NMOSD, MOGAD** angeklickt. Das Qualitätshandbuch wird als benutzerfreundliche Online-Applikation bzw. als App zur Verfügung gestellt.



- ⇒ Im Qualitätshandbuch kann z. B. als Diagnose MS ausgewählt werden. Man findet sowohl nach den einzelnen Medikationen sortierte Empfehlungen zur Therapie, weiterhin Patientenaufklärungen, Skalen und Scores sowie Schemata und Tabellen. Werden konkrete Medikationen ausgewählt (z.B. Teriflunomid) kann hier gezielt nach Informationen, z.B. zur Kontraindikation gesucht werden.

- ⇒ Alternativ können unter Spezialsituationen auch gezielt nach Informationen z.B. zu „Schwangerschaft“ gesucht werden bzw. die Suchfunktion verwendet werden.



Kompetenznetz  
Multiple Sklerose

MS Stand 03/2024

# SPEZIALSITUATIONEN

SCHWANGERSCHAFT UND KINDERWUNSCH  
IMPFUNGEN  
REISEN  
NARKOSE / OP  
EPILEPSIE  
SPORT  
ERNÄHRUNG  
RAUCHEN  
AUTOREN

## SCHWANGERSCHAFT UND KINDERWUNSCH

### 1. Verlauf der MS in der Schwangerschaft und postpartum

- ▶ Die Schubrate nimmt in der Schwangerschaft kontinuierlich ab (bis zu 80% im letzten Drittel).
- ▶ Neueren Daten zufolge ist der Schubanstieg nach der Geburt deutlich geringer ausgeprägt als die früher beschriebenen 30% in den ersten sechs Monaten und übersteigt das Vorschwangerschaftsniveau bei milder Multiple Sklerose (MS) möglicherweise nicht.

### 1.3 Diagnosis Related Groups: Dokumentation zur Leistungsabrechnung

In Deutschland werden klinische Leistungen nach einem Fallpauschalensystem, den **Diagnosis Related Groups (DRGs)**, vergütet. Die Existenz von Kliniken oder sogar ganzer Krankenhäuser hängt inzwischen davon ab, ob die medizinischen Leistungen im Sinne dieses Vergütungssystems kostendeckend erbracht werden können und davon, dass diese Leistungen gemäß der vorgeschriebenen Dokumentationsvorgaben erfasst und an die Kostenträger gemeldet werden. Um einen Fall nach den DRG-Fallpauschalen abzurechnen, müssen Angaben zu den Diagnosen dieses Falles mit Angaben zu therapeutischen Maßnahmen kombiniert werden. Außerdem sind bestimmte Eigenschaften des Patienten/der Patientin, u. a. Alter oder Länge des Krankenhausaufenthalts, zu erfassen.

Die Angaben zu den Diagnosen werden in Form von ICD-Kodes eingegeben; die Angaben zu den therapeutischen Maßnahmen werden nach OPS kodiert: Die von der Weltgesundheitsorganisation WHO koordinierte **International Statistical Classification of Diseases and Related Health Problems (ICD)** in Ihrer aktuell verwendeten Version 10 dient der Klassifikation und Kodierung von Diagnosen. Das bedeutet, dass z. B. einer Erstmanifestation einer Multiplen Sklerose ein Kürzel (ICD-Kode) zugeordnet wird (hier z. B. G35.0). Die ICD ist als umfangreiche **Hierarchie** organisiert. Ausgehend von in Kapiteln zusammengefassten Krankheitsgruppen finden sich immer genauere Angaben bis hin zu spezifischen Diagnosen. Zur Dokumentation sollte der ICD-Kode verwendet werden, der am genauesten zur aktuellen Diagnose passt, ohne dabei zu speziell zu sein.

Ein ähnliches Klassifikationssystem sind die **Operationen- und Prozedurenschlüssel (OPS)**. Hier geht es um die Erfassung von diagnostischen und therapeutischen Maßnahmen. Auch die OPS sind hierarchisch organisiert. Und auch die OPS ordnen den Maßnahmen und Prozeduren Kürzel zu: die OPS-Kodes.

Beide Klassifikationssysteme sind für das deutsche **Fallpauschalensystem** wichtig. Sie dokumentieren damit die im Falle des Patienten vorliegenden Diagnosen und Maßnahmen. Das DRG-System ermittelt aus diesen Angaben eine **Fallgruppe**. Einer solchen Fallgruppe sind dann Entgelte für das Krankenhaus zugeordnet: Die Kostenträger zahlen das sog. **effektive Entgelt**.

**Aufgabe 1.3:** Nehmen Sie in Anlehnung an unser Fallbeispiel eine solche Fallgruppierung vor. Nutzen Sie dazu das im Internet frei verfügbare Computerprogramm DRG-Grouper der DRG-Research Group (<https://www.drg-research-group.de>).

#### Lösungsschritte:

- ⇒ Klicken Sie auf der genannten Webseite auf den Reiter **Grouper**. Machen Sie sich vertraut mit den verfügbaren Eingabefeldern. Im oberen Bereich sind einige Angaben zum Aufenthalt zu machen. Wenn Sie versuchen, das Gewicht von Erwachsenen als Aufnahmegewicht einzutragen, wird schnell klar, dass dieses Feld sich nur für Neugeborene eignet. Ansonsten wird dort nichts eingetragen. Gleich mehrere Felder dienen der Erfassung von ICD- und OPS-Kodes. Bei den Diagnosen ist zu beachten, dass im ersten Feld (links oben neben der Überschrift „Diagnosen:“) die Hauptdiagnose eingetragen wird, also der hauptsächliche Grund für den klinischen Aufenthalt. (Hinweis: Wenn Sie auf die blaue Schrift „Diagnosen:“ bzw. „Prozeduren:“ klicken öffnet sich jeweils ein Suchfenster für die ICD- bzw. OPS-Kodes.)
- ⇒ Nutzen Sie die Verknüpfungen zur Suche von ICD-Kodes (**Diagnosen**, blaue Schrift) und Prozedurenkodes (**Prozeduren**, wieder blaue Schrift). Suchen Sie geeignete Codes für obiges Fallbeispiel heraus und tragen Sie diese in die Erfassungsmaske des Webgroupers ein. (Hinweis: Die im DRG Grouper angebotene ICD-Suche liefert Ergebnisse, die von der Groß- bzw. Kleinschreibung abhängen. Probieren Sie im Zweifelsfalle beides.)
- ⇒ Klicken Sie nach Eingabe der Angaben auf die Taste **Gruppieren** Sehen Sie sich die Ergebnisse an. Klären Sie weiter unten auf der Ergebnisseite, ab welchem Aufenthaltstag die Fallpauschale für diesen Fall wieder ansteigt (sehen Sie sich das Diagramm dazu an).



DRG-RESEARCH GROUP Über uns Kodierung Grupper Kontakt Impressum Datenschutzerklärung

## Webgrupper

**Achtung:** Die Adresse dieser Webseite hat sich geändert. Bitte aktualisieren Sie Ihre Favoriten sofern diese noch auf drg.uni-muenster.de verweisen!

Für die Angabe der Seitenlokalisierung bei den Prozeduren ergänzen Sie ggf. den jeweiligen Prozedurencode mit :l (links), :r (rechts) oder :b (beidseits).

Geschlecht: weiblich Alter: 29 Jahre

Aufnahmegew.: 0 Gramm Behandlung: freiwillig

Verweildauer: 3 Tage Beatmungszeit: 0 Stunden (max. 9999)

Abteilungstyp: Hauptabteilung

Aufnahmeart: Krankenhausbehandlung, vollstationär

Aufnahmearbeit: Einweisung durch einen Arzt

Entlassungsart: Behandlung regulär beendet

Diagnosen: G35.0

Prozeduren: 1-208.4 am 01.10.2019

Basisfallwert: 3544,97 €

G-DRG 2019 Gruppieren Neuer Fall

**Gruppierungsergebnis (G-DRG 2019)**

MDC 01 Krankheiten und Störungen des Nervensystems

DRG B68D Multiple Sklerose und zerebellare Ataxie, ein Belegungstag oder ohne äußerst schwere CC, Alter > 15 Jahre, ohne komplexe Diagnose

PCCL 0 Status: normales Grouping (GetDRG-Grupper 2019) (GetDRG V18.1.2)

Verweildauer  
 aktuelle VWD: 3 1. Tag Abschlag: 1 1. Tag Zuschlag: 15  
 mittl. VWD (arith.): 6,6 untere GVWD: 2 obere GVWD: 14

Kostengewicht  
 Basiskostengewicht: 0,752 eff. Kostengewicht: 0,752 (Zuschlag: 0)

Entgelt  
 Basisentgelt: 2665,82 € eff. Entgelt: 2665,82 € (Zuschlag: 0,00 €, Basisfallpreis: 3544,97 €)

Diagnosen (ICD-10-GM 2019)	Kode	Bezeichnung	Verwendet	CCL	
	G35.0	Erstmanifestation einer multiplen Sklerose	J	2	gültig

Prozeduren (OPS Version 2019)	Kode	Bezeichnung	Verwendet	OR/NOR	Amtlich
	1-208.4	Registrierung evozierter Potentiale: Motorisch [MEP]	N		1-208.4

**DRG-Grafik**

DRG-Grafik für GDRG2019 / B68D (Basisfallpreis: 3544.97 Euro)  
 (c) DRG Research Group, Angaben ohne Gewähr.

Verweildauer [Tage]	Entgelt [Euro]	Annotation
1	1500.26	1. Tag Zuschlag
2	2665.82	untere GVWD
3	2665.82	mittlere GVWD
7	2665.82	mittlere GVWD
14	2665.82	obere GVWD
15	2945.87	1. Tag Zuschlag
16	3225.02	2. Tag Zuschlag
17	3505.98	3. Tag Zuschlag
18	3786.03	Basisfallpreis

## Anlage 1.1: Erläuterung der Abkürzungen in Klinischen Leitlinien

### Entwicklungsstufen

Nach dem System der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) werden Leitlinien in vier **Entwicklungsstufen von S1 bis S3** entwickelt und klassifiziert, wobei S3 die höchste Qualitätsstufe der Entwicklungsmethodik ist.

- **S1:** von einer Expertengruppe im informellen Konsens erarbeitet
- **S2k:** eine formale Konsensfindung hat stattgefunden
- **S2e:** eine systematische „Evidenz“-Recherche hat stattgefunden
- **S3:** Leitlinie mit zusätzlichen/allen Elementen einer systematischen Entwicklung (Logik-, Entscheidungs- und „Outcome“-Analyse, Bewertung der klinischen Relevanz wissenschaftlicher Studien und regelmäßige Überprüfung)

Die methodische Qualität einer S3-Leitlinie ist dementsprechend höher einzuschätzen als die einer S2- oder S1-Leitlinie. Die überwiegende Mehrheit (knapp 70 %) aller Leitlinien der wissenschaftlichen medizinischen Fachgesellschaften sind S1-Leitlinien.

Zusatz **IDA:** Interdisziplinärer Abgleich.

### Evidenzklasse

Mit Hilfe von *Evidenzklassen* (synonym *Evidenzebenen* oder *Evidenzlevel*), erfasst man die wissenschaftliche Aussagefähigkeit klinischer Studien. Dabei unterscheidet man nach den Empfehlungen des AHRQ (*Agency for Healthcare Research and Quality*) die Evidenzklassen I bis IV. Studien der Klasse Ia haben die höchste Evidenz, Studien der Klasse IV die geringste. Je höher die Evidenzklasse, desto besser ist die wissenschaftliche Begründbarkeit für eine Therapieempfehlung.

- **Klasse Ia:** Evidenz durch wenigstens eine Meta-Analyse auf der Basis methodisch hochwertiger randomisierter, kontrollierter Studien.
- **Klasse Ib:** Evidenz aufgrund von mindestens einer ausreichend großen, methodisch hochwertigen randomisierten, kontrollierten Studie.
- **Klasse IIa:** Evidenz aufgrund von mindestens einer gut angelegten (hochwertigen), jedoch nicht randomisierten und kontrollierten Studie.
- **Klasse IIb:** Evidenz aufgrund von mindestens einer gut angelegten, quasi-experimentellen Studie.
- **Klasse III:** Evidenz aufgrund gut angelegter (methodisch hochwertiger), nicht-experimenteller deskriptiver Studien wie etwa Vergleichsstudien, Korrelationsstudien oder Fall-Kontroll-Studien.
- **Klasse IV:** Evidenz aufgrund von Berichten der Experten-Ausschüsse oder Expertenmeinungen bzw. klinischer Erfahrung (Meinungen und Überzeugungen) anerkannter Autoritäten; beschreibende Studien.
- **Klasse V:** Fallserie oder eine oder mehrere Expertenmeinungen

### Empfehlungsgrad

Auf der Basis der Evidenzklassen werden Behandlungsempfehlungen mit einem bestimmten Empfehlungsgrad (synonym: Empfehlungsstärke) entwickelt. Dazu werden verschiedene Klassifizierungssysteme verwendet; verbreitet ist folgende Unterteilung:

- **Grad A:** „Soll“-Empfehlung: zumindest eine randomisierte kontrollierte Studie von insgesamt guter Qualität und Konsistenz, die sich direkt auf die jeweilige Empfehlung bezieht und nicht extrapoliert wurde (Evidenzebenen Ia und Ib).

- **Grad B:** „Sollte“-Empfehlung: Gut durchgeführte klinische Studien, aber keine randomisierten klinischen Studien, mit direktem Bezug zur Empfehlung (Evidenzebenen II oder III) oder Extrapolation von Evidenzebene I, falls der Bezug zur spezifischen Fragestellung fehlt.
- **Grad C:** „Kann“-Empfehlung: Berichte von Expertenkreisen oder Expertenmeinung und/oder klinische Erfahrung anerkannter Autoritäten (Evidenzkategorie IV) oder Extrapolation von Evidenzebene IIa, IIb oder III; diese Einstufung zeigt an, dass direkt anwendbare klinische Studien von guter Qualität nicht vorhanden oder nicht verfügbar waren.
- **Good Clinical Practice** Wenn es für eine Behandlungsmethode keine experimentellen wissenschaftlichen Studien gibt, diese nicht möglich sind oder nicht angestrebt werden, das Behandlungsverfahren aber dennoch allgemein üblich ist und innerhalb der Konsensusgruppe eine Übereinkunft über das Verfahren erzielt werden konnte, so erhält diese Methode die Empfehlungsstärke *Good Clinical Practice* GCP (synonym: KKP = *Klinischer Konsensuspunkt*).

Aus:

[http://de.wikipedia.org/wiki/Evidenzbasierte\\_Medizin](http://de.wikipedia.org/wiki/Evidenzbasierte_Medizin) und

[http://de.wikipedia.org/wiki/Medizinische\\_Leitlinie](http://de.wikipedia.org/wiki/Medizinische_Leitlinie)



## 2 Literaturverwaltung und –zitierung mit EndNote

Die Nutzung des Internets zur Recherche in Online-Datenbanken gehört bereits zu den alltäglichen Routineabläufen in der wissenschaftlichen Arbeit. Das gezielte Suchen und Auswerten von Literatur ist bei der steigenden Zahl publizierter Studien und Übersichten ohne rechentechnische Hilfsmittel nicht mehr denkbar. Daher existiert eine Reihe von Programmen, mit denen Literatur sowohl gesucht und verwaltet, als auch direkt zur Zitierung in Textdokumenten (z. B. Microsoft Word) benutzt werden kann. Das (kommerzielle) Programm EndNote® fasst z. B. alle Aufgaben zusammen, die mit der Recherche, Verwaltung und Nutzung von Literaturdokumenten verbunden sind (All-in-One-Lösung):

- Automatische Erstellung und Auswertung von eigenen Publikationsverzeichnissen (Zusammenstellungen von Publikationen = Literaturdatenbanken)
- Recherche in Online-Datenbanken und Integration der Publikationsdaten in eigene Publikationsverzeichnisse
- Einbindung von Literaturdaten (Referenzen) in Publikationen (automatisierte Zitierung)

EndNote (Version 20) kann mehr als 50 unterschiedliche Dokumententypen (Artikel, Buch, elektronische Quelle etc.) verwalten und individuell anpassen. In einer einzelnen Datenbank kann eine sehr große Anzahl von Refenzen (Literaturdokumente) gespeichert und verwaltet werden. Bei der Recherche werden alle bekannten Online-Datenbanken berücksichtigt (rund 6000 vordefinierte Verbindungsdateien), vorhandene Full-Texte können ebenfalls in die Datenbank integriert werden. Das Schreiben von Publikationen wird durch im EndNote hinterlegte Formatierungen (ca. 4000 vordefinierte Styles) unterstützt. Styles strukturieren die Literaturangaben aus den eigenen Datenbanken so, wie dies von den einzelnen Zeitschriften vorgeschrieben ist.

### **Bemerkung:**

Für die Arbeit mit EndNote im Home-Use steht unter

<https://wiki.uni-jena.de/pages/viewpage.action?pageId=22453199> unter Endnote 20 eine Version für Studierende der Universität Jena zur Verfügung. Bitte beachten Sie, dass dafür das FSU-Netz erforderlich ist (ggf. via VPN verbinden).

Ebenfalls angeboten wird dort die Literaturverwaltungssoftware Citavi, es existieren noch weitere Lösungen wie z. B. Zotero. Wir nutzen für unsere Lehre EndNote, weil wir den Studierenden dort einen vorgefertigten Style-File für die Zitierung innerhalb der Dissertation an unserer Medizinischen Fakultät liefern können.

Welche Angaben (bibliographische Daten) enthält ein Literaturdokument?

Sowohl die konkreten Publikationen als auch die entsprechenden Datenbanken folgen einheitlichen Standards.

**ORIGINAL**

Open access Protocol

**BMJ Open Hospital-wide EElectronic medical record evaluated computerised decision support system to improve outcomes of Patients with staphylococcal bloodstream infection (HELP): study protocol for a multicentre stepped-wedge cluster randomised trial**

Stefan Hagel,<sup>1</sup> Julia Gantner,<sup>2</sup> Cord Spreckelsen,<sup>2</sup> Claudia Fischer,<sup>2</sup> Danny Ammor, Susanne Müller, Florian Rübner, Marc Fabian J Henning Trawinski, Margarethe Konik, Jonas Schulte-Peter-Michael

**DATENBANK**

Randomized Controlled Trial > BMJ Open. 2020 Feb 10;10(2):e033391.  
doi: 10.1136/bmjopen-2019-033391.

**Hospital-wide EElectronic medical record evaluated computerised decision support system to improve outcomes of Patients with staphylococcal bloodstream infection (HELP): study protocol for a multicentre stepped-wedge cluster randomised trial**

Stefan Hagel<sup>1</sup>, Julia Gantner<sup>2</sup>, Cord Spreckelsen<sup>2</sup>, Claudia Fischer<sup>2</sup>, Danny Ammor<sup>3</sup>, Kutaiba Saleh<sup>3</sup>, Lo An Phan-Vogtmann<sup>2</sup>, Andrew Heide<sup>3</sup>, Susanne Müller<sup>2</sup>, Alexander Helthorn<sup>3</sup>, Henner Kruse<sup>3</sup>, Eric Thomas<sup>3</sup>, Florian Rübner<sup>4</sup>, Silke Haferkamp<sup>5</sup>, Jens Vorwerk<sup>5</sup>, Saskia Deffge<sup>6</sup>, Marc Fabian J Henning Trawinski<sup>7</sup>, Norman Lippmann<sup>8</sup>, Christoph Lübbert<sup>9</sup>, Henning Trawinski<sup>9</sup>, Sebastian Wendt<sup>9</sup>, Thomas Wendt<sup>10</sup>, Andreas Dürschmid<sup>10</sup>, Margarethe Konik<sup>11</sup>, Stefan Moritz<sup>12</sup>, Daniel Tiller<sup>13</sup>, Rainer Röhrig<sup>14</sup>, Jonas Schulte-Coerne<sup>15</sup>, Jonas Fortmann<sup>14</sup>, Stephan Jonas<sup>15</sup>, Oliver Witzke<sup>16</sup>, Peter-Michael Rath<sup>17</sup>, Mathias W Pletz<sup>18</sup>, André Scherag<sup>18</sup>

**Abstract**

**Introduction:** Staphylococci are the most commonly identified pathogens in bloodstream infections. Identification of *Staphylococcus aureus* in blood culture (SAB) requires a prompt and adequate clinical management. The detection of coagulase-negative staphylococci (CoNS), however, corresponds to contamination in about 75% of the cases. Nevertheless, antibiotic therapy is often initiated, which contributes to the risk of drug-related side effects. We developed a computerised clinical decision support system (HELP-CDSS) that assists physicians with an appropriate management of patients with *Staphylococcus* bacteraemia. The CDSS is evaluated using data of the *Data Integration Centers* (DIC) established at each clinic. DICs transform heterogeneous primary clinical data into an interoperable format, and the HELP-CDSS displays information according to current best evidence in bacteraemia treatment. The overall aim of the HELP-CDSS is a safe but more efficient allocation of infectious diseases specialists and an improved adherence to established guidelines in the treatment of SAB.

Wichtige Angaben bei Journal-Artikeln (Felder der EndNote-Datenbank) sind:

Titel der Publikation; Autoren; Zeitschrift; Band (Volume); Erscheinungsjahr; Seitenangaben; Adressen der Autoren; Zusammenfassung des Artikels (Abstract); Schlüsselworte (key words); Dokumententyp (Artikel, Übersicht, Kurzzusammenfassung...); Veröffentlichungssprache

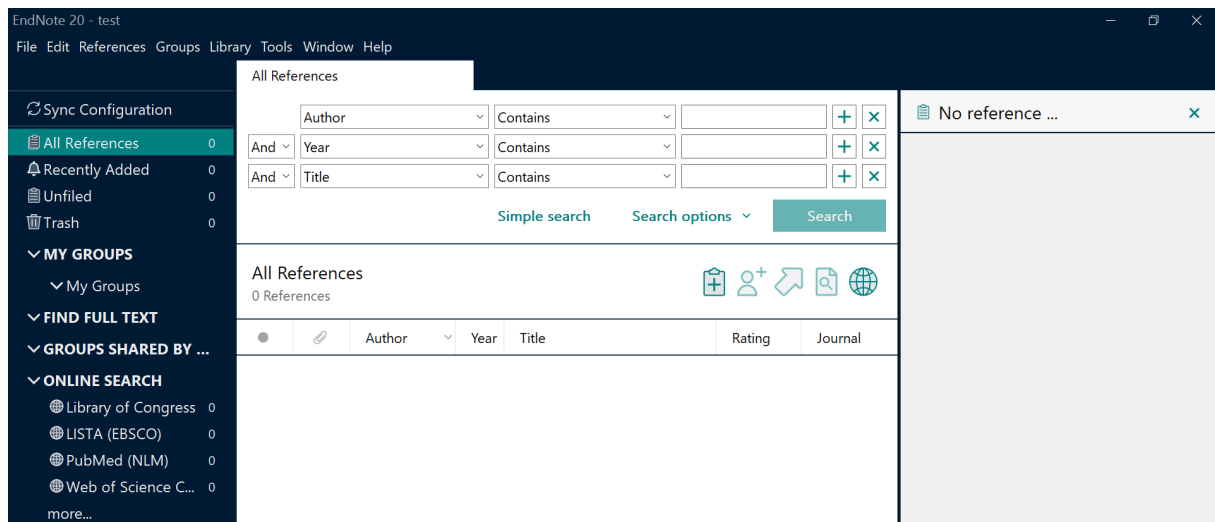
Bei Büchern und Monographien werden teilweise andere Daten benötigt und die Datenstruktur ist von der eines Artikels abweichend, so dass diese als andere Dokumententypen geführt werden müssen. Wesentlich ist, dass mit Hilfe dieser Daten jedes Dokument eindeutig gekennzeichnet und auffindbar ist sowie alle wichtigen Informationen enthält.

## 2.1 Anlegen einer Literaturdatenbank

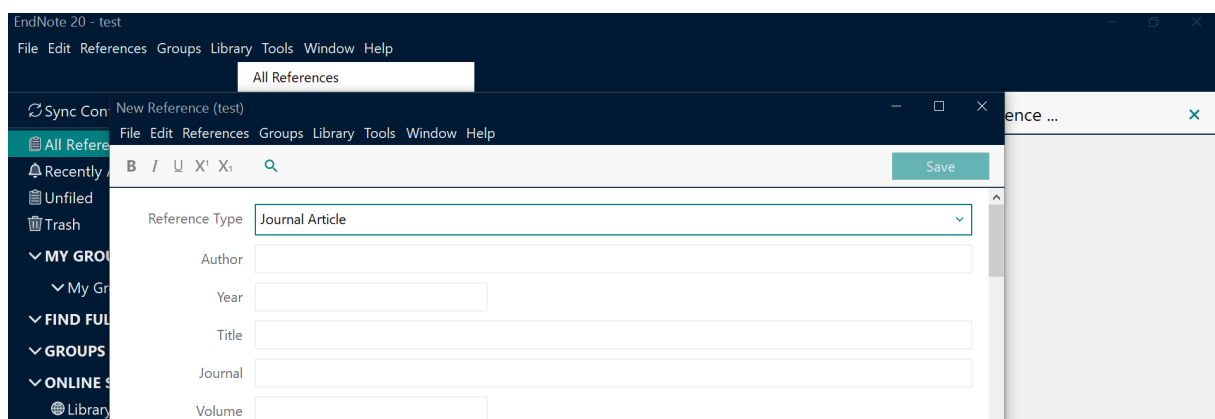
**Aufgabe 2.1:** Es soll eine neue Bibliographie durch das Anlegen einer Literaturdatenbank und die Eingabe erster Datensätze aufgebaut werden.

### Lösungsschritte:

⇒ Das Anwendungsprogramm EndNote wird von der Desktop-Bedienoberfläche aus gestartet (Doppelklick oder **Öffnen**). Es erscheint die Arbeitsoberfläche von EndNote mit der Aufforderung, entweder eine existierende Bibliothek zu öffnen oder eine neue anzulegen. Alternativ kann im Pull-Down-Menü von **File** (Datei) **New ...** angeklickt und der gewünschte Name der neuen Datenbank eingetragen werden. Nach **Speichern** erscheint die Anzeigestruktur der neuen Datenbank, wobei als Standardspalten die Autoren, das Erscheinungsjahr, der Titel, die Zeitschrift und der Referenztyp angezeigt werden. Die Ansicht der folgenden Datenbankeinträge kann dann entsprechend nach diesen Attributen sortiert werden.



⇒ Über den Menüpunkt **References** und **New Reference** (oder über Symbolleiste: New References) wird je nach gewähltem Dokumenttyp (Reference Type) ein Eingabefenster angezeigt, so dass die Eingabe sofort vorgenommen werden kann. Als Standarddokumententyp wird **Journal Article** angeboten. Die Eingabe erfolgt durch das Einklicken in die entsprechenden Eingabezeilen.



⇒ Geben Sie einige der folgenden Literaturstellen als Journal Article in die Datenbank ein:

- [1] Reuken PA, Scherag A, Stallmach A: "Postcoronavirus Disease Chronic Fatigue Is Frequent and Not Only Restricted to Hospitalized Patients", *Crit Care Med*, vol. 49(10), pp. e1052-e1053, 2021.
- [2] Spreckelsen C, Schemmann U, Phan-Vogtmann LA, Scherag A, Winter A, Schneider B: "Health Informatics Learning Objectives on an Interoperable, Collaborative Platform", *Stud Health Technol Inform*, vol. 281, pp. 1019-1020, 2021.
- [3] Schiecke K, Schumann A, Benninger F, Feucht M, Baer KJ, Schlattmann P: "Brain-heart interactions considering complex physiological data: processing schemes for time-variant, frequency-dependent, topographical and statistical examination of directed interactions by convergent cross mapping", *Physiol Meas*, vol. 40(11), pp. 114001, 2019.
- [4] Pester B, Lehmann T, Leistriz L, Witte H, Ligges C: "Influence of imputation strategies on the identification of brain functional connectivity networks", *J Neurosci Methods*, vol. 309, pp. 199-207, 2018.

⇒ Achten Sie darauf, dass jeder Autor in eine separate Zeile geschrieben wird. Ist ein neu eingetragener Autor noch nicht in der Datenbank enthalten, wird er rot markiert. Bereits in der Datenbank enthaltene Autoren werden schwarz geschrieben und von EndNote bei der Eingabe vorgeschlagen. Dasselbe gilt für Zeitschriftennamen.

⇒ Nach der Eingabe (Eingabefenster für einzelne Reference schließen und durch Speichern bestätigen) werden die eingegebenen Literaturstellen in der oben beschriebenen Ansicht in der Datenbank sichtbar.

**Aufgabe 2.2:** *Es ist eine Recherche in einer Literatur-Datenbank über das Internet vorzunehmen und das Ergebnis in einer EndNote-Datenbank zu speichern. Möglicherweise vorhandene PDF-Files des gesuchten Artikels (Full-Text) sollen dabei direkt in die Datenbank eingebunden werden.*

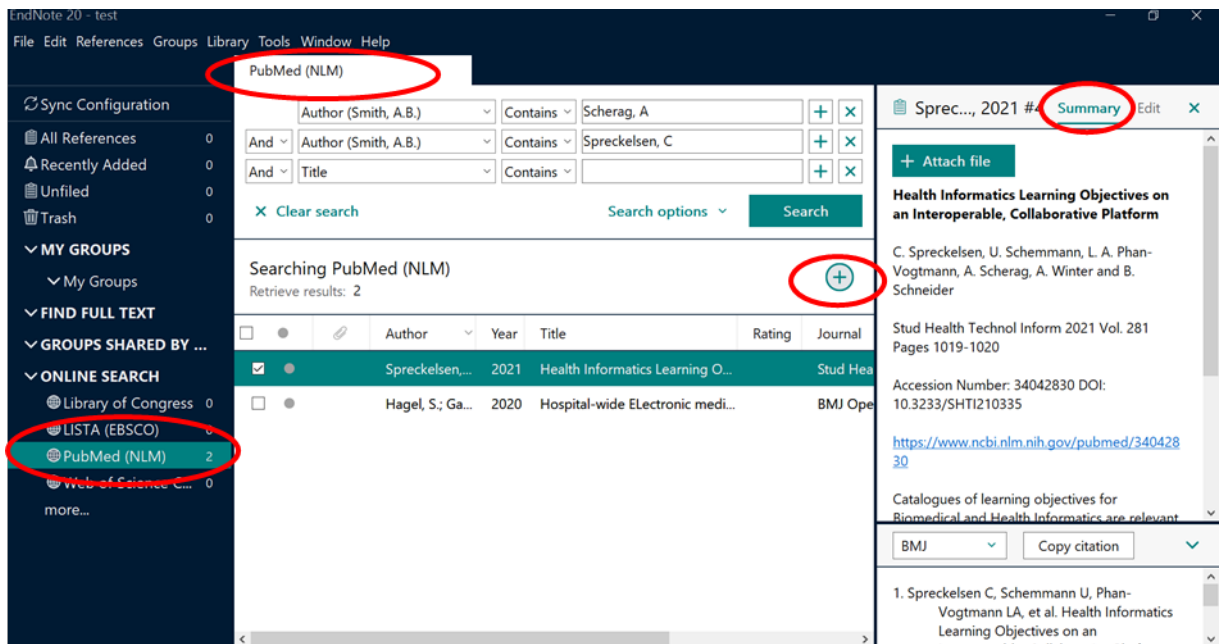
### **Lösungsschritte:**

⇒ Die Verbindung zu externen Datenbanken wird über Verbindungsdateien (-protokolle), die so genannten **Connection Files**, direkt aus der EndNote-Datenbank heraus aufgebaut. In der Datenbankansicht ist auf der linken Seite eine Auswahl unter **ONLINE SEARCH** zu finden, alternativ ist das Auswahlfenster aller möglichen Verbindungen über die Menüpunkte **Tools** und **Online Search** zu erreichen.

⇒ Aus der Liste auf der linken Seite wird PubMed (NLM) gewählt, eine frei zugängliche Literaturdatenbank der National Library of Medicine. Die Nutzung ist unentgeltlich. Bei Nutzung der Auswahlfenster über die Menüpunkte wird mit **Choose** EndNote und PubMed verbunden.

⇒ Das Suchfenster (**Search**) von EndNote ist im oberen Bereich der Datenbankansicht zu finden. Dort ist auch sichtbar, mit welcher Online-Datenbank verbunden wurde (hier: PubMed(NLM), *Achtung: Wird keine Verbindung ausgewählt, erfolgt die Suche über dieses Fenster nur in der eigenen Datenbank!*)

⇒ Bei der Suche werden für die bibliographischen Merkmale Kriterien eingegeben und mit logischen Operatoren (AND, OR, NOT) miteinander verknüpft. Kriterien können mittels eines Pull-Down-Menüs ausgewählt werden. Gesucht werden kann z. B. nach **Author** „Scherag, A“ AND **Author** „Spreckelsen, C“.



- ⇒ Alle Suchergebnisse werden im mittleren Teil der Datenbank aufgelistet, auf der rechten Seite kann unter Summary auf weitere Informationen zu der gefundenen Referenz zugegriffen werden (z. B. Abstract, direkter Link zu PubMed usw. – *Achtung: Der rechte Teil des Fensters ist geteilt und bietet auch noch eine Vorschau auf den aktuell verwendete EndNote-Style, ggfs. kann die Linie zwischen beiden Teilen zur verschoben werden!*) und dementsprechend relevante Referenzen ausgewählt werden. Wurden eine oder mehrere Referenzen ausgewählt, können diese mittels des „+“-Zeichens direkt über den Suchergebnissen zur eigenen Datenbank hinzugefügt werden (**add selected online records to your local library**).
- ⇒ Eine neue Suche (z. B. nach **Author** „Scherag, A“ AND **Year** „2021“ oder auch nach **Author** „Schiecke, K“ AND **Author** „Leistritz, L“ AND **All Fields** „EEG“) kann vorgenommen werden. Bei der Suche kann alternativ anstatt auf die Online-Datenbank PubMed (NLM) auch auf Web of Science zugegriffen werden (ebenfalls in der Auswahl unter **ONLINE SEARCH** auf der linken Seite). Es ist sinnvoll, zum Abschluss noch nach Duplikaten zu suchen. Dies kann über **Library** ⇒ **Find Duplicates** realisiert werden.
- ⇒ Alle nun in der Datenbank vorhandenen Einträge oder nur eine Auswahl an Einträgen werden markiert und über den entsprechenden Icon die Full-Text-Suche gestartet. Das Fortschreiten und das Ergebnis der Suche wird auf der linken Seite unter **FIND FULL TEXT** angezeigt.
- ⇒ Gefundene Full-Text-Dokumente werden durch eine Büroklammer vor dem Autor symbolisiert und sind (z. B. als PDF) für die entsprechende Referenz unter der Summary auf der rechten Seite sichtbar. Neue/andere Dateien können dort für den entsprechenden Datenbankeintrag auch mittels des Icons **Attach File** angehängt werden.



The screenshot shows the EndNote 20 interface. On the left sidebar, the 'FIND FULL TEXT' button is circled in red. The top search area has three filters: 'Author' containing 'Schiecke, K', 'Author' containing 'Leistriz, L', and 'Any Field' containing 'EEG'. The main panel displays a list of 8 references. The reference by Spreckelsen et al. (2021) is circled in red. On the right, a preview of the article is shown, with a red circle around the title 'Spreckelsen-2021-Health Inf...'.

**Aufgabe 2.3:** Innerhalb der vorhandenen Literaturdatenbank sollen Gruppen (z. B. nach einzelnen Autoren, bestimmten Anwendungen oder Jahren) gebildet werden.

**Lösungsschritte:**

⇒ In der Kategorie **MY GROUPS** auf der linken Seite können neue Gruppen mittels rechter Maustaste ⇒ **Create Group** gebildet und benannt werden (z. B. Einschränkung auf das Jahr 2021). Aus der Gruppe **All References** können dann einzelne oder mehrere Einträge mittels gedrückter linker Maustause in diese Gruppe gezogen werden.

The screenshot shows the 'Groups' menu in EndNote 20. The 'Create Group' option is circled in red. In the left sidebar, the 'My Groups' section shows a group named '2021' circled in red. The main panel shows a list of references, with the reference by Spreckelsen et al. (2021) circled in red.

⇒ Genauso können **Smart Groups** gebildet werden, wobei hier Kriterien (z. B. Jahr, Autor, Journal oder ähnliches in OR oder AND Verknüpfungen) eingegeben werden können. Hier sollen z. B. aus den vorhandenen Einträgen alle Referenzen zum Autor Scherag ausgewählt werden. Dazu wird in das Feld **Smart Group Name** „Autor Scherag“ und in das **Suchfenster** „Author - Contains - Scherag“ eingegeben. Die entsprechenden Referenzen erscheinen automatisch in der erstellten Gruppe.

The screenshot shows the EndNote 20 interface. A 'Smart Group' dialog box is open, with the 'Smart Group Name' field containing 'Autor Scherag'. The criteria are set to 'Author' 'Contains' 'Scherag'. Below the dialog, the 'MY GROUPS' list shows 'Autor Scherag' with 5 references. The main window displays a list of references for the 'Autor Scherag' group.

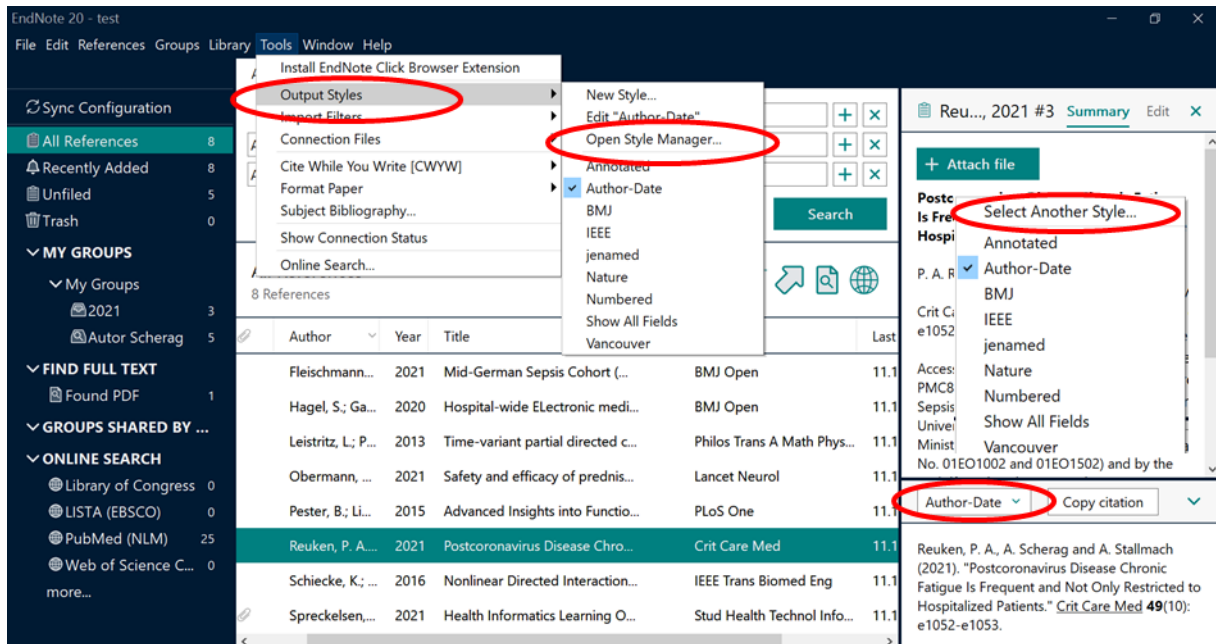
Author	Year	Title	R...	Journal
Fleischmann...	2021	Mid-German Sepsis Cohort (...)		BMJ Open
Hagel, S.; Ga...	2020	Hospital-wide Electronic medi...		BMJ Open
Obermann, ...	2021	Safety and efficacy of prednis...		Lancet Neurol
Reuken, P. A...	2021	Postcoronavirus Disease Chro...		Crit Care Med
Spreckelsen,...	2021	Health Informatics Learning O...		Stud Health Technol Info.

⇒ Eine Umbenennung der Gruppe erfolgt mittels der rechten Maustaste über **Rename Group**. Änderungen in den Auswahlkriterien für die Smart Group können mittels **Edit Group** vorgenommen werden.

**Aufgabe 2.4:** Es sollen die Aktivierung und Nutzung anderer EndNoteStyles betrachtet werden. Die Zitationsvorschrift für Dissertationsarbeiten an der Medizinischen Fakultät ist in einem speziellen EndNote Style umgesetzt worden. Dieser Style ist von der Webseite des Dekanats zu laden und in EndNote einzubinden.

#### Lösungsschritte:

⇒ Die Liste der Zeitschriften (Styles) erreicht man über das Pull-Down-Menü **Tools** und weiter **Open Style Manager**. Dort müssen aus der Vielzahl vorhandener Styles diejenigen aktiv ausgewählt werden, die man benutzen möchte. Im unteren Teil der Vorschau auf der rechten Seite wird dann die jeweilige Referenz in der vorgeschriebenen Zitationsvorschrift dieser Zeitschrift dargestellt. Hier ist der aktive Style (**current style**) sichtbar und kann gewechselt werden. Auch hier ist über **select another style** die Auswahl noch nicht aktivierter Styles möglich.



⇒ Der benötigte Style-File *jenamed.ens* kann von der Webseite des Dekanats <https://www.uni-klinikum-jena.de/Uniklinikum+Jena/Medizinische+Fakult%C3%A4t/Akademische+Verfahren/Promotion/Gestaltungshinweise.html> über den Link **Filter für Literatur (EndNote u. a.)** heruntergeladen werden. Die Datei *jenamed.ens* muss in das Unterverzeichnis *Styles* in Ihrem EndNote-Verzeichnis gespeichert werden! Bei den meisten Systemen handelt es sich dabei um das Verzeichnis: *C:\Programme\EndNote\*\Styles*. Wenden Sie sich bitte im Zweifelsfall an Ihren Systemadministrator.

⇒ Der so eingebundene Style *jenamed* muss nun analog zum ersten Lösungsschritt eingebunden werden (Pull-Down-Menü **Current Style** und weiter **Select Another Style** – *jenamed* aktivieren).

**Aufgabe 2.5:** Nutzung von EndNote bei der Erstellung eines Artikels in MS Word. Als Aufgabe wird das Einfügen von Zitationen aus EndNote vorgegeben.

**Lösungsschritte:**

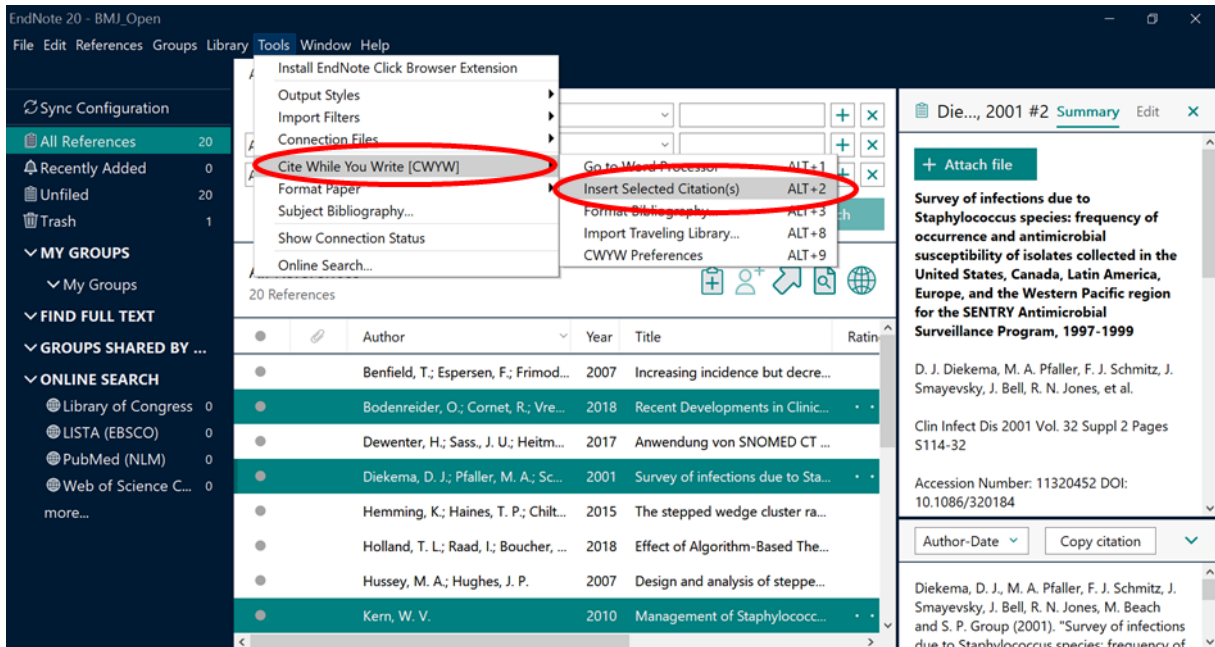
⇒ Im Textverarbeitungsprogramm MS Word wird das Dokument *BMJ\_open.docx* (Beispieltext einer konkreten Publikation) geöffnet. In den schon vorbereiteten Text sollen jetzt die Literaturstellen aus der EndNote Datenbank *BMJ\_Open.enl* eingesetzt werden.

⇒ Dazu wird durch die Menüfunktionen im Word **EndNote 20** ⇒ **GoToEndNote** das Literaturdatenbankprogramm EndNote geöffnet und die Datenbank *BMJ\_Open.enl* ausgewählt. EndNote kann über die Minimalisierungstaste als Schalter in die untere Bedienleiste verlegt werden. Durch Anklicken wird EndNote wieder auf den Bildschirm geholt. So kann zwischen dem Textverarbeitungsprogramm und dem Literaturdatenbankprogramm gewechselt werden.

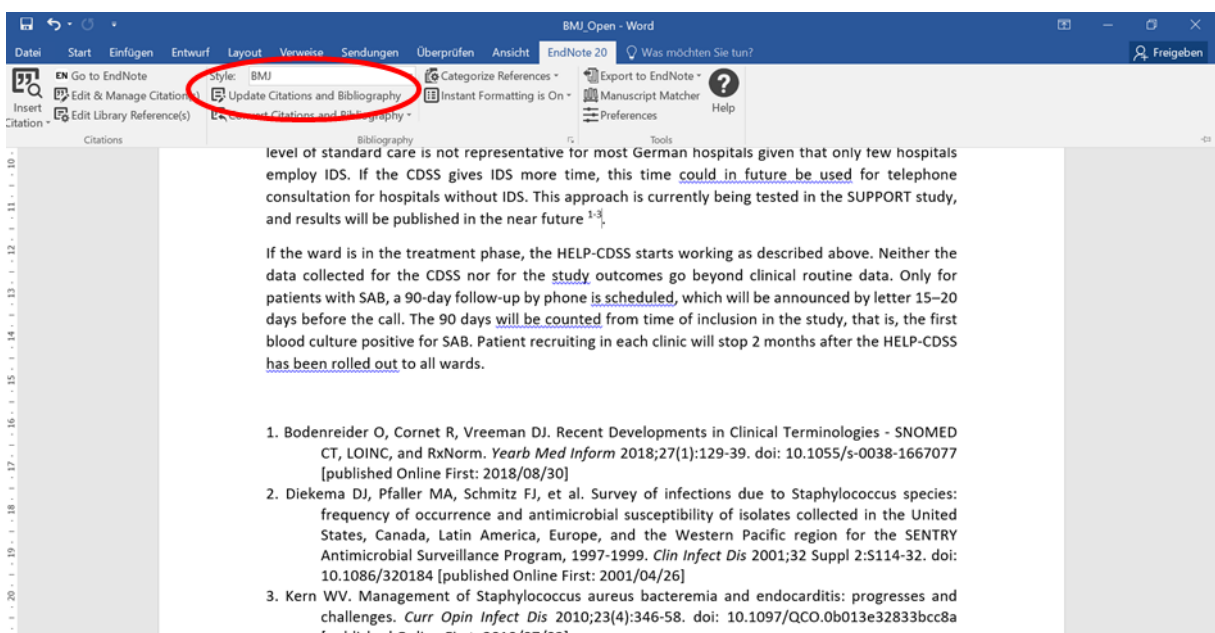
⇒ Nun wird im Wordtext die Stelle angeklickt, an der die Zitation eingefügt werden soll. Es wird zu **EndNote 20** (*BMJ\_Open.enl*) gewechselt und der gewünschte Datensatz (Referenz) markiert. Eine Mehrfachauswahl ist mittels der Strg-Taste möglich. Über die Menüpunkte **Tools** ⇒ **CiteWhileY-**



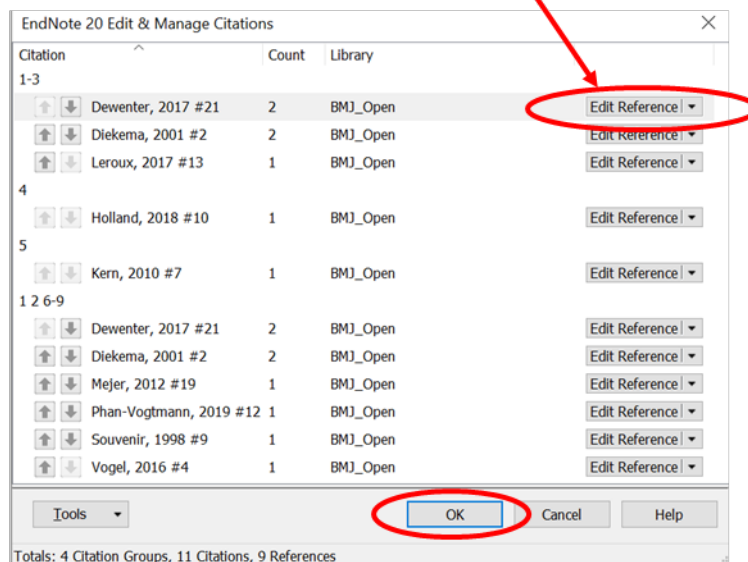
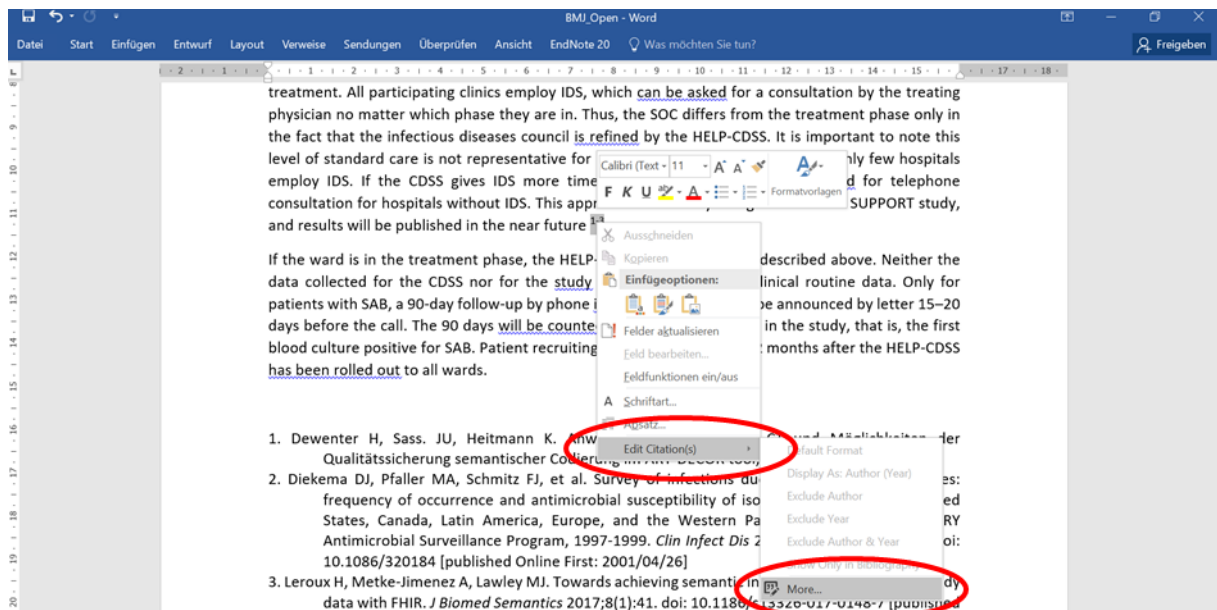
**ouWrite** ⇒ **InsertSelectedCitation(s)** bzw. die Tastenkombination **Alt+2** wird das Literaturzitat automatisch in den Wordtext integriert und der Datensatz an das Ende des Textes angehängt (Literaturverzeichnis). Klickt man die entsprechenden Stellen an, sind diese grau hinterlegt (Feldverweis). Das Format entspricht dem in Word für EndNote aktivierten bibliographischen Stil (**Current Style**) und damit den Zitationsvorschriften der ausgewählten Zeitschrift.



⇒ Mit der Änderung der Zeitschrift bzw. des Stils ändern sich sowohl der Zitationsstil im Wordtext als auch die Formatierung und Sortierordnung der bibliographischen Angaben des Artikels im Literaturverzeichnis am Ende des Textes. Diese Formatierung kann auch aktiv mittels **EndNote 20** ⇒ **Update Citations and Bibliography** aktualisiert werden.



- ⇒ Bei Löschung der Zitation im Text oder bei Veränderung der Literaturliste (Sortieren) wird wechselseitig eine automatische Korrektur bzw. Aktualisierung vorgenommen, so dass Zitationen im Text und im Literaturverzeichnis immer aufeinander abgestimmt und damit konsistent sind.
- ⇒ Die saubere Löschung von Zitationen im Text erfolgt dadurch, dass man durch einen Rechtsklick auf die zu löschende(n) Zitation(en) (grau hinterlegt) das Kontextmenü mit dem Punkt **Edit Citation(s)** – **More...** aufruft. Mit dem Schalter **Edit Reference** können Referenzen entfernt bzw. eingefügt werden. Der Vorgang muss mit **OK** bestätigt werden.



## Anlage 2.1: Beispieltext einer konkreten Publikation

### **Hospital-wide eElectronic medical record evaluated computerised decision support system to improve outcomes of patients with staphylococcal bloodstream infection (HELP): study protocol for a multicentre stepped-wedge cluster randomised trial**

Stefan Hage<sup>1</sup>, Julia Gantner<sup>2</sup>, Cord Spreckelsen<sup>2</sup>, Claudia Fischer<sup>2</sup>, Danny Ammon<sup>3</sup>, Kutaiba Saleh<sup>3</sup>, Lo An Phan-Vogtmann<sup>2</sup>, Andrew Heide<sup>3</sup>, Susanne Müller<sup>2</sup>, Alexander Helhorn<sup>3</sup>, Henner Kruse<sup>3</sup>, Eric Thomas<sup>3</sup>, Florian Reißner<sup>4</sup>, Silke Haferkamp<sup>5</sup>, Jens Vorwerk<sup>5</sup>, Saskia Deffge<sup>6</sup>, Marc Fabian Juzek-Küpper<sup>7</sup>, Norman Lippmann<sup>8</sup>, Christoph Lübbert<sup>9</sup>, Henning Trawinski<sup>9</sup>, Sebastian Wendt<sup>9</sup>, Thomas Wendt<sup>10</sup>, Andreas Dürschmid<sup>10</sup>, Margarethe Konik<sup>11</sup>, Stefan Moritz<sup>12</sup>, Daniel Tiller<sup>13</sup>, Rainer Röhrig<sup>14</sup>, Jonas Schulte-Coerne<sup>15</sup>, Jonas Fortmann<sup>14</sup>, Stephan Jonas<sup>15</sup>, Oliver Witzke<sup>16</sup>, Peter-Michael Rath<sup>17</sup>, Mathias W Pletz<sup>1</sup>, André Scherag<sup>2</sup>

#### **ABSTRACT**

**Introduction** Staphylococci are the most commonly identified pathogens in bloodstream infections. Identification of *Staphylococcus aureus* in blood culture (SAB) requires a prompt and adequate clinical management. The detection of coagulase-negative staphylococci (CoNS), however, corresponds to contamination in about 75% of the cases. Nevertheless, antibiotic therapy is often initiated, which contributes to the risk of drug-related side effects. We developed a computerised clinical decision support system (HELP-CDSS) that assists physicians with an appropriate management of patients with *Staphylococcus bacteraemia*. The CDSS is evaluated using data of the Data Integration Centers (DIC) established at each clinic. DICs transform heterogeneous primary clinical data into an interoperable format, and the HELP-CDSS displays information according to current best evidence in bacteraemia treatment. The overall aim of the HELP-CDSS is a safe but more efficient allocation of infectious diseases specialists and an improved adherence to established guidelines in the treatment of SAB.

**Methods and analysis** The study is conducted at five German university hospitals and is designed as a stepped-wedge cluster randomised trial. Over the duration of 18 months, 135 wards will change from a control period to the intervention period in a randomised stepwise sequence. The coprimary outcomes are hospital mortality for all patients to establish safety, the 90-day disease reoccurrence-free survival for patients with SAB and the cumulative vancomycin use for patients with CoNS bacteraemia. We will use a closed, hierarchical testing procedure and generalised linear mixed modelling to test for non-inferiority of the CDSS regarding hospital mortality and 90-day disease reoccurrence-free survival and for superiority of the HELP-CDSS regarding cumulative vancomycin use.

**Ethics and dissemination** The study is approved by the ethics committee of Jena University Hospital and will start at each centre after local approval. Results will be published in a peer-reviewed journal and presented at scientific conferences.

**Trial registration number** DRKS00014320.

#### **INTRODUCTION**

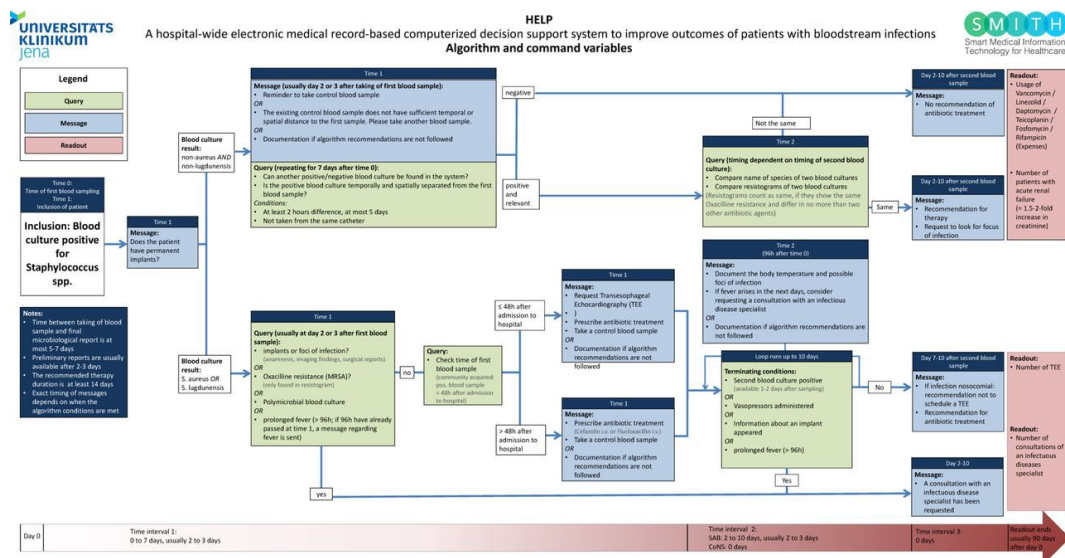
Staphylococci are the most commonly identified pathogens in both hospital-acquired and community-onset bloodstream infections. Coagulase-negative staphylococci (CoNS), followed by *Staphylococcus aureus*, are most commonly detected. *S. aureus* bloodstream (SAB) infection is a serious medical condition. If treated insufficiently, mortality rates can reach up to 40% and recurrences are frequent. Outcomes of SAB can be improved by strict adherence to treatment guidelines, including choice of appropriate antibiotic agent. In CoNS bloodstream infection, the clinical significance (except for *S. lugdunensis*, as this pathogen is associated with the same pathogenicity as *S. aureus*), however, is less clear. CoNS are part of the normal skin flora and identification in blood cultures corresponds to contamination in up to 80% of the cases. Nevertheless, in clinical practice, antibiotic therapy is often initiated, which in turn fosters the development of antibiotic resistance, increases the risk of drug-related side effects and the cost of therapy. Souvenir et al, for example, reported the use of antibiotics to treat 41% of patients with false-positive blood cultures, with vancomycin used for 83% of the treated pseudo-bacteraemic patients.

The use of algorithm-based therapy can improve the treatment of patients with staphylococcal bacteraemia as recently shown by Holland et al. Here, we examine the effects of a computerised clinical decision support system (CDSS), abbreviated HELP-CDSS, that supports attending physicians of patients with staphylococcal bacteraemia with regard to the implementation of best practice recommendations. This includes recommendations regarding follow-up blood cultures (FUBCs), early source control, early intravenous cloxacillin or cefazolin for methicillin-susceptible isolates and appropriate duration of therapy in patients with *S. aureus* or *S. lugdunensis* bloodstream infection. In patients with CoNS bloodstream infection, recommendations will particularly encourage adherence to FUBCs to differentiate whether the CoNS bacteraemia reflects true infection or contamination only. The HELP-CDSS is evaluated using data that are part of clinical routine documentation. It aims at a more efficient allocation of infectious diseases specialists (IDS) without worsening patient-related outcomes such as mortality and relapse rates and also aims at improving adherence to best practice recommendations in treatment of blood stream infections. The HELP-CDSS operates on the structure of so-called 'Data Integration Centers' (DIC), which will be established at each study site by means of the 'Smart Medical Information Technology for Healthcare' (SMITH) consortium, which is one of four consortia funded by the 'Medical Informatics Initiative' of the German Federal Ministry of Education and Research. Two of the main goals of this initiative are to establish interoperability standards for medical data and to enable the integration of heterogeneous clinical routine data to make it available for clinical and scientific use in order to improve patient care.

**METHODS AND ANALYSIS**

**HELP-computerised clinical decision support system**

A simplified schematic representation of the HELP-CDSS is presented in figure 1. The HELP-CDSS assigns cases to one of two arms: one for CoNS bloodstream infection and one for *S. aureus* and *S. lugdunensis* bloodstream infection. In the CoNS arm, it is first checked whether there are two independent blood cultures available (drawn at different time points or different loci). If there is only one blood culture, the CDSS advises to take a FUBC before starting the antibiotic therapy to rule out a contamination. If there are two positive blood cultures, and they are positive for the same CoNS (ie, species and antimicrobial susceptibility testing), the CDSS recommends checking for clinical significance of the result, including identification of a possible source of infection. It also recommends starting an antibiotic therapy if there are signs for a CoNS infection. If the FUBC is negative (ie, no detection of CoNS or a different CoNS species), the CDSS informs the physician about the high probability that the first blood culture, which is positive for CoNS, represents a contamination and recommends reconsidering the necessity of an antibiotic treatment (if already initiated) or discourages starting a therapy, respectively.



**Figure 1** Schematic representation of the HELP-CDSS. CDSS, computerised clinical decision support system.

In the case of *S. aureus* or *S. lugdunensis* bloodstream infection, the CDSS gives the following recommendations:

- FUBCs should be drawn at intervals of 2–4 days until negative conversion.
- Early source identification and source control.
- Early use of beta-lactams for methicillin-susceptible infections (within 24 hours of culture).
- Treatment duration of at least 14 days for uncomplicated and at least 28 days for complicated SAB.
- Perform transoesophageal echocardiogram in complicated SAB (intravenous drug abuse, persistent bacteraemia, embolisation, unknown source of infection, community-acquired SAB, presence of prosthetic valves or cardiovascular implantable devices).

In case of indications for a complicated course (eg, endocarditis, positive FUBCs or prolonged fever), a request for an infectious diseases consultation should be initiated.

Necessary patient-related information on which the CDSS is evaluated will be extracted from different primary data sources of the clinical routine (patient data management system, laboratory information system and so on). These data will be represented in an interoperable syntactic format (HL7 FHIR) and semantically annotated with standard terminologies (LOINC, SNOMED CT and so on). Annotated HL7 FHIR resources will be stored in a 'Health Data Storage' of the respective DIC, enabling analytical methods to query all data based on internationally consented code systems or value sets.

Results of the modelling of relevant information, semantic annotation and the application of HL7 FHIR have been made publicly available through tools such as ART-DECOR or Simplifier.net, corresponding to processes carried out by standards developing organisations to develop interoperability standards. They can be accessed at <https://art-decor.org/art-decor/decor-datasets--help->.

### Study design

The study is designed as a stepped-wedge cluster randomised trial (SW-CRT) with a preceding pilot phase and will be conducted at five German university hospitals, namely the hospitals of Jena (JUH), Leipzig, Aachen, Halle and Essen with a total of 135 participating wards. At the beginning of the study, which is scheduled for September 2019, the piloting phase starts. In this phase, the data extraction and technical implementation will be tested. When all technical aspects are functioning, the actual SW-CRT phase begins. According to the SW-CRT design, all wards start in the control phase (standard of care, SOC) and enter the treatment phase (application of the HELP-CDSS) in a stepwise fashion. The time of cross-over from control to treatment phase is assigned to the wards by stratified randomisation. A schematic overview of the stepped-wedge design is provided in figure 2. There will be nine randomisation steps with 2 months in between each step, and 15 clusters will enter the treatment phase per step. Including a 2-month control phase when no ward receives the algorithm, this results in a study duration of 20 months plus about 3 months of follow-up. In the participating wards, all adult patients with a blood culture positive for *S. aureus*/*S. lugdunensis* or CoNS are included in the study. If the ward is in the control phase when the patient is included, they get the SOC treatment. All participating clinics employ IDS, which can be asked for a consultation by the treating physician no matter which phase they are in. Thus, the SOC differs from the treatment phase only in the fact that the infectious diseases council is refined by the HELP-CDSS. It is important to note this level of standard care is not representative for most German hospitals given that only few hospitals employ IDS. If the CDSS gives IDS more time, this time could in future be used for telephone consultation for hospitals without IDS. This approach is currently being tested in the SUPPORT study, and results will be published in the near future.

If the ward is in the treatment phase, the HELP-CDSS starts working as described above. Neither the data collected for the CDSS nor for the study outcomes go beyond clinical routine data. Only for patients with SAB, a 90-day follow-up by phone is scheduled, which will be announced by letter 15–20 days before the call. The 90 days will be counted from time of inclusion in the study, that is, the first blood culture positive for SAB. Patient recruiting in each clinic will stop 2 months after the HELP-CDSS has been rolled out to all wards. ...





### 3 Praktischer Datenschutz

Die Rolle elektronischer Dokumente bzw. digitaler Daten in der Medizin nimmt stetig zu. Auch Sie werden bei Ihren ärztlichen Tätigkeiten oder in der Forschung mit Daten umgehen. Viele dieser Daten beziehen sich erkennbar auf Personen bzw. Patienten. Solche Daten sind besonders schützenswert. **Es ist Ihre Verantwortung, mit solchen Daten so vorsichtig umzugehen, dass diese nicht unbefugt genutzt oder verändert werden.**

#### *Hintergrundwissen: Vorbemerkung zum rechtlichen Rahmen*

Zum Datenschutz existieren strenge **gesetzliche Regelungen**. Diese sind **nicht** Gegenstand dieses Kurses.

Vorab ist allerdings wichtig: Die für den Umgang mit personenbezogenen Daten einschlägige europäische Gesetzgebung (**EU Datenschutzgrundverordnung, DS-GVO**) stellt fest: Die Speicherung und Verarbeitung personenbezogener Daten ist **grundsätzlich verboten – es sei denn, Sie können belegen, dass Ihnen der Umgang mit diesen Daten erlaubt ist**. Hierzu müssen Sie a) die Gesetze und Normen nennen, die die Datenverarbeitung in Ihrem Falle ausdrücklich erlauben, b) oft auch eine Einverständniserklärung der Betroffenen vorweisen können, c) den/die zuständige Beauftragte/n für den Datenschutz informiert und in die Datenschutzkonzeption einbezogen haben, d) eine sogenannte Datenschutzfolgen-Abschätzung aufgestellt und dokumentiert haben und e) geeignete Maßnahmen für Datenschutz und Datensicherheit konzipiert und umgesetzt haben. Hier ist also erheblicher Aufwand erforderlich.

Dieser Aufwand ist auch dann nötig, wenn für Namen (oder andere direkt eine Person identifizierenden Daten) Ersatzdaten (**Pseudonyme**) eingesetzt wurden (z. B. Patientennummern). Solange – z. B. durch eine Tabelle, in der echte Namen zu den Pseudonymen eingetragen sind – die Möglichkeit besteht, den Bezug zu den Personen zu rekonstruieren, gelten alle strengen datenschutzrechtlichen Bestimmungen. Das gilt selbst dann, wenn Sie selbst die Pseudonyme nicht rückbeziehen können.

Erst Daten, aus denen der Personenbezug gründlich und unumkehrbar gelöscht wurde (**Anonymisierung**), können frei genutzt werden. Beispiel: Mittelwerte, die aus vielen Datensätzen berechnet wurden, sind anonym und dürfen z.B. in einer Publikation erscheinen.

Dass Sie sich rechtlich absichern, wenn Sie mit (Patienten-)Daten umgehen, ist das eine. Mindestens ebenso wichtig ist aber, dass Sie **tatsächlich dazu in der Lage sind, Daten zu schützen**. Hierzu gibt Ihnen der Kurs wichtige grundlegende Hilfen und Hinweise.

#### *Lernziele*

Die folgenden Hauptpunkte stehen auf dem Programm. Nach diesem Kurs können Sie ...

- 1) Einen Passwort-Manager nutzen
- 2) Daten auf Datenträgern (Festplatten, USB-Stick, o. ä.) verschlüsseln
- 3) Den Nachrichten-/Datenaustausch mit Public-Private-Key-Verfahren absichern
- 4) Sicherstellen, dass unautorisierte Änderungen an Daten entdeckt werden können (Digitale Signatur).

### *Hintergrundwissen: Nutzen Sie starke Passwörter und einen Passwort-Manager*

#### *Warum Passwörter so problematisch sind*

Schlechte bzw. schlecht gesicherte **Passwörter sind eine der Hauptquellen für Datenschutzprobleme**. Nutzen Sie niemals einfach zu ratende Passwörter! Nutzen Sie niemals dasselbe Passwort für viele Internetdienste (immer wieder kommt es vor, dass einer dieser Dienste Ihr Passwort schlecht beaufsichtigt – wenn Sie verschiedene Passwörter verwenden, sind wenigstens die anderen Dienste nicht betroffen). Achtung: Verschicken Sie **Passwörter niemals lesbar in unverschlüsselten E-Mails** – dadurch teilen Sie Ihre Passwörter einer breiten und interessierten Öffentlichkeit mit.

Inzwischen gibt es detaillierte Erkenntnisse darüber, welche Passwörter wir uns gern ausdenken. Und diese Kenntnisse sind offen für die Nutzung durch Hacker – und werden gern und häufig genutzt: Es existieren riesige Passwortsammlungen zum Durchprobieren. In diesen befinden sich Passwörter, die zuvor veruntreut oder gestohlen wurden. Beispiele bekannter Passwörter sind: Ein beliebiges **Wort**, das es (in einem Wörterbuch) gibt oder – **ganz schlimm** – **Namen**. Auch **nicht gut**: Ein **Wort gefolgt von 1 - 2 Ziffern**. Machen Sie sich klar: Wenn keine Maßnahmen zur Verzögerung/Beschränkung der Eingabe ergriffen werden, ist es überhaupt kein Problem, die „wenigen“ (ca. 200.000) Wörter einer normalen Sprache digital rasend schnell durchzuprobieren. Und zwei beliebige Ziffern am Ende machen daraus auch nur 100-mal so viele Möglichkeiten – immer noch kein technisches Problem. In den verfügbaren Passwortsammlungen, die regelmäßig durchprobiert werden, sind auch **alle einschlägigen Zitate** enthalten.

#### *Praktischer Hinweis: Eine einigermaßen akzeptable Methode, Passwörter zu erfinden*

Ein etwas besseres Schema ist, unter Beibehaltung der Groß-/Klein-Schreibung immer nur die Anfangsbuchstaben der Wörter eines längeren Satzes zu verwenden, der eine Zahlenangabe enthält und mit ein paar Sonderzeichen zu garnieren. **Verwenden Sie dazu unbedingt einen Satz, den Sie sich selbst ausgedacht haben (kein Zitat!)**. Beispiel: „Vor meinem Fester stehen zwei große, grüne Bäume und sonnen sich“ daraus wird: **VmFs2ggBuss**.

#### *Starke Passwörter*

**Wirklich starke Passwörter** erkennen Sie daran, **dass Sie sich diese zunächst überhaupt nicht merken können** (ein Beispiel: mYONA2s6d5aTwiefb). Es sieht so aus, als wären solche Passwörter deswegen nicht nutzbar. Damit unterschätzen Sie sich aber. **Sie sind in der Lage dazu**, sich ein solches Passwort zu merken. Ein **Trick**: Verwenden Sie dieses Passwort zum Login auf Ihrem Rechner. Stellen Sie die Zeit für die Bildschirmsperre für einen Tag auf ein sehr kurzes Intervall (20 sec), schreiben Sie Ihr Passwort sicherheitshalber auf einen Zettel, den Sie umdrehen, aber nicht aus den Augen lassen, und arbeiten Sie an dem Tag möglichst oft, aber mit kurzen Unterbrechungen an Ihrem Rechner.

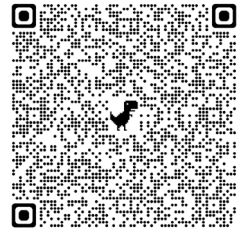
#### *Papier-Merkliste*

Das Bundesamt für Sicherheit in der Informationstechnologie (BSI) gibt einen sinnvollen **Tipps zum Notieren von Passwörtern**. Schreiben Sie **auf Papier** eine Tabelle mit drei Spalten: In Spalte 1 notieren Sie jeweils, um welche Anmeldung es geht, Spalte 2 enthält dann den Anmeldenamen dafür und Spalte 3 ein individuelles Passwort. Dieses Passwort ist aber immer nur die **zweite Hälfte des wirklichen Passworts**. Sie merken sich zusätzlich ein hinreichend schweres Passwort (**im Kopf!**). An dieses Passwort hängen Sie dann jeweils das individuelle Passwort aus der Tabelle an. Also: Ein **Grundpasswort im Kopf** und individuelle **Passwortfortsetzungen in der Tabelle**. Fällt die Liste Unbefugten in die Hand, nutzt sie nichts ohne das auswendig gelernte Grundpasswort.



Hier der Tipp des BSI (inkl. Vorlage für die Tabelle):

[\[https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-Verbraucher/Informationen-und-Empfehlungen/Cyber-Sicherheitsempfehlungen/Accountschutz/Sichere-Passwoerter-erstellen/Umgang-mit-Passwoertern/umgang-mit-passwoertern\\_node.html\]](https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-Verbraucher/Informationen-und-Empfehlungen/Cyber-Sicherheitsempfehlungen/Accountschutz/Sichere-Passwoerter-erstellen/Umgang-mit-Passwoertern/umgang-mit-passwoertern_node.html)



### Passwortmanager

Mit einem starken Passwort gesichert können Sie auch viele weitere Passwörter in einem Passwortmanager verwalten.

Es gibt kleinere Unterschiede in der Bedienung je nach Version und Betriebssystem.

**Aufgabe 3.1:** *Einrichten einer Passwort-sammlung. Richten Sie eine neue Passwortsammlung in KeePass ein. Sichern Sie diese mit einem starken Passwort. Tragen Sie neue Passwörter in Ihre Passwortsammlung ein. Lassen Sie sich von KeePass starke Passwörter vorschlagen.*

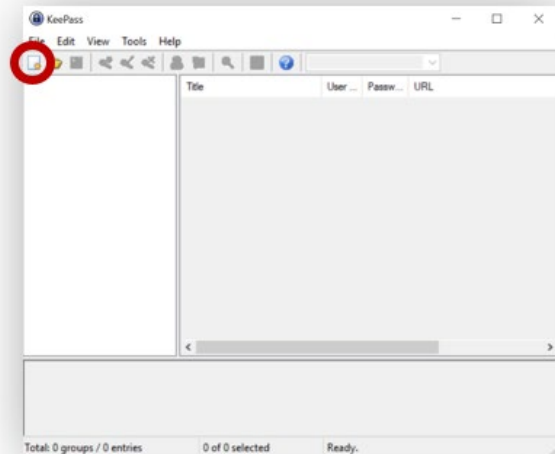


Abbildung 1: Startbildschirm von KeePass

### Lösungsschritte: KeePass einrichten

- 1) Starten Sie **KeePass**. Legen Sie als erstes eine **neue Passwortsammlung** an (Klick auf das Symbol für eine neue Datei oben links). Geben Sie anschließend ein **neues Passwort** ein.
- 2) **Wiederholen Sie das Passwort** noch einmal in der nachfolgenden Eingabemaske.
- 3) **Speichern** Sie dann als erstes die neu erstellte, leere Passwortsammlung (Diskettensymbol).
- 4) Sorgen Sie nun dafür, dass Ihre Passwortsammlung Passwörter enthält: Klicken Sie auf das Symbol für einen **neuen Eintrag** und füllen Sie dann die Eingabemaske so aus wie im Screenshot (Abbildung 4) gezeigt.

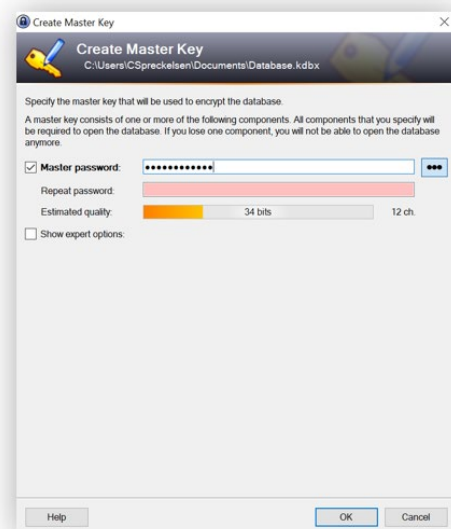


Abbildung 2: Festlegen des Masterpassworts

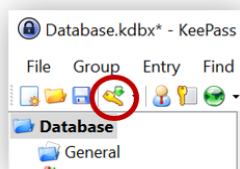


Abbildung 3: Eingabe eines neuen Eintrags starten

5) KeePass erzeugt Ihnen **zufallsgesteuert starke Passwörter**. Durch Klick auf den Button mit den drei Punkten können Sie sich das Passwort im Klartext anzeigen lassen. Durch Klick auf das Schlüsselsymbol können Sie sich ein neues Passwort erzeugen lassen.

Natürlich können Sie auch ein **eigenes Passwort** eintragen (dann bitte doppelt: in der Zeile Passwort und in der Zeile Repeat).

Achten Sie zu Übungszwecken darauf, dass Sie das Feld für die Web-Adresse (URL) mit einer URL füllen, die sich im Browser öffnen lässt. Verwenden Sie aber bitte keine echten persönlichen Zugangsdaten.

Wenn Sie mit dem Eintrag Ihres neuen Passworts fertig sind, speichern Sie am besten die ganze Sammlung (wie zu Beginn).

### **Aufgabe 3.2: KeePass nutzen**

*Öffnen Sie eine Webseite direkt aus KeePass heraus. Überzeugen Sie sich davon, dass KeePass Zugangsdaten automatisch in Login-Fenster eintragen kann: Lassen Sie hier zu Übungszwecken KeePass die Zugangsdaten in eine Word-Datei schreiben.*

### **Lösungsschritte:**

Mit dem Testeintrag können Sie nun die (recht praktische) Nutzung des Passwort-Managers üben:

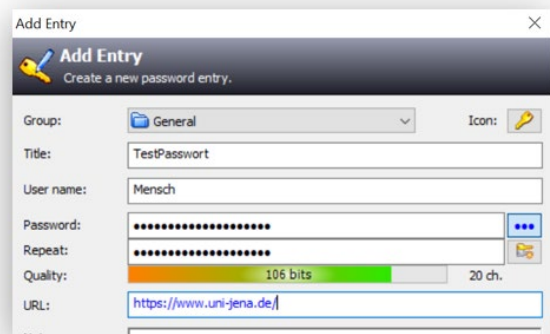


Abbildung 2: Anzeigen des Passworts im Klartext und Start der automatischen Erzeugung eines zufälligen Passworts

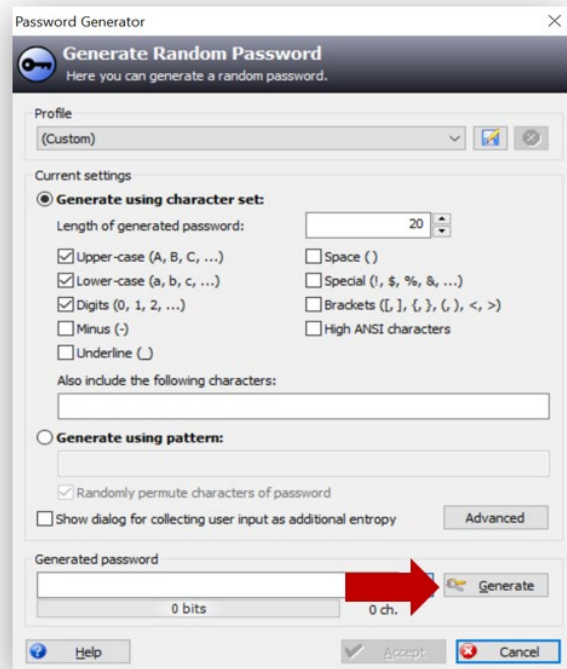


Abbildung 3: Abschluss der automatischen Passwörterzeugung

1. Wählen Sie mit einem Einfachklick Ihren neuen Eintrag aus. Öffnen Sie dann mit einem Rechtsklick das Kontextmenü. Wählen Sie „**Open URL(s)**“ aus. (Statt mit dem Kontextmenü zu arbeiten, können sie auch einfach die Tastenkombination „Strg“ und „U“ verwenden).
2. Normalerweise tragen Sie genau die Seite ein, auf der sich das Login-Fenster für das Passwort befindet. Anschließend würden Sie wieder Im Kontextmenü „**Perform Auto-Type**“ anklicken (oder einfach die Tasten „Strg“ + „V“ kombinieren) und KeePass würde Sie **automatisch einloggen**.

**Hinweis:** Hier haben wir ja keine gültigen Zugangsdaten hinterlegt. Sie können sich aber dennoch die Login-Daten ansehen, die „Perform Auto-Type“ erzeugt.

3. **Öffnen Sie einfach Microsoft Word.** Klicken Sie dort auf eine leere Seite. Gehen Sie dann **zurück zu KeePass** und starten „**Perform Auto-Type**“ im Kontextmenü (oder eben durch „Strg“ + „V“). In Ihrem Word-Dokument steht jetzt etwas wie „Mensch IOa4jwVdLVLX6ooL3lrm“, Die Zeichen hinter „Mensch“ werden anders sein, denn sie entsprechen dem automatisch von KeePass erzeugten Passwort.

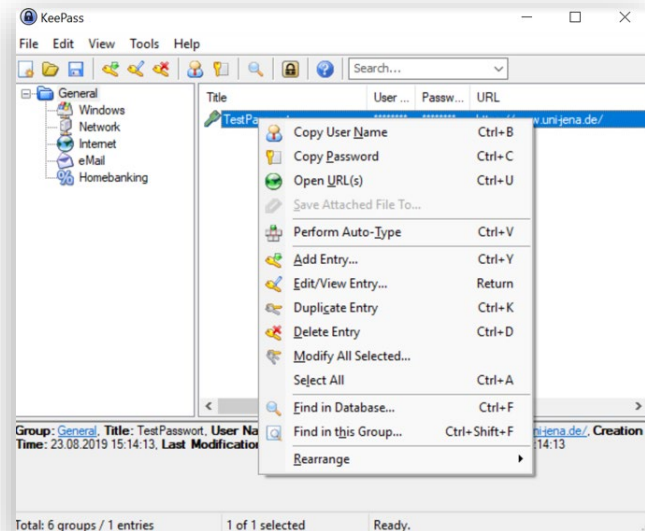


Abbildung 4: Automatisches Öffnen der Webseite, für die das Passwort hinterlegt ist

### Ergänzung: Zwei-Faktor-Authentifizierung

Als **Ergänzung von Passwörtern** dient die **Zwei-Faktor-Authentifizierung (2FA)**. Sie hilft, Online-Konten deutlich besser zu schützen – ist aber leider auch lästig. Man kann sich 2FA vorstellen, wie ein zusätzliches Sicherheitsschloss an der Tür. Normalerweise öffnen wir die Tür mit einem Schlüssel, aber hier fügen wir eine zweite Sicherheitsebene hinzu.

Mit 2FA benötigen wir zwei "Schlüssel", um auf ein Konto zuzugreifen. Der erste Schlüssel ist in der Regel das Passwort, das nur dem Kontoinhaber bekannt sein sollte. Der zweite Schlüssel ist eine Art Einmalcode. Er wird auf einem **vorgegebenen Extragerät** zur Verfügung gestellt – meist wird der Einmalcode auf das eigene Mobilgerät gesendet. Dieser Code ändert sich ständig und ist nur für kurze Zeit gültig. Wichtig hier: der **Besitz genau dieses einen Geräts** ist Voraussetzung dafür, den Schlüssel zu erhalten (**Vorsicht beim Wechsel** auf ein neues Smartphone! Erst das neue Gerät mit **Hilfe des alten Geräts** als neuen „2FA-Empfänger“ anmelden).

Bei der Anmeldung in einem Konto geben wir **zuerst das Passwort** ein (Schlüssel 1). Danach müssen wir den **Einmalcode** eingeben, den wir auf das Mobilgerät erhalten (Schlüssel 2). Dies macht es für Hacker viel schwieriger, auf Konten zuzugreifen, selbst wenn sie das Passwort kennen.

## Hintergrundwissen: Daten bzw. Datenträger verschlüsseln (Containerbasierte Verschlüsselung)

Eine recht sichere Methode, Daten vor unbefugtem Zugriff zu schützen, ist die containerbasierte Verschlüsselung mit VeraCrypt. VeraCrypt ist kostenlos erhältlich und quelloffen, d. h. Interessierte können sich ansehen, wie es programmiert wurde. Mit VeraCrypt erzeugen Sie zuerst besonderen Speicherplatz – den Container – und können dort Dateien hinterlegen. Solange der Container nicht mittels Passwort geöffnet wird, sind die Dateien darin nicht lesbar. Von außen sieht ein Container aus wie eine ganz gewöhnliche - allerdings nicht sinnvoll lesbare – Datei, die Sie auch kopieren oder verschicken können. Sie können auch Container so anlegen, dass ein kompletter Datenträger (Externe Festplatte oder USB-Stick) auf diese Weise Platz für verschlüsselte Daten bietet.

**Aufgabe 3.3:** Legen Sie den (fiktiven) Entlassungsbrief für Frau Erna Diestel in einem verschlüsselten VeraCrypt-Container ab. Verschieben Sie den Container im Dateisystem (z. B. zu Übungszwecken einfach in einen neuen Unterordner - aber das könnte im echten Leben natürlich auch ein USB-Stick sein). Öffnen Sie den Entlassungsbrief vom neuen Speicherort aus.

**Lösungsschritte:** Anlegen eines VeraCrypt Containers

1. Starten Sie VeraCrypt und klicken Sie auf „**Volume erstellen**“.
2. Lassen Sie die Auswahl wie sie ist („Eine verschlüsselte Containerdatei erstellen“) und klicken Sie einfach auf „**Weiter**“.
3. Auch die nächste Auswahl lassen Sie einfach, wie sie ist – und „**Weiter**“.
4. Teilen Sie dann VeraCrypt mit, **wo die Datei für den Container gespeichert werden soll**. Am einfachsten hier im Kurs unter „Dokumente“. Gebe Sie Ihrer Containerdatei einen beliebigen Namen. Sie können die Datei noch etwas besser verstecken, indem Sie so tun als wäre es z. B. eine Word-Datei. Dazu geben Sie einfach die Endung einer Word-Datei (.docx) mit ein z. B. „vctext.docx“. Schließen Sie diesen Schritt mit Klick auf „**Öffnen**“ ab.
5. Klicken Sie dann auf „**Weiter**“.

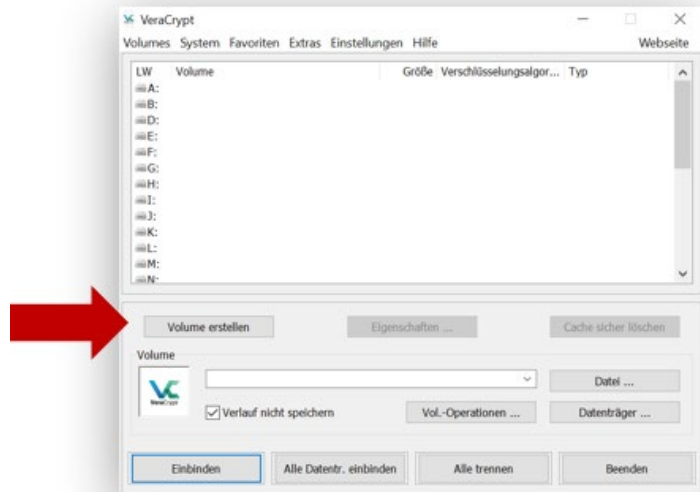


Abbildung 5: VeraCrypt: Start der Erzeugung eines Containers



Abbildung 6: Auswahl des Speicherorts

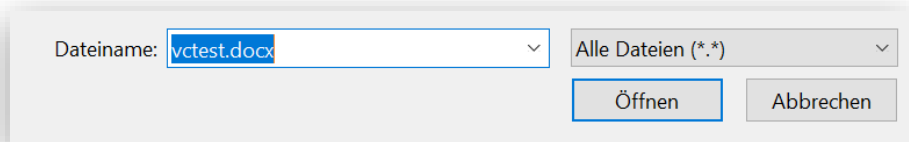


Abbildung 7: Containerdatei anlegen



Abbildung 8: Auswahl des Container-Typs



Abbildung 9: Auswahl der Verschlüsselung – kann in der Regel so bleiben, wie voreingestellt.

6. Auch die Auswahl der **Verschlüsselungseinstellungen** können Sie einfach **übernehmen** („Weiter“).
7. Als nächstes müssen Sie angeben, wieviel Platz Sie in ihrem Container brauchen. Bei der **Wahl der Größe** sollten Sie im Blick haben, ob Sie viele oder große Dateien verschlüsseln wollen (dann ggf. mehrere Gigabyte, GB eintragen) oder ob sie den Container häufig kopieren oder gar verschicken wollen (dann nur wenige Megabyte, MB). Die Container-Datei ist immer von Anfang an so groß, wie Sie es hier auswählen. **Hier im Kurs bitte einen kleinen Container anlegen**, sonst dauert es zu lange: Geben Sie bitte ein 2 MB.



Abbildung 10: Festlegen der Containergröße



- Anschließend wählen Sie ein neues, starkes **Passwort**. Ohne Kenntnis des Passworts kommen Sie nie mehr an die Daten. Nutzen Sie hier im Kurs gleich KeePass. Erzeugen Sie dort ein neues Passwort und kopieren Sie es hier in die Maske. Klicken Sie anschließend auf „Weiter“.

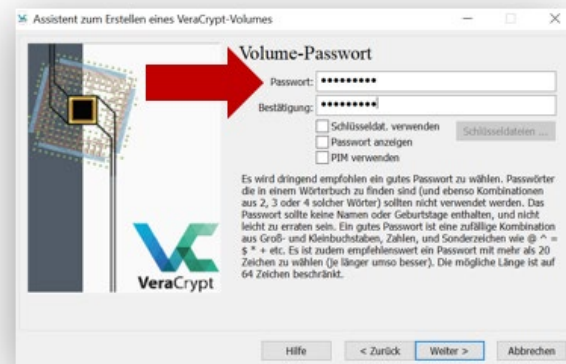


Abbildung 11: Passwordeingabe

- Als nächstes haben Sie Gelegenheit zu einem sportlichen **Fitnessstraining**: Wackeln und fahren Sie so lange mit der Maus auf dem Bildschirm hin und her, bis der Balken auf dem Bildschirm von roter Farbe nach Grün wechselt. Das Programm kann so dafür sorgen, dass die zum Verschlüsseln verwendeten (**Pseudo**-)Zufallszahlen **nicht reproduzierbar** sind.
- Klicken Sie abschließend auf „Formatieren“. Nun wird der Container erzeugt. Abschließend können Sie entweder noch einen Container (Volume) erstellen oder VeraCrypt beenden.

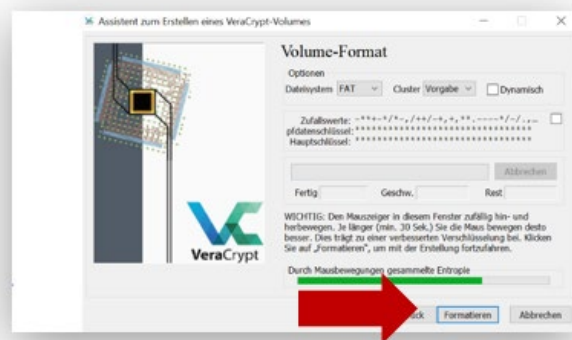


Abbildung 12: Sammeln von Entropie und Formatieren

Stellen Sie es sich bitte so vor: Auf der Festplatte Ihres Rechners ist nun eine Datei entstanden. Diese ist wie ein verschlossener, noch **leerer Container** und wartet darauf, geöffnet und befüllt zu werden.

**Lösungsschritte:** Öffnen (Einbinden bzw. Mounten) des Containers

- Wenn nicht schon geschehen gehen Sie zurück auf das Hauptfenster von VeraCrypt. Wählen Sie in der Mitte einen **Laufwerksbuchstaben** aus (anklicken).
- Nun müssen Sie Ihre Containerdatei auf dem Computer suchen und auswählen. Klicken Sie dazu auf „Datei“ und zuletzt dann auf „Einbinden“.

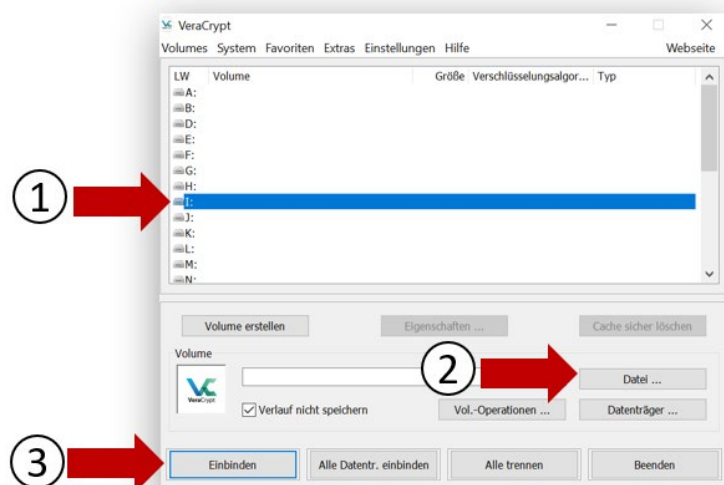


Abbildung 13: Einbinden des Containers



3. Sie werden aufgefordert, Ihr **Passwort** einzugeben. Tun Sie das. Falls das Passwort stimmt, öffnet VeraCrypt nach etwas Zeit den Container (dass das etwas dauert, ist ein Sicherheitsmerkmal von VeraCrypt, das gegen automatisches Durchprobieren von Passwörtern hilft). Wenn der Container eingebunden ist, erscheint hinter dem von Ihnen gewählten Laufwerksbuchstaben der volle Dateiname.
4. Was das gebracht hat, sehen Sie, wenn Sie den Windows-Explorer öffnen oder **einfach auf die blaue Zeile mit dem Laufwerk dop-**

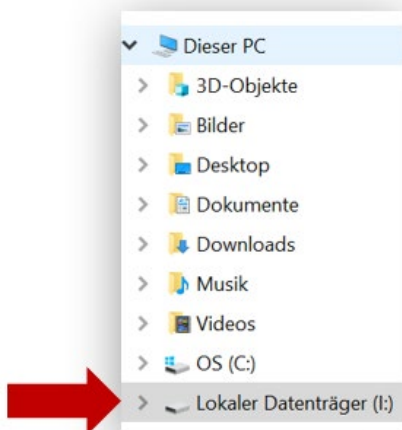


Abbildung 15: Fertig eingebundener Container: Kein Unterschied zu einem normalen Laufwerk

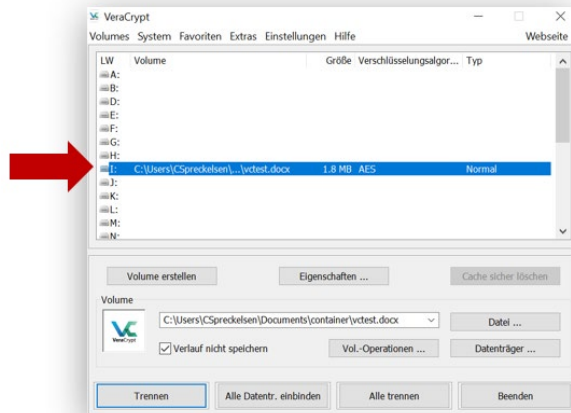


Abbildung 16: Öffnen des Containers

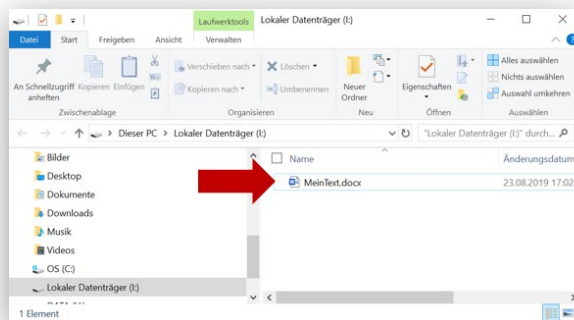


Abbildung 14: Geöffneter Container mit eigener Datei.

- pelklicken.** Im Dateisystem Ihres Rechners ist ein neues Laufwerk sichtbar (so als hätten Sie einen USB-Stick eingesteckt). Sie können damit alles machen, was Sie mit einem normalen Laufwerk (oder USB-Stick) auch machen können: Dateien anlegen, löschen, kopieren, verschieben etc. Sie können mit einem **Rechtsklick in das noch leere Verzeichnis** des neuen Laufwerks das Kontextmenü aufrufen und z. B. ein **neues Word Dokument** erzeugen. Legen Sie in dem Laufwerk den Entlassungsbrief ab.
5. Gehen Sie anschließend zurück zu VeraCrypt und klicken Sie auf „**Trennen**“. Ihr Container schließt sich nun, und das Laufwerk verschwindet aus Ihrem Dateisystem.
  6. Verschieben/kopieren Sie ihre Containerdatei an einen anderen Ort in Windows (z. B. Desktop). Merken Sie sich diesen Ort.
  7. Wiederholen Sie die Schritte zum **Einbinden** des Containers (s. o. 1. – 3.) **aber nutzen Sie dabei den verschobenen/kopierten Container**. So wird das Laufwerk wieder sichtbar und Sie finden darin auch Ihre Datei wieder.

**Anleitung: Container als Favoriten nutzen**

Sie können einige **Arbeitsschritte abkürzen** (nämlich die Auswahl des Laufwerksbuchstabens und das Suchen nach Ihrer Containerdatei), wenn Sie einen Container zu Ihren **Favoriten** hinzufügen. Achten Sie darauf, dass Sie zuvor Ihren Container eingebunden haben. Klicken Sie dann in VeraCrypt auf „**Favoriten**“ und wählen Sie „**Eingebundenes Volume zu den Favoriten hinzufügen**“

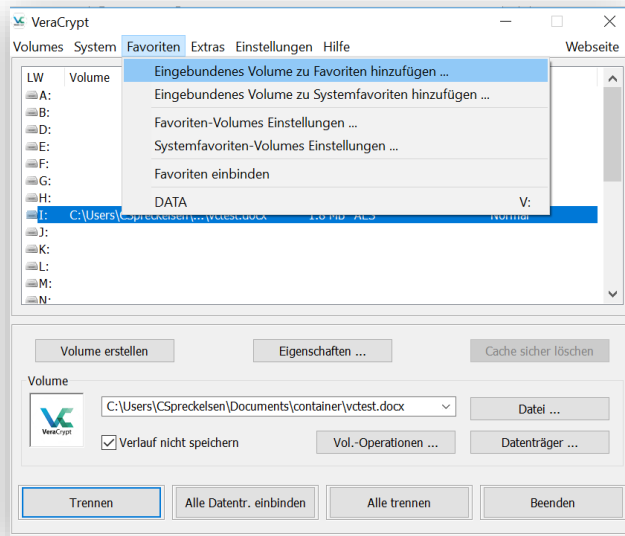


Abbildung 17: Eingebundenen Container zu Favoriten hinzufügen

Die folgende Maske bietet zusätzlich die folgende – sehr nützliche – Option: Sie können VeraCrypt dazu bringen, Sie **gleich bei der Anmeldung** am Rechner nach dem Passwort für den Container zu fragen und diesen dann **sofort einzubinden**. Auf diese Weise sind Ihre verschlüsselten Daten verfügbar, solange Sie arbeiten und verschlüsselt, wenn Sie den Rechner ausschalten.

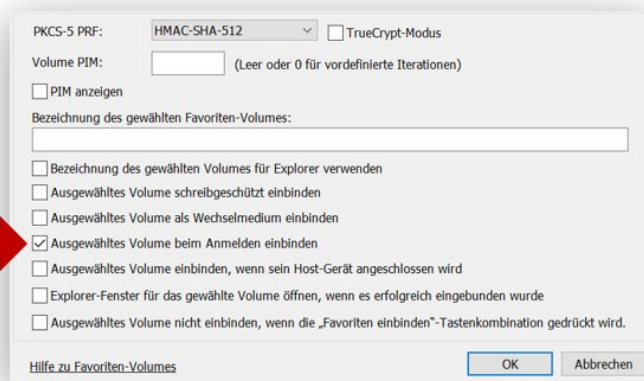


Abbildung 18: Favoriten immer beim Login einbinden

**Aufgabe 3.4:** *Richten Sie Ihren Container als Favoriten ein und sorgen Sie dafür, dass der Container gleich bei der Rechneranmeldung eingebunden werden kann. (Lösungsschritte: Siehe Anleitung oben).*

**Hintergrundwissen: Public-Private-Key Verfahren zur Absicherung der Kommunikation**

Zunächst eine grundsätzliche Warnung: Die übliche Nutzung von E-Mail ist aus Datenschutzsicht meist unakzeptabel und die wahrscheinlich **häufigste Ursache schwerwiegender Datenschutzverstöße**. Eine **normale E-Mail** ist nichts anderes als eine für alle Interessierten **jederzeit lesbare digitale Postkarte**. Der Aufwand, E-Mails abzufangen und zu lesen ist minimal. **Patientendaten mit ungesicherten E-Mails** zu verschicken, ist **völlig unakzeptabel**.

Die gute Nachricht: **Ein Schutz ist mit vertretbarem Aufwand möglich**.

Die zweite gute Nachricht: Anders als der E-Mail-Versand ist der Nachrichten-Austausch in Messenger-Diensten oft besser abgesichert. Wenn ein Messenger-Dienst eine Ende-Zu-Ende-Verschlüsselung anbietet, lassen sich Nachrichten nicht wie offene Postkarten lesen. Wie Ende-Zu-Ende-Verschlüsselung funktioniert, lernen Sie nun kennen.

### *Das Konzept der asymmetrischen Verschlüsselung*

Viele Jahrhunderte lang – bis in die 1970er Jahre – haben Menschen sich verschlüsselte Nachrichten geschickt, aber bei allen technischen Verbesserungen dasselbe Prinzip benutzt: **Symmetrische Verschlüsselung**. Symmetrisch heißt: Derselbe Weg/Mechanismus/Schlüssel führt von der Klartextnachricht zur verschlüsselten Nachricht, wie umgekehrt von der verschlüsselten Nachricht zum Klartext. Symmetrische Verschlüsselung hat ein **zentrales Problem**: Wie lässt sich der Schlüssel über Distanz sicher zum Empfänger/zur Empfängerin der Nachricht bringen, wenn diese/dieser nicht vor Ort ist. Weiteres symmetrisches Verschlüsseln des Schlüssels hebt das Problem nur auf eine höhere Ebene, ohne es zu lösen.

**Asymmetrische Verschlüsselung** ist eine geniale mathematische Erfindung. Hier werden Verschlüsselung und Entschlüsselung mit unterschiedlichen Schlüsseln durchgeführt. Diese **gehören immer exakt paarweise zusammen**: ein **Schlüsselpaar** aus einem „**Ver-Schlüssel**“ und einem genau zu diesem passenden „**Ent-Schlüssel**“. Mathematisch genial ist es, dass sich diese Paare so herstellen lassen, dass es praktisch unmöglich ist, aus dem einen Schlüssel des Paares den anderen zu rekonstruieren: Die Schlüssel gehören zwar paarweise zusammen, aber Kenntnis des einen hilft nicht weiter, um den anderen Schlüssel zu bekommen.

Wie löst die asymmetrische Verschlüsselung das Problem der Schlüsselverteilung? Einfach dadurch, dass für geheime Nachrichten an mich mein **Ver-Schlüssel allgemein bekannt** sein darf. Es gibt kein Problem mit der Geheimhaltung: Mit dem Ver-Schlüssel bekommt man eine Nachricht verschlüsselt, aber nicht wieder geöffnet. Solange ich der/die Einzige bin, der/die den zugehörigen Ent-Schlüssel besitzt, sind meine Nachrichten sicher. Wer möchte, verschlüsselt eine Nachricht für mich und schickt sie mir. Mein Ent-Schlüssel ist die einzige Möglichkeit, die Nachricht zu lesen. Es gibt also **den allgemein bekannten Schlüssel, den Public Key** (hier: der Ver-Schlüssel) und den jeweils **ausschließlich mir bekannten Schlüssel, den Private Key** (hier: der Ent-Schlüssel).

Damit mehrere Personen kommunizieren können, **muss jeder/jede ein eigenes Schlüsselpaar besitzen**. Alle behalten jeweils ihren Private Key für sich und veröffentlichen ihren Public Key.

**Hinweis:** Im Folgenden lernen Sie, wie Sie z. B. Ihre E-Mail-Kommunikation besser schützen können. Wichtig: Das hier verwendete Werkzeug „Mailvelope“ ist **nicht die beste Form der E-Mail-Absicherung**. Das liegt daran, dass Mailvelope als Plugin in den Web-Browser (hier: Firefox) eingebunden wird und diese Einbindung informationstechnisch eine Flickschusterei ist. Allerdings: Das Programm hat sich in letzter Zeit stark verbessert und ist eingehend auf Probleme überprüft worden. Für eine bessere Absicherung nutzen Sie ähnliche Techniken wie hier vorgestellt in einem auf dem Rechner installierten Mailprogramm (Outlook, Thunderbird o. ä.).

Mailvelope wird hier vorgestellt, weil es 1) bei Nutzung von E-Mail-Diensten über den Browser (Webmail) eine der wenigen überhaupt möglichen Schutzmaßnahmen ist, 2) besonders klar das Prinzip der **asymmetrischen bzw. Public-Private-Key-Verschlüsselung** demonstrieren lässt. Und dieses Prinzip ist die Basis für ein Verständnis sehr vieler wichtiger Datenschutzmaßnahmen.

**Aufgabe 3.5:** Erzeugen Sie mit Mailvelope ein Schlüssel-paar (Public-Private Key-Pair).

**Lösungsschritte:** Schlüssel-paare verwalten mit Mailvelope

1. Öffnen Sie den Browser. Oben rechts befindet sich das Icon zum Start von Mailvelope. Auf Klick öffnet sich ein Menü.
2. Klicken Sie als nächstes auf „Übersicht“. Hierdurch öffnet sich die Vollansicht von Mailvelope.

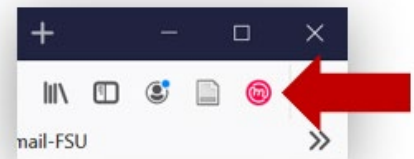


Abbildung 19: Mailvelope öffnen

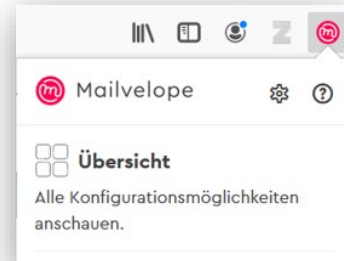


Abbildung 20: Geöffnetes Seitenmenü von Mailvelope im Browser

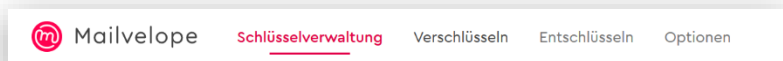


Abbildung 21: Hauptmenü mit Aufruf der Schlüsselverwaltung

**Hinweis:** Bei der allerersten Benutzung der Schlüsselverwaltung sieht diese wie im nächsten Screenshot aus. Der nächste Schritt ist es dann, einen Schlüssel zu erstellen. Dies ist nur ein Hinweis für den Fall, dass Sie Mailvelope frisch auf Ihrem Rechner installieren.

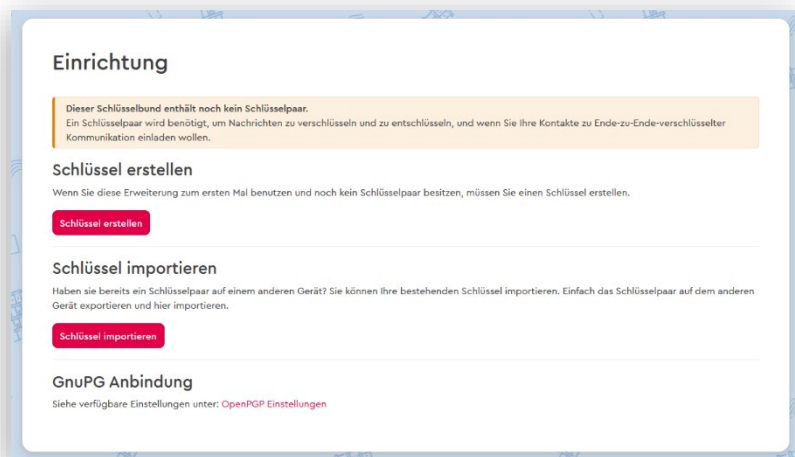


Abbildung 22: Konfigurationsansicht bei Erstinstallation von Mailvelope

Im Kurs ist Mailvelope bereits in Gebrauch gewesen. Daher zeigt sich dort die Schlüsselverwaltung in der Routineform.



Abbildung 23: Ansicht der Schlüsselverwaltung mit bereits existierenden Schlüssel-paaren

- Falls Sie im Computer-Pool arbeiten: Falls Sie bereits Schlüssel vorfinden, löschen Sie diese bitte als erstes (die Schlüssel stammen ggf. aus vorhergehenden Kursen). Fahren Sie dazu mit der Maus über die jeweilige Zeile mit dem Schlüssel und klicken Sie dann rechts auf das Mülleimersymbol.

**Hinweis:** Im Kurs arbeiten Sie nur zur Übung mit einem Schlüsselpaar. Sie dürfen das im Computer-Pool erzeugte Paar **nicht dauerhaft verwenden**. Wenn Sie ihre eigenen E-Mails verschlüsseln wollen, erzeugen Sie sich **später** einfach **eigene Schlüssel auf Ihrem eigenen Rechner**.

- Klicken Sie danach auf „Erstellen“ um sich ein eigenes Schlüsselpaar aus privatem und öffentlichem Schlüssel (Ent-Schlüssel und Ver-Schlüssel) zu erzeugen. Nutzen Sie an dieser Stelle zunächst eine frei erfundene E-Mail-Adresse (damit nicht unfreiwillig eine echte entsteht: Verwenden Sie nach einem frei gewählten Namen „@wbs-med.org“). **Und geben Sie sich einen frei erfundenen Namen**. Das **Passwort** Ihres Schlüsselpaars denken Sie sich bitte neu aus und **schreiben Sie es sich auf**. Bei Schlüsseln im echten Leben verwenden Sie bitte immer ein starkes Passwort!
- Im Kurs: **Entfernen Sie bitte das grüne Häkchen** vor „Öffentlichen Schlüssel zum Mailvelope Server hochladen“ (siehe Pfeil). Wenn Sie im Kurs regulär mit Mailvelope arbeiten ist es stattdessen durchaus sinnvoll, Ihren richtigen öffentlichen Schlüssel über den Mailvelope Server bekannt zu machen. Sie erhalten dann eine (verschlüsselte) Mail um Ihre (echte) E-Mail zu bestätigen. Hier im Kurs macht das keinen Sinn.
- Wenn Sie alle **Angaben zum Schlüsselpaar eingetragen** haben, klicken Sie abschließend auf „Erzeugen“. Dann sollte Ihr neues Schlüsselpaar in der Liste erscheinen.

Abbildung 24: Angaben zum Schlüsselpaar (bitte entfernen Sie im Kurs den grünen Haken vor „Öffentlichen Schlüssel ...“ – siehe Pfeil)

Abbildung 25: Neu eingetragenes Schlüsselpaar

**Aufgabe 3.6:** Nun können Sie ausprobieren, ob das Verschlüsseln in Mailvelope grundsätzlich funktioniert: Sie verschlüsseln einfach eine **Nachricht an sich selbst** und entschlüsseln diese wieder.

Abbildung 26: Text verschlüsseln

**Lösungsschritte:** Überprüfen Sie, ob die Verschlüsselung für Sie selbst funktioniert

1. Klicken Sie im **Hauptmenü auf „Verschlüsseln“**. Fangen Sie dann an, in der Empfängerzeile genau **die E-Mailadresse** einzugeben, **für die Sie eben den Schlüssel erzeugt** haben. Sobald der passende Eintrag aus Ihrer Schlüsselverwaltung angezeigt wird: Einfach durch Anklicken übernehmen.
2. Klicken Sie unten auf **„Möchtest Du auch noch einen Text verschlüsseln“** und tragen Sie dann Ihren Text ein. Klicken Sie zuletzt oben rechts auf **„Verschlüsseln“**.

Von Ihrem Text ist nun nichts mehr zu sehen. Er ist in den vielen Zeichen der sogenannten PGP-Mesage (PGP: Pretty Good Privacy) versteckt.

3. **Kopieren Sie nun den verschlüsselten Text in die Zwischenablage**. Klicken Sie dazu das **„Kopie“-Symbol** an (siehe roter Pfeil in der Abbildung).

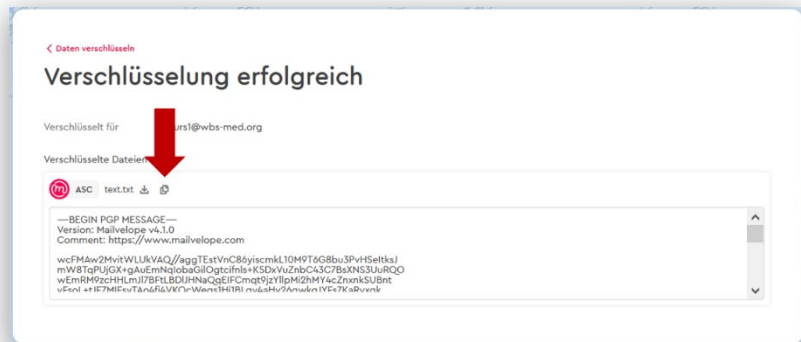


Abbildung 27: Erfolgsmeldung mit dem verschlüsselten Text

4. Gehen Sie nun im Hauptmenü auf **„Entschlüsseln“** und klicken Sie dort auf **„Möchtest Du auch noch einen Text entschlüsseln“**.

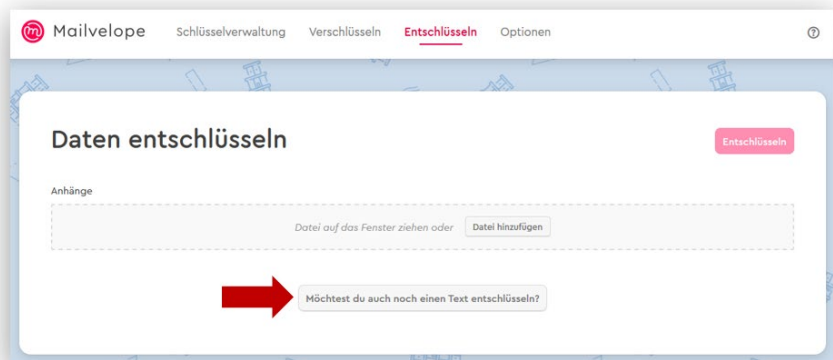


Abbildung 28: Start der Entschlüsselung

5. Klicken Sie in das Texteingabefeld und setzen Sie den Inhalt der Zwischenablage mittels der **Tastenkombination „Strg“ + „V“** dort ein. Es sollte dann wie abgebildet aussehen.

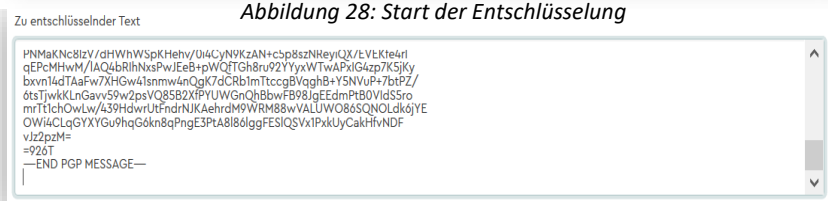


Abbildung 29: Kopierter verschlüsselter Text (die ersten Zeilen sind hier verdeckt).



6. Wenn Sie anschließend oben rechts auf „**Entschlüsseln**“ klicken, werden Sie zur Eingabe Ihres Schlüsselpassworts aufgefordert. Geben Sie das **Passwort** ein und sehen Sie sich das Ergebnis an.

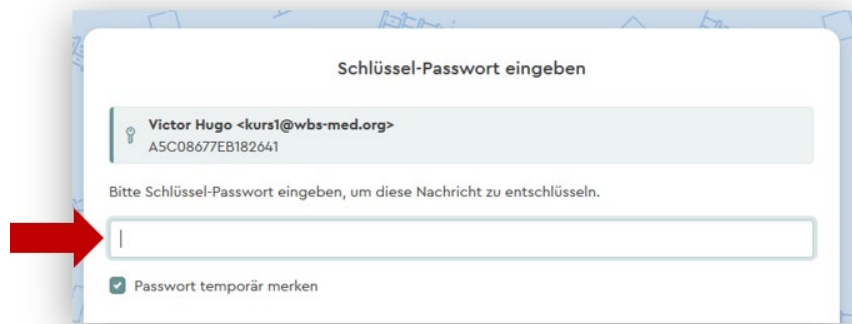


Abbildung 30: Aufforderung zur Passworteingabe vor Entschlüsselung

7. Wenn das Passwort stimmt, bekommen Sie eine Erfolgsmeldung und können durch **Anklicken von „TXT“ oder „text“** ihre Nachricht lesen.

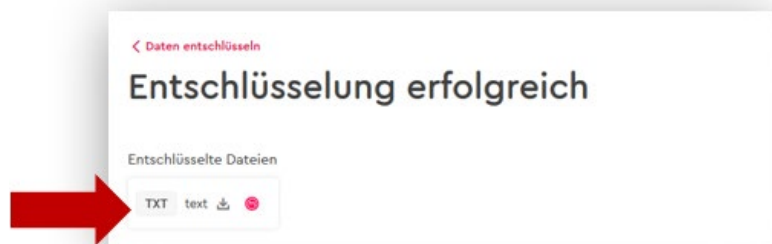


Abbildung 31: Erfolgsmeldung zur Entschlüsselung

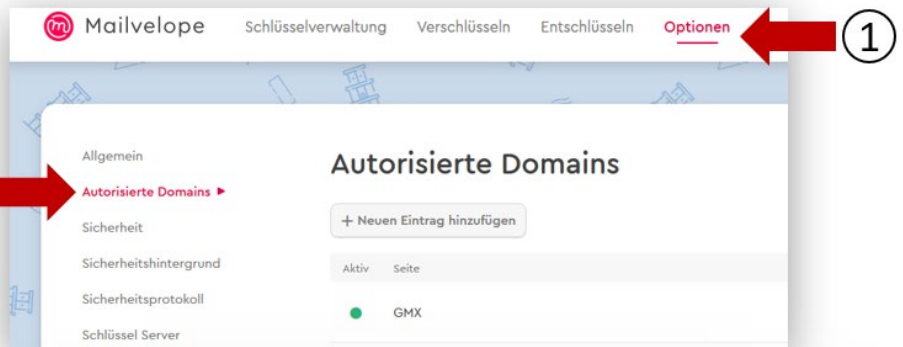
**Kurze Erinnerung:** Passiert ist hier Folgendes: Sie haben mit dem **öffentlichen Ver-Schlüssel** (hier Ihrem eigenen) eine Nachricht verschlüsselt und dann mit dem **privaten Ent-Schlüssel** wieder lesbar gemacht. Der private Schlüssel (Ent-Schlüssel) ist dabei geheim, deshalb mussten Sie erst im zweiten Schritt das **Passwort** eingeben, das Ihren privaten Schlüssel schützt.

**Hinweis:** Diese erste Verschlüsselungsübung war nur ein **Probelauf**, ob der Mechanismus funktioniert. Schließlich haben Sie sich nur selbst eine Nachricht überbracht (noch dazu ohne E-Mail durch bloßes Kopieren). Der Umgang mit den Mailvelope-Eingabemasken ist außerdem unkomfortabel, hat aber den Vorteil, unabhängig vom verwendeten E-Mail-Dienst immer zu funktionieren. Es geht aber viel komfortabler.

Dazu ist eine **Vorbereitung** nötig. Sie müssen Mailvelope mitteilen, dass der von Ihnen verwendete Webmail-Dienst unterstützt werden soll (viele Dienste sind voreingestellt, aber diesen Punkt sollten Sie kontrollieren).

**Anleitung:** Autorisierung des zu nutzenden Webmaildienstes

Gehen Sie dazu im Hauptmenü auf „Optionen“ und dort auf „autorisierte Domains“



Sehen Sie in der Liste nach, ob Ihr Webmail-Dienst dort aufgeführt ist.

**Nur, falls nicht:** Klicken Sie bitte auf „Neuen Eintrag hinzufügen“ und tragen Sie die Startseite Ihres Web-Maildienstes in das erste Feld ein. Schreiben Sie darunter das Domänenmuster, d. h. die beiden Zeichen \*. gefolgt vom hinteren Teil der Adresse (sogenannte erste Subdomäne – hier „wbs-med“ und Hauptdomäne – hier „.org“)

Abbildung 32: Autorisierung des von Ihnen später genutzten Webmail-Dienstes

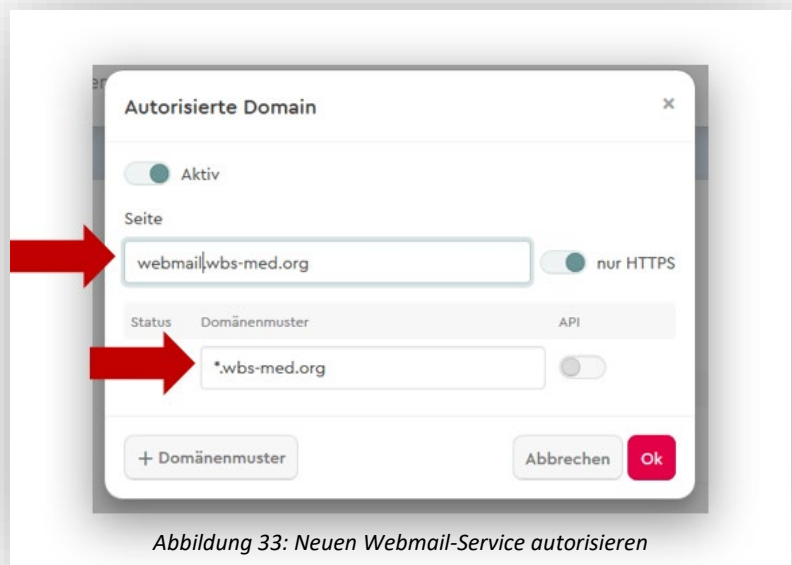


Abbildung 33: Neuen Webmail-Service autorisieren

**Aufgabe 3.7:**

Sorgen Sie dafür, dass ein von Ihnen selbst genutzter Webmail-Dienst für die Verschlüsselung mit Mailvelope autorisiert wird.

(Lösungsschritte: Siehe Anleitung oben)

**Aufgabe nach dem Kurs oder optional, wenn noch Zeit ist:** Verschlüsselte E-Mails verschicken

Ihre Aufgabe ist es, einer Teilnehmerin/einem Teilnehmer eine verschlüsselte E-Mail zu senden, die sie/er dann auch lesen kann (während des Kurses gern auch an ihre Kursbetreuer).

### Lösungsschritte: Verschlüsselte E-Mails im Kurs verschicken - Vorbereitung

Öffnen Sie wie eingangs beschrieben Mailvelope in Ihrem Browser. Klicken Sie auf „Schlüssel verwalten“. Erstellen Sie einen neuen Schlüssel. Die ausführliche Anleitung dazu können Sie ggf. oben noch einmal ansehen.

Erstellen Sie dieses Mal einen Schlüssel für Ihre eigene echte E-Mail-Adresse. Achten Sie auch diesmal darauf, den Haken vor „Öffentlichen Schlüssel zum Mailvelope Schlüssel Server hochladen“ **zu entfernen**. (Es ist aber kein Problem, wenn Sie das versehentlich vergessen).

#### *Hinweis:* Schlüsselaustausch

Noch kann es keine sichere Kommunikation geben: Sie kennen ja die öffentlichen Schlüssel Ihrer Adressanten noch nicht. Aber machen Sie sich klar: die öffentlichen Schlüssel dürfen alle kennen. Es ist also völlig in Ordnung, wenn Sie im Kurs **öffentliche Schlüssel** mit **unverschlüsselten E-Mails** austauschen.

1. Gehen Sie dazu wieder ins **Mailvelope** (Logo im Bowser oben rechts). Gehen Sie zur **Schlüsselverwaltung**. Klicken Sie als erstes Ihren **eigenen Schlüssel** an. Klicken Sie dann auf **exportieren**.

Screenshot of the 'Schlüssel erstellen' (Create Key) form in Mailvelope. The form includes fields for Name (Victor Hugo), E-Mail (kurs1@wbs-med.org), and Password. A red arrow points to the checkbox 'Öffentlichen Schlüssel zum Mailvelope Schlüssel Server hochladen', which is currently unchecked.

Abbildung 34: Angaben zum Schlüsselpaar (bitte entfernen Sie im Kurs den grünen Haken vor „Öffentlichen Schlüssel ...“ – siehe Pfeil)

Screenshot of the 'Schlüsselverwaltung' (Key Management) page in Mailvelope. It shows a key for 'Victor Hugo' with a green dot indicating it is 'gültig' (valid). A red arrow points to the 'Exportieren' (Export) button. Below the key details, there is a message: 'Der Schlüssel ist nicht mit dem Mailvelope Schlüssel-Server synchronisiert.' and a 'Synchronisieren' (Synchronize) button.

Abbildung 35: Start des Schlüsselexports

2. Für den nächsten Schritt ist es **absolut entscheidend, dass Sie nur Ihren öffentlichen Schlüssel exportieren** (der private bleibt, wo er ist: der ist nur für Sie). Klicken Sie „**In die Zwischenablage kopieren**“ an.
3. **Verabreden** Sie (per E-Mail, Messenger, Telefon oder mündlich) **einen Schlüsselaustausch** in einer kleinen Gruppe. Schicken Sie sich gegenseitig ihre öffentlichen Schlüssel. Nutzen Sie dazu Ihr E-Mail-Programm: Setzen Sie eine E-Mail an die anderen auf. Kopieren Sie Ihren öffentlichen Schlüssel aus der Zwischenablage in die E-Mail (erst: siehe vorherigen Schritt: Schlüssel in die Zwischenablage kopieren, dann rechte Maustaste klicken, dann „Einfügen“). Versenden Sie die E-Mail.
4. Sie sollten nun von den anderen, mit denen Sie sich verabredet haben, ebenfalls eine E-Mail bekommen, die den jeweiligen öffentlichen Schlüssel enthält.
5. Sobald die Nachricht eintrifft, **öffnen** Sie diese. Kopieren Sie den kompletten öffentlichen Schlüssel in dieser Nachricht in die Zwischenablage: Markieren Sie dazu alles zwischen „-----BEGIN PGP PUBLIC KEY BLOCK-----“ und „-----END PGP PUBLIC KEY BLOCK-----“ (inklusive dieser Zeilen selbst) zum Kopieren. Dann Rechtsklick und „kopieren“.

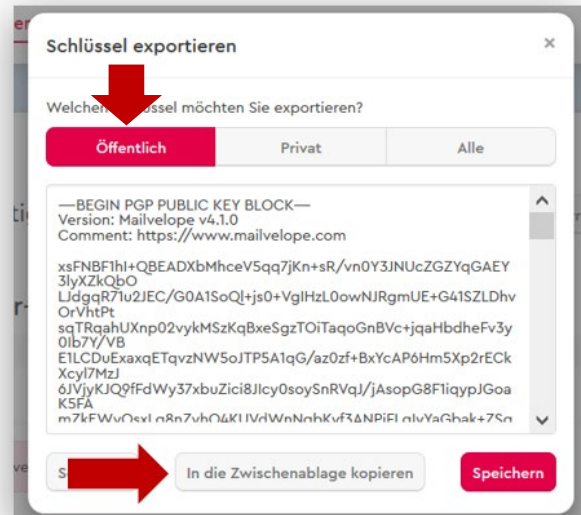
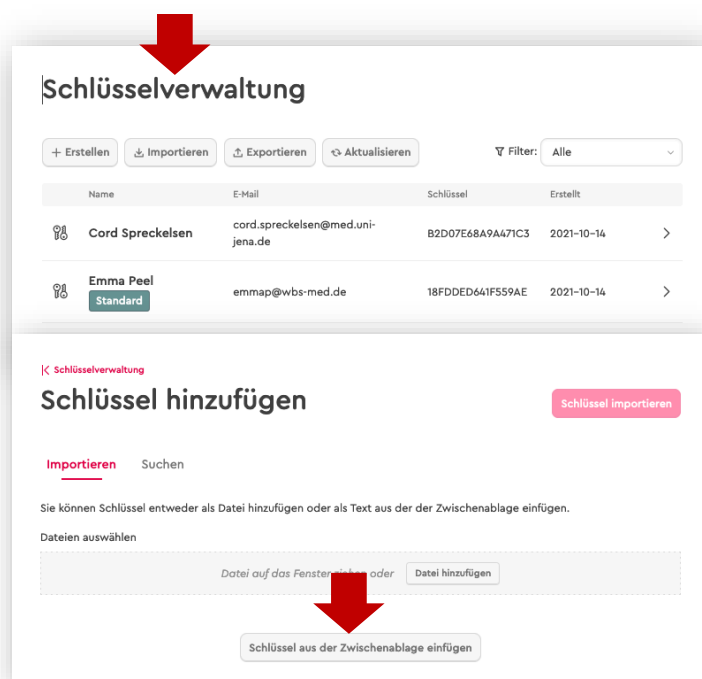


Abbildung 36: Exportieren des öffentlichen Schlüssels

6. Öffnen Sie nun wieder Mailvelope. Gehen Sie dort zur Schlüsselverwaltung. Klicken Sie auf Importieren. Klicken Sie in dem Fenster, das sich dann öffnet, auf „Schlüssel aus der Schlüsselablage einfügen“.



7. Nachdem Sie die Schlüssel der Mitstudierenden, mit denen Sie sich verabredet haben, in Ihren Schlüsselbund importiert haben, können Sie verschlüsselte Nachrichten austauschen: Das erfolgt immer in zwei Schritten: Erstens **verschlüsseln Sie in Mailvelope den Text**, den Sie versenden wollen, zweitens **versenden Sie** den verschlüsselten Text **einfach per E-Mail** an den Adressaten/die Adressatin. **Das Verschlüsseln erfolgt exakt so, wie Sie es am Anfang für eine Nachricht an sich selbst geübt haben** (Siehe Abschnitt „Lösungsschritte: Überprüfen, ob die Verschlüsselung für Sie selbst funktioniert“). Der einzige Unterschied: Wählen Sie beim Verschlüsseln in Mailvelope nicht ihre eigene (Übungs-)Adresse, sondern eine der eben importierten Adressen zu den Schlüsseln der anderen in Ihrem Team.
8. **Wichtig: Die Verschlüsselung erfolgt immer individuell für eine Person: Versenden Sie die verschlüsselte Nachricht nur an die Person, für die Sie sie verschlüsselt haben** (andere können diese Nachricht nicht entschlüsseln, da stört sie nur).
9. Wenn Sie selbst dann von anderen eine für Sie verschlüsselte Nachricht erhalten, **kopieren Sie den E-Mail-Inhalt** von „-----BEGIN PGP MESSAGE-----“ bis „-----END PGP MESSAGE-----“ (wieder inklusive dieser beiden Zeilen). Wechseln Sie zu **Mailvelope**. Gehen Sie zu **Entschlüsseln**. Klicken Sie **„Möchtest Du auch noch einen Text entschlüsseln“**. Fügen Sie den Text aus der Zwischenablage dort ein (Rechter Mausklick und „Einfügen“) und lassen Sie ihn entschlüsseln.
10. Lesen Sie die für Sie bestimmte Nachricht  
... .



### **Zusammenfassung** zur Verschlüsselten E-Mail-Kommunikation.

- Das Konzept der (asymmetrischen) Verschlüsselung mit Public-Private-Keys löst das Problem der **Schlüsselverteilung auf Distanz**.
- Schlüssel treten immer **paarweise** auf.
- Alle, die verschlüsselt kommunizieren wollen, brauchen jeweils ein **eigenes Schlüsselpaar**.
- **Wer Ihnen eine verschlüsselte Nachricht schicken will, braucht Ihren öffentlichen Schlüssel zum Verschlüsseln.**
- Ihren **öffentlichen Schlüssel** können Sie **offen kommunizieren**
- Sie halten Ihren **privaten Schlüssel immer geheim**.
- Nachrichten, die mit Ihrem öffentlichen Schlüssel verschlüsselt wurden, können **nur Sie** mit Ihrem privaten Schlüssel **wieder öffnen**.

### Digitale Signatur

Elektronisch zu unterschreiben (signieren) bedeutet: Es ist beweisbar klar, wer das signierte Dokument verfasst und autorisiert hat. Spätere Änderungen am Dokument fallen sofort auf. Damit ist für einen elektronisch unterschriebenen Arztbrief klar, wer für den Inhalt verantwortlich ist. Das Signaturgesetz von 2001 stellte eine weitgehende rechtliche Gleichwertigkeit von elektronischen Signaturen (unter klar beschriebenen Voraussetzungen) und handschriftlicher Unterschrift her. Es wurde inzwischen abgelöst von der **EU-VERORDNUNG (EU) Nr. 910/2014** und ihrer Umsetzung in nationales Recht durch das Vertrauensdienstegesetz. Erklärtes Ziel: die Gleichwertigkeit von elektronischer und handschriftlicher Signatur.

### Aufgabe (optional): E-Mail elektronisch unterschreiben

Nutzen Sie Mailvelope, um eine elektronische Unterschrift für Ihre Nachricht zu erzeugen: Signieren Sie Ihre Nachricht.

**Lösungsschritte:** Elektronische Signatur

1. **Verschlüsseln** Sie eine neue abgesicherte Botschaft mit **Mailvelope**. Gehen Sie genau so vor, wie eben eingeübt. **Bevor** Sie den Text aus Mailvelope in den Webmailer übernehmen: Klicken Sie neben „Die Daten werden beim Verschlüsseln nicht signiert“ auf „Ändern“. Wählen Sie dann Ihren eigenen Schlüssel zum Signieren.
2. Versenden Sie ruhig probierhalber eine signierte E-Mail. Wie Sie sehen, war kaum Vorbereitung nötig. **Die elektronische Signatur und die Public-Private-Key-Verschlüsselung nutzen dieselbe technische Grundlage:** Ihr Schlüsselpaar.

#### Daten verschlüsseln

Empfänger  
cord.spreckelsen@med.uni-jena.de x

Die Daten werden beim Verschlüsseln nicht signiert **Ändern**

Anhänge

### Hintergrundwissen: Konzept der elektronischen Signatur

Verwirrenderweise **tauschen** bei der elektronischen Signatur die beiden Schlüssel eines Schlüsselpaares **ihre Rollen**, wobei natürlich der **private Schlüssel immer privat** - und bei Ihnen - bleibt und nur der öffentliche Schlüssel in die Welt gelangen darf. Aber der **private Schlüssel** dient bei der Signatur **zum Verschlüsseln** und der öffentliche zum Entschlüsseln. Mathematisch ist das möglich. Und es funktioniert wie eine Spielplatz-**Wippe**: Dient Schlüssel A zum Verschlüsseln, dann kann nur der andere Schlüssel des Schlüsselpaares (Schlüssel B) entschlüsseln, verwendet man stattdessen den Schlüssel B zum Verschlüsseln, dann kann nur Schlüssel A entschlüsseln. Niemals kann aber derselbe Schlüssel das wieder aufschließen, was er verschlüsselt hat: das kann nur der jeweils andere. Wie bei der Wippe: einer hängt immer in der Luft, wenn der andere am Boden ist, es sind nie beide gleichzeitig unten.

Bei der Signatur dient der **private** Schlüssel also zum Verschlüsseln. Verschlüsselt wird ein sogenannter **digitaler Fingerabdruck** der signierten Daten. Aus Daten lässt sich immer ein solcher Fingerabdruck berechnen: selbst kleinste Änderungen an den Daten führen dazu, dass der Fingerabdruck (auch **Hash** genannt) sich bei Neuberechnung mit den geänderten Daten auch ändert.

Die Signatur besteht aus dem – zunächst verschlüsselten, aber bei Überprüfungsbedarf dann entschlüsselten – Fingerabdruck der Daten. Aus den Daten selbst (z. B. der E-Mail) wird ein **frischer Fingerabdruck** berechnet und mit dem in der Signatur versendeten Fingerabdruck **verglichen**. Stimmen beide überein, ist klar: **Die erhaltenen Daten stammen so, wie sie angekommen sind, vom Unterzeichner/von der Unterzeichnerin.**



Wenn Sie elektronisch signierte Daten (z. B. E-Mails) erhalten und mittels Signatur **verifizieren** wollen, passiert folgendes: Der öffentliche Schlüssel des Unterzeichners/der Unterzeichnerin wird gesucht. Wenn dieser Schlüssel passt, ist klar: Die Signatur stammt vom Absender/von der Absenderin, denn nur der paarweise zu diesem öffentlichen Schlüssel gehörende private Schlüssel kann die Signatur erzeugt, d. h. verschlüsselt haben. Wichtig ist hier: Es muss öffentlich **beglaubigt**, d.h. zertifiziert sein, dass ein bestimmtes Schlüsselpaar zu einer bestimmten Person gehört (**Schlüsselauthentizität**). Dieses Problem ist organisatorisch lösbar (zentral durch Zertifizierungsstellen oder dezentral durch sogenannte Networks of Trust).

Rechtlich ist die sogenannte **qualifizierte elektronische Signatur** wichtig. Die Schlüsselauthentizität muss hier mittels Zertifizierung durch eine öffentliche Stelle sichergestellt werden und die Signatur wird ausschließlich durch ein Gerät (inkl. Computerprogramm) erstellt, das als „qualifizierte Signaturerstellungseinheit“ freigegeben wurde.

### **Zusammenfassung: Digitale Signatur**

- 1) Grundlage der digitalen Signatur sind Public-Private-Keys (d. h. Schlüsselpaare).
- 2) Sie halten Ihren privaten Schlüssel immer geheim!
- 3) Die Schlüssel Ihres Schlüsselpaars tauschen die Rollen: die digitale Signatur ist mit Ihrem privaten Schlüssel verschlüsselt und kann nur mit Ihrem allseits bekannten öffentlichen Schlüssel aufgeschlossen werden.
- 4) Weil die Schlüssel paarweise zusammengehören, ist klar, von wem die Signatur erstellt wurde.
- 5) Wenn der in der Signatur enthaltene Fingerabdruck der Daten zum aktuell berechneten Fingerabdruck der empfangenen Daten passt, ist deren unveränderte Herkunft vom Autor der Signatur gesichert.

### **Ergänzung: Einloggen ohne Passwörter – FIDO2 basierte Anmeldung**

Die Public-Private-Keys sind äußerst vielseitig nutzbar. Eine weitere wichtige Anwendung ist die **Anmeldung ohne Passwort**:

Mithilfe von **FIDO2** (die Abkürzung steht für Fast ID Online) ist es nicht mehr erforderlich, sich Passwörter zu merken, die oft vergessen oder gestohlen werden können. Stattdessen kommt ein spezielles Schlüsselpaar zum Einsatz. Wenn jemand ein Benutzerkonto anlegt, wird der öffentliche Schlüssel des Schlüsselpaares auf dem Server des Diensteanbieters gespeichert. Der private Schlüssel bleibt sicher auf dem eigenen Gerät.

Wer sich an seinem Konto anmelden möchte, sendet eine Nachricht an den Diensteanbieter. Ein Teil der Nachricht wird mit dem **eigenen privaten Schlüssel verschlüsselt**. Der Anbieter versucht jetzt, den verschlüsselten Teil mit dem hinterlegten öffentlichen Schlüssel zu öffnen. Gelingt das, ist klar, dass der Zugriff freigegeben werden kann. FIDO2 funktioniert im Prinzip genau, wie die **elektronische Signatur**. Das Verfahren stellt sicher, dass der Zugriff nur der Person freigegeben wird, deren öffentlicher Schlüssel hinterlegt wurde. Und diese Person wird mittels einer Botschaft authentifiziert, die mit dem eindeutig zugeordneten privaten Schlüssel verschlüsselt wurde.



## 4 Einführung in Excel

*Welches Programm sollte für die Datenhaltung und -verwaltung verwendet werden?*

Die Beantwortung dieser wichtigen Frage steht am Beginn einer jeden Promotionsarbeit, kann aber nicht allgemeingültig beantwortet werden. Die Wahl eines Programms zur Datenhaltung und -verwaltung hängt von vielen Faktoren ab, wobei in der Medizin die Tabellenkalkulation häufig bevorzugt wird. Das liegt u. a. daran, dass die Alternative, die Nutzung eines Datenbanksystems, als zu kompliziert angesehen wird. Ein Grund dafür ist, dass die einführende Literatur in der Regel mit Theorie überfrachtet ist. Es gibt aber einige wichtige Kriterien, die bei der Wahl des Programms helfen.

Für eine Tabellenkalkulation sprechen:

- Ihnen reicht eine einfache Datenverwaltung, die ohne größere Anforderungen an die Datensicherheit (z. B. Sicherung der Datensatzstruktur) auskommt.
- Sie möchten die Daten zunächst nur in Tabellenform bringen und **zwischenspeichern**, um sie später in eine Datenbank oder in ein Statistik-Programm (z. B. SPSS, R) zu exportieren.
- Der Umfang Ihrer **Datenmenge ist überschaubar**.

Argumente für die Verwendung von Datenbanken sind:

- Sie verfügen über sehr **große Datenmengen**.
- Sie beziehen Daten aus **unterschiedlichen Quellen** und müssen diese zusammenführen.
- Sie erwarten eine **variable Anzahl von Messwiederholungen** je Individuum.
- Sie haben hohe Anforderungen an die **Datenkonsistenz**.
- Sie arbeiten mit weiteren Personen (evtl. mit Zugriffsbeschränkungen) an den Daten.

Die Tabelle ist eine Organisations- und Speicherungsform für Daten, die übersichtlich und einfach auswertbar ist. Ein Excel-Tabellenblatt (als Bestandteil einer Excel-Datei) ist ein Kalkulationsblatt, in dem u. a. Text, Zahlen und Formeln eingegeben und verwaltet werden können.

*Zum Seminar Tabellenkalkulation:*

Das Seminar „Tabellenkalkulation“ dient der Vermittlung von Grundwissen zur Strukturierung bei Dateneingabe, -formatierung und -austausch, Organisation einer Mastertabelle in Untertabellen, Erstellung von Formeln für Datenauswertung, bedingte Formatierung, Programmierung von Logik- und Datumsfunktionen und Erstellung von Diagrammen.

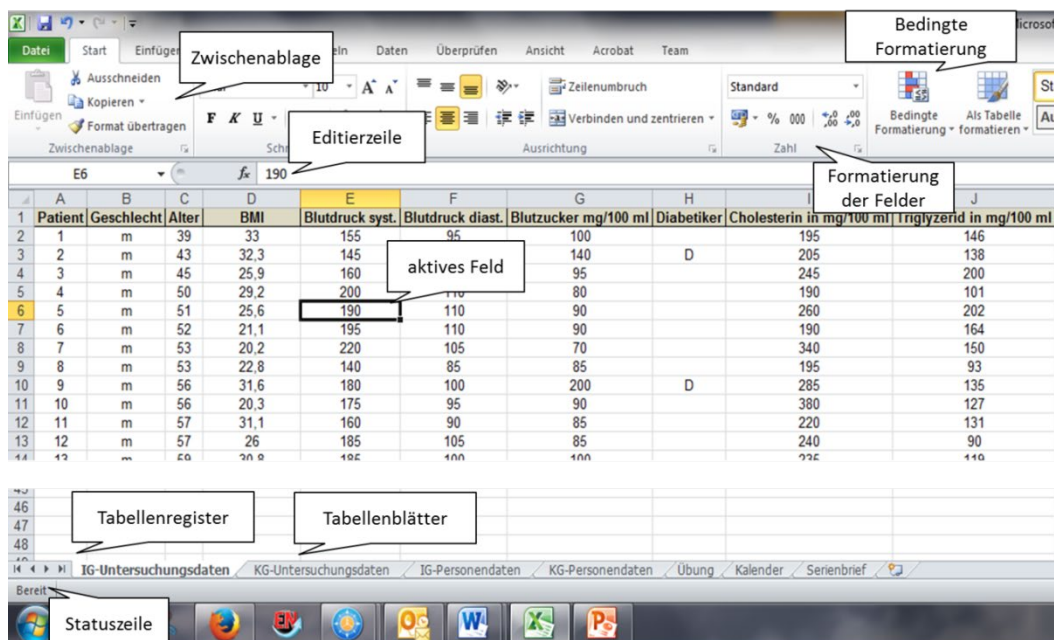
Im Seminar werden Daten einer Studie genutzt. Die Excel-Datei „**Infarktstudie.xlsx**“ umfasst sowohl Studiendaten als auch alle Übungstabellen. Die Datei enthält die Daten von 40 Patienten (**IG** = Infarkt-Gruppe) und 40 Kontrollpersonen (**KG** = Kontroll-Gruppe). Die Tabellenblätter „**ÜbungsTab\***“ dienen zur Unterstützung weiterer Übungen.

### Allgemeines zum Tabellenaufbau und zur Eingabe von Daten

Das Programm Excel wird gestartet und die Datei „Infarktstudie.xlsx“ geöffnet.

Das Tabellenblatt ist in einzelne Felder (Zellen) aufgeteilt, deren Lage durch die Spalte (Buchstaben in alphabetischer Ordnung) und die Zeile (Zahlen in aufsteigender Ordnung) gekennzeichnet sind. Die Position der Zelle nennt man Adresse. So hat die aktive Zelle in der Abbildung die Adresse E6.

Die Zelle, in die eingegeben werden soll, wird mit der Maus angeklickt und eine stärkere Umrandung zeigt, dass diese Zelle aktiviert ist. Die Eingabe kann in der Bearbeitungs- oder Editierzeile vorgenommen werden. Mit dem Anklicken der Zelle wird die Adresse (hier E6) in den linken Teil der Editierzeile übernommen. Mit der Maus kann man sich in die Bearbeitungszeile "einklicken". Ein blinkender Textcursor (senkrechter Balken) zeigt an, dass man im Textmodus eingeben kann. Das Editieren (Eingabe, Ändern, Löschen) erfolgt so, wie Sie es von der Textverarbeitung kennen. Alternativ können auch direkt nach dem Anklicken der Zelle Eingaben über die Tastatur vorgenommen werden. Diese werden automatisch in die Editierzeile übernommen.



Alle über die Navigationsleisten und Maustasten ausführbaren Prozeduren werden immer für den gesamten markierten Adressbereich ausgeführt. Die Markierung wird z. B. mit gedrückter linker Maustaste vorgenommen.

**Markierung eines Feldes:** Klicken auf das Feld (fette Umrandung erscheint, rechte untere Ecke mit kleinem Quadrat gekennzeichnet)

**Markierung eines Bereichs:** Der Bereich wird hellblau-transparent ausgefüllt, das Feld in der linken oberen Ecke bleibt weiß, die Umrandung umfasst den gesamten markierten Bereich.

## 4.1 Eingabe von Daten in eine Tabelle

**Aufgabe 4.1:** *Eingabe von Daten in eine Tabelle. Das Tabellenblatt „ÜbungsTab1“ wird geöffnet. Die Tabelle ist unvollständig und soll ergänzt werden.*

### Lösungsschritte:

- ⇒ Das Tabellenblatt „ÜbungsTab1“ der Datei „Infarktstudie.xlsx“ wird geöffnet (Tabellenregister unter der Tabelle)
- ⇒ Die Eingabe der nicht ausgefüllten Felder soll entsprechend der vorgegebenen Mustertabelle erfolgen.

	A	B	C	D	E	F	G
1	Patient	BlutDrdia	BlutDrsys		Patient	BlutDrdia	BlutDrsys
2	101	95	170		101	95	170
3	102	90	140		102	90	140
4	103	90	145		103	90	145
5	104		140		104	95	140
6	105	80	150		105	80	150
7	106	90			106	90	155
8	107	95	160		107	95	160
9	108	90	145		108	90	145
10	109	95	160		109	95	160
11	110	85	135		110	85	135

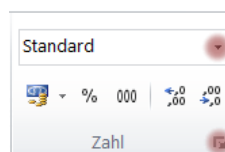
## 4.2 Festlegung des Datenformats

**Aufgabe 4.2:** *Die Festlegung von verschiedenen Datenformaten soll geübt und ihre Wirkung auf Dateneingaben deutlich gemacht werden. Dabei ist die folgende Beispieltabelle zu bearbeiten:*

	A	B	C	D	E
1		Aber	Zahl: 178	PLZ: 07740	Station: 2.1
2	Format: Standard				
3	Format: Zahl				
4	Format: Text				

### Lösungsschritte:

- ⇒ Das Tabellenblatt „ÜbungsTab2“ wird geöffnet.
- ⇒ Die Zeile „Standard“ wird nicht explizit formatiert; die Zeile „Zahl“ soll mit dem Format **Zahl** und die Zeile „Text“ mit dem Format **Text** formatiert werden.
- ⇒ Die Formatierung geschieht durch Markierung der Zellen B3 bis E3 bzw. B4 bis E4 und über die Formatierungsfelder für Datenformate



- ⇒ Mit der Wahl eines Datenformats sind die markierten Zellen formatiert.

⇒ Geben Sie die vorgegebenen Daten wie in den Spaltenüberschriften angegeben (ohne Kopieren/Einfügen) ein und diskutieren Sie die Effekte.

### 4.3 Statischer Datenaustausch über die Zwischenablage

**Aufgabe 4.3:** *Es soll der statische Datentransport über die Zwischenablage wiederholt werden.*

**Lösungsschritte:**

- ⇒ Das Tabellenblatt „**ÜbungsTab3**“ wird geöffnet.
- ⇒ Auswahl des Quellbereichs: Mit gedrückter linker Maustaste soll ein zusammenhängender Feldbereich einer Tabelle markiert werden (hier A1 bis C11 bzw. A1:C11). Ein Einzelfeld ist automatisch durch Anklicken markiert. Unterbrochene Bereiche (Felder) können markiert werden, indem die „Strg“-Taste während der Auswahl der Felder gedrückt wird.
- ⇒ Die im markierten Bereich befindlichen Daten können über die bekannten Varianten für **Kopieren** in die Zwischenablage gebracht werden.
- ⇒ Auswahl des Zielbereichs: Danach ist das Feld anzuklicken, das die linke obere Ecke des neuen Tabellenbereichs für die kopierten Daten werden soll (z. B. E1).
- ⇒ Durch die bekannten Varianten von **Einfügen** gelangen die kopierten Daten in den neuen Bereich. Es werden sowohl Zellinhalte und Formate übertragen.
- ⇒ Ändern Sie Einträge in den Zellen des Quellbereichs und diskutieren Sie eventuelle Änderungen im Zielbereich. Da es sich um ein statisches Kopieren handelt, erfolgt keine Aktualisierung im Zielbereich

### 4.4 Dynamischer Datenaustausch mittels Adressen (Zellbezügen)

**Aufgabe 4.4:** *Erstellung von dynamischen Verknüpfungen zwischen Feldern in und zwischen Tabellen(blättern).*

Dieser Übungsteil bezieht sich auf den dynamischen Datenaustausch zwischen unterschiedlichen Tabellen durch die Verbindung von Feldern über so genannte Zellbezüge (Adressen). Der Datenaustausch kann innerhalb einer Tabelle, zwischen Tabellenblättern einer Excel-Datei („Erstellen einer Untertabelle“) und zwischen verschiedenen Excel-Dateien erfolgen. Der Einfachheit halber werden nur Bezüge innerhalb eines Tabellenblatts geübt.

Jeder Befehl zum Datenaustausch beginnt mit einem Gleichheitszeichen, sollte nur einmal, d. h. für ein Beispiel als Prototyp geschrieben und mit **Kopieren/Einfügen** vervielfältigt werden.

#### Fall 1 – Relative Zellbezüge

Es soll eine dynamische Kopie des Bereichs A1:C11 im Bereich A13:C23 erstellt werden.

**Lösungsschritte:** Realisierung als Formel

- ⇒ Formeln beginnen mit einem Gleichheitszeichen
- ⇒ In der Zelle A13 (linke obere Ecke des Zielbereichs) wird die Formel „=A1“ eingetragen.



- ⇒ Die Formel „=A1“ in der Zelle A13 bewirkt eine Verlinkung mit der Zelle A1, die im speziellen Fall in derselben Spalte (A vs. A) aber 12 Zeilen weiter oben (13 vs. 1) steht. Diese Art von Bezug ist also relativ.
- ⇒ Kopieren Sie die Zelle A13 auf den Bereich A13 bis C23.
- ⇒ Klicken Sie in Zellen des neu gefüllten Tabellenbereichs und beobachten Sie die Einträge in der Editierzeile. Die Bezüge sind immer „gleiche Spalte, 12 Zeilen darüber“).
- ⇒ Ändern Sie Einträge in den Zellen des Quellbereichs und diskutieren Sie eventuelle Änderungen im Zielbereich. Da es sich um ein dynamisches Kopieren handelt, erfolgt eine Aktualisierung im Zielbereich. Formatierungen werden nicht mit kopiert.

### Fall 2 – Absolute Zellbezüge

**Lösungsschritte:** Realisierung als Formel

- ⇒ In der Zelle E13 (linke obere Ecke des Zielbereichs) wird die Formel „=\$A\$1“ eingetragen. Die Ausführung dieser Formel bewirkt, dass der Inhalt der Zelle E13 gleich dem Inhalt der Zelle A1 ist.
- ⇒ Die Formel „=\$A\$1“ in der Zelle E13 bewirkt eine absolute Verlinkung mit der Zelle A1 unabhängig von der Position der Zielzelle.
- ⇒ Kopieren Sie die Zelle E13 auf den Bereich E13:G23.
- ⇒ Weiter wie bei Fall 1.

### Fall 3 – Mischung von absoluten und relativen Zellbezügen (1)

**Lösungsschritte:** Realisierung als Formel

- ⇒ In der Zelle I13 (linke obere Ecke des Zielbereichs) wird die Formel „=\$A1“ eingetragen. Die Ausführung dieser Formel bewirkt, dass der Inhalt der Zelle I13 gleich dem Inhalt der Zelle in Spalte A sowie 12 Zeilen oberhalb ist. Das gilt unabhängig davon, welchen Typ die Daten haben.
- ⇒ Die Formel „=\$A1“ in der Zelle I13 bewirkt eine absolute Verlinkung mit der Spalte A.
- ⇒ Kopieren Sie die Zelle I13 auf den Bereich I13:K23.
- ⇒ Weiter wie bei Fall 1.

### Fall 4 – Mischung von absoluten und relativen Zellbezügen (2)

**Lösungsschritte:** Realisierung als Formel

- ⇒ In der Zelle I1 (linke obere Ecke des Zielbereichs) wird die Formel „=A\$1“ eingetragen. Die Ausführung dieser Formel bewirkt, dass der Inhalt der Zelle I1 gleich dem Inhalt der Zelle in Zeile 1 sowie 8 Spalten weiter links ist.
- ⇒ Die Formel „=A\$1“ in der Zelle I1 bewirkt eine absolute Verlinkung mit der Zeile 1.
- ⇒ Kopieren Sie die Zelle I1 auf den Bereich I13:K11.
- ⇒ Weiter wie bei Fall 1.

	A	B	C	D	E	F	G	H	I	J	K
1	Patient	BlutDrdia	BlutDrsys		Patient	BlutDrdia	BlutDrsys		Patient	BlutDrdia	BlutDrsys
2	101	95	170		101	95	170		Patient	BlutDrdia	BlutDrsys
3	102	90	140		102	90	140		Patient	BlutDrdia	BlutDrsys
4	103	90	145		103	90	145		Patient	BlutDrdia	BlutDrsys
5	104	95	140		104	95	140		Patient	BlutDrdia	BlutDrsys
6	105	80	150		105	80	150		Patient	BlutDrdia	BlutDrsys
7	106	90	155		106	90	155		Patient	BlutDrdia	BlutDrsys
8	107	95	160		107	95	160		Patient	BlutDrdia	BlutDrsys
9	108	90	145		108	90	145		Patient	BlutDrdia	BlutDrsys
10	109	95	160		109	95	160		Patient	BlutDrdia	BlutDrsys
11	110	85	135		110	85	135		Patient	BlutDrdia	BlutDrsys
12											
13	Patient	BlutDrdia	BlutDrsys		Patient	Patient	Patient		Patient	Patient	Patient
14	101	95	170		Patient	Patient	Patient		101	101	101
15	102	90	140		Patient	Patient	Patient		102	102	102
16	103	90	145		Patient	Patient	Patient		103	103	103
17	104	95	140		Patient	Patient	Patient		104	104	104
18	105	80	150		Patient	Patient	Patient		105	105	105
19	106	90	155		Patient	Patient	Patient		106	106	106
20	107	95	160		Patient	Patient	Patient		107	107	107
21	108	90	145		Patient	Patient	Patient		108	108	108
22	109	95	160		Patient	Patient	Patient		109	109	109
23	110	85	135		Patient	Patient	Patient		110	110	110

**Bemerkung:** Datei- und tabellenblattübergreifend ist die Syntax:  
 „=[Dateiname]'Tabellenblattname'!Zelladresse“

### 4.5 Arbeiten mit Formeln und Zellbezügen

**Aufgabe 4.5:** Es soll der mittlere arterielle Blutdruck  $p_m$  für jede Person aus dem diastolischen ( $p_d$ ) und systolischen Blutdruck ( $p_s$ ) mit folgender Näherungsformel berechnet werden:

$$p_m = \frac{(b-1) \cdot p_d + p_s}{b} \quad \text{mit dem Parameter } b = 2,65 \dots 3$$

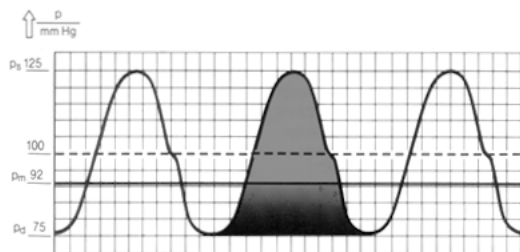


Abb. 3.4: Bestimmung des arteriellen Mitteldrucks  $p_m$  (nähere Erläuterungen im Text)

Der arterielle Mitteldruck wird bei fortlaufender invasiver Blutdruckmessung durch Integration über die gesamte Fläche unter einem Druckpuls dividiert durch die Pulsdauer ermittelt. Das nebenstehende Beispiel zeigt, wie der arithmetische Mittelwert (100 mmHg) abweicht (aus Morr-Strathmann & Tillmann, Grundlagen des invasiven Kreislaufmonitoring, 1982).

**Bemerkung:** Bei der Arbeit mit Formeln sind die unterschiedlichen Arten von Zellbezügen (Adressen) von Bedeutung. Ziel ist es, Formeln nur für einen Datensatz als Prototyp zu entwerfen, um sie dann durch Kopieren/Einfügen auf andere Zellen, Tabellen oder Datensätze anzuwenden.

#### Lösungsschritte:

- ⇒ Zuerst wird das Tabellenblatt „Übungstab3“ wieder in den Originalzustand gebracht. Dazu markiert man die Zeilen 13 bis 23. Nach Rechtsklick auf die Zeilennummern nutzt man **Zellen löschen** aus dem Kontextmenu. Analog verfährt man mit den Spalten E bis K.
- ⇒ Die Näherungsformel muss entsprechend der erforderlichen Syntax für Formeln in Excel umgeformt werden. Da der mittlere arterielle Blutdruck für jede Person (Datensatz) ermittelt werden soll, ist die Berechnung über eine Formel durch Verwendung von Zellbezügen (Adressen) am effektivsten zu realisieren. Ziel ist es, die Formel nur für den ersten Datensatz zu schreiben und danach für alle anderen Datensätze zu übertragen. Diese Formel soll schließlich in Spalte D neben

den anderen Blutdruckwerten (Spalten B und C) stehen, angefangen mit dem Feld D2. Der variable Faktor  $b$  soll dabei für alle Formeln der Spalte D verfügbar sein. Zur Speicherung des Wertes von  $b$  soll hier die Zelle G2 vorgesehen sein, damit später einfach mit verschiedenen Werten für  $b$  gearbeitet werden kann. Die Formel lautet für das Feld D2:  $=((\$G\$2-1)*B2+C2)/\$G\$2$ . In G2 wird ein beliebiger Zahlenwert zwischen 2,63 und 3 für  $b$  eingetragen. Die Eingabe der Formel wird zuerst im Feld D2 vorgenommen (Prototyp). Die relative bzw. absolute Schreibweise der Adresse G2 mit  $\$G\$2$  soll anschließend diskutiert werden. Ist die Formel eingegeben, wird der Wert des mittleren arteriellen Blutdrucks berechnet und angezeigt.

⇒ Durch Vervielfältigung der Formel aus D2 in die anderen Felder der Spalte D wird die Berechnung der anderen Werte des mittleren arteriellen Blutdrucks erreicht.

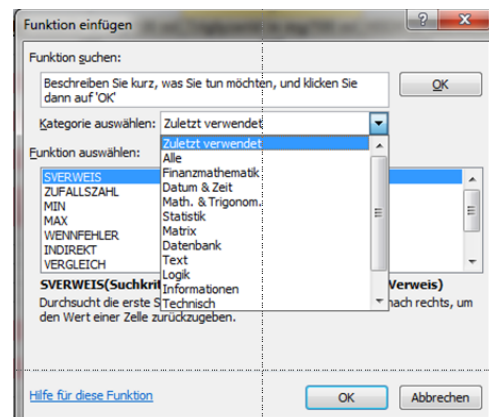
## 4.6 Nutzung der Tabellenfunktionen

**Aufgabe 4.6:** Es sollen Tabellenfunktionen genutzt werden, um weitere Berechnungen in der Tabelle vorzunehmen. Eine Vielzahl von allgemeinen Funktionen ist über den Funktionsassistenten (Symbol links neben Editierzeile) verfügbar.

Tabellenfunktionen sind bereits implementierte Formeln, deren Argumente nur noch geeignet ausgefüllt werden müssen. Es steht eine Vielzahl von Excel-Funktionen aus verschiedenen Kategorien zur Verfügung.

Kategorien von Tabellenfunktionen sind:

- Statistik
- Mathematik & Trigonometrie
- Finanzmathematik
- Datenbanken
- Textfunktionen
- Logik
- u. a. m.



**Teilaufgabe 1:** Die Werte des mittleren arteriellen Blutdrucks sollen gerundet werden (0 Stellen nach dem Komma). Dafür steht die Spalte E zur Verfügung.

*Mathematik-Funktion:* RUNDEN (Zahl, Anzahl\_Stellen); Zahl und Anzahl\_Stellen heißen Argumente

**Teilaufgabe 2:** Die arithmetischen Mittelwerte der Spalten B bis E sind zu ermitteln.

*Statistik-Funktion:* MITTELWERT(Zahl1; Zahl2;...)

Immer wenn für zusammenhängende Spalten bzw. Zeilenbereiche formuliert werden soll (z. B. von B2 bis B11 - von ... bis ...), kann der Doppelpunkt zur Adressierung des verwendeten Datenbereichs genutzt werden (z. B. B2:B11). Beispiel =MITTELWERT(B2:B11)

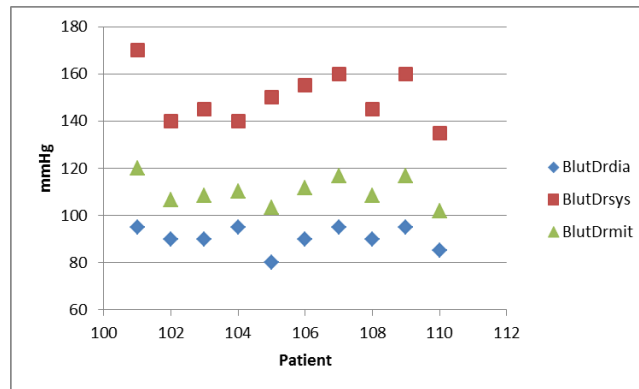
**Lösungsschritte (für beide Teilaufgaben):**

⇒ Nutzen Sie das Tabellenblatt „ÜbungsTab3“. Durch Anklicken wird die Zelle E2 aktiviert, welche die Tabellenfunktion enthalten soll.

- ⇒ Über **den Funktionsassistenten** erreicht man das oben abgebildete Dialogfenster, aus dem die entsprechende Tabellenfunktion durch Anklicken der Funktion und der OK-Taste in die Bearbeitungszeile übernommen wird.
- ⇒ Die Tabellenfunktion ist allgemein geschrieben, sie muss konkret angepasst werden. Für **Zahl** wird der aktuelle Zellbezug D2 (als relativer Zellbezug) eingegeben.
- ⇒ Die einmalig als Beispiel vollständig ausgefüllte Tabellenfunktion wird über die Zwischenablage in die anderen Felder kopiert.
- ⇒ Analog verfährt man in Zelle B12 mit der Funktion **Mittelwert** über die darüber liegenden Daten. Diese Formel kann dann wieder auf die Spalten C bis E übertragen werden.

## 4.7 Grafische Darstellung der Daten

**Aufgabe 4.7:** Die Tabelle mit den Blutdruckwerten soll in einem Diagramm für eine wissenschaftliche Präsentation dargestellt werden. Dabei ist zu beachten, dass die drei Werte  $p_d$ ,  $p_s$  und  $p_m$  für jede Person darzustellen sind. Folgender Darstellungsvorschlag soll umgesetzt werden.



### Lösungsschritte:

- ⇒ Nutzen Sie das Tabellenblatt „**Übungst3**“.
- ⇒ Der Tabellenbereich mit den Spalten **Patient** bis **BlutDrmit** muss markiert werden.
- ⇒ Danach wird über **Einfügen** ⇒ **Diagramme** ⇒ **Punkt** ein unformatiertes Diagramm eingefügt.



- ⇒ Die Beschriftung der Achsen kann in der schrittweisen Bearbeitung des Diagramms mittels der Funktionen unter **Diagrammtools** ⇒ **Layout** erfolgen.
- ⇒ Ändern Sie den Wert der Variablen  $b$  und diskutieren Sie die Werte in den Spalten D und E sowie das Verhalten der Graphik.

## 4.8 Nutzung von Logik-Funktionen und bedingten Formatierungen

### Allgemeines zur Nutzung von Logikfunktionen:

**WENN**(Prüfung;Dann\_Wert;Sonst\_Wert)

**WENN()** nimmt eine Wahrheitsprüfung vor, d. h. wenn ein Kriterium WAHR (erfüllt) ist (Argument **Prüfung**), DANN wird ein einzugebender Wert (Argument **Dann\_Wert**) ausgegeben, SONST (wenn das Kriterium FALSCH (nicht erfüllt) ist) kann ein anderer Wert ausgegeben werden (Argument **Sonst\_Wert**). Dann- und Sonst-Werte können Zahlen bzw. Text sein (Text in Anführungsstrichen eingeben). Wenn für das Argument **Prüfung** eine Funktion eingesetzt wird (Verschachtelung), dann müssen diese den Wert WAHR bzw. FALSCH ausgeben (**ODER()**, **UND()**), d. h. entsprechende Dann- bzw. Sonst-Werte können ausgegeben werden:

(=**WENN**(**ODER**(Wahrheitswert1; Wahrheitswert2);Dann\_Wert;Sonst\_Wert))


**ODER**(Wahrheitswert1; Wahrheitswert2; ...)

**ODER()** gibt das Ergebnis WAHR aus, wenn mindestens eines der eingegebenen Kriterien WAHR (erfüllt) ist. Ansonsten wird FALSCH ausgegeben.

**UND**(Wahrheitswert1; Wahrheitswert2; ...)

**UND()** gibt das Ergebnis WAHR aus, wenn alle der eingegebenen Kriterien WAHR (erfüllt) sind. Ansonsten wird FALSCH ausgegeben.

**Aufgabe 4.8:** Bestimmen und kennzeichnen Sie auf der Grundlage des Cholesterinwertes und weiterer Risikofaktoren für die koronare Herzerkrankung die Personen der Kontrollgruppe, die ein Risiko dafür besitzen.

	ohne weitere Risikofaktoren	mit weiteren Risikofaktoren	bei Vorerkrankungen
Gesamtcholesterin	< 240 mg/dl	< 200 mg/dl	< 180 mg/dl
	< 6,2 mmol/l	< 5,2 mmol/l	< 4,7 mmol/l
LDL-Cholesterin	< 160 mg/dl	<130 mg/dl	< 100 mg/dl
	< 4,1 mmol/l	< 3,4 mmol/l	< 2,6 mmol/l
Triglyzeride	< 150 mg/dl	< 150 mg/dl	< 150 mg/dl
	< 1,7 mmol/l	< 1,7 mmol/l	< 1,7 mmol/l
HDL-Cholesterin	> 40 mg/dl	> 40 mg/dl	> 40 mg/dl
	> 1,0 mmol/l	> 1,0 mmol/l	> 1,0 mmol/l
<b>Risikofaktoren</b>			
u.a. Übergewicht, Rauchen, Bluthochdruck, Zuckerkrankheit, Herz-Kreislaufkrankung in der Familie, Alter			
<b>Vorerkrankungen</b>			
u.a. Herzinfarkt, Schlaganfall, Angina pectoris 			

Alters- und geschlechtsunabhängige Risikofaktoren für eine Erkrankung der Koronararterien sind Nikotinkonsum, Diabetes mellitus, Bluthochdruck, familiäre Belastung (früh auftretende Herz-Kreislaufkrankungen wie Infarkt oder Schlaganfall) und ererbte oder erworbene Störungen des Fettstoffwechsels.

Hierbei sind vor allem ein hoher Cholesterinwert und erhöhte Triglyzeride problematisch.

Die Tabelle<sup>1</sup> gibt an, ab welchen Cholesterinwerten ein Risiko besteht (zum Zwecke der Vereinfachung sollen die LDL- und HDL-Cholesterinwerte nicht in die Entscheidung einbezogen werden). Bei der Entscheidung muss das Vorhandensein weiterer Risikofaktoren bzw. einer Vorerkrankung beachtet werden.

<sup>1</sup> Fehler! Linkreferenz ungültig. blutfettwerte/blutfettwerte.php



Zuerst wird die Aufgabe mit Logikausdrücken formuliert, wobei wir zur Vereinfachung vier Schritte vorgesehen haben. Die Prüfung erfolgt anhand des Cholesterinwertes.

**Schritt 1:** Prüfung Cholesterinwert  $> 240$  ohne weitere Risikofaktoren. Wenn vorhanden, dann Kennzeichnung mit dem Wert WAHR. Die Prüfung des logischen Ausdrucks *Risiko1*

$$\{\text{Cholesterin} > 240\}$$

liefert WAHR oder FALSCH.

**Schritt 2:** Ermittlung eines Extrarisikos. Wenn vorhanden, dann Kennzeichnung mit dem Wert WAHR. Die Prüfung des logischen Ausdrucks *RisikoExtra*

$$\{\text{Zigarettenverbrauch} > 5\} \text{ ODER } \{\text{BMI} > 30\} \text{ ODER } \{\text{Diabetiker} = 1\}$$

liefert WAHR oder FALSCH.

**Schritt 3:** Ermittlung eines weiteren Risikofaktors (*Risiko2*) in Verbindung von Extrarisiko mit Cholesterinwert  $> 200$ . Wenn vorhanden, dann Kennzeichnung mit dem Wert WAHR. Die Prüfung des logischen Ausdrucks *Risikofaktor2*

$$\{\text{Cholesterin} > 200\} \text{ UND } \{\text{Extrarisiko} = \text{WAHR}\}$$

liefert WAHR oder FALSCH.

**Schritt 4:** Verbindung von *Risiko1* oder *Risiko2* durch farbige Kennzeichnung der Datensätze von Personen mit erhöhtem Risiko für Erkrankungen der Koronararterien.

### Lösungsschritte:

Erstellen Sie zuerst eine Kopie des Tabellenblatts **KG\_Untersuchungsdaten** (Rechtsklick auf das Tabellenblatt und Nutzung des Kontextmenüs (**Kopieren oder verschieben...**)).

**Schritt 1:** Prüfung Cholesterinwert  $> 240$

⇒ Als Spalte für die Kennzeichnung, wird die Spalte P gewählt und mit **Risiko1** (in der Zelle P1) beschriftet.

⇒ Die in P2 einzutragende Formel für das Beispiel lautet:  $=J2>240$

⇒ Das Beispiel wird für den Rest der Spalte P vervielfältigt.

**Schritt 2:** Ermittlung mindestens eines weiteren Risikofaktors

⇒ Als Spalte für die Kennzeichnung, dass mindestens ein weiterer Risikofaktor vorliegt, wird die Spalte Q gewählt und mit **RisikoExtra** (in der Zelle Q1) beschriftet.

⇒ Die in Q2 einzutragende Formel für das Beispiel lautet:  $=\text{ODER}(O2>5;E2>30;I2=1)$

⇒ Das Beispiel wird für den Rest der Spalte Q vervielfältigt.

**Schritt 3:** Prüfung der Verbindung Extrarisiko mit Cholesterinwert  $> 200$

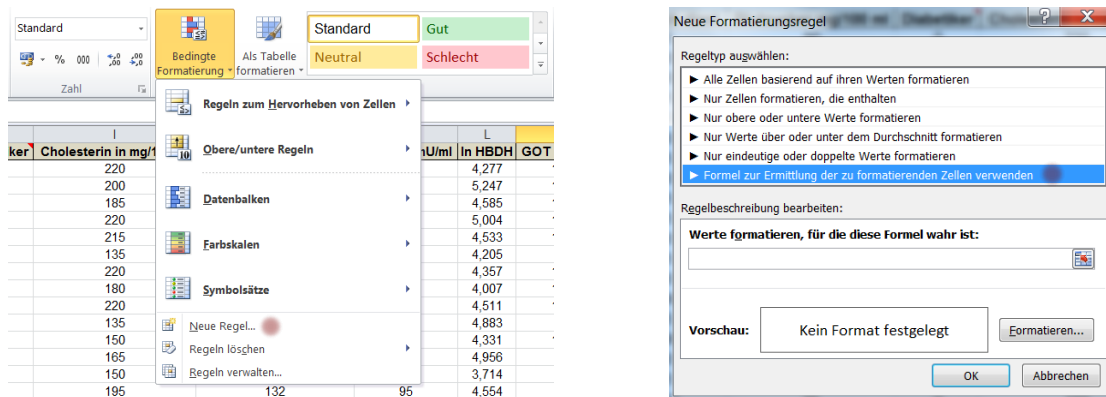
⇒ Als Spalte für die Kennzeichnung, dass Risikofaktor 2 vorliegt, wird die Spalte R gewählt und mit **Risiko2** (in der Zelle R1) beschriftet.

⇒ Die in R2 einzutragende Formel für das Beispiel lautet:  $=\text{UND}(J2>200;Q2)$ . Hinweis: logische Zellinhalte müssen nicht extra auf WAHR oder FALSCH getestet werden, weil das Ergebnis der Testung wieder „nur“ WAHR oder FALSCH ergeben würde. Damit ist  $=\text{UND}(J2>200;Q2=\text{WAHR})$  gleichbedeutend mit der kürzeren Schreibweise  $=\text{UND}(J2>200;Q2)$ .

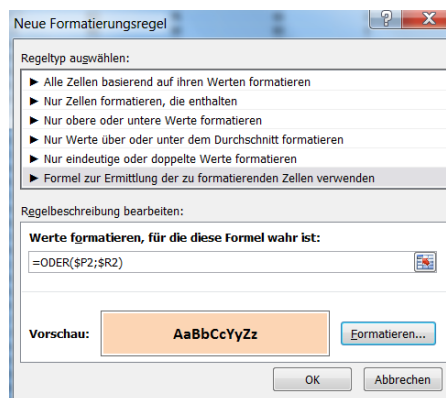
⇒ Das Beispiel wird für den Rest der Spalte R vervielfältigt.

**Schritt 4:** Bedingte Formatierung in Abhängigkeit von Risiko1 und Risiko2.

- ⇒ Der gesamte Tabellenbereich (ohne Überschriften) A2:Q41 wird markiert.
- ⇒ Danach wird in der Start-Navigationsleiste **Bedingte Formatierung, Neue Regel...** und **Formel zur Ermittlung der zu formatierenden Zellen verwenden** ausgewählt.



- ⇒ Als Formel zur Prüfung wird **=ODER(\$P2;\$R2)** eingetragen. Wichtig sind die Mischformen zwischen absoluten und relativen Zellbezügen auf die zuvor erstellten Spalten P und R (absoluter Verweis auf die Spalten, relativer Verweis auf Zeilen). Die Formatierung der Datensätze kann individuell gestaltet werden (Schalter **Formatieren...**):



- ⇒ Alle *bedingten* Formatierungen kann man nach Markierung des betreffenden Tabellenbereiches unter **Regeln verwalten ...** anschauen, verändern und löschen.

## 4.9 Nutzung von Datumsfunktionen

### Allgemeines zur Nutzung von Datumsfunktionen:

**DATEDIF**(Anfangsdatum; Enddatum; Kennung) ist eine häufig benötigte Datumsfunktion zur Berechnung von Datumsdifferenzen in verschiedenen Einheiten. Sie nimmt aber eine Sonderstellung ein, da sie durch den Funktionsassistenten nicht unterstützt und in der Excel-Hilfe nicht aufgeführt wird. Die Kennung ist als Abkürzung der englischen Worte für Tag (D) und Jahr (Y) einzugeben.

Differenz in Tagen	=DATEDIF(A1;A2;"D")
Differenz in Monaten	=DATEDIF(A1;A2;"M")
Differenz in Jahren	=DATEDIF(A1;A2;"Y")
Differenz in Tagen im selben Jahr	=DATEDIF(A1;A2;"YD")
Differenz in Monaten im selben Jahr	=DATEDIF(A1;A2;"YM")
Differenz in Tagen im selben Jahr/Monat	=DATEDIF(A1;A2;"MD")

**Datumsfunktion und Datumsformatierung:** Datumsfunktionen können Datumsangaben erzeugen, zerlegen sowie damit rechnen. Dabei wird mit Adressen gearbeitet bzw. das Datum mit Anführungsstrichen ("11.04.2016") in die Funktion oder Formel eingegeben. Bei der Verwendung von Datumsfunktionen ist zu beachten, dass diese sehr häufig im Zusammenhang mit einer benutzerdefinierten Datumsformatierung eingesetzt werden. Dies bedeutet, dass die Felder mit den Datumsfunktionen zusätzlich formatiert werden müssen, um das gewünschte Ergebnis zu erhalten (siehe Aufgabe 3.2).

**Aufgabe 4.9:** Für das aktuelle Tabellenblatt sind das Alter in Jahren bei der Untersuchung und ein Bestelltermin in 12 Wochen nach der Untersuchung zu berechnen.

**Schritt 1: Alter bei Untersuchung (in Jahren):**

- ⇒ Die Spalte S wird in der Zelle S1 mit **Alter** beschriftet.
- ⇒ Die Funktion **DATEDIF** kann zur Bestimmung der Altersdifferenz verwendet werden und wird exemplarisch in die Zelle S2 geschrieben. Dabei sind das Geburtsdatum (Spalte B) als Anfangsdatum und das Untersuchungsdatum (Spalte D) als Enddatum zu verwenden. Durch die Wahl der Kennung "Y" wird die Differenz in Jahre umgerechnet: **=DATEDIF(B2;D2;"Y")**
- ⇒ Das Beispiel wird für den Rest der Spalte S vervielfältigt.

**Schritt 2: Bestelltermin in 12 Wochen:**

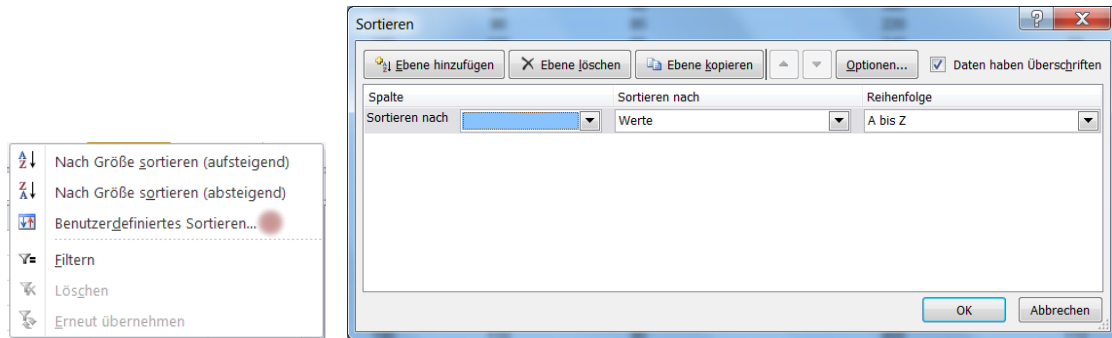
- ⇒ Die Spalte T wird in der Zelle T1 mit Bestelltermin beschriftet.
- ⇒ Die exemplarische Berechnung des Bestelltermins erfolgt in der Zelle T2 wieder mittels einer Formel. Bei Addition einer Zahl zu einem Datum wird diese als Anzahl von Tagen addiert: **=D2+7\*12**. Alternativ kann die Anzahl der Wochen auch analog zu Aufgabe 3.5 variabel gestaltet werden. Hierzu müsste man diese in einer Zelle eines Tabellenblatts speichern und in der Formel zur Bestimmung des Bestelltermins mit einem absoluten Zellbezug auf den Inhalt dieser Zelle zugreifen.
- ⇒ Formatierung der Zelle als **Datum**.
- ⇒ Vervielfältigung der Formel auf den Rest der Spalte T.

## 4.10 Datensätze sortieren

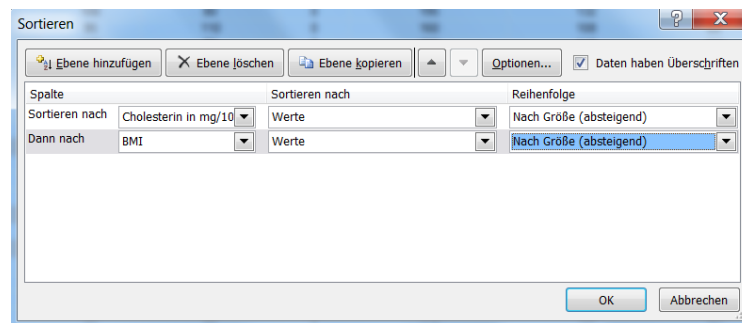
**Aufgabe 4.10:** Die Datensätze sollen nun absteigend nach Cholesterin (primäres Sortierkriterium) und BMI (sekundäres Sortierkriterium) sortiert werden. Dabei ist sicherzustellen, dass die Zuordnung aller Beobachtungen zu den einzelnen Personen erhalten bleibt.

**Lösungsschritte:**

- ⇒ Empfehlenswert ist es, den gesamten Datenbereich (A1:R41) inklusive Spaltenüberschriften zu markieren. Es ist dabei wichtig, dass nicht nur die Spalten J (Cholesterin) und E (BMI) markiert werden, da sonst beim Sortieren die Zuordnung aller Beobachtungen zu den einzelnen Patienten zerstört wird. Alternativ kann auf die Markierung der ersten Zeile (Spaltenüberschriften) verzichtet werden. Dann erscheinen die Spaltenüberschriften jedoch nicht mit bei der nachfolgenden Definition der Sortierregeln, was insbesondere bei Tabellen mit vielen Spalten von Nachteil ist.
- ⇒ Danach wird über **Start** ⇒ **Sortieren und Filtern** ⇒ **Benutzerdefiniertes Sortieren...** der Dialog zu Sortiereinstellungen geöffnet.



- ⇒ Man sollte in diesem Dialog prüfen, ob die Option **Daten haben Überschriften** richtig voreingestellt ist.
- ⇒ Da zwei Sortierkriterien verwendet werden sollen, muss eine Sortierebene hinzugefügt werden: **Ebene hinzufügen**.
- ⇒ In der ersten Sortierebene wird **Cholesterin... / Werte / nach Größe (absteigend)** eingestellt.
- ⇒ In der zweiten Sortierebene wird **BMI / Werte / nach Größe (absteigend)** ausgewählt.



### 4.11 Datenimport aus Textdateien

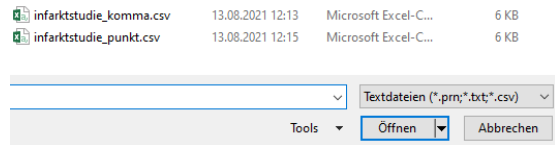
In der Praxis liegen tabellarisch organisierte Daten häufig bereits in Textdateien vor und müssen nur noch in Tabellenblätter importiert werden. Je nachdem, welche Dateitypen (z. B. csv, txt) mit dem Tabellenkalkulationsprogramm des Computers verknüpft sind, kann das bereits mit einem Doppelklick auf die jeweilige Textdatei im Explorer erfolgen. Dieses Vorgehen ist jedoch nicht zu empfehlen, da es abhängig vom Dateityp zu fehlerhaften Importen kommen kann. Insbesondere hängen die Ergebnisse dieser Methode von der Ländereinstellung des Betriebssystems des Computers ab. Sicherer ist das Öffnen von Textdateien aus dem Tabellenkalkulationsprogramm heraus, da sich dann (bei Excel abhängig vom Dateityp) ein Importassistent öffnet, mit dem sich formatspezifische Adjustierungen vornehmen lassen.

***Aufgabe 4.11:*** *Strukturierte Daten in Form von kommaseparierten Werten (typische Dateierweiterung: csv) sollen importiert werden.*

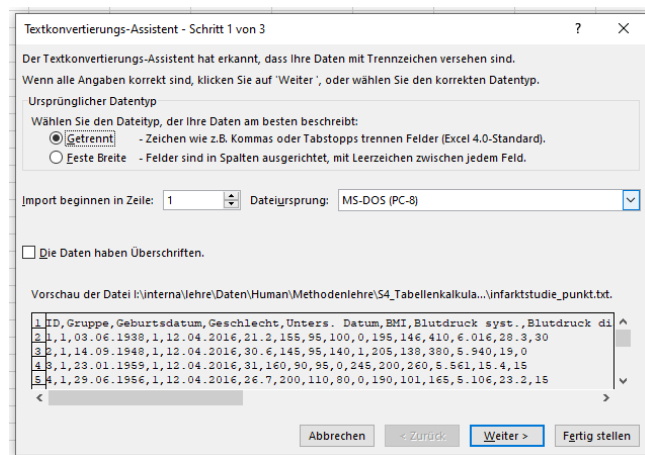
#### **Lösungsschritte:**

- ⇒ Die bisher verwendeten Studiendaten liegen auch als Textdateien in zwei Varianten vor. In der Datei „**infarktstudie\_komma.csv**“ ist das Dezimaltrennzeichen ein Komma und der (Spalten-) Separator ein Semikolon; in der Datei „**infarktstudie\_punkt.csv**“ wurden dafür Punkt bzw. Komma verwendet. Öffnen Sie die Datei „**infarktstudie\_punkt.csv**“, wenn die Ländereinstellung des

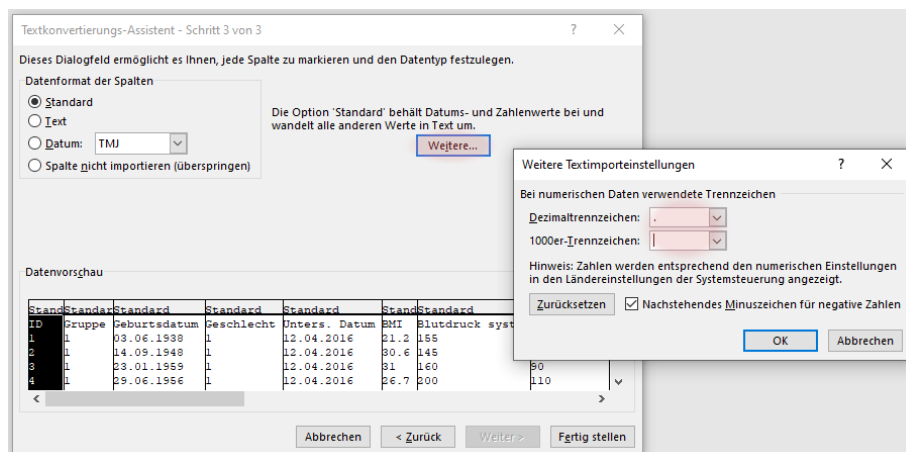
Computers „Deutsch“ ist, sonst „**infarktstudie\_komma.csv**“. Wählen Sie dafür in der Dropdown-liste ‚Dateityp‘ im Dialogfeld ‚Öffnen‘ die Option Textdateien aus.



- ⇒ Wenn in Excel eine csv-Datei geöffnet wird, werden automatisch die aktuellen Standardeinstellungen für Datenformate entsprechend der aktuellen Ländereinstellung verwendet. Der Textimport-Assistent wird nicht gestartet. Wenn Sie die Verwendung des Assistenten zum Importieren von Text erzwingen möchten (z. B. weil das Importergebnis sonst fehlerhaft ist), können Sie vor dem Öffnen der Textdatei die Dateierweiterung von ‚csv‘ in z.B. ‚txt‘ ändern. Dieses Vorgehen wird vom Microsoft Office-Support empfohlen. Eine alternative Methode wird am Ende dieses Abschnittes beschrieben.
- ⇒ Nennen Sie die csv-Datei in \*.txt um und öffnen Sie die umbenannte Datei „infarktstudie\_\*.txt“. Der Textimport-Assistent öffnet sich.



- ⇒ Wählen Sie im ersten Schritt des Import-Assistenten, wie die Spalten in der Textdatei separiert sind (hier: durch Kommas „getrennt“)
- ⇒ Wählen Sie im zweiten Schritt das/die Spaltentrennzeichen (hier: Komma)
- ⇒ Im dritten Schritt kann man bei Bedarf Trennzeichen von numerischen Daten anpassen (hier: Punkt als Dezimaltrennzeichen)



**Bemerkung 1:** Nach dem Import empfiehlt es sich, insbesondere Spalten mit Zahlen- und Datumsangaben zu überprüfen.

**Bemerkung 2:** Der Textimport-Assistent kann auch manuell gestartet werden, nachdem Daten importiert bzw. eingefügt worden sind. Das setzt allerdings voraus, dass sich alle zu konvertierenden Daten in einer einzelnen Spalte des Tabellenblatts befinden, da die Konvertierung nur für eine Spalte auf einmal durchgeführt werden kann.

Um den Import-Assistenten zu starten, markiert man zuerst die zu konvertierende Spalte und wählt danach im Menüband **Daten** den Eintrag **Text in Spalten**. Alles Weitere folgt dann entsprechend der zuvor beschriebenen Schritte.

The screenshot shows the Microsoft Excel interface with the 'Daten' (Data) ribbon selected. The 'Text in Spalten' (Text to Columns) button is highlighted. Below the ribbon, a data table is displayed with the following content:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
1	ID,Gruppe	Geburtsdatum	Geschlecht	Unters.	Datum	BMI	Blutdruck syst.	Blutdruck diast.	Blutzucker mg/100 ml	Diabetiker	Cholesterin in mg/100 ml	Triglyzerid in mg/100 ml	HBDH in mU/ml	In HBDH	GOT in mU/ml	Zigaretten pro Tag		
2	1	1.03.06.1938	1	12.04.2016	21	2,155	95	100	0	195	146	410	6	916	28	3	30	
3	2	1.14.09.1938	1	12.04.2016	30	6,145	95	140	1	205	138	380	5	940	19	0		
4	3	1.23.01.1939	1	12.04.2016	31	169	90	95	0	245	200	260	5	561	15	4	15	
5	4	1.29.06.1945	1	12.04.2016	26	7	200	110	80	0	190	101	165	5	106	23	2	15
6	5	1.07.08.1942	1	12.04.2016	24	2	190	110	90	0	260	202	300	5	704	20	6	5
7	6	1.15.03.1943	1	16.04.2016	26	195	110	90	0	190	164	290	5	670	18	7	20	
8	7	1.05.10.1937	1	16.04.2016	30	8	220	105	70	0	340	150	605	6	405	33	1	0



## 5 Datenmanagement mit R

### *Motivation*

Im Seminar „Tabellenkalkulation“ werden Funktionalitäten des Excel-Tabellenblattes vorgestellt und im Seminar „Datenbanken“ werden Grundideen und Prinzipien von Datenbanken umrissen. In der biomedizinischen Forschung entstehen jedoch zunehmend so große Datensätze, dass die Darstellung und Bearbeitung mittels Excel-Tabellenblatt oft nicht mehr funktioniert. Datenbanksysteme können zwar gut zur Speicherung und Zusammenführung sehr großer Datenmengen genutzt werden, aber die eigentliche Analyse der Daten ist damit nur bedingt möglich. An dieser Stelle möchten wir Ihnen daher einen Einstieg in die freie Programmiersprache R bieten, die es ermöglicht, genau diese Lücke zu schließen. R erlaubt den Umgang mit größeren Datenmengen als Excel und bietet weit mehr Möglichkeiten zur professionellen Datenanalyse, Statistik und Datenvisualisierung. R ist kostenfrei nutzbar.



### *Lernziele*

Nach dem Kurs können Sie

- Daten für die anschließende Verarbeitung in R laden und Daten aus R speichern
- Sich in RStudio (einem komfortablen Zusatzprogramm zur Nutzung von R) orientieren und dort grundlegende Funktionen nutzen
- Erklären, wie ein in R geschriebenes Computerprogramm aufgebaut ist
- Eigene Funktionen deklarieren und daran erklären, was eine Funktion ist
- Kleine R-Programme formulieren
- Erklären, was ein R-Package ist und R-Packages zur Erweiterung von R nutzen
- Datentabellen in R verwalten, filtern und verknüpfen

### *Verfügbarkeit von R*

R entstand ab 1992 an der Universität Auckland (Ross Ihaka & Robert Gentleman) und wird inzwischen von einer größeren Gruppe von Software-Entwicklern (R-Core Team) weltweit weiterentwickelt. R ist **kostenfrei nutzbar** und in weiten Teilen **quelloffen**, d. h. Interessierte können den Programmcode von R ansehen und weiterentwickeln (lizenziert unter der **GNU General Public License**). R ist in **allen gängigen Betriebssystemen** (Windows, MacOS, Linux) nutzbar.

### *Paketweise Erweiterbarkeit: R-Packages*

R lässt sich um viele weitere Funktionen erweitern. Auch diese Erweiterungen entstehen durch weltweite Arbeit von Freiwilligen. Erweiterungen werden als **Pakete (R-Packages)** angeboten. Solche Pakete enthalten Dateien mit den zusätzlichen Funktionen sowie elektronische Dokumentation dazu. Sie finden große Sammlungen von R-Erweiterungen im Internet: Am wichtigsten sind das „Comprehensive R Archive Network“ (CRAN) sowie „Bioconductor“.

Für das **Laden und Einbinden von R-Packages** gibt es **eigene Befehle in R**, die alles Nötige automatisch erledigen, wenn Ihr Rechner mit dem Internet verbunden ist.

### R-Interpreter

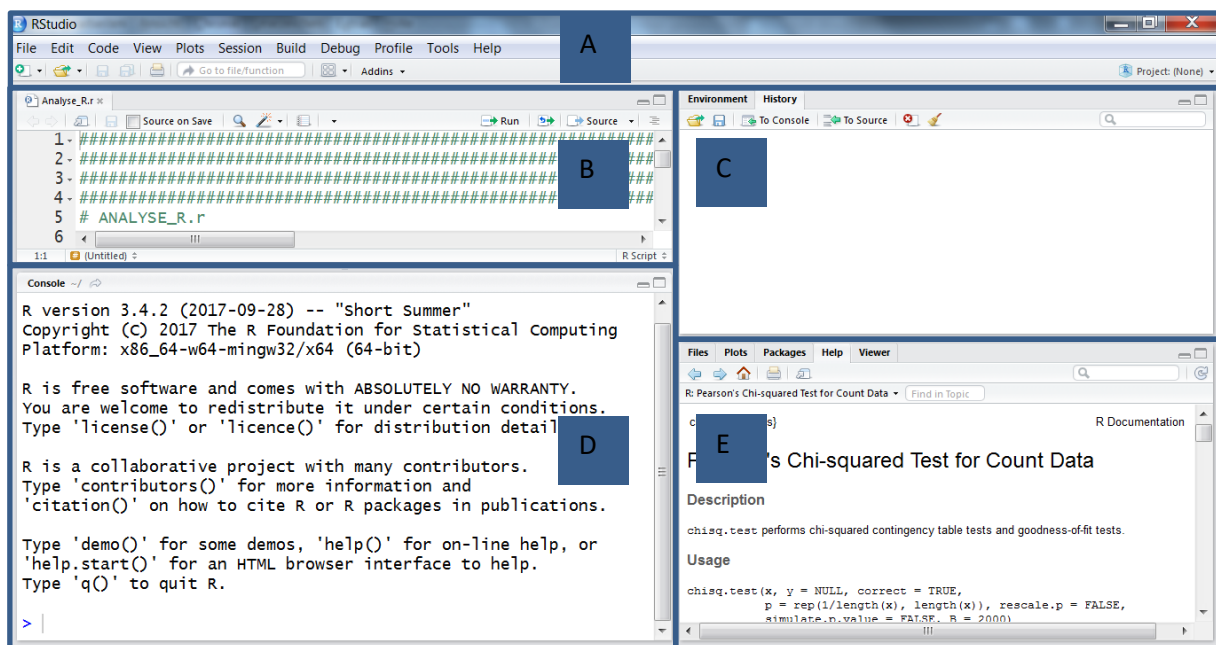
R ist (wie hier schon mehrfach geschrieben) eine **Programmiersprache**. Damit ein Computer diese verstehen kann, gibt es eigene Computerprogramme, die als Übersetzer dienen. Diese übersetzen, was Sie in der Programmiersprache schreiben in (für Menschen äußerst schlecht lesbare) Befehle, die der Rechenkern des Computers versteht. Das kann quasi wie bei einer **Simultanübersetzung** geschehen, dann heißen die Übersetzer **Interpreter**. Übersetzer, die stattdessen alles auf einmal übersetzen und am Ende eine Übersetzung abliefern, welche später zum Abarbeiten des Programms genutzt werden kann, nennt man stattdessen Compiler.

Im **Falle von R** werden Ihre Programme bzw. Befehle durch einen Interpreter, den **R-Interpreter**, verarbeitet.

### R Studio

Sie lernen im Kurs das Programm RStudio kennen. RStudio eine **Entwicklungsumgebung**. Diese erleichtert das Erstellen von Programmen in R. RStudio hilft auch bei der Verwaltung der Dateien, in denen Programme und Daten gespeichert sind. Vor allem aber ist RStudio sehr bequem, um Ihr R-Programm an den "Simultanübersetzer", den **Interpreter von R**, zu übergeben und sich Ergebnisse anzusehen.

RStudio ist folgendermaßen aufgebaut:



- A (Kopfzeile): **Hauptmenü**
- B (links oben): **Script-Editor** (d. h. Schreibbereich) für R-Skripte, d. h. Programme, die aus vielen R-Befehlen bestehen und in Dateien zur Wiederverwendung gespeichert werden können
- C (rechts oben): Umgebungsbereich („Environment“): Auflistung der eingegebenen Daten, Funktionen und Werte; „History“: die zuletzt ausgeführten Befehle
- D (links unten): **R-Konsole**: Sie können dort einzelne Befehle eingeben und sehen R-Ausgaben

- E (rechts unten): „Files“: Ordnerstruktur, „Plots“: GrafikbildschirmAusgabe, „Packages“: Installieren von CRAN Paketen, „Help“: Hilfetextanzeige zu z. B. Funktionen, „Viewer“: Anzeigen von geladenen Grafiken

### *Hinweis: So bearbeiten Sie die Übungen*

Die Übungen kombinieren Anleitungstext und Bereiche, in denen Sie R-Befehle eintragen und ausführen können. Bereiche mit R-Code (sog. “Chunks”) sind grau hinterlegt und lassen sich ausführen, indem Sie auf das kleine grüne Dreieck in der rechten oberen Ecke des Chunks klicken.

Diese Anleitung existiert in einer Lang- und einer Kurzfassung. Die Langfassung enthält ausführliche Erklärungen und ermöglicht es, den Stoff zu Hause selbstständig durchzuarbeiten. Die Kurzfassung wird während des Kurstermins bearbeitet. Dort bekommen Sie die nötigen Erläuterungen jeweils durch den/die Kursdozenten/in, bevor Sie die Aufgabe bearbeiten.

## 5.1 Tabellen, Variablen und Spalten

In statistischen Anwendungen geht es in der Regel um Analysen von Daten, die aus mehreren Beobachtungen verschiedener Variablen bestehen. In dieser Übung werden wir beispielsweise einen Datensatz aus einer Studie zu fetalen Herztönen<sup>1</sup> betrachten, der fetale Herztonmessungen sowie einige demographische Variablen von 60 schwangeren Frauen enthält.

Daten wie diese werden in einer Tabelle zusammengefasst, in der jede Zeile eine Beobachtung (= eine Frau) und jede Spalte eine Eigenschaft (z.B. Alter, Gewicht etc.) darstellt. Unsere Übungstabelle befindet sich in der Datei “fetalheartsound.csv”, die wir erst einlesen müssen. Zunächst wird die import Funktion aus dem Menü zu diesem Zweck genutzt: *File -> Import Data Set -> From text (readr)*. Im Popup-Fenster wählen sie die einzulesende Datei aus: *Browse -> Datei im Explorer suchen und anklicken -> Gewünschten Name unten links angeben (hier: fetalheartsound)*.

Mittels des vergebenen Namens (Variable) kann von nun innerhalb von R auf die Daten zugegriffen werden. Außerdem findet sich die Variable auch im “Environment”-Fenster auf der rechten Seite. Führen Sie den folgenden Code-Chunk (grau hinterlegten Bereich) aus, indem Sie auf das grüne Dreieck in der rechten oberen Ecke klicken um zu sehen, was sich hinter `fetalheartsound` verbirgt.

```
print(fetalheartsound)
```

Statt über das Menü, kann das Einlesen von Daten kann auch direkt im Code erfolgen: Um mit der Tabelle arbeiten zu können, haben wir sie der Variablen `fetalheartsound` zugewiesen. Wenn wir dies im Code machen, verwenden für die sog. *Zuweisung* ein Pfeilsymbol (`<-`). Es wird aus dem “kleiner“-Symbol (`<`) und dem Minus (`-`) zusammengesetzt und zeigt *in umgekehrter Leserichtung*, d.h. von den Daten auf die Variable, der die Daten zugewiesen werden [Beachten Sie, dass diese Zuweisung auch mit einem einfachen “=” zwischen Variablenname (links) und Objekt (rechts) funktioniert und in vielen Tutorials, Dokumentationen etc. so genutzt wird]. Sobald ein Objekt einer Variablen zugewiesen ist, können wir die Variable oben rechts im “Environment” sehen.

---

<sup>1</sup> Bhaskaran, A., & Arora, M. (2022). Indian Institute of Science Fetal Heart Sound Database (IIScFHSDB) (version 1.0). *PhysioNet*.

Der folgende Codeblock liest den fetalheartsound-Datensatzes ein und speichert ihn in der Variablen fetalheartsound:

```
library("readr")
fetalheartsound <- read_csv("fetalheartsound.csv", show_col_types = FALSE)
```

Für konkrete Analysen arbeiten wir oft nicht mit der ganzen Tabelle, sondern mit einzelnen Spalten. In unserer Tabelle könnten wir uns beispielsweise für die Spalte Age interessieren, für die wir z.B. einen Mittelwert oder eine Standardabweichung berechnen könnten. Eine einzelne Spalte einer Tabelle wird in R als *Vektor* bezeichnet. Ein Vektor ist im Grunde einfach eine Reihe von Zahlen, in unserem Beispiel die Reihe aller Altersangaben der 60 Probandinnen. Da Vektoren ein Grundbaustein der meisten statistischen Verfahren sind, ist R darauf spezialisiert mit solchen Vektoren zu rechnen, wie wir im Folgenden sehen werden.

### *Aufgabe 5.1.1: Zugriff auf einzelne Spalten einer Tabelle*

Um mit einzelnen Spalten aus der Tabelle rechnen zu können, müssen wir sie zunächst aus der Tabelle extrahieren. Die Auswahl einer Spalte funktioniert in R mit dem Dollar-Zeichen (\$). Führen Sie den untenstehenden Chunk aus um zu sehen, dass Sie einen Vektor mit Werten (den Altersangaben) der Probandinnen erhalten. Ändern Sie dann den Spaltennamen Age zu einem der anderen Spaltenbezeichnungen, um auf eine andere Spalte zuzugreifen und testen sie das Verhalten von R bei der Eingabe eines unbekanntem Spaltennamens.

```
fetalheartsound$Age
```

## 5.2 Rechnen mit Tabellenspalten

Auf einem Vektor aus Zahlen können Sie in R die Grundrechenoperationen ausführen wie auf Skalaren (= einzelnen Zahlen). Das Rechnen mit Vektoren funktioniert dabei üblicherweise elementweise.

### *5.2.1 Aufgabe: Grundrechenarten*

Ändern Sie den Rechenoperator von + zu - oder / und führen Sie den Chunk aus.

```
fetalheartsound$Age + 3
```

### *5.2.2 Aufgabe: Umrechnen von Einheiten*

Lesen Sie die Spalte Height aus unserer Tabelle aus, und rechnen Sie alle Größen in Zoll um (Hinweis: 2,54 Zentimeter sind etwa 1 Zoll). Achtung: Zahlen werden im Englischen durch einen Dezimalpunkt getrennt statt wie im Deutschen durch ein Komma.

```
# hier die Lösung einfügen und ausprobieren
fetalheartsound$Height / 2.54
```

### *5.2.3 Aufgabe: Kombination von Spalten*

Nutzen Sie die Tabelle, um für alle Patienten den BMI zu berechnen. (Der einfache BMI ist definiert als Masse (in kg) durch Größe (in m) zum Quadrat). Weisen sie den Ergebnisvektor der Variable mit dem Namen bmi zu.

```
# hier die Lösung einfügen und ausprobieren
```

```
bmi <- fetalheartsound$Weight / ((fetalheartsound$Height/100) * (fetalheartsound$Height/100))
```

Ein einzelnes Element aus einem Vektor können wir mit eckigen Klammern auswählen (also [ und ]):

```
bmi[2] # gibt das zweite Element aus
```

### 5.2.4 Aufgabe: Definition neuer Spalten

Um eine neue Spalte in einer Tabelle anzulegen wird auch der Zuweisungsoperator (<-) verwendet. Anstatt eines neuen Variablennamens wird jedoch ein neuer Spaltenname mit der \$-Notation verwendet. ACHTUNG: Wenn Sie einen bereits verwendeten Spaltennamen verwenden, wird diese Spalte überschrieben. Beispiel: Die Spalte `InchHeight` wird neu definiert.

```
# hier die Lösung einfügen und ausprobieren
```

```
fetalheartsound$InchHeight <- fetalheartsound$Height / 2.54
```

### 5.2.5 Aufgabe: Definition neuer Spalten in einer Tabelle

Schreiben Sie das Ergebnis aus Ausgabe 2.3 in eine neue Spalte "BMI" der Tabelle `fetalheartsound`.

```
fetalheartsound$BMI <- bmi
```

Lassen Sie sich den Datensatz erneut anzeigen um sich davon zu überzeugen, dass die beiden neuen Variablen tatsächlich angelegt wurden:

```
print(fetalheartsound)
```

## 5.3 Funktionen

Meist werden mehrere aufeinander folgende Befehle benötigt, um reale Probleme zu lösen. Falls eine solche Verkettung von Befehlen häufiger verwendet werden soll, ist es sinnvoll diese als *Funktion* zusammenzufassen. Eine Funktion gleicht dabei einem Automaten, in den etwas eingeworfen wird und der dann etwas auswirft. Dem Einwurf entsprechen die *Argumente der Funktion*, ausgeworfen wird die *Rückgabe*.

Sie haben oben schon einige Funktionen angewendet (`print`, `c`), welche bereits in R vorhanden sind. Wir werden später noch mehr kennen lernen.

### Beispiel:

Im Folgenden definieren wir eine Funktion, die das Quadrat einer Zahl berechnet. Dazu vergeben wir einen neuen Namen (hier: *quadrat*), dem wir im Anschluss eine Funktion mit *function* zuweisen. Der Variablenname in den anschließenden runden Klammern ( ) wird auch *Übergabeparameter* genannt. Er ist zunächst ein Platzhalter, dem wir erst bei Aufruf der Funktion einen konkreten Wert (Argument) zuweisen. In den darauffolgenden geschweiften Klammern { } wird dann definiert, was die Funktion mit dem Übergabeparameter machen soll. Im untenstehenden Beispiel wird der Parameter `x` mit sich selbst multipliziert (quadriert) und in der Variable `res` gespeichert. Anschließend wird das Ergebnis `res` mit dem Befehl `return` ausgegeben. *Achtung*: Bei der Definition wird die Berechnung noch nicht

ausgeführt, sondern wie der Name schon sagt, erstmal nur definiert, dementsprechend erzeugt das Ausführen des Chunks noch keinen sichtbaren Output:

```
quadrat <- function(x)
{
  res <- x * x
  return(res)
}
```

Erst wenn wir die Funktion mit einem konkreten Wert für `x` aufrufen, erhalten wir ein Ergebnis:

```
quadrat(x = 4)
```

### 5.3.1 Aufgabe: Funktionsaufruf mit unterschiedlichen Argumenten

Ändern Sie das Argument und führen Sie den obenstehenden Chunk erneut aus. Da im Funktionskörper (also allem was innerhalb der geschweiften Klammern steht) nur Operatoren vorkommen, die auch auf Vektoren ausgeführt werden können, kann die Funktion auch mit einem Vektor als Argument ausgeführt werden. Übergeben sie den `bmi` Vektor als Argument und führen Sie den Chunk erneut aus.

### 5.3.2 Aufgabe: Funktionsdeklaration

Unten sehen Sie eine unvollständige Funktionsdefinition. Die Funktion soll auf Basis der beiden Parameter `weight` und `height` den BMI berechnen. Vervollständigen Sie den Funktionskörper (Die Lösung von 5.2.3 kann hier wiederverwendet werden). Gehen Sie davon aus, dass das Gewicht in kg und die Größe in cm angegeben wird.

```
berechne_bmi <- function(weight, height)
{
  h2 <- quadrat(height/100)
  res <- weight/h2
  return(res)
}
```

Rufen Sie nun die Funktion mit den beiden Spalten aus der Tabelle auf

```
berechne_bmi(fetalheartsound$Weight, fetalheartsound$Height)
```

## 5.4 Funktionen aus dem Bestand von R

R bringt sehr viele Funktionen mit, die andere Programmierer bereits geschrieben haben. Diese müssen Sie lediglich mit den von Ihnen gewünschten Daten aufrufen.

In der praktischen Anwendung müssen Sie daher selten Funktionen selbst programmieren, sondern meist einfach bestehende Funktionen/Befehle in sinnvoller Reihenfolge ausführen. Ihr Code ist dann praktischerweise gleich eine vollständige Dokumentation Ihrer statistischen Auswertung, die von anderen (z.B. einem Betreuer oder Berater) leicht nachvollzogen werden kann.

### 5.4.1 Aufgabe: Mathematische Funktionen

Verwenden Sie die in R bereits vorhandene Funktion `sqrt` um die Wurzel der Zahl 42 zu berechnen und ausgeben zu lassen.



```
sqrt(42)
```

### 5.4.2 Aufgabe: Hilfe zur Funktionsverwendung

Nutzen sie den Befehl `?<Funktionsnamen>` vor einem Funktionsnamen, um dessen Verwendung erklärt zu bekommen. Das Ergebnis erscheint rechts im Fenster unter "Help". Wenden Sie den Befehl auf die Funktion mit Namen `mean` an.

```
?mean
```

### 5.4.3 Aufgabe: Einfache Statistische Funktionen

Berechnen Sie Mittelwert und Varianz aller (Mütter-) Alterswerte aus der Beispieltabelle `fetalheartsound`.

Lösungsschritte:

1. Rufen Sie die Funktion `mean` mit der Variable `Age` aus dem `fetalheartsound`-Datensatz als Argument auf. Erinnern Sie sich dabei an die `$`-Notation um die Variable aus dem Datensatz zu extrahieren.
2. Rufen Sie die Funktion `var` ebenfalls mit diesem Argument auf

```
mean(fetalheartsound$Age)
```

```
var(fetalheartsound$Age)
```

### 5.4.4 Aufgabe: Stichprobe aus einer Verteilung ziehen

Lassen Sie sich mit Hilfe der Funktion `rnorm` fünf zufällige Zahlen aus einer Menge normalverteilter Zahlen ausgeben. (Hinweis: Wenden Sie zuvor die Hilfe auf `rnorm` an, um zu erfahren, wie die Funktion zu verwenden ist.)

```
?rnorm
```

```
rnorm(100, mean=0, sd=1)
```

### 5.4.5 Aufgabe: Einfache Grafiken

Grafiken helfen, Ergebnisse besser darzustellen und Zusammenhänge intuitiv begreifen zu können. Nutzen Sie die Funktion `hist`, um die Verteilung von 1000 zufälligen normalverteilten Zahlen zu veranschaulichen.

```
sample = rnorm(100)
```

```
hist(sample)
```

## 5.5 Bedingte Anweisungen

Programme sollten nicht nur stur ein paar Rechenoperationen nacheinander ausführen, sondern flexibler auf Daten reagieren können. Der erste Schritt dazu ist die Nutzung sogenannter bedingter Anweisungen: Ein Befehl wird nur dann ausgeführt, wenn eine bestimmte Bedingung erfüllt ist, sonst nicht. Wir wollen in unserer Tabelle nun die Probandinnen analysieren, die aufgrund ihres Alters eine Risikoschwangerschaft haben, das sind alle Schwangeren ab 35 Jahren. Dafür definieren wir eine Funk-

tion `risiko`, die prüft, ob eine Probandin mindestens 35 Jahre alt ist und in diesem Fall die Information "Risikoschwangerschaft" zurückgibt. Liegt das Alter unter 35 Jahren, gibt die Funktion den Wert "Keine Risikoschwangerschaft" zurück.

Sie sehen in der Funktion eine etwas unübersichtliche Klammerstruktur der Form: `if(...){...} else {...}`: Zwischen den runden Klammern nach `if` wird die *Bedingung* geschrieben, zwischen das erste Paar geschweiften Klammern die Befehle, die nur ausgeführt werden, *wenn die Bedingung wahr ist* und zwischen das zweite Paar geschweiften Klammern die Befehle, die ggf. auszuführen sind, *wenn die Bedingung nicht zutrifft*.

### 5.5.1 Aufgabe: Bedingte Anweisung

Ergänzen Sie unten die Funktion, so dass die Bedingung `alter >= 35` geprüft wird.

```
risiko <- function(alter){  
  if(alter >= 35){  
    out <- "Risikoschwangerschaft"  
  }  
  else {  
    out <- "Keine Risikoschwangerschaft"  
  }  
  return(out)  
}
```

### 5.5.2 Aufgabe: Test der Funktion mit bedingter Anweisung

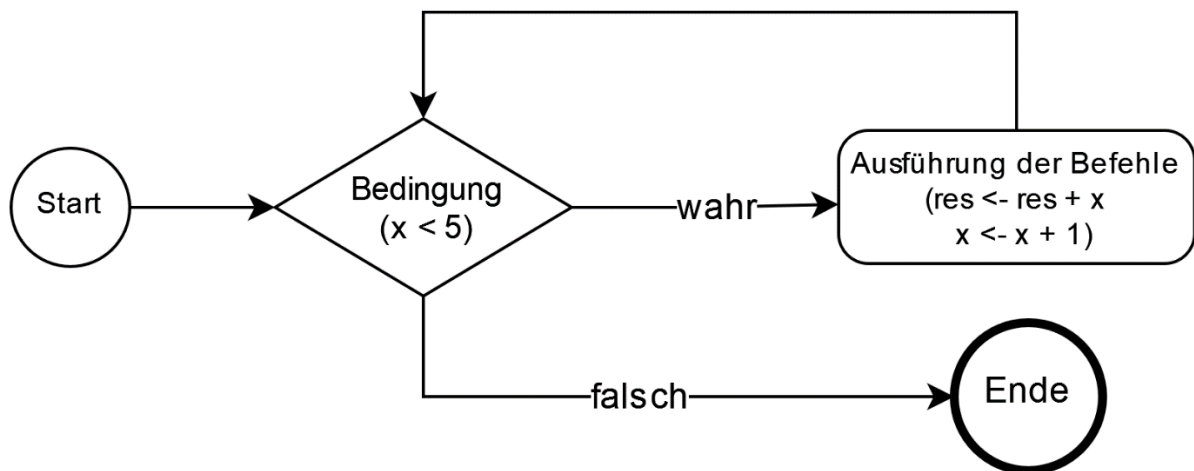
Überprüfen Sie mittels der oben definierten Funktion für verschiedene Werte von `alter`, ob es sich dabei um ein Alter für eine Risikoschwangerschaft handelt oder nicht.

```
risiko(32)
```

## 5.6 Wiederholschleifen

Schleifen erlauben es, einen Block von Anweisungen wiederholt ausführen zu lassen, bis eine bestimmte Bedingung *nicht* (mehr) erfüllt ist. Dies ermöglicht Problemlösungen, die auf tausende oder eine vorab unbekannte Anzahl an Wiederholungen angewiesen sind und somit unmöglich "per Hand" auszuführen sind. In diesem Kurs wird beispielhaft die "While"-Schleife eingeführt (es gibt allerdings auch noch weitere Schleifenarten). Im Folgenden finden Sie ein Beispiel einer solchen Schleife und die grafische Zusammenfassung ihres Ablaufs.

```
x <- 1  
res <- 0  
  
while (x < 5){  
  res <- res + x  
  x <- x + 1  
}
```



### 5.6.1 Aufgabe: Zufälliges Ziehen einer Person mit Risikoschwangerschaft

Der Befehl `sample_n(fetalheartsound, 1)` gibt Ihnen eine Tabelle mit nur einer zufällig aus `fetalheartsound` ausgewählten Zeile zurück.

```
library(dplyr)
sample_n(fetalheartsound, 1)
```

Passen Sie die folgende While-Schleife so an, dass nach der Schleifenausführung in der Variable `resWhile` eine solche einzeilige Tabelle steht, die zu einer *zufällig* ausgewählten Person *mit Risikoschwangerschaft* gehört. Wiederholen Sie also der Aufruf `sample_n(fetalheartsound, 1)` dafür so oft, bis die zurückgegebene Tabelle nicht der Bedingung (vgl. 5.5.1) entspricht.

```
resWhile <- sample_n(fetalheartsound, 1)
# Füge Sie die Bedingung in die runden Klammern ein.
while (resWhile$Age < 35){
  print("Neues Sample, das vorheriges nicht passt")
  resWhile <- sample_n(fetalheartsound, 1)
}
```

Wiederholen Sie die Ausführung und beachten Sie die unterschiedliche Anzahl der Durchläufe. Überprüfen Sie die Richtigkeit der Ausgabe durch Ausführung des folgenden Chunks.

```
print(resWhile)
```

## 5.7 R-Pakete nachladen

Um die Ladezeit gering zu halten, sind nicht alle denkbaren Funktionen direkt erreichbar. Reichen die Standardfunktionen aus R nicht aus, so kann man Funktionen nachladen. Diese sind thematisch in sogenannten Paketen zusammengefasst und beinhalten sogar die Ergebnisse aktueller Forschungsprojekte. Der folgende Befehl macht ein installiertes Paket in Ihrem Programm zugänglich:

```
library("PAKETNAME")
```

(Hinweis: Sie haben das oben bereits mit `"readr"` gemacht, um csv-Dateien lesen zu können.)

Ist ein Paket noch nicht installiert, so können Sie dies mit dem Befehl `install.packages("Paketname")` nachholen. In RStudio ist dies alternativ über das Menü möglich. Klicken Sie auf "Tools" -> "Install Packages", geben Sie den Paketnamen ein (Sie erhalten Suchvorschläge) und klicken Sie anschließend auf "Install Packages". Die Paketinstallation ist nur einmal nötig.

### 5.7.1 Aufgabe: Pakete laden

Sorgen Sie dafür, dass Sie die Pakete `tibble` und `dplyr` nutzen können. Eine Installation ist nicht nötig. Das ist bereits geschehen und würde hier zu lange dauern. (Lassen Sie sich nicht von der entstehenden Warnmeldung irritieren.)

```
library(tibble)
library(dplyr)
```

## 5.8 Auswahl in Tabellen

Als nächstes wollen wir zeigen, wie man Tabellen in R filtert, Spalten auswählt und mehrere Tabellen zusammenfügt. Als Beispiel dafür haben wir zwei Datensätze vorbereitet, die Informationen zu Phänotyp und Genexpression von 44 Probanden enthalten.<sup>2</sup> Sie finden diese Datensätze in den Dateien „Phaenotypdatei.csv“ und „Expressionsdatei.csv“. *Achtung:* Versuchen Sie nicht, diese Dateien außerhalb von R zu öffnen. Expressionsdatei.csv ist so groß, dass Programme wie Excel sich daran aufhängen können.

Wir beginnen mit der Auswahl bestimmter Spalten aus einem Datensatz. Hierfür wird die Funktion `select` verwendet.

### 5.8.1 Aufgabe: Spalten auswählen

Lesen Sie die Datei „Phaenotypdatei.csv“ als Variable `phaenoDaten` ein. Erzeugen Sie aus dieser Tabelle eine neue, welche nur die Spalten `Sample`, `Genotyp` und `Alter` enthält.

Lösungsschritte:

1. Lesen Sie die Daten wie bereits gelernt ein (`read_csv`).
2. Verwenden Sie die Funktion `select`, um Spalten auszuwählen.
3. Tragen Sie zwischen die runden Klammern erst den Namen `phaenoDaten` der Ausgangstabelle ein und danach (per Komma abgetrennt) einen mittels `c`-Funktion erzeugten Vektor mit den Namen der auszuwählenden Spalten (jeweils in Anführungszeichen).

```
phaenoDaten <- read_csv("Phaenotypdatei.csv", show_col_types = FALSE)
select(phaenoDaten, c("Sample", "Genotyp", "Alter"))
```

---

<sup>2</sup> Sturm D, Witt H, Hovestadt V, Khuong-Quang DA et al. Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* 2012;22(4):425-37.

### 5.8.2 Aufgabe: Zeilen auswählen

Filtern Sie aus `phaenoDaten` alle Zeilen, in denen in der Spalte `Alter` eine Zahl steht die größer als 15 ist.

Lösungsschritte:

1. Verwenden Sie die Funktion `filter` und tragen als erstes Argument den Namen des Datensatzes ein (hier: `phaenoDaten`).
2. Formulieren Sie als Bedingung `Alter > 15` und tragen Sie diese als zweites Argument ein. In der Bedingung sollten keine Anführungszeichen auftreten.

```
filter(phaenoDaten, phaenoDaten$Alter > 15)
```

## 5.9. Joins von Tabellen

Zur Demonstration von Joins wollen wir einen Ausschnitt des zugehörigen Genexpressionsdatensatzes laden.

### 5.9.1 Aufgabe: Daten einlesen

Ergänzen Sie den folgenden Chunk um einen Aufruf des eingelesenen Datensatzes `expressionsDaten` und führen Sie ihn dann aus. Wundern Sie sich nicht, wenn das Einlesen einen Moment dauert, dieser Datensatz hat 10.000 Spalten und ist damit recht groß.

```
expressionsDaten <- read_csv("Expressionsdatei.csv", show_col_types = FALSE)
expressionsDaten
```

*Tabellenverknüpfungen (Joins)* gehören zu den nützlichsten Funktionen, die Datenbanksysteme mitbringen. In Excel lassen sich Tabellen mit vertretbarem Aufwand nicht verknüpfen. In R ist das aber sehr wohl möglich. Bei *Joins* geht es darum, Zeilen der einen Tabelle mit Zeilen der anderen zu kombinieren. Eine Verknüpfung erfolgt dabei über eine oder mehrere Spalten, die in beiden Tabellen vorkommen. Hier kommt u.a. die Spalte namens `Sample` in beiden Tabellen vor. Eine Verknüpfung achtet dann darauf, bei welchen Zeilen `Sample` übereinstimmt und kombiniert diese zu einer Gesamtzeile der neuen Tabelle. Wenn die neue Tabelle nur solche Zeilen enthält, für die sich in beiden Ausgangstabellen eine passende Zeile gefunden hat, heißt die Verknüpfung *Inner Join*. In der Funktion zur Verknüpfung sollten die Spalten mit angegeben werden, mittels derer die Zeilen verknüpft werden (R lässt zwar zu, diese wegzulassen z.B. wenn es nur eine in Frage kommende Spalte gibt. Eine explizite Nennung verbessert aber die Nachvollziehbarkeit).

```
inner_join(phaenoDaten, expressionsDaten, by='Sample')
```

Dieser Befehl ergibt die Kombination solcher Zeilen aus `phaenoDaten`, für die es in `expressionDaten` Zeilen mit derselben `Sample` - ID gibt.

Weitere Verknüpfungsmöglichkeiten ergeben sich, wenn Sie die erste oder die zweite in der `join`-Funktion genannte Tabelle komplett reproduzieren und jeweils Werte aus der anderen Tabelle in den passenden Zeilen anhängen. Dabei bleiben oft Felder der entstehenden Tabelle leer – weil sich keine passende Zeile in der anderen Tabelle gefunden hat.

Der Befehl `left_join(phaenoDaten, expressionsDaten, by='Sample')` reproduziert die komplette erste Tabelle (`phaenoDaten`) und hängt immer wo passende `Sample`-Werte vorhanden sind die passende Zeile der zweiten Tabelle an. Umgekehrt sorgt der Befehl `right_join(phaenoDaten, expressionDaten, by='Sample')` dafür, dass die zweite Tabelle vollständig reproduziert wird und passende Zeilen der ersten Tabelle angefügt werden.

### 5.9.2 Aufgabe: Joins

Probieren Sie einen der beiden genannten Joins aus und überprüfen Sie, was sich im Vergleich zum Inner Join geändert hat.

```
left_join(phaenoDaten, expressionsDaten, by = "Sample")
```



## 6 Datenbankstrukturierung und -abfrage mit Access

### *Datenhaltung und -verwaltung in Datenbanken*

Wie im Seminar „Tabellenkalkulation“ gesehen, können Tabellenkalkulationsprogramme für die Verwaltung von einfach zu strukturierenden, tabellarischen Daten mit geringen Anforderungen an die Datensicherheit eingesetzt werden. Generell sind jedoch Datenbanksysteme besser geeignet, um große Datenmengen effizient, konsistent und dauerhaft zu speichern. Darüber hinaus bieten Datenbanken vielfältige Möglichkeiten zum schnellen Suchen bzw. Filtern von Datensätzen, Berechnen, Zusammenfassen bzw. Überprüfen von Daten, sowie zum Automatisieren von Datenverwaltungsaufgaben. Wenn eines der folgenden Kriterien auf die zu verwaltenden Daten zutrifft, sollten Datenbanksysteme unbedingt bevorzugt werden:

- Sie verfügen über **große Datenmengen**.
- Sie beziehen Daten aus **unterschiedlichen Quellen** und müssen diese zusammenführen.
- Sie erwarten eine **variable Anzahl von Messwiederholungen** je Individuum.
- Sie haben hohe Anforderungen an die **Datenkonsistenz**.
- Datensätze enthalten **Information aus vielen verschieden Themenbereichen**, welche inhaltlich bzw. logisch klar voneinander abgrenzbar sind.
- Analysen von Daten erfolgen auf (verschiedenen) **Teilmengen** des gesamten Datensatzes.
- Sie **arbeiten gemeinsam mit weiteren Personen** (evtl. mit Zugriffsbeschränkungen) an den Daten.
- Sie wollen **multimediale Daten** (z. B. Signale, Bilder, Videos) verwalten.

Es existieren verschiedene Datenbankmodelle. Eines davon ist das sogenannte relationale Datenbankmodell (z. B. Access, Db2, MySQL, Oracle Database), bei dem Datensätze zeilenweise in Tabellen verwaltet werden und Daten aus unterschiedlichen Tabellen miteinander in Beziehung stehen können.

### *Zum Seminar „Datenbanken“*

Das Seminar dient der Vermittlung von Grundwissen zur Strukturierung und Verknüpfung von Tabellen in relationalen Datenbanken sowie dem Entwurf von Datenbankabfragen zum Suchen, Filtern, Abrufen, Berechnen und Zusammenfassen von Daten.

Für die Übungen verwenden wir ausgewählte und aufbereitete Daten einer am UKJ durchgeführten Studie mit mehreren beteiligten Kliniken und Instituten.

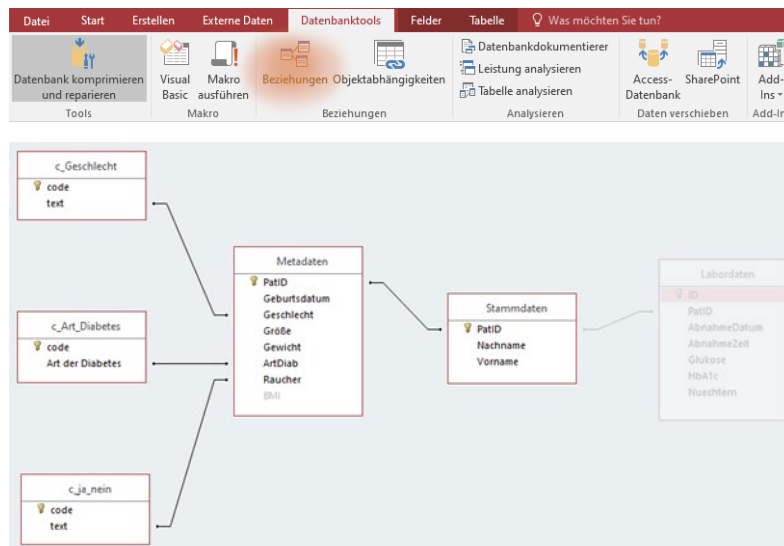
### 6.1 Strukturierung von Tabellen

In relationalen Datenbanken werden Informationen in Tabellen organisiert. Dabei ist es empfehlenswert, Daten themenbasiert in einzelne Tabellen aufzunehmen, um möglichst keine redundanten Daten zu speichern. Dadurch ergibt sich jedoch die Notwendigkeit, Zusatzinformationen mit in die Tabellen aufzunehmen, die erforderlich sind, um Daten aus verschiedenen Tabellen bei Bedarf zu verknüpfen. Innerhalb einer Tabelle wird jede Zeile als Datensatz und jede Spalte als Feld bezeichnet. Ein Datensatz beschreibt hierbei eine fest verbundene Kombination von verschiedenen Informationen (Feldern).

**Aufgabe 6.1.1:** Die für das Seminar vorbereitete Datenbank „ML\_Datenbanken.accdb“ enthält bereits einige Tabellen. Zuerst soll sich ein Überblick über die Tabellen und die Beziehungen zwischen ihnen verschafft werden.

**Lösungsschritte:**

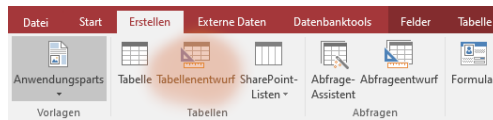
1. Die Datenbank „ML\_Datenbanken.accdb“ wird mittels Microsoft Access geöffnet. Die Datei ist passwortgeschützt (Passwort: Se3%um92!), dieses Passwort muss beim Öffnen eingegeben werden. Nach dem Öffnen wird in der linken Navigationsleiste die Gruppe *Tabellen* ausgewählt (sofern nicht bereits voreingestellt). Man erhält eine komplette Übersicht aller in der Datenbank enthaltenen Tabellen.
2. Durch Doppelklick auf eine Tabelle wird diese geöffnet. Man beachte, dass neben den eigentlich zu speichernden Daten alle Tabellen weitere (Schlüssel)spalten enthalten. Auf die Bedeutung dieser Spalten wird später genauer eingegangen.
3. Im Menüband *Datenbanktools* gibt es einen Schalter *Beziehungen*, mit dem die Beziehungen zwischen den Tabellen graphisch angezeigt werden können.



**Aufgabe 6.1.2:** Eine Tabelle „Labordaten“ soll als neue Tabelle in die Datenbank aufgenommen werden. Die Tabelle ist Spalte für Spalte zu strukturieren. Gültigkeitsregeln und -meldungen werden implementiert und Verknüpfungen zu anderen Tabellen erstellt.

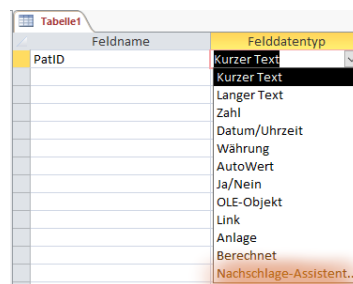
**Lösungsschritte:**

1. Die fertig strukturierte Tabelle soll schließlich die Spalten *Patienten ID*, *Abnahmedatum der Probe*, *Abnahmezeit der Probe*, *HbA1c*, *Glukosespiegel*, und eine Information darüber, ob die Blutprobe im *nüchternen Zustand* abgenommen wurde, enthalten. Generell kann eine neue Tabelle mittels des Schalters *Tabellenentwurf* des Menübands *Erstellen* erzeugt werden. Es öffnet sich die sogenannte Entwurfsansicht, in welcher die die Spalten der Tabellen und deren Eigenschaften definiert werden können. In der Entwurfsansicht werden keine Daten in die Tabelle eingetragen, sondern die Tabellenstruktur entwickelt.



- Die erste Spalte der Tabelle soll die Patientenkennung enthalten. Hierbei ist es wichtig, dass in der Tabelle „Labordaten“ nur Daten von Personen eingetragen werden können, die bereits in der Stammdatentabelle erfasst worden sind. Das bedeutet schließlich, dass die Patientenkennung in der Tabelle *Labordaten* aus der Stammdatentabelle bezogen werden muss.

Neue Spalten einer Tabelle werden erzeugt, indem man Feldnamen (Spaltennamen/-überschriften) für jede Spalte vergibt. Diese Namen sollten kurz und prägnant sein und den Inhalt des Feldes (der Spalte) widerspiegeln. Für die Patientenkennung könnte das z. B. *PatID* sein. Darüber hinaus muss für jedes Feld ein Datentyp festgelegt werden. Dieser kann aus dem Drop-Down-Menü rechts neben dem Feldnamen ausgewählt werden. Wenn, wie im Fall der Patientenkennung, Werte aus anderen Tabellen übernommen werden sollen, kann der *Nachschlage-Assistent* verwendet werden, mit dessen Hilfe Beziehungen zwischen Tabellen einfach aufgebaut werden können. Nachschlageprüfungen dienen als Hilfsmittel bei der Dateneingabe und können die Dateneingabemöglichkeiten drastisch einschränken. Sie zielen auf eine Standardisierung der Eingaben und damit einer Vermeidung von Fehlern während der Dateneingabe ab.

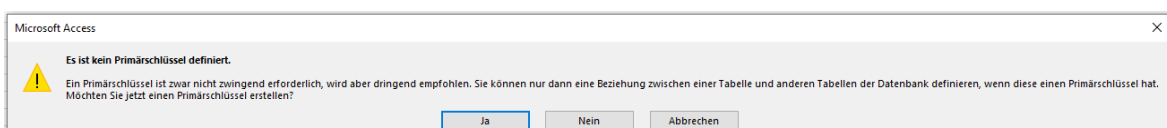


- Mit Hilfe des Nachschlage-Assistenten kann man Schritt für Schritt Beziehungen zwischen Tabellen aufbauen. Zuerst muss ausgewählt werden, aus welcher Quelle die nachzuschlagenden Daten stammen sollen (Tabelle, Abfrage, oder selbstdefinierte Liste).

Danach wird die konkrete Tabelle bzw. Abfrage ausgewählt (oder die Liste manuell eingegeben). In unserem Fall ist das die Tabelle *Stammdaten*.

Im nächsten Schritt des Nachschlage-Assistenten können die Felder der ausgewählten Tabelle ausgewählt werden, welche später bei der Dateneingabe in die Tabelle *Labordaten* als Drop-Down-Menü angezeigt werden. Die hierbei ausgewählten Felder sind also wichtig, um die richtige Person zu identifizieren, der die Labordaten zuzuordnen sind. Um die Identifikation zu erleichtern, kann die Sortierung des Drop-Down-Menüs im folgenden Schritt des Assistenten individuell angepasst werden.

Bevor der Nachschlage-Assistent beendet werden kann, muss der Tabelle ein Name zugewiesen werden (z. B. *Labordaten*). Danach erscheint ein Hinweis, dass noch kein Primärschlüssel definiert worden ist.



In Datenbanken werden ein Feld oder eine Gruppe von Feldern, welche einen Datensatz eindeutig kennzeichnen, als Schlüssel bezeichnet. Da jeder Datensatz (jede Zeile) einen anderen Wert für den Schlüssel hat, können Werte für den Schlüssel innerhalb der Datenbank zum Verweisen auf ganze Datensätze verwendet werden. Jede Tabelle kann nur einen Primärschlüssel haben, welchen Access automatisch beim Erstellen einer Tabelle erstellt (Schalter: *Ja*). Alternativ kann man das Feld (bzw. die Felder) selbst angeben, welche man als Primärschlüssel verwenden möchte (Schalter: *Nein*). Access verwendet Primärschlüsselfelder, um u. a. Daten aus mehreren Tabellen schnell zu lesen und sinnvoll zu kombinieren. Gute Felder für Primärschlüssel haben die Eigenschaft, dass sie jede Zeile der Tabelle eindeutig kennzeichnen, niemals leer sind und sich bestenfalls nie ändern.

Entscheidet man sich für die Auswahl *Ja*, wird in der Tabelle automatisch ein neues Feld *ID* angelegt, welches vom Typ *AutoWert* ist. Das bedeutet, dass für jeden Datensatz der Tabelle automatisch eine fortlaufende Nummer generiert wird. Diese Nummern haben alle geforderten Eigenschaften eines Primärschlüssels.

Labordaten		
	Feldname	Felddatentyp
🔑	ID	AutoWert
	PatID	Kurzer Text

- Nachdem der Nachschlageassistent beendet worden ist, findet man eine Reihe von durch den Assistenten aktualisierten Feldeigenschaften am unteren Bildrand unter dem Reiter *Nachschlagen*. Möglicherweise ist der Wert für *Nur Listeneinträge* auf *Nein* eingestellt, was zur Folge hätte, dass bei der Eingabe einer Patientenkenung zwar die Werte aus der Nachschlagetabelle angezeigt und übernommen werden können aber auch frei gewählte Werte eintragbar sind. Da das im vorliegenden Fall nicht erwünscht ist, muss die Option *Nur Listeneinträge* auf *Ja* geändert werden. Schließlich können in der Tabelle nur noch Datensätze für bereits in den Stammdaten erfasste Personen gespeichert werden.

Allgemein	Nachschlagen
Steuerelement anzeigen	Kombinationsfeld
Herkunftstyp	Tabelle/Abfrage
Datensatzherkunft	SELECT [Stammdaten].[PatID] FROM Stammdaten ORDER BY [PatID];
Gebundene Spalte	1
Spaltenanzahl	1
Spaltenüberschriften	Nein
Spaltenbreiten	2,54cm
Zeilenanzahl	16
Listenbreite	2,54cm
Nur Listeneinträge	Nein
Mehrere Werte zulassen	Nein
Wertlistenbearbeitung zu	Ja
Bearbeitungsformular für	
Nur Datensatzherkunftsw	Nein

- Mit den aktuellen Einstellungen besteht jetzt noch die Möglichkeit, Datensätze ohne *PatID* in die Tabelle der *Labordaten* einzufügen. Das kann unterbunden werden, indem die Eigenschaften des Feldes *PatID* angepasst werden. Dazu ist im Reiter *Allgemein* bei der Option *Eingabe Ja* und bei der Option *Leere Zeichenfolge Nein* einzustellen.

Allgemein	Nachschlagen
Feldgröße	255
Format	
Eingabeformat	
Beschriftung	
Standardwert	
Gültigkeitsregel	
Gültigkeitsmeldung	
Eingabe erforderlich	Ja
Leere Zeichenfolge	Nein
Indiziert	Ja (Duplikate möglich)
Unicode-Kompression	Ja
IME-Modus	Keine Kontrolle
IME-Satzmodus	Keine
Textausrichtung	Standard

- Die nächste Spalte der Tabelle soll für die Erfassung des Datums der Probenentnahme strukturiert werden. Dazu wird ein Feldname vergeben (z. B. *AbnahmeDatum*) und als Datentyp *Datum/Uhrzeit* gewählt.
- Analog dazu wird eine Spalte *AbnahmeZeit* von Datentyp *Datum/Uhrzeit* erstellt. Standardmäßig wird bei Feldern vom Type *Datum/Uhrzeit* bei der Dateneingabe ein Kalender angezeigt. Bei Zeitvariablen ist daher die Einstellung eines Zeitformates erforderlich. Das geschieht bei den Feldeigenschaften unter dem Reiter *Allgemein* in den Zeilen *Format* und *Datumsauswahl anzeigen*.

Allgemein	Nachschlagen
Format	Zeit, 24Std
Eingabeformat	
Beschriftung	
Standardwert	
Gültigkeitsregel	
Gültigkeitsmeldung	
Eingabe erforderlich	Nein
Indiziert	Nein
IME-Modus	Keine Kontrolle
IME-Satzmodus	Keine
Textausrichtung	Standard
Datumsauswahl anzeigen	Nie

- In der fünften Spalte der Tabelle soll der Glukosespiegel im Blut als numerischer Wert eingetragen werden.

Labordaten			
	Feldname	Felddatentyp	
▼	ID	AutoWert	Datensatznummer
	PatID	Kurzer Text	Patientenkennung
	AbnahmeDatum	Datum/Uhrzeit	Datum der Probenentnahme
	AbnahmeZeit	Datum/Uhrzeit	Uhrzeit der Probenentnahme
	Glukose	Zahl	Glukosespiegel (mmol/l)

- Hierbei bietet sich die Verwendung einer Gültigkeitsprüfung an, mit der nicht plausible Daten schon bei der Dateneingabe verhindert werden können. Spezielle Eigenschaften und Gültigkeitsregeln können bei den Feldeigenschaften unter dem Reiter *Allgemein* eingestellt werden. Hier bietet sich an, nur positive Zahlen zu akzeptieren. Das kann mit dem Ausdruck  $> 0$  als Gültigkeitsregel erreicht werden. Wenn Gültigkeitsregeln definiert werden, sollte man auch eine Gültigkeitsmeldung angeben, die immer dann angezeigt wird, wenn Nutzer unzulässige Daten eingeben.

Bei der Definition von Gültigkeitsprüfungen können u. a. die üblichen Vergleichsoperatoren wie  $=$ ,  $>$ ,  $<$ ,  $>=$ ,  $<=$  sowie UND, ODER, NICHT, ZWISCHEN verwendet werden. Für eine vollständige Liste sowie Besonderheiten für verschiedene Datentypen sei an dieser Stelle auf Dokumentation des verwendeten Datenbanksystems verwiesen.

**Wichtig:** Wenn eine Tabelle bereits Daten enthält, kann das Erstellen oder Ändern von Gültigkeitsregeln zu nicht korrigierbaren Widersprüchen in der Tabelle führen. Es muss daher vor dem Erstellen oder Ändern von Gültigkeitsregeln überprüft werden, ob alle bereits in der Tabelle enthaltenen Daten die Gültigkeitsregeln erfüllen. Eine entsprechende Warnung wird vor dem Speichern der Tabelle ausgegeben.

Allgemein	Nachschlagen
Feldgröße	Double
Format	
Dezimalstellenanzeige	1
Eingabeformat	
Beschriftung	
Standardwert	
Gültigkeitsregel	$> 0$
Gültigkeitsmeldung	Es sind nur positive Werte erlaubt.
Eingabe erforderlich	Nein
Indiziert	Nein
Textausrichtung	Standard

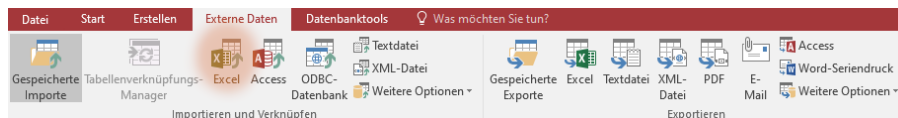
- Analog dazu wird eine Spalte *HbA1c* von Datentyp *Zahl* erstellt.

11. In der letzten Spalte der Tabelle soll erfasst werden, ob die Blutprobe im nüchternen Zustand entnommen wurde. Dazu wird ein Feld mit dem Namen *Nuechtern* von Datentyp *Ja/Nein* erstellt. Hier bietet sich die Vergabe eines Standardwertes an (Feldeigenschaften, Allgemein, Zeile Standardwert).
12. Wenn die Tabelle fertig strukturiert worden ist, kann sie geschlossen werden. Sie erscheint dann dauerhaft in der Tabellenliste links.

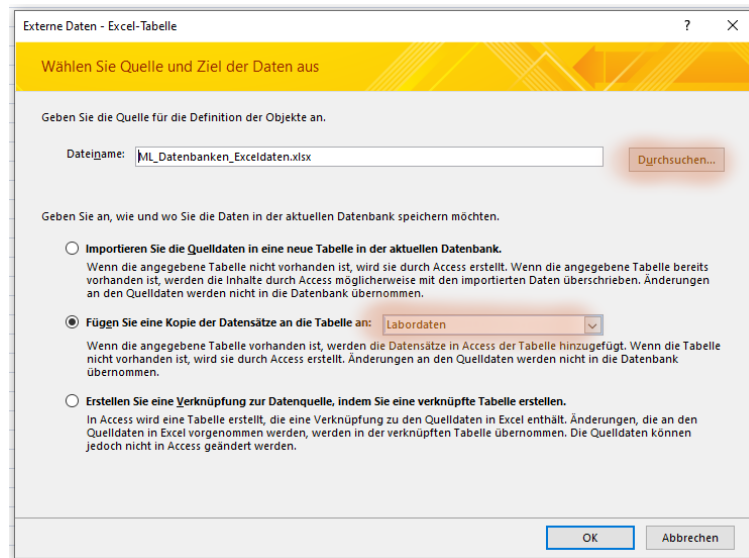
**Aufgabe 6.1.3:** Die Tabelle „*Labordaten*“ soll mit Daten gefüllt werden, die bereits als Excel-Tabelle vorliegen.

### Lösungsschritte:

1. Der Import und Export von Daten aus bzw. in andere Dateiformate kann mit verschiedenen Aktionen des Menübands *Externe Daten* durchgeführt werden. Der Schalter *Excel* des Bereiches *Importieren und Verknüpfen* startet einen Import-Assistenten.



2. Zuerst werden die entsprechende Datei und der Speicherort der zu importierenden Daten gewählt. Im vorliegenden Fall sollen die importierten Datensätze an die bereits existierende Tabelle *Labordaten* eingefügt werden, da die Struktur der vorhandenen Excel-Tabelle mit der Struktur der zu füllenden Datenbank-Tabelle übereinstimmt.



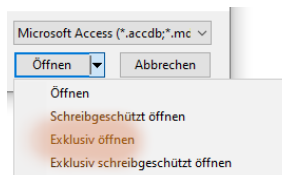
3. Es wird automatisch erkannt, ob die erste Zeile der Excel-Tabelle Spaltenüberschriften enthält. Falls Daten in Tabellenkalkulationsprogrammen erfasst wurden, sollte immer darauf geachtet werden, dass in der ersten Zeile Spaltenüberschriften eingetragen sind, welche später in anderen Programmen als Variablen- bzw. Feldnamen verwendet werden.
4. Beim Import wird geprüft, ob die Feldnamen der Datenbank-Tabelle mit den Spaltenüberschriften der Excel-Tabelle übereinstimmen (die Reihenfolge spielt keine Rolle). Falls nicht, erscheinen entsprechende Warnungen und der Import wird abgebrochen.

**Wichtig:** Feldnamen und Spaltenüberschriften müssen beim Import von Datensätzen in existierende Datenbank-Tabellen übereinstimmen!

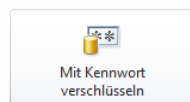
**Aufgabe 6.1.4:** Um die unbefugte Verwendung von Daten zu verhindern, sollten Sie die Datenbank durch Festlegen eines Kennworts schützen.

### Lösungsschritte:

1. Die Datenbank „ML\_Datenbanken.accdb“ wird geschlossen und Access neu gestartet.
2. Um eine Datenbank mit einem Passwort zu schützen, muss diese im so genannten Exklusivmodus geöffnet werden. Zum Öffnen der Datenbank darf diese nicht über die Liste der zuletzt geöffneten Datenbanken geöffnet werden, sondern es muss die Option *Weitere Dateien öffnen* und anschließend *Durchsuchen* verwendet werden. Danach wird im bekannten Dialog zum Öffnen einer Datenbank die Option *Exklusiv öffnen* gewählt.



3. Das bestehende Passwort *Se3%um92!* muss vor dem Öffnen eingegeben werden.
4. Im Menüband *Datei* finden sie in der Kategorie *Information* den Schalter zum Vergeben eines Passwortes. Falls die Datenbank bereits mit einem Kennwort versehen ist, muss dieses zuerst (Schalter: *Datenbank entschlüsseln*) entfernt werden.
5. Danach kann ein neues Passwort vergeben werden.



**Mit Kennwort verschlüsseln**  
Verwenden Sie ein Kennwort, um den Zugriff auf die Datenbank zu beschränken. Dateien, die im Dateiformat von Microsoft Access 2007 oder höher vorliegen, werden verschlüsselt.

6. Vergeben Sie im darauffolgenden Dialogfeld ein möglichst sicheres Passwort. **Wenn Sie das Kennwort vergessen, kann es von Microsoft nicht wieder ermittelt werden.**

## 6.2 Abfragen

In einer Datenbank befinden sich die Daten, die z. B. für eine statistische Auswertung nötig sind, in der Regel in verschiedenen Tabellen, und man benötigt weder alle Felder (Spalten) dieser Tabellen noch alle Datensätze (Zeilen). Solche Teilmengen der gesamten Datenbasis kann man mit Abfragen extrahieren. Außerdem lassen sich mit Abfragen Berechnungen ausführen, Daten zusammenfassen, hinzufügen, ändern oder löschen. Dementsprechend gibt es verschiedene Abfragetypen, wie z. B. Auswahl-, Kreuztabellen-, Tabellenerstellungs-, Aktualisierungs- und Löschartabfragen. Je nachdem, ob Daten nur abgerufen oder auch verändert werden, unterscheidet man prinzipiell zwischen Auswahl- und Aktionsabfragen.

**Aufgabe 6.2.1:** Aus der Tabelle *Stammdaten* sollen alle Patienten ermittelt werden, die männlich sind und mit Nachnamen *Meier* in verschiedenen Schreibweisen heißen. Die Übung dient dem Kennenlernen von Platzhaltern bei Auswahlabfragen.

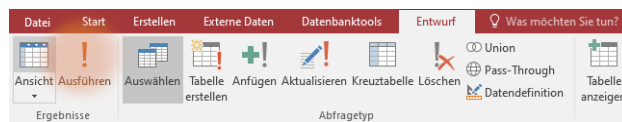


**Lösungsschritte:**

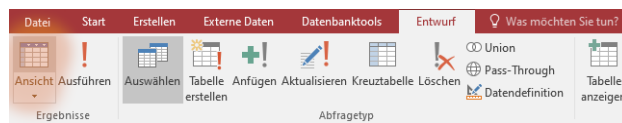
1. Ziel ist die Erstellung einer Auswahlabfrage. Im Menüband *Erstellen* wird dazu mit dem Schalter *Abfrageentwurf* eine neue Abfrage in der Entwurfsansicht erstellt.
2. Im ersten Schritt müssen die Tabellen ausgewählt werden, welche für die Abfrage notwendig sind. Im vorliegenden Beispiel sind das die Tabellen *Stammdaten* und *Metadaten*. Sind alle Tabellen ausgewählt, kann das Auswahlfenster geschlossen werden.
3. Nun werden alle benötigten Felder ausgewählt. Am einfachsten erfolgt das durch Doppelklick auf diese. Hier: *PatID*, *Nachname*, *Vorname* (aus Stammdaten) und *Geschlecht* (aus Metadaten).
4. Bevor wir die geplante Abfrage erstellen, sollen schrittweise einige verschiedene Ansätze umgesetzt werden, die schließlich nicht zum gewünschten Ergebnis führen werden, jedoch den wichtigen Umgang mit Platzhaltern verdeutlichen. Man kann annehmen, dass eine Gemeinsamkeit in der Schreibweise des Familiennamens Meier ist, dass alle mit einem „M“ beginnen und mit „er“ enden. Das legt die Verwendung des Zeichens „\*“ nahe, welches als Platzhalter für eine beliebig lange (auch leere) Zeichenfolge verwendet werden kann. Man könnte also als *Kriterium* in dem Feld *Nachname* den Suchbegriff (Ausdruck) M\*er eintragen, um diejenigen Datensätze zu extrahieren, bei denen der Familienname mit einem „M“ beginnt und mit „er“ endet. Wenn nicht explizit angegeben, wird bei der Verwendung von Platzhaltern der Operator *Wie* von System ergänzt, und der Suchbegriff wird in Anführungszeichen (für Datentyp *Text*) eingeschlossen.

Feld:	PatID	Nachname	Vorname	Geschlecht
Tabelle:	Stammdaten	Stammdaten	Stammdaten	Metadaten
Sortierung:				
Anzeigen:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Kriterien:		Wie "M*er"		
oder:				

5. Mit dem Schalter *Ausführen* im Menüband *Entwurf* wird die Abfrage ausgeführt. Man sieht, dass der Ausdruck M\*er nicht spezifisch genug war und zu viele Datensätze aus der Abfrage resultieren.



6. Die noch nicht zufriedenstellende Abfrage kann verändert werden, indem man wieder in die Entwurfsansicht wechselt (Menüband *Entwurf*, Schalter *Ansicht*), um das Abfragekriterium zu ändern.



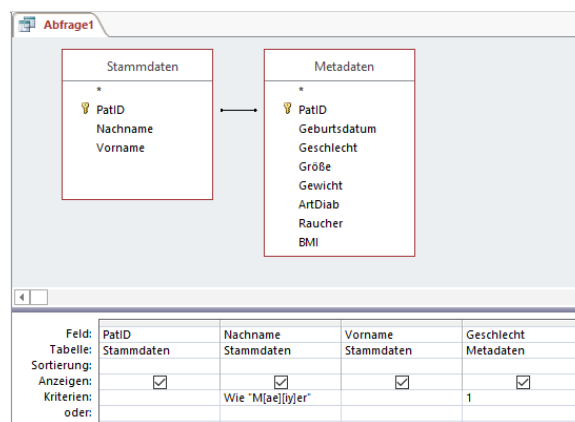
7. Man kann die Gemeinsamkeit der verschiedenen Schreibweisen des Namens dahingehend genauer spezifizieren, dass alle Schreibweisen genau fünf Buchstaben haben. Das führt zur Verwendung des Zeichens „?“ , welches als Platzhalter für genau ein Zeichen benutzt werden kann. Damit wäre ein entsprechender Suchbegriff: M??er. Die Abfrage wird ausgeführt und erhält ein verbessertes, jedoch noch nicht exaktes Ergebnis, da es noch weitere Namen gibt, welche die soeben aufgestellten Bedingungen erfüllen (z. B. Moser, Maler).
8. Weitere Gemeinsamkeiten der Schreibweisen sind, dass der zweite Buchstabe entweder „a“ oder „e“ sowie der dritte Buchstabe „i“ oder „y“ lauten. Diese Bedingungen lassen sich mittels eckigen Klammern umsetzen, da sich damit eine Auswahl von erlaubten Zeichen angeben lässt. Das einzutragende Kriterium wäre dann M[ae][iy]er. Führt man die Abfrage aus, erhält man eine Tabelle, welche vier Schreibweisen des Familiennamens einschließt.

9. Zuletzt wird in der Spalte *Geschlecht* das Kriterium 1 eingetragen und zwar in derselben Zeile wie zuvor M[ae][iy]er, da die Bedingungen für Namen und Geschlecht durch ein logisches *UND* verbunden werden müssen.

Die Geschlechterinformation in der Tabelle *Metadaten* wird aus der Tabelle *c\_Geschlecht* bezogen, woraus sich der Suchbegriff 1 als Codierung für männlich ergibt.

c_Geschlecht	
code	text
1	männlich
2	weiblich
3	divers

Generell werden die Kriterien innerhalb einer Zeile durch *UND* verknüpft; Einträge innerhalb einer Spalte durch *ODER*.



Feld:	PatID	Nachname	Vorname	Geschlecht
Tabelle:	Stammdaten	Stammdaten	Stammdaten	Metadaten
Sortierung:				
Anzeigen:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Kriterien:		Wie "M[ae][iy]er"		1
oder:				

10. Die Abfrage wird letztmalig ausgeführt.

**Aufgabe 6.2.2:** *Alle Datensätze der Labordatentabelle sollen durch das Alter in Jahren der jeweiligen Personen bei der Probenentnahme ergänzt werden.*

**Lösungsschritte:**

1. Zuerst wird eine neue Abfrage in der Entwurfsansicht erstellt.
2. Da für die Altersbestimmung das Geburtsdatum und das Datum der Probenentnahme benötigt werden, müssen die Tabellen *Metadaten* und *Labordaten* zur Abfrage hinzugefügt werden.
3. Um alle Felder der Tabelle *Labordaten* zu übertragen, müssen sie nicht einzeln übernommen werden. Stattdessen kann man den Eintrag „\*“ in der ersten Zeile der Tabellenansicht von *Labordaten* auswählen (z. B. mit Doppelklick).
4. Access verfügt über viele integrierte Funktionen, die man zum Berechnen von neuen Daten verwenden kann. Bei der Verwendung von Funktionen muss man einen Funktionsnamen, eine geöffnete Klammer, eventuell durch Semikolons (bei deutschen Ländereinstellungen) getrennte Argumente und eine schließende Klammer eingeben.

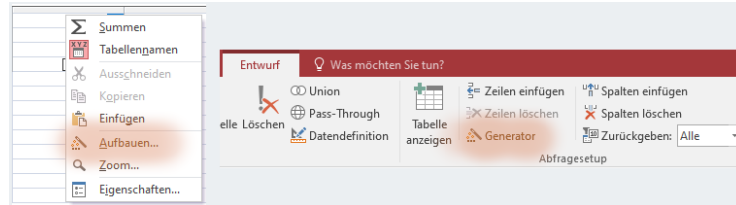
z. B. *Funktionsname(Argument1; Argument2; Argument3)*

Neben der manuellen Eingabe (setzt Kenntnis der Funktionen und der Syntax voraus) stellt Access einen Assistenten für die Benutzung von Funktionen zur Verfügung, welchen wir im Folgenden verwenden werden. Für die Berechnung des Alters kann die integrierte Funktion *DatDiff* verwendet werden. Diese hat mindestens drei Argumente,

*DatDiff (Intervall; DatumVon; DatumBis)*

wobei *Intervall* eine Zeichenfolge ist, die das Zeitintervall zwischen den Datumsangaben *DatumVon*, *DatumBis* darstellt. Hierfür gibt es viele verschiedene Möglichkeiten, die hier nicht einzeln betrachtet werden können (siehe Dokumentation des verwendeten Datenbanksystems). Der Ausdruck für Jahre bei deutschen Ländereinstellungen ist „jjjj“.

- Um ein neues Feld zu berechnen, wird in eine leere Spalte in der Zeile *Feld* mit Rechtsklick das Kontextmenu aufgerufen und in diesem der Eintrag *Aufbauen* gewählt oder es wird der entsprechende Schalter *Generator* im Menüband *Entwurf* gedrückt.



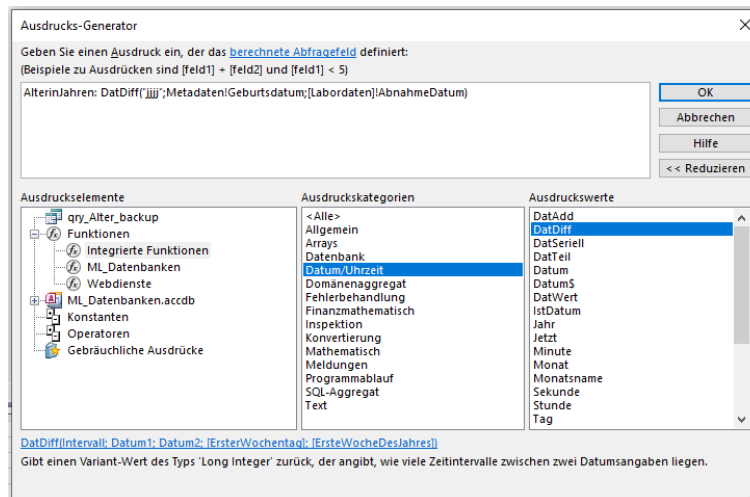
- Der Ausdrucks-Generator öffnet sich und man wählt die Funktion *DatDiff* aus den Blöcken Ausdruckselemente: Integrierte Funktionen, Ausdruckskategorien: Datum/Uhrzeit und Ausdruckswerte. Im Editierfeld erscheint die allgemeine Syntax der Funktion, in der die Argumente spezifiziert werden müssen. Das sind für das *Intervall* „jjjj“; für *Datum1* das Feld *Geburtsdatum* aus der Metadaten-tabelle; und für *Datum2* das Feld *AbnahmeDatum* aus der Labordaten-tabelle. Beide Datumsangaben lassen sich aus den Blöcken Ausdruckselemente und Ausdruckskategorien auswählen.

Schließlich entsteht der Ausdruck:

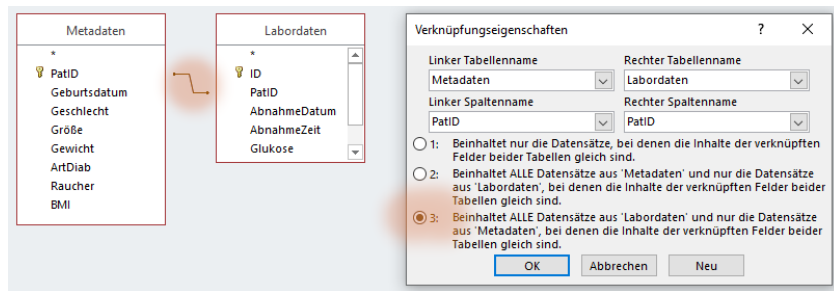
Ausdr1: `DatDiff("jjjj";[Metadaten]![Geburtsdatum];[Labordaten]![AbnahmeDatum]),`

bei dem Ausdr1 noch durch einen geeigneten Feldnamen ausgetauscht werden sollte, z. B.

AlterInJahren: `DatDiff("jjjj";[Metadaten]![Geburtsdatum];[Labordaten]![AbnahmeDatum]).`



- Es existieren noch keine Beziehungen zwischen Feldern der Meta- und Labordaten-tabelle. Das erkennt man graphisch daran, dass es keine Verbindungslinien zwischen den beiden Tabellen gibt. Um aber Datensätze aus den beiden Tabellen mit derselben Patienten-kennung miteinander in Beziehung zu setzen, müssen die Felder *PatID* beider Tabellen verknüpft werden. Das lässt sich erreichen, indem man die Zeile *PatID* einer Tabelle anklickt und mit gedrückter linker Maustaste zu der *PatID* der anderen Tabelle zieht. Die Verknüpfung ist damit aufgebaut. Als letzten Schritt muss man noch die Art der Verknüpfung festlegen: dazu klickt man die soeben entstandene Verbindungslinie doppelt an und wählt im darauffolgenden Dialog die passende Verbindungseigenschaft aus.



8. Die Abfrage wird abschließend ausgeführt.
9. Die Abfrage wird geschlossen und unter dem Namen *qry\_Alter* gespeichert, da die Ergebnisse in späteren Aufgaben weiterverwendet werden sollen.

**Aufgabe 6.2.3:** Für jede Person in der Datenbank soll der BMI bei Rekrutierung berechnet und in der Metadaten-Tabelle abgelegt werden (sofern Körpergröße und -gewicht zur Verfügung stehen).

**Lösungsschritte:**

1. Diese Aufgabe könnte generell mittels einer Abfrage analog zu Aufgabe 6.2.2 bearbeitet werden. Im Unterschied zu 6.2.2 befinden sich aber die für den BMI nötigen Informationen in einer einzigen Tabelle (*Metadaten*). Damit ergibt sich einerseits eine einfachere Möglichkeit der Berechnung eines neuen Feldes (hier *BMI*) und andererseits die zusätzliche Funktionalität, dass für neu eingegebene Daten von Körpergröße und -gewicht sofort der BMI berechnet und in die Tabelle eingetragen wird.
2. Die Tabelle *Metadaten* wird in der Entwurfsansicht geöffnet.
3. Ein neues Feld *BMI* vom Typ *Berechnet* wird erstellt.

BMI	Berechnet	Body Mass Index (kg/m <sup>2</sup> )
-----	-----------	--------------------------------------

4. Die Eigenschaften des Feldes *BMI* müssen angepasst werden: Datentyp: *Double*, Ausdruck: zur Berechnung des BMI  $[Gewicht]/[Größe]^2$  und ggf. die Anzahl der Dezimalstellen. Alternativ kann man zur Eingabe des Ausdrucks auch den Ausdrucks-Generator verwenden (Schalter ... am rechten Rand der Zeile *Ausdruck*).

Allgemein	Nachschlagen
Ausdruck	[Gewicht]/[Größe]^2
Ergebnistyp	Double
Format	00,00
Dezimalstellenanzeige	2
Beschriftung	
Textausrichtung	Standard

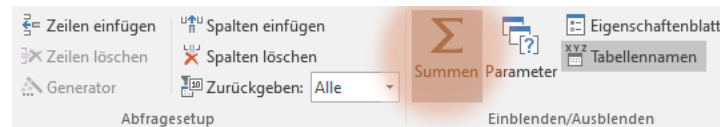
5. Die Metadaten-Tabelle enthält nun eine neue Spalte *BMI* (Datenblattansicht), deren Zelleninhalte sich dynamisch bei Änderungen in den Spalten *Größe* oder *Gewicht* anpassen.

**Aufgabe 6.2.4:** Es soll für jede Person in der Datenbank der Altersbereich aufgelistet werden, in dem Probenentnahmen in der Datenbank erfasst sind.

**Lösungsschritte:**

1. Eine neue Abfrage wird in der Entwurfsansicht erstellt.
2. Die in Aufgabe 6.2.2 entworfenen Abfrage *qry\_Alter* wird zur neu erstellten Abfrage hinzugefügt und das Fenster *Tabelle anzeigen* geschlossen.

- Die Felder *PatID* und *AlterInJahren* (2x) werden zur Abfrage hinzugefügt. Es empfiehlt sich, dass Feld *AlterInJahren* zweimal hinzuzufügen, um gleichzeitig das kleinste und höchste Alter bei Probenentnahme zu bestimmen und anzuzeigen.
- Die Aufgabenstellung erfordert, dass für jede Person die Altersangaben separat betrachtet werden müssen und personenbezogen minimales und maximales Alter zu bestimmen sind. Mit anderen Worten, die Daten müssen personenbezogen gruppiert werden (Gruppierungsabfrage). Dazu muss die Funktion *Gruppierung* im Abfrageentwurf zuerst eingeblendet werden (Menüband *Entwurf*, Schalter *Summen* oder rechter Mausklick auf Feldnamen und Auswahl von *Summen* im Kontextmenü)



- Als Funktionen werden für die Spalte *PatID* ‚*Gruppierung*‘ (personenbezogenes Zusammenfassen) und in den beiden Spalten bzgl. des Alters die Funktionen ‚*Min*‘ bzw. ‚*Max*‘ ausgewählt.
- Schließlich wird die Abfrage ausgeführt.

**Aufgabe 6.2.5:** *Generell muss mit Fehlern bei der Eingabe von Daten gerechnet werden, so dass über streng formulierte Gültigkeitsregeln hinaus auch Plausibilitätstests durchgeführt werden sollten, um fehlerhafte bzw. fragwürdige Daten frühzeitig zu erkennen. Beispielhaft sollen die Labordaten dahingehend überprüft werden, dass pro Probenentnahmedatum mindestens ein Wert erfasst worden ist und ob die HbA1c-Werte zwischen 0,1 und 20 mmol/l liegen.*

**Lösungsschritte:**

- Eine neue Abfrage wird in der Entwurfsansicht erstellt.
- Die Tabelle *Labordaten* wird zur Abfrage hinzugefügt.
- Die Felder *PatID*, *Abnahmedatum*, *Glukose* und *HbA1c* werden zur Abfrage hinzugefügt. Die ersten beiden Felder sind notwendig, um auffällige Datensätze (menschenslesbar) identifizierbar zu machen; die letzteren dienen zur Definition der Auswahlkriterien.
- Als erstes Kriterium wird bei den Feldern *Glukose* und *HbA1c* das Kriterium Ist Null in dieselbe Zeile (UND Verknüpfung) eingetragen. Dabei ist es auch ausreichend, nur Null einzugeben; der Rest wird vom System ergänzt. Der Begriff *Null* gibt hier an, dass eine Zelle einer Tabelle keine gültigen Daten hat.
- Für den zweiten Teil der Aufgabenstellung kann der sich selbst erklärende Abfrageausdruck Zwischen ... Und ... verwendet werden, welcher für alle sortierbaren Daten (z. B. Zahl, Datum) anwendbar ist. Somit wird in eine zweite Zeile der Abfragekriterien (ODER Verknüpfung) das Kriterium Nicht Zwischen 0,1 Und 20 in der Spalte *HbA1c* eingefügt.

Feld:	PatID	AbnahmeDatum	Glukose	HbA1c
Tabelle:	Labordaten	Labordaten	Labordaten	Labordaten
Sortierung:				
Anzeigen:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Kriterien:			Ist Null	Ist Null
oder:				Nicht Zwischen 0,1 Und 20

**Aufgabe 6.2.6:** Es sollen alle Personen ohne bekannten Diabetes mellitus aufgelistet werden, die bei mindestens einer Untersuchung einen nüchtern Glukosewert größer als 5,6 mmol/l aufwiesen. Falls für eine Person mehrere auffällig erhöhte Glukosewerte beobachtet worden sind, soll der größte Wert aufgeführt werden.

**Lösungsschritte:**

1. Eine neue Abfrage wird in der Entwurfsansicht erstellt.
2. Die Tabellen *Metadaten* (enthält die Information über den Diabetesstatus) und *Labordaten* werden zur Abfrage hinzugefügt.
3. Die Felder *PatID*, *Glukose* und *ArtDiab* werden zur Abfrage hinzugefügt.
4. Als Kriterium werden in dieselbe Zeile (UND Verknüpfung) in die Spalte *Glukose*  $\geq 5,6$  sowie in die Spalte *ArtDiab*  $= 3$  eingetragen (die Codierung für „keine Diabetes“ ist gleich 3 (siehe Tabelle *c\_Art\_Diabetes*)).
5. Analog zu Aufgabe 6.2.4 wird die Funktion *Gruppierung* im Abfrageentwurf eingeblendet.
6. Als Funktionen werden gewählt: Spalte *PatID* – Gruppierung, Spalte *Glukose* – Max, Spalte *ArtDiab* – Bedingung.
7. In der jetzigen Form würde die resultierende Spalte *ArtDiab* der Abfrage konstant den Wert ‚keine‘ enthalten, womit diese redundant wäre. Deswegen muss sie nicht angezeigt werden: die Checkbox der Zeile Anzeigen wird deaktiviert.

Feld:	PatID	Glukose	ArtDiab
Tabelle:	Labordaten	Labordaten	Metadaten
Funktion:	Gruppierung	Max	Bedingung
Sortierung:			
Anzeigen:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Kriterien:		> 5,6	3
oder:			

8. Die Abfrage wird ausgeführt.

### 6.3 Formulare

Die Eingabe von und das Navigieren durch Daten erfolgt in der Regel nicht direkt in den Tabellen. Um die Nutzerfreundlichkeit, Übersichtlichkeit und Sicherheit zu erhöhen, werden Formulare genutzt.

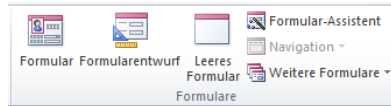
Formulare sind graphische Benutzeroberflächen, die meistens direkt mit Tabellen und Abfragen verbunden sind und zum Eingeben, Bearbeiten oder Anzeigen von Daten verwendet werden (sogenannte gebundene Formulare). Darüber hinaus lässt sich mittels Formularen der Zugriff auf Daten steuern, indem man Nutzern in Formularen nur ausgewählte Felder oder Datensätze von Tabellen anbietet. Da gebundene Formulare mit Datenquellen verbunden sind, wirken sich Eingaben in Formularen direkt auf verknüpfte Tabellen oder Abfragen aus – Datenquelle und Formulardaten sind synchronisiert.

Alternativ gibt es noch ungebundene Formulare (nicht Gegenstand des Seminars), welche nicht direkt mit Datenquellen verknüpft sind, sondern Befehlsschaltflächen, Beschriftungen und andere Steuerelemente enthalten und somit vorrangig zum Bedienen der Datenbankanwendung verwendet werden.

**Aufgabe 6.3:** Es soll ein Formular zur Eingabe von Daten in die Tabelle „Metadaten“ erstellt werden.

**Lösungsschritte:**

1. Öffnen Sie zuerst die Tabelle *Metadaten*.
2. Es gibt verschiedene Möglichkeiten, Formulare zu erstellen. Diese finden sich im Menüband *Erstellen*.



3. Mittels des Schalters *Formular* wird ein einfaches, funktionsfähiges Formular für die Tabelle *Metadaten* mit einem vordefinierten Layout erstellt. In der Entwurfsansicht können noch individuelle Anpassungen (Formatierungen, Anordnungen, Hinzufügen oder Entfernen von Feldern) vorgenommen werden. Alternativ besteht die Möglichkeit, über den Schalter *Formularentwurf* Formulare völlig frei zu gestalten.

4. Im Menüband *Start* kann man zwischen der Entwurfsansicht und dem Formular wechseln. In der Formularansicht kann das Formular genutzt werden (Eingeben, Ändern, Suchen von Datensätzen). In der Layout- oder Entwurfsansicht werden Änderungen an Struktur, Inhalt und Layout des Formulars vorgenommen.

