

Der Kinder- und Jugendgesundheits-survey (KiGGS): Datenmanagement

Der Kinder- und Jugendgesundheits-survey (KiGGS) ist ein Befragungs- und Untersuchungssurvey, der von Mai 2003 bis Mai 2006 vom Robert Koch-Institut (RKI) in 167 für die Bundesrepublik repräsentativen Städten und Gemeinden unter der Zielsetzung durchgeführt wurde, erstmals umfassende und bundesweit repräsentative Daten zum Gesundheitszustand von Kindern und Jugendlichen im Alter von 0–17 Jahren zu erheben [1].

Für die Planung, Organisation und Durchführung der KiGGS-Studie ergab sich die Notwendigkeit, die datentechnischen Voraussetzungen nicht allein für die Erfassung und Analyse der Erhebungsdaten, sondern auch für die Unterstützung der umfangreichen und komplexen Geschäftsabläufe im Projekt zu schaffen. Das Datenmanagement für die Erhebungsdaten basierte auf den Statistiksystemen SPSS und SAS. Für diese Prozesse konnte auf bewährte Verfahren zur Datenaufbereitung, Prüfung und Bereinigung zurückgegriffen werden. Für die Entwicklung von Benutzeroberflächen und Prozesssteuerungskomponenten waren diese Programmpakete jedoch aufgrund ihrer fachlichen Ausrichtung nicht geeignet. Daher wurde eigens für diese speziellen Anforderungen ein IT-Teilprojekt initiiert. Langfristiges Ziel dieses IT-Projekts war die Erarbeitung einer standardisierten, erweiterbaren und allgemeingültigen Lösungsarchitektur, eines Frameworks, zur verbesserten

Studienbetreuung, zur Erhöhung der Datenqualität und zur Flexibilisierung der Datenanalyse. Die im Rahmen der KiGGS-Studie entwickelte erste Fassung des Frameworks, im Folgenden auch Basisframework genannt, hatte die Aufgabenstellung,

- für alle Beteiligten eine umfassende Datentransparenz zu schaffen,
- Bearbeitungsschritte zu standardisieren,
- die Bearbeitungsdauer zu reduzieren,
- Fehler zu minimieren und
- den Datenaustausch zu vereinfachen.

Das Basisframework

Das Basisframework ist eine Sammlung einzelner, weitestgehend unabhängiger Komponenten oder auch fachlicher Anwendungen, die auf einer Reihe von Basisdiensten aufsetzen. Die Basisdienste sind Standardfunktionen, wie z. B. Benutzerauthentifizierung, Zugriffssteuerung oder Ereignisprotokollierung, die von allen Komponenten nach Bedarf genutzt werden können. Jede Komponente ist als separate Datenbankapplikation realisiert und enthält einen Set fachlich zusammenhängender Anwendungsdienste bzw. fachspezifischer Funktionen (■ **Abb. 1**). Das Basisframework besteht aus folgenden Kernkomponenten:

- Stammdatenverwaltung (Probanden- und Sample-Point-Daten),
- Lager- und Logistikverwaltung,

- Labordatenverwaltung inklusive Qualitätssicherung und Befundung,
- Datenqualitätssicherung,
- formale Datenprüfung,
- Datenimport/-export.

Die technische Umsetzung des Basisframeworks

Der Lösungsansatz basiert auf dem Prinzip der 2-Schichten-Architektur (Client-Server). Dies bedeutet, dass die Ausführung der Geschäftsprozesslogik getrennt von der Datenhaltung erfolgt. Zur Verwaltung und Speicherung von Stamm- und Verwaltungsdaten dienen – ebenso wie für einen Teil der Erhebungsdaten – Microsoft SQL-Server-Datenbanken (Server). Die Entwicklung der Benutzerschnittstelle (Client) erfolgte mit Microsoft Access. Die Auswahl des Datenbankmanagementsystems sowie der Softwareentwicklungstools orientierte sich an den im RKI vorhandenen Softwarestandards. Für die Verwaltung von Benutzerzugriffsrechten sowie für die Datensicherung wurde die bestehende Infrastruktur (Software, Hardware) des RKI genutzt.

Die anonymisierten Untersuchungsdaten werden aus Gründen des Datenschutzes getrennt von den Personendaten (Name und Anschrift der Probanden) verwahrt. Die Zuordnung der Untersuchungs- zu den Personendaten ist nur über eine zentrale Stelle möglich und laut Einverständniserklärung auch nur

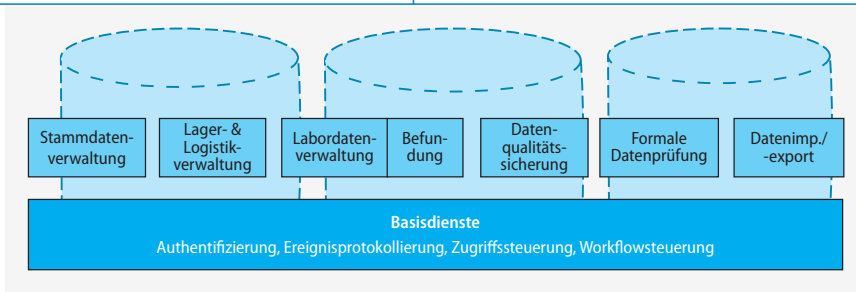


Abb. 1 ▲ **Übersicht Basisframework**

im Falle einer Wiederholungsuntersuchung erlaubt. Eine probandenbezogene Auswertung von Untersuchungsdaten ist somit nur in anonymisierter Form (Identifikationsnummer) durchführbar.

Nachfolgend werden exemplarisch die wichtigsten Funktionen zweier Aufgabenschwerpunkte des Datenmanagements, d. h. die Stammdatenverwaltung und die Qualitätssicherung der Erhebungsdaten, kurz vorgestellt.

Probanden- und Sample-Point-Stammdatenverwaltung

Bevor ein Proband durch das Feldteam befragt und untersucht werden kann, muss im Vorlauf eine Vielzahl von Prozessen durch die Mitarbeiter der Survey-Geschäftsstelle bearbeitet werden [2]. Diese Tätigkeiten dienen vor allem der Vorbereitung und der organisatorischen Begleitung der Feldphase. Die zu bearbeitenden Einzelprozesse unterscheiden sich, bezogen auf die Datenhaltung, durch ih-

re Bezugsgröße, z. B. Sample Point oder Proband. Zu den wichtigsten Prozessen zählen die

- Stichprobenbearbeitung,
- Maßnahmen für die Öffentlichkeitsarbeit,
- Probandendatenübernahme von Datenlieferanten,
- Probandendatenaktualisierung und -pflege,
- Probandenkontaktaufnahme und Terminverwaltung,
- Unterstützung der Feldlogistik,
- Koordinierung der Feldmitarbeiter.

Zur Unterstützung dieser vor allem organisatorischen Arbeitsprozesse innerhalb der Geschäftsstelle wurden diverse Softwarebausteine entwickelt, die in der zentralen Datenbankapplikation Verwaltungsdatenbank zusammengefasst sind (Abb. 2).

Die Verwaltungsdatenbank ist die zentrale Benutzerschnittstelle zur Terminkontrolle, zur Erfassung und Verwaltung

der Probandendaten sowie zur Bereitstellung verschiedener Kontrolllisten für die Feldmitarbeiter (Abb. 3).

Zur Steuerung und Überwachung komplexer oder zeitkritischer Arbeitsabläufe wurden in der Verwaltungsdatenbank aufgabenbezogenen Workflowkomponenten integriert. Diese kontrollieren und steuern unter anderem den termingerechten Versand von Serienbriefen (Einladung, Erinnerung und Terminbestätigung). Außer den Sample-Point- und den Probandendaten sind keine weiteren externen Daten in das System eingegangen [3]. Datenimport/-exportfunktionen konnten nur durch ausgewählte Geschäftsstellenmitarbeiter ausgelöst werden. Tabelle 1 stellt eine Auswahl der häufigsten durch die Verwaltungsdatenbank unterstützten Einzelaufgaben und der damit verbundenen Datenmengen dar.

Datenmanagement und Qualitätssicherung der Erhebungsdaten

Den Großteil der durch das Datenmanagement zu verwaltenden Daten stellen die Erhebungsdaten dar. Dazu zählen die Angaben aus den Fragebögen, das computergestützte ärztliche Interview, die Mess- und Laborwerte sowie die bei einigen Motoriktests direkt vom Messgerät elektronisch abgespeicherten Ergebnisse [2]. Der Datenfluss und die damit verbundenen Prozesse zur Aufbereitung, Prüfung und Bereinigung, die pro Erhe-

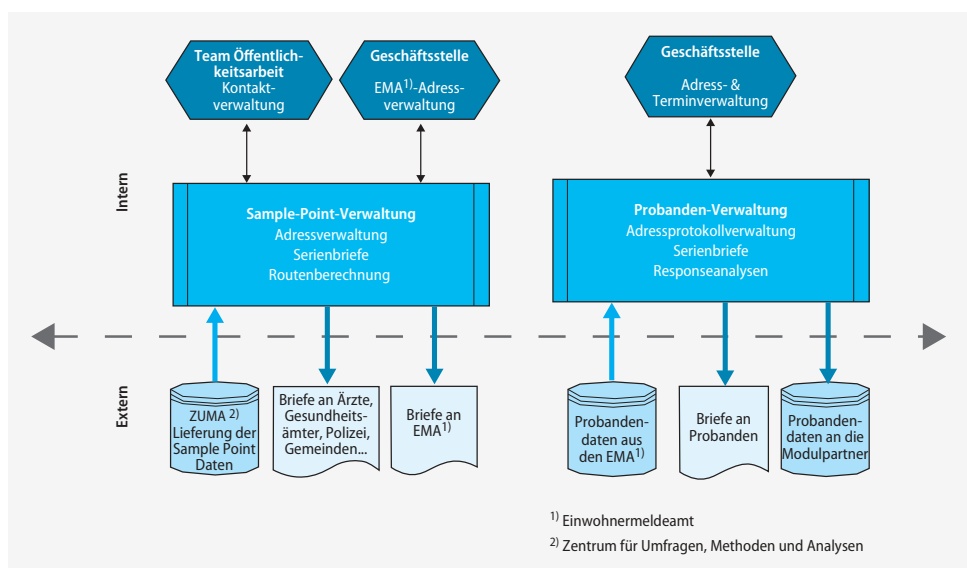


Abb. 2 ► **Übersicht Verwaltungsdatenbank**

¹) Einwohnermeldeamt

²) Zentrum für Umfragen, Methoden und Analysen

Bundesgesundheitsbl - Gesundheitsforsch - Gesundheitsschutz 2007 · 50:567–572
DOI 10.1007/s00103-007-0217-7
© Springer Medizin Verlag 2007

R. Dölle · A. Schaffrath Rosario · H. Stolzenberg

Der Kinder- und Jugendgesundheitsurvey (KiGGS): Datenmanagement

Zusammenfassung

Im Rahmen des Kinder- und Jugendgesundheitsurveys (KiGGS) wurden erstmals umfassend und bundesweit repräsentative Daten zum Gesundheitszustand von Kindern und Jugendlichen in Deutschland erhoben. Im Laufe der 3-jährigen Erhebungsphase wurden 17.641 Probanden untersucht und befragt und ca. 1.500 Items erfasst. Das Datenmanagement beschränkte sich nicht allein auf die Erfassung, Verwaltung und Qualitätssicherung der Erhebungsdaten, sondern ebenso auf die Bereitstellung von Werkzeugen zur Verwaltung und Kontrolle von Prozessdaten sowie zur Steuerung surveyspezifischer Geschäftsabläufe. Im Rahmen der KiGGS-Studie wurde eine Vielzahl von Komponenten zur Unterstützung der umfangreichen und komplexen Prozessabläufe für die Stu-

dienmitarbeiter entwickelt. Dies betraf in erster Linie die Aufgabenbereiche Probanden- und Sample-Point-Verwaltung, Terminplanung, Stichprobenziehung, Berichterstattung, Feldlogistik und Labordatenverwaltung. Die computergestützte Bearbeitung von Routineaufgaben im Rahmen der Feldorganisation führte zu einer deutlichen Arbeitserleichterung sowie zu einer verbesserten Projektablaufkontrolle. Teilweise konnten die KiGGS-spezifischen Komponenten mit minimalem Anpassungsaufwand bereits auf andere Studien übertragen werden. Beim Datenmanagement der Erhebungsdaten stand die Standardisierung der Verfahren ihrer Aufbereitung, Prüfung und Bereinigung im Vordergrund. Hier konnte auf einen breiten Erfahrungsschatz aufgebaut werden. Die eta-

blierten Methoden zur Qualitätssicherung wurden weitgehend standardisiert und teilweise automatisiert und durch Datenbanktools zur Verwaltung und Dokumentation von Erhebungsinstrumenten und Qualitätssicherungsmaßnahmen ergänzt. Die Summe aller Maßnahmen ermöglichte es, den Datennutzern relativ schnell einen geprüften und bereinigten Enddatensatz einschließlich einer ausführlichen Dokumentation zur Verfügung zu stellen.

Schlüsselwörter

Gesundheitsurvey · Kinder · Jugendliche · Datenmanagement · Datenqualitätssicherung · Datenerhebung · Ablauforganisation

The German Health Interview and Examination Survey for Children and Adolescents (KiGGS): Data management

Abstract

In the German Health Interview and Examination Survey for Children and Adolescents (KiGGS), comprehensive, nationwide representative data on the state of health of children and adolescents were collected for the first time. During the 3-year data collection phase of the survey, 17,641 subjects were examined and interviewed and approx. 1,500 items were recorded. Data management was not limited to survey data collection, administration and quality assurance alone, but also comprised the provision of tools for the management and control of process data, as well as for managing survey-specific business processes. In the context of the KiGGS study, numerous components for supporting the extensive

and complex processes were developed for the study staff. Here the primary focus was on subject and sample point administration, scheduling, sampling, reporting, field logistics and laboratory data management. Thanks to the computer-based processing of routine tasks involved in the organisation of the field work, ease of work and project progress control were enhanced significantly. To some extent, KiGGS-specific components have already been used in other studies and only minor adaptations were needed for the transfer. The main emphasis with regards to survey data management was on the standardisation of methods for data processing, data control and data cleaning. Here, a wealth of previous experiences was

available as a starting point. The established quality assurance methods were standardised to a large extent and partly automated and complemented by data base tools for the management and documentation of survey instruments and quality assurance measures. All these measures combined made it possible to provide data users with a controlled and cleaned final data set, including a detailed documentation.

Keywords

Health survey · Children · Adolescents · Data management · Data quality assurance · Data collection · Workflow organisation

Abb. 3 ► Benutzeroberfläche zur Verwaltung von Probandendaten

bungswelle abzarbeiten sind, werden in **Abb. 4** dargestellt.

Datenerfassung

Die auf Papier erhobenen Daten, d. h. der Eltern- und Kinderfragebogen, das Messblatt mit den Untersuchungsergebnissen etwa zu Körpermaßen und Blutdruck sowie der Ernährungsfragebogen wurden mittels SPSS-DataEntry von geschulten Datentypistinnen erfasst. Die Qualität der Datenerfassung wurde im Rahmen

der externen Qualitätssicherung über eine 10-%ige Doppeleingabe überprüft, wobei sich lediglich eine Gesamtfehlerquote von 0,03% ergab [4]. Freitexte wurden zunächst im Original erfasst und anschließend kodiert.

Datentransformation

Für Daten, die computergestützt erhoben wurden (ärztliches Interview/CAPI) oder bereits in elektronischer Form vorlagen (Labordaten, Daten motorischer

Funktionstests) wurden gegebenenfalls Datentransformationen automatisch vorgenommen. Dazu zählen beispielsweise Format- und Werteanpassungen sowie die Generierung neuer Variablen.

Abgleich der Identitätsmerkmale

Anhand einer Kontrolldatei wurde zunächst die Integrität der Stammdaten (Brutto- und Nettonummer, Geburtsdatum, Geschlecht, Alter, Untersuchungsdatum, Zuordnung zum Sample Point) durch Abgleich der verschiedenen Datenquellen (Fragebogen, ärztliches Interview, Verwaltungsdatenbank) sichergestellt. Dabei wurden auch die Vollständigkeit der Instrumente und die Eindeutigkeit der Probandenidentifikationsnummern überprüft. Ein ebensolcher Abgleich fand mit den Daten der Modulpartner statt.

Formale und inhaltliche Datenprüfung

In einem zusätzlichen Schritt der Datenqualitätsprüfung wurden ausgewählte Variablen aus dem Messbogen zeitnah auf Interviewer- und Geräteeffekte, Endziffernpräferenzen und andere Auffällig-

Tabelle 1

Mengengerüst der KiGGS-Verwaltungsdatenbank (Auswahl)

Arbeitsvorgang	Typ/Adressat/Lieferant	Datensatzanzahl
Serienbriefverarbeitung (probandenbezogen)	Elternansreiben	28.300
	Jugendansreiben	6270
	Erinnerungsansreiben	19.660
	Adressprotokolle	28.300
	Befundansreiben	17.640
	Ausländerspezifische Ansreiben	3530
Serienbriefverarbeitung (Sample-Point-bezogen)	Pressedienste, Bürgermeister, Apotheken, Ärzte, Kirchen ...	7566
	Einwohnermeldeämter	169
Importierte Personendaten der	Einwohnermeldeämter	534.720

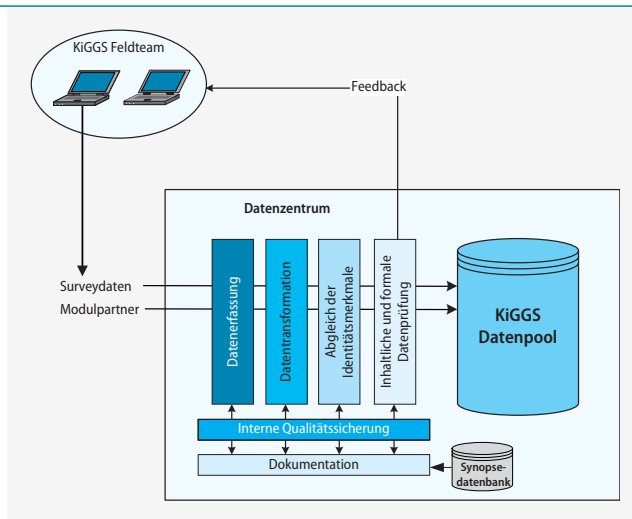


Abb. 4 ► Prozesse der Datenaufbereitung und Prüfung

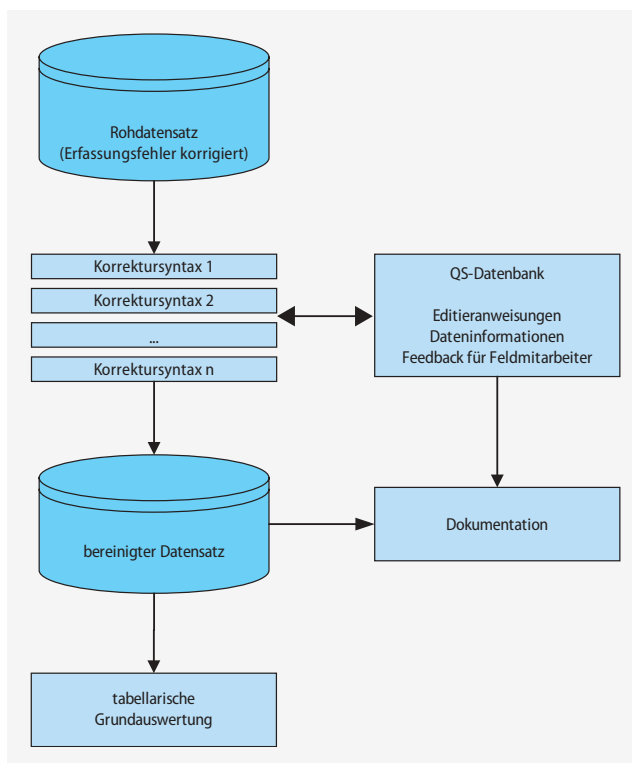


Abb. 5 ► Datenflussplan der Erhebungsdaten

keiten geprüft. Dadurch war es möglich, gegebenenfalls Rückmeldungen an das Feldteam zu geben und so die Qualität der erhobenen Daten kontinuierlich zu verbessern [2].

Schließlich fand eine umfassende Plausibilitätsprüfung aller Fragebogen- und Interviewdaten statt. Wegen der Größe des Gesamtdatensatzes, sowohl was die Zahl der Probanden (N=17.641) als auch die Zahl der Variablen (ca. 2000) angeht, war ein Team von 6 Personen mit der Datenprüfung befasst. Diese beinhaltete so-

wohl formale als auch inhaltliche Aspekte. Entscheidungen bezüglich der Bereinigung von Implausibilitäten wurden unter Einbeziehung der fachlich verantwortlichen Mitarbeiter vorgenommen. Besondere Aufmerksamkeit bei der Prüfung erforderte auch die Tatsache, dass die Erhebungsinstrumente je nach Altersgruppe unterschiedlich aufgebaut waren.

Zur Gewährleistung eines einheitlichen Vorgehens bei Datenänderungen wurden innerhalb des Teams allgemeine Regeln zur Datenprüfung und -korrektur

vereinbart. Hierzu wurden alle Editieranweisungen im Team diskutiert und in einer Datenbank dokumentiert. Der Dokumentationsaufwand ist damit vergleichsweise hoch, aber nur so sind Transparenz, Einheitlichkeit, Wiederholbarkeit und Korrigierbarkeit von Datenkorrekturen zu gewährleisten. Zudem wurden im Rahmen der Dokumentation alle während des Aufbereitungsprozesses gewonnenen Erkenntnisse zur fachgerechten Interpretation der Variablen festgehalten. Die elektronische Dokumentation erlaubt es außerdem, für die Auswertung relevante Editieranweisungen zu selektieren und als sog. „Dateninfo“ den Datennutzern zur Verfügung zu stellen.

Bei der Datenaufbereitung wurde der Rohdatensatz, der nur Korrekturen von Erfassungsfehlern enthält, vom aufbereiteten, voll editierten Datensatz getrennt gehalten (■ Abb. 5). Datenänderungen erfolgten über eine Reihe programmierter Korrekturanweisungen, sodass die Änderungen nachvollziehbar und gegebenenfalls selbst wieder korrigierbar sind. Zudem ist es so möglich, bei Bedarf festzustellen, wie häufig gewisse Änderungen vollzogen wurden.

Das Ergebnis der Datenqualitätssicherung ist ein bereinigter, aber nicht vollständig widerspruchsfreier Datensatz. So wurden beispielsweise keine Abgleiche zwischen verschiedenen Fragebogenteilen durchgeführt, etwa zwischen Eltern- und Kinderfragebogen; schließlich stellen die unterschiedlichen Sichtweisen von Kindern und Eltern selbst einen Aspekt der im KiGGS untersuchten Phänomene dar. Bei stetigen Variablen ist im Hinblick auf statistische Auswertungen weiterhin zu beachten, dass nicht alle Extremwerte („Ausreißer“) aus den Daten eliminiert wurden, sondern nur solche Werte, die als unplausibel galten.

Dokumentation und Enddatensatz

Das zentrale Tool zur Verwaltung und Dokumentation aller Erhebungsinstrumente stellt die sogenannte „Synopsis-Datenbank“ dar, die neben einem „Data Dictionary“ eine wesentlich erweiterte Funktionalität bietet. Diese Spezialdatenbank enthält alle studienbezogenen Erhebungsinstrumente und stellt somit ein Abbild der

Gesamtheit der Erhebungsinstrumente dar. Neben der rein dokumentarischen Funktion ermöglicht sie auch die Steuerung der Auswahl und Exportreihenfolge von Variablen. Somit lässt sie sich auch zur automatisierten Erzeugung des bereinigten Enddatensatzes einsetzen, der zurzeit sowohl im SPSS- als auch im SAS-Format vorliegt. Die Bereitstellung des Datensatzes einschließlich Dokumentation wird ergänzt durch eine tabellarische Grundausswertung aller Variablen.

Erfahrungen mit dem KiGGS-Datenmanagement

Bei der Konzeption des Basisframeworks zur Unterstützung der Arbeitsprozesse standen die Aspekte Kostenreduzierung, Anpassbarkeit und Übertragbarkeit im Vordergrund. Das Datenmanagement und die Qualitätssicherung der Erhebungsdaten orientierten sich vorrangig an den Prinzipien der Transparenz und Nachvollziehbarkeit sowie der Einheitlichkeit und der Dokumentation von Datenkorrekturen.

Der Praxisbeweis und die Alltags-tauglichkeit des KiGGS-Basisframeworks wurden in der 3-jährigen Studienphase ausführlich unter Beweis gestellt. Die Nutzung von Standardsoftware für die Entwicklung von Frontendapplikationen hat sich als vorteilhaft erwiesen, da auf kostspielige Softwareentwicklungen verzichtet werden konnte. In der KiGGS-Studie standen für Softwareentwicklung und Datenbankmanagement lediglich 1,5 Mitarbeiter zur Verfügung.

Das KiGGS-Basisframework ist gekennzeichnet durch die KiGGS-Studiencharakteristika. Dennoch ist eine Übertragbarkeit einzelner Module gegeben, wodurch sich deutliche Rationalisierungseffekte einstellen können. Beispielsweise konnte nach geringfügiger Modifikation die KiGGS-Verwaltungsdatenbank erfolgreich bei weiteren RKI-Studien wiederverwendet werden. Dadurch reduzierte sich der Zeitraum bis zur Bereitstellung der IT-Basisstruktur in der Vorbereitungsphase für diese Studien um ca. 70%.

Die Qualitätssicherung der Erhebungsdaten war ein zeit- und personalintensiver Prozess. Dieser ist jedoch unerlässlich, um eine hohe Datenqualität zu erreichen. Das

Ergebnis ist ein bereinigter Datensatz, der unmittelbar für statistische Auswertungen nutzbar ist und dementsprechend eine hohe Akzeptanz bei den Datennutzern erzielt. Für einfache deskriptive Fragestellungen liefert zudem die ausführliche tabellarische Grundausswertung die notwendigen Informationen.

Bewährt hat sich die enge und direkte Zusammenarbeit zwischen dem Datenmanagement und den Anwendern und Datennutzern über den ganzen Erhebungsprozess hinweg sowohl in Bezug auf die Verwaltungsdaten als auch auf die Erhebungsdaten.

Fazit und Ausblick

Die Einbindung von IT-Komponenten in den gesamten empirischen Forschungsprozess, beginnend mit der Stichprobenakquise bis zur Datenaufbereitung, verspricht ein enormes Problemlösungs- und Optimierungspotenzial, das letztlich auch zur Sicherung der Datenqualität und Weiterentwicklung methodisch-inhaltlicher Standards beiträgt. In Zukunft wird das KiGGS-Basisframework derart weiterentwickelt werden, dass es die notwendige technische Unterstützung für longitudinale Untersuchungen bietet, wie etwa für den Aufbau einer KiGGS-Kohorte [1] und die Etablierung eines bundesweiten Health Monitoring am RKI.

Auch die bei der Qualitätssicherung der Erhebungsdaten für den KiGGS erarbeiteten Standards bilden eine wertvolle Ressource für die Datenprüfung in künftigen Erhebungen. Ziel ist es, eine Vereinheitlichung der Vorgehensweisen innerhalb des Instituts über Studien hinweg zu erreichen. Die im Laufe des Qualitätssicherungsprozesses erarbeitete Dokumentation bietet darüber hinaus eine gute Grundlage für die Dokumentation zum KiGGS-Public-Use-File, der im Herbst 2008 der wissenschaftlichen Öffentlichkeit zur Verfügung gestellt werden soll.

Korrespondierender Autor

Rüdiger Dölle

Robert Koch-Institut
Abteilung für Epidemiologie und
Gesundheitsberichterstattung
Postfach 650261
13302 Berlin, BRD
E-Mail: DoelleR@rki.de

Literatur

1. Kurth B-M (2007) Der Kinder- und Jugendgesundheits survey (KiGGS): Ein Überblick über Planung, Durchführung und Ergebnisse unter Berücksichtigung von Aspekten eines Qualitätsmanagements. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz 50:533–546
2. Hölling H, Kamtsiuris P, Lange M et al. (2007) Der Kinder- und Jugendgesundheits survey (KiGGS): Studienmanagement und Durchführung der Feldarbeit. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz 50:557–566
3. Kamtsiuris P, Lange M, Schaffrath Rosario A (2007) Der Kinder- und Jugendgesundheits survey (KiGGS): Stichprobendesign, Response und Nonresponse-Analyse. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz 50:547–556
4. Filipiak-Pittroff B, Wölke G (2007) Externe Qualitätssicherung im Kinder- und Jugendgesundheits survey (KiGGS). Vorgehensweise und Ergebnisse. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz 50:573–577