

Sekundärdaten in der pharmakoepidemiologischen Forschung: Können moderne statistische Verfahren Limitationen der Daten ausgleichen?

Iris Pigeot, Vanessa Didelez, Dirk Enders, Bianca Kollhorst

Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS, Achterstraße 30, 28359 Bremen

Im Idealfall werden zur Beantwortung epidemiologischer Fragestellungen Daten im Rahmen von Studien erhoben, die speziell auf die jeweilige Fragestellung zugeschnitten sind. Die qualitätsgesicherte Erhebung von Primärdaten ist aber mit erheblichem Kosten- und Zeitaufwand verbunden und hat zudem mit Schwierigkeiten zu kämpfen, die sich z.B. aus Erinnerungsfehlern der Probanden, einer hohen Non-response oder einer hohen Drop-out Rate ergeben. Daher lohnt es sich u.U., für die Beantwortung bestimmter Fragestellungen bereits existierende Datenquellen, sogenannte Sekundärdaten, zu nutzen. Speziell in der Pharmakoepidemiologie sind zur Untersuchung von seltenen Arzneimittelrisiken oder zum Monitoring eines Arzneimittels nach Marktzulassung sehr große Datensätze erforderlich, um statistisch valide Aussagen treffen zu können, so dass in solchen Fällen häufig auf Sekundärdaten wie z.B. Abrechnungsdaten der Krankenkassen zurückgegriffen wird. Da diese Daten jedoch ursprünglich nicht zu Forschungszwecken erhoben worden sind, sind typischerweise nicht alle notwendigen Informationen in den jeweiligen Datensätzen vorhanden oder sie liegen nur in eingeschränkter Qualität vor.

In diesem Vortrag werden anhand der am BIPS seit 2004 aufgebauten deutschen pharmakoepidemiologischen Forschungsdatenbank GePaRD mit Abrechnungsdaten von vier gesetzlichen Krankenversicherungen mit insgesamt ca. 20 Millionen Versicherten Limitationen dieser Daten und mögliche Ansätze zum Umgang mit diesen Limitationen in praktischen Beispielen aufgezeigt. Dabei werden zum einen statistische Lösungsansätze vorgestellt, aber auch Möglichkeiten diskutiert, fehlende oder nur lückenhaft vorhandene Informationen durch die Verlinkung mit anderen Datenbanken oder mit Primärdaten aufzufüllen. Es wird jedoch nicht nur gezeigt, wie z.B. Primärdaten Lücken in Sekundärdaten füllen können, sondern auch am Beispiel der NAKO Gesundheitsstudie das Potenzial von Sekundärdaten demonstriert, Primärdaten sinnvoll zu ergänzen.