

# Stellungnahme zur Datennutzung in der Gesundheitsforschung zur Sitzung vom 5. Juli 2023 des Bundestagsausschusses für Bildung, Forschung und Technikfolgenabschätzung

Wörter: 1489

## 1. Alles aus Datenbanken?

Bis vor wenigen Jahren galt bei Arzneimittelbehörden und Methodikern, dass Wirksamkeit nur durch interventionelle klinische Studie belegt werden könne. Die Welt dreht sich aber weiter, so dass sich dieser Rigorismus mittlerweile aufgeweicht zu sein scheint. Und zwar zu Gunsten von - allgemein gesprochen - nicht-interventionellen Studien, also Beobachtungsstudien, Registerstudien und Datenbankauswertungen, oder den Schlagworten *Big Data* oder *Real World Data*.

Nach wie vor wesentliche Bedenken von *Real World Data* Auswertungen gegenüber klinischen Studien sind:

- Geringere Datenqualität
- Fehlender Quelldatenabgleich
- Hohes Risiko für *Bias* (Neigung)

Diese Probleme können bei *Real World Data* durch hohe Fallzahlen, die bisweilen die Grundgesamtheit (z.B. alle Erwachsene in Deutschland) umfassen, einigermaßen aufgewogen werden. Dazu sollte man den p-Wert für die Anerkennung einer Signifikanz deutlich tiefer ansetzen (z.B. 0.005) und sich sorgfältig um Identifizierung, Kontrolle und Diskussion von *Bias* bemüht. Die Qualitätsprobleme wird man wohl kaum angehen können.

Geht man nach den Zielen von Gesundheitsforschung, dann sollte man unbedingt mehr sogenannte **pragmatische randomisierte Studien** nach dem Vorbild der britischen *Recovery Trials* durchführen. Dazu sollten Dokumentvorlagen und ein Netzwerk ausgebaut werden, so dass rasch relevante neue Fragen überprüft werden können.

## 2. Forschung mit Gesundheitsdaten ist sinnvoll und nötig!

Diese Aussage gilt völlig unabhängig vom oben erwähnten Rigorismus. Es kann hier nicht um ein entweder - oder gehen, **sondern um sowohl - als auch**.

Die größten Sammlungen von Gesundheitsdaten dürften in Deutschland bei den gesetzlichen Kassen liegen. Allerdings gibt es noch viele andere Quellen.

Es gibt in vielen Ländern sehr umfassende Datensammlungen, insbesondere in den USA und Großbritannien. Oft stellen diese neben den Berichten auch auswertbare Tabellen zur Verfügung. Dies sollte das Vorbild sein.

### 3. Datenschutz

Ich jedenfalls sehe diese Fragen durch die Regeln der Pseudonymisierung sowie einer gewissen Weitergabebeschränkung (keine offenen persönlichen Daten, Ersatz durch „<3“ bei sehr kleinen Fallzahlen) was die eigentliche Gesundheitsforschung angeht, als gelöst an.

Im Folgenden gehe ich davon aus, dass die Daten bereits vorgefertigt wurden, also in pseudonymisierter bzw. in aggregierter Form vorliegen; dies bezeichne ich hier als **Rohdaten**, welche z.B. in Tabellenform vorliegen könnten. Die Vorstufe wären dann **individuelle Daten**, also identifizierbare Gesundheitsdaten, die in Datenbanken vorliegen. Nachfolgend gäbe es dann **Analysen, Auswertungen und Berichte**.

Inzwischen kommt aber durch den WHO-Pandemie-Vertrag und ähnliche Bestrebungen ein ganz neuer Aspekt hinzu. Diese Initiativen nähren nämlich den dringenden Verdacht, dass demnächst die WHO Zugriff auf individuelle Daten wie z.B. Impfstatus, Testergebnisse oder Krankengeschichte bekommt und dann in der Lage sein wird, z.B. individuelle Grundrechte einzuschränken. Dem wird man kaum einen technischen Riegel vorschieben können. Umso wichtiger wäre ein unüberwindbarer gesetzlicher Riegel. **Die WHO, deren Mitgliedsstaaten oder davon abhängige Organisationen dürfen keinesfalls irgendeinen Zugriff auf individuelle Gesundheitsdaten bekommen! Denn anderenfalls wären massivste Grundrechtseinschränkungen möglich.**

### 4. Verfügbarkeit

Ich bin der Auffassung, dass alle Rohdaten, welche die Gesundheitsdaten betreffen, stets allgemein, d.h. öffentlich und kostenlos verfügbar sein sollten. Die gilt für mich ausdrücklich auch für Rohdaten aus Verknüpfungen von zwei oder mehr Datenbanken.

Was auch immer mit solchen Rohdaten dereinst gemacht wird: Stets sollte die Nachvollziehbarkeit und die Plausibilität, wenigstens auf aggregierten Ebene, grob überprüfbar bleiben. Es muss erkennbar bleiben, ob Daten unterdrückt wurden oder gewisse Daten fehlinterpretiert wurden.

Ein treffendes Beispiel dafür waren die zeitweisen Behauptungen des RKI und von einigen Ministerpräsidenten, es hätte eine „Pandemie der Ungeimpften“ gegeben. Bei Überprüfung der Aussagen ließ sich aber leicht feststellen, dass 70-90% der Daten, aus welchen Gründen auch immer, gar nicht in die Berechnungen einfließen. Dieser riesige Batzen hätte die Ergebnisse auf jeden Fall in die komplett gegenteilige Richtung drücken können.

Da auch bei wissenschaftlichen Veröffentlichungen immer mehr auch Rohdaten eingefordert werden: Warum diese nicht gleich veröffentlichen?

Dann stellt sich natürlich die Frage: Wer übernimmt die Kosten für die Aufbereitung der individuellen Daten in Rohdaten? Hier geht es um einen gewissen, aber m.E. sehr überschaubaren Kostenanteil. Denn die Daten liegen ja vor und müssen nur noch aufbereitet werden. Ein qualifizierte Datenbankprogrammierer sollte dies leicht

bewerkstelligen. Etwas aufwändiger könnte es allerdings bei der Integration von Daten aus zwei verschiedenen Datenbanken geben. Hier müssten vermutlich Testläufe gemacht werden, bevor man die Rohdaten freigibt.

Umgekehrt muss man auch bedenken: Letztendlich „gehören“ die Daten z.B. der KBV den Versicherten. Will man dafür vielleicht eine Art Börsenhandel einführen? Sind die Daten derart wertvoll? Hätten diese einen nennenswerten Wert, dann würden mit Sicherheit Grenzen gezogen werden, um die weitere Monetarisierung zu ermöglichen bzw. zu verbessern. Das kann dann nur in prohibitiver Datenbankverwaltung münden. Ist das aber wirklich im Sinne der Versicherten oder der Bürger? Im Sinne von Gesundheitsforschung, die dann doch wieder veröffentlicht werden muss? Ich jedenfalls würde diese Fragen mit einem klaren Nein beantworten.

Schließlich soll auf den allgemeinen Trend in der Forschung hingewiesen werden: Der geht in Richtung von freiem Online-Zugang für alle Forschungsberichte und weg vom klassischen Journal-Modell, wo der Verlag die Rechte übernimmt – an sich schon ein Unding – und dieser sich per Abonnements oder *pay per view* bezahlen lässt. Übrigens vor allem von Universitäten, also i.d.R. der öffentlichen Hand. Dies ist schon ein spezielles Geschäftsmodell: Die öffentliche Hand finanziert i.d.R. Forschung, aber wenn die Ergebnisse jahrelanger Arbeit vorliegen, dann nur hinter einer Bezahlschranke eines Verlags, der nur eine kleine Dienstleistung anbietet? Immerhin trat die letzte Bundesregierung schon diesen Praktiken entgegen und verfügte, dass öffentlich geförderte Forschungen immer frei verfügbar sein sollen.

Wenn aber Rohdaten bei Veröffentlichung des Berichts allgemein zur Verfügung stehen: Warum dann nicht gleich? Wenn die Rohdaten sofort zur Verfügung stehen, weiß jeder, dass er explizit und im Detail überprüft werden könnte. Allein schon diese „Drohung“ dürfte zu einem Qualitätssprung führen.

**Ich plädiere daher für völlige und freie Verfügbarkeit der Rohdaten zur Gesundheitsforschung.**

**Weiterhin plädiere ich für freie Verfügbarkeit von allen Primär-Artikeln zur Gesundheitsforschung.**

Primär-Artikel meint hier: Mit Daten, die damit erstmalig ausgewertet wurden. Bei Sekundär-Daten, also vor allem Reviews, spielt das keine essentielle Rolle.

## **5. Überprüfung der Plausibilität**

Durch die völlige und freie Verfügbarkeit von Rohdaten wäre Plausibilität weitgehend überprüfbar.

Die letzten drei Jahre zeigten aber, dass gerade Behörden wie PEI und RKI in dieser Hinsicht mauerten. Sie gaben nämlich überhaupt keine Rohdaten preis und selbst in Berichten gaben sie nur in sehr begrenztem Umfang absolute Zahlen heraus. Dadurch waren Nachberechnungen und Plausibilitätschecks oft unmöglich.

Daher ist die unabdingbare Forderung: Es müssen stets absolute und relative Zahlen angegeben werden. Die alleinige Angabe von relativen Zahlen darf nur dann

vorkommen, wenn im öffentlich zugänglichen Hauptteil auch die absoluten Zahlen präsentiert werden.

Die obigen Beispiele von RKI oder auch den Sicherheitsberichten zu C19-Impfungen des PEI zeigen, wie essentiell korrekte und verlässliche individuelle Daten wie auch Rohdaten sind. Da der Datenanalyst ja nur die Rohdaten zu Gesicht bekommt, muss die Verantwortung für die Korrektheit der individuellen Daten und deren Verknüpfungen mit anderen Datenbanken eindeutig geregelt sein.

Hier muss man unbedingt über unabhängige, stichprobenartige Kontrollen nachdenken, also Audits. Solche Kontrollen sollten in gewissem Umfang routinemäßig erfolgen, erst recht aber, wenn der begründete Verdacht für Fehler, also eine Implausibilität, vorliegt. Derartige Audits werden in fast allen Bereichen der Industrie seit Jahren vorgeschrieben. Also auch hier ein Muss!

Außerdem muss darüber nachgedacht werden, solche Audit-Berichte zu veröffentlichen.

**Es muss Audits für den Weg von individuellen Daten zu Rohdaten geben. Ich plädiere für die Veröffentlichung von Audit-Berichten.**

## **6. Konkrete Anregungen**

Die C19-Krise zeigte, gerade gegenüber britischen Gesundheitsdaten, erschreckende Schwächen bei deutschen Gesundheitsdaten.

Mindestens in England (als Nation von Großbritannien) gab es die Bemühung, Information zu C19-Impfungen aus dem *National Immunisation Management System* (NIMS) mit z.B. Sterbefällen zusammen zu führen. Sicher war das noch nicht perfekt [1], aber immerhin besser, als alle deutschen Zahlen zum Thema.

Auch in Deutschland wurden die C19-Impfungen penibel registriert. **Es spricht nichts dagegen, diese Information zu nutzen!** Im Gegenteil, derartige Analysen wären sogar unabdingbar zur Beantwortung der derzeit drängendsten Fragen zu Sicherheit der C19-Impfungen.

Solche Analysen wären aber nicht auf Todesfälle zu beschränken, sondern sollten auf alle Indikationen, welche als „Signal“ aus Pharmakovigilanzdatenbanken, vor allem der US-VAERS-Datenbank, detektiert wurden (siehe z.B. [2]),. Also mindestens für :

- Alle onkologischen Erkrankungen, also Krebs (siehe auch [3])
- Alle Formen von thrombotischen und embolischen Erkrankungen sowie Blutungen
- Infektionen und Entzündungen
- Rheumatologische Erkrankungen, inklusive multiple Sklerose u.ä.
- Neurologische Erkrankungen, inklusive alle Formen von Demenz

Diese Erkrankungen könnte und sollte man anhand der KBV-Daten analysieren, eben unter Hinzunahme von Informationen zum Impfstatus. **Diese würde eine sinnvolle Abschätzung erlauben, ob und wie die Inzidenzen solcher Ereignisse aufgrund der**

## **Impfkampagnen erhöht sind. Das wäre höchst sinnvolle Forschung mit Gesundheitsdaten.**

Ähnliches gilt auch für die neueste Panik-Attacke : **Die Hitzetoten**. Hier sollte man Landkreis-genau Extremtemperaturen mit Sterbefällen, dann möglichst auch mit genauem Sterbeort verknüpfen und auswerten, ob es tatsächlich Hitzetote in nennenswertem Umfang gibt. Selbstverständlich sollte man so auch allgemein untersuchen, ob bestimmte Erkrankungen nach Hitzewellen vermehrt auftreten.

## **7. Referenzen**

1. Kremer HJ. Englische Todeszahlen zu Geimpften und Ungeimpften zeigen : Es war die C19-Impfung. <https://tkp.at/2023/06/21/englische-todeszahlen-zu-geimpften-und-ungeimpften-zeigen-es-war-die-c19-impfung/>
2. Kremer HJ. CDC Analyse : Massenhaft schwerste Nebenwirkungen – ohne Einschreiten der Behörden. <https://tkp.at/2023/02/14/cdc-analyse-massenhaft-schwerste-nebenwirkungen-ohne-einschreiten-der-behoerden/>
3. Kremer HJ. VAERS-Daten belegen Turbokrebs (6) : Metastasen und Übersicht. <https://tkp.at/2023/03/15/vaers-daten-belegen-turbokrebs-6-metastasen-und-uebersicht/>