



^b
**UNIVERSITÄT
BERN**

**Interfaculty Centre for Educational
Research (ICER)**

Philosophisch-humanwissenschaftliche Fakultät
Institut für Erziehungswissenschaft
Wirtschafts- und Sozialwissenschaftliche Fakultät
Institut für Soziologie

www.icer.unibe.ch

Evaluation von Datenverknüpfungen mit Daten aus dem Scientific Use File der ÜGK und Registerdaten des BFS

Tobias Ackermann, Andrea B. Erzinger

Bern, 09.08.2023 (Version 2)

Zitiervorschlag: Ackermann, Tobias; Erzinger, Andrea B. (2023). *Evaluation von Datenverknüpfungen mit Daten aus dem Scientific Use File der ÜGK und Registerdaten des BFS*. Universität Bern: Interfaculty Centre for Educational Research. <https://doi.org/10.48350/182873>

Inhalt

1.	Ausgangslage.....	3
2.	Einleitung	4
3.	Dokumentation und Evaluation des Verknüpfungsprozesses.....	6
3.1.	Projekt 1: Verknüpfung des SUF aus den ÜGK-Erhebungen 2016 und 2017	6
3.2.	Projekt 2: Verknüpfung der Daten aus der ÜGK H4 Präpilotierung	10
3.3.	Projekt 3: Verknüpfung der Daten aus der ÜGK H4 Piloterhebung	13
3.4.	Fazit der Dokumentation und Evaluation des Verknüpfungsprozesses mit einem Fokus auf die Nutzung von verknüpften Daten für den Erstberichterstattungsprozess.....	14
4.	Validierung von Konstrukten der ÜGK durch Registerdaten.....	15
4.1.	Messfehler und fehlende Angaben bei Antworten von Schülerinnen und Schülern	15
4.2.	Abdeckung der Registerdaten	18
5.	Zusammenfassung und Bewertung des Nutzens einer Verknüpfung für die ÜGK	20
6.	Fazit und Ausblick	24
7.	Literatur	25
8.	Datensätze:	27

1. Ausgangslage

Das ICER beantragte beim Koordinationsstab für die Umsetzung der interkantonalen Vereinbarung über die Harmonisierung der obligatorischen Schule (Kosta HarmoS; Steuerungsgremium der *Überprüfung des Erreichens der Grundkompetenzen*) seit November 2020 Verknüpfungen von drei verschiedenen Datensätzen aus ÜGK-Erhebungen mit Daten des Bundesamts für Statistik (BFS). Neben wissenschaftlichem Forschungsinteresse motivierten folgende Punkte die mit den Verknüpfungsanträgen in Verbindung stehenden Projekte:

- Dokumentation und Evaluation des Verknüpfungsprozesses als Information für die Projektleitung ÜGK (PL ÜGK)
- Nutzung von Registerdaten für die Validierung von Informationen aus dem Schülerfragebogen, die zur Generierung von projekt- und forschungsrelevanten Konstrukten (wie z.B. dem sozioökonomischen Status) genutzt werden
- Beitrag zur „once-only“-Strategie des Bundes und Prüfung potenzieller Kosteneinsparungen für die ÜGK
- Ausloten einer möglichen Nutzung von Registerdaten im Rahmen der Datenaufbereitung und –verwertung sowie zur erweiterten Kontextualisierung der Erhebungsdaten (zusätzliche Informationen aus anderen Quellen)
- Datenaufbewahrung und –nutzung für externe Forschende (von einzelnen oder aggregierten Merkmalen aus administrativ Daten im Rahmen des SUF)

Der vorliegende Bericht umfasst die Ergebnisse der bisher durchgeführten Projekte und leistet einen Beitrag zur Klärung der ersten drei Punkte. Das Eruiieren von Potenzialen der Nutzung von Registerdaten für die Datenaufbereitung und –verwertung sowie von Möglichkeiten zur Datenaufbewahrung und –nutzung einzelner oder aggregierter Merkmale aus den Registerdaten für externe Forschende ist Gegenstand laufender Projekte. Diesbezügliche Erkenntnisse werden zu einem späteren Zeitpunkt in einem weiteren Bericht zusammengefasst.

2. Einleitung

In Large-Scale Assessments (LSA) wie beispielsweise der Überprüfung des Erreichens der Grundkompetenzen (ÜGK) oder dem Programme for International Student Assessment (PISA) nehmen Schülerinnen und Schüler (SuS) einerseits an Leistungstests teil. Andererseits beantworten sie zahlreiche Fragen in einem Fragebogen, einige davon zu ihrer sozialen Herkunft. Diese Angaben aus dem Fragebogen für Schülerinnen und Schüler (SFB) dienen der Kontextualisierung der schulischen Leistung, da davon ausgegangen wird, dass sie im direkten Zusammenhang mit den Ergebnissen aus den Leistungstests stehen. Neben einer inhaltlichen Kontextualisierung werden diese Merkmale auch für die Imputation von fehlenden Angaben und für die Schätzung der Plausible Values¹ verwendet. Letzteres dient im Falle von LSA der Leistungsmessung, weswegen die Vollständigkeit sowie die Qualität der Kontextualisierungsmerkmale auch für Auswertungen der Leistungstests von zentraler Bedeutung sind.

Die Kontextualisierung von schulischen Leistungen durch Informationen aus dem SFB sind in LSA gängig, jedoch ist dieses Vorgehen nicht unproblematisch. Die Literatur zeigt, dass Proxy-Angaben von SuS – etwa Aussagen des Kindes über den Beruf der Mutter – fehlerhaft sein können (u.a. Ensminger et al., 2000; Kreuter et al., 2010; Engzell & Jonsson, 2015; Hovestadt & Schneider, 2021). Die Problematik liegt gemäss dieser Studien darin, dass Zusammenhänge zwischen der kognitiven Leistung von SuS und der Qualität ihrer Antworten bestehen: SuS mit höheren Testleistungen sind in der Lage, validere Aussagen zu machen. Kritisch für eine adäquate Kontextualisierung der Leistungsmessung wird diese Problematik insbesondere dann, wenn systematische Verzerrungen in den Antworten (Messfehler) oder Muster bei fehlenden Angaben vorliegen. Je jünger die SuS sind, desto grösser ist das Risiko, dass sich die Qualität ihrer Antworten weiter reduziert (Ensminger et al., 2000; Kreuter et al., 2010; Ridolfo & Maitland, 2011; Wittrock et al., 2017), was insbesondere die geplante ÜGK-H4-Erhebung bei achtjährigen SuS vor besondere Herausforderungen stellt.

Vor diesem Hintergrund lancierte das ICER mehrere aufeinander aufbauende Projekte mit dem Ziel, Potenziale von Datenverknüpfungen zur Erhöhung der Datenqualität zu eruieren. Ein (möglicher) Ansatz besteht darin, auf bestimmte Items im SFB zu verzichten und stattdessen Informationen aus Registerdaten des Bundesamts für Statistik (BFS) auf Individualebene herbeizuziehen. Dieser Ansatz ist vielversprechend, zumal das BFS über verschiedene Informationen auf Individualebene verfügt, welche den Kontextmerkmalen in der ÜGK entsprechen. Vorausgesetzt, dass die meisten SuS

¹ Um die zeitliche Belastung der SuS bei Leistungstests zu reduzieren, erhalten die SuS systematisch ausgewählte Aufgaben. Anhand deren Bearbeitung kann für das einzelne Kind eine plausible Leistung geschätzt werden, sog. Plausible Values. Im Fall der ÜGK wird diese in eine binäre Variable umgewandelt, die angibt, ob die Grundkompetenzen in diesem Leistungsbereich erreicht wurden oder nicht.

in den Registerdaten identifizierbar sind und dass die Daten des BFS über eine ausreichende Qualität verfügen, vermag dieser Ansatz nicht nur zur Erhöhung der Datenqualität beizutragen, sondern ermöglicht es auch, in Sinne des „once-only“-Prinzips die Befragungslast bei den SuS zu verringern und Ressourcen einzusparen.

Konkret bedeutet eine Datenverknüpfung, dass Informationen über dieselbe Person aus unterschiedlichen Datenquellen in einem Datensatz zusammengeführt werden. Dazu muss eine Person in beiden Datenquellen eindeutig identifiziert werden können, was über Personenmerkmale wie Namen, Geschlecht, Geburtstag und Wohnort geschieht. Nur wenn die Identifizierung eindeutig ist, können Identifikationsnummern² verschiedener Datenquellen einander zugeordnet werden. In der Folge können alle weiteren Datenquellen herbeigezogen werden, sofern sie eine der zugeordnete Identifikationsnummern verfügen.

Der Fokus des vorliegenden Berichts liegt primär auf den eingangs erwähnten Punkten. Dazu werden in einem ersten Schritt (Kapitel 3: Dokumentation und Evaluation des Verknüpfungsprozesses) die bis dato durchgeführten drei Verknüpfungsprojekte beschrieben und die jeweils zentralen Erkenntnisse bezüglich des Verknüpfungsprozesses festgehalten. In einem zweiten Schritt (Kapitel 4: Validierung von Konstrukten der ÜGK durch Registerdaten) liefern die Ergebnisse aus dem ersten Forschungsprojekt zu den Daten der ÜGK-Erhebungen von 2016 und 2017 Auskünfte über die Datenqualität in Bezug auf Messfehler, fehlende Angaben sowie über die Abdeckung von Registerdaten. In einem dritten Schritt (Kapitel 5: Zusammenfassung und Bewertung des Nutzens einer Verknüpfung) werden die gewonnen Erkenntnisse synthetisiert und mit einem Blick auf einen möglichen Beitrag der Verknüpfung mit Registerdaten zur „once-only“-Strategie des Bundes und zur Prüfung potenzieller Kosteneinsparungen für die ÜGK beurteilt. Im letzten Kapitel (Kapitel 6: Fazit und Ausblick) werden die Erkenntnisse zusammengefasst und ein Ausblick skizziert.

² Identifikationsnummern sind anonymisierte Nummern, welche in den Datensätzen hinterlegt sind, so dass alle Daten, die vom gleichen Kind stammen (beispielsweise Daten aus dem Leistungstest und dem SFB), verbunden werden können.

3. Dokumentation und Evaluation des Verknüpfungsprozesses

Dieser Abschnitt befasst sich mit drei Verknüpfungsprojekten, die seit September 2020 erfolgt sind. In Projekt 1 wurden Daten aus den ÜGK-Erhebungen aus den Jahren 2016 und 2017 mit Registerdaten verknüpft. Dabei handelt es sich um den Fall, dass das Scientific Use File (SUF) bei Projektstart *verfügbar ist und bereits eine Zuordnung der ÜGK-Identifikationsnummer der SuS zur AHV-Nummer vorhanden ist (Standardfall bei Verknüpfungen der ÜGK-Daten)*. Projekt 1 zielte darauf ab, das Verfahren des Verknüpfungsprozesses kennenzulernen und zu prüfen, ob ÜGK-Konstrukte validiert werden können (siehe Kapitel 4) und ob sich Verknüpfungen für künftige Erhebungen bewähren, damit sie allenfalls in weitere Prozesse der ÜGK (wie z.B. den Erstberichterstattungsprozess) integriert werden können.

In Projekt 2 werden die Daten der ÜGK H4 Präpilottierung verknüpft. Dabei erfolgt die Zuordnung der ÜGK-Identifikationsnummer der SuS zur AHV-Nummer *vor der Datenerhebung, also früher als im Prozess standardmässig vorgesehen*. In Projekt 3 werden Daten der ÜGK H4 Piloterhebung verknüpft. Ein Spezifikum lag hierbei darin, dass die *Verknüpfung über einen Vertragszusatz geregelt* werden konnte. Bei beiden Vorhaben war das Ziel zu evaluieren, wie der Prozess optimiert werden könnte, um Registerdaten für eine Erstberichtserstattung zu gebrauchen.

Mit jedem Antrag versuchte das Team am ICER bereits Erkenntnisse aus vorangegangenen Projekten umzusetzen, was zu dynamischen Veränderungen der Prozesse geführt hat. Als Fazit werden jeweils die zentralen Erkenntnisse aus den einzelnen Projekten zusammengefasst.

3.1. Projekt 1: Verknüpfung des SUF aus den ÜGK-Erhebungen 2016 und 2017

Für eine Verknüpfung des ÜGK-SUF mit den Registerdaten des BFS sind verschiedene Voraussetzungen zentral:

Aufseiten des BFS gilt es, den Identifikationsnummern der SuS aus den ÜGK-Daten der Jahre 2016 und 2017 eine eindeutige AHVN13-Nummer zuzuordnen. Spezifisch bedeutet das, dass die Merkmale Vorname, Name, Wohnort, Geschlecht und Geburtsdatum der Teilnehmenden der ÜGK eindeutig in beiden Datenquellen identifiziert werden (Voraussetzung 1 für eine Datenverknüpfung). Der Schlüssel für diese Zuordnung (Schlüssel 1) wird gemäss Datennutzungskonzept (EDK 2021; Punkt 5.3.2) durch das BFS sicher aufbewahrt. Bei der Aufbereitung des ÜGK-SUF durch die von der EDK definierte datensichernde Institution zur Hinterlegung auf SWISSUbase, ersetzt diese die Identifikationsnummer der SuS aus den ÜGK-Daten durch eine neue Nummer und ein zweiter Schlüssel (Schlüssel 2) wird generiert, der die neue mit der ursprünglichen ÜGK-Identifikationsnummer verbindet (EDK 2021, Kapitel 4.3). Ziel dieses Vorgehens ist, dass es zukünftig nur mit beiden Schlüsseln

möglich ist, die Daten aus der ÜGK 2016 und ÜGK 2017 mit Registerdaten des BFS zu verknüpfen. Während einer davon bei der datensichernden Institution liegt, wird der andere beim BFS aufbewahrt (EDK 2021).

Das SUF lässt sich über SWISSUbase beantragen, was administrativ wenig Aufwand darstellt und an gewisse, hier nicht weiter dargelegte Voraussetzungen geknüpft ist. Wird jedoch eine Verknüpfung mit Registerdaten angestrebt, ist diese beim Kosta HarmoS zu beantragen. Dieser Antrag umfasst die detaillierte Beschreibung des Projekts, ob eine Publikation beabsichtigt wird und welche Datenbestände beim BFS bestellt werden. Die positive Evaluation dieses Antrags durch den Kosta HarmoS ist Voraussetzung 2 für eine Datenverknüpfung (EDK 2021). Erst dann wird der bei der datensichernden Institution hinterlegte Schlüssel an das BFS herausgegeben, was die Verknüpfung erst ermöglicht. Wenn bei Beantragung der Daten auf SWISSUbase keinen vom Kosta HarmoS bewilligten Datenverknüpfungsantrag vorgelegt wird, wird die datensichernde Institution den Schlüssel 2 nicht ans BFS herausgeben und die Daten können nicht mit den Registerdaten verknüpft werden.

Voraussetzung 3 für die Verknüpfung stellt weiter ein bewilligter Antrag zur Verknüpfung beim BFS dar. Dieser beschreibt den Zweck des Projekts und beinhaltet eine genaue Auflistung aller Variablen und Datenquellen, die verknüpft werden sollen. Im Falle der Dateneigentümerschaft bei Drittparteien wird eine schriftliche Erlaubnis ebendieser vorausgesetzt (was hier der im vorangegangenen Absatz beschriebene Beschluss des Kosta HarmoS bezüglich der beantragten Datenverknüpfung darstellt). Weiter gilt es, das Publikationsvorhaben zu beschreiben.

Zusammenfassend gilt, dass *bereits bei der Durchführung eines Projekts erste Voraussetzungen (Generieren der für eine Verknüpfung relevanten Informationen im Samplingprozess, Beantragung der Verknüpfung und der Hinterlegung des Verknüpfungsschlüssels (Schlüssel 1) beim BFS) geschaffen werden müssen, damit später eine Verknüpfung mit Registerdaten erfolgreich möglich ist.* Bei der ÜGK sind zudem *das Generieren des Schlüssels 2 durch die datensichernde Institution (EDK 2021), eine positive Evaluation des Verknüpfungsantrags durch den Kosta HarmoS sowie ein erfolgreicher Antrag auf Verknüpfung beim BFS* relevante Voraussetzungen für eine erfolgreiche Verknüpfung.

Die schematische Darstellung (Abbildung 1: Verknüpfung des SUF der ÜGK-Erhebungen 2016 und 2017) zeigt, welche Schritte im Falle des Antrags des ICER auf Datenverknüpfung wann erfolgt sind und beziffert damit auch den ungefähren Zeitbedarf, welche die Verknüpfung in diesem Fall benötigt hat.

Das Projekt begann im September 2020 damit, dass dem Kosta HarmoS ein Antrag unterbreitet wurde, welcher das Projekt umschrieb. Parallel dazu wurde bereits damit begonnen, den Antrag zuhanden des BFS zu erstellen. Allein das Erstellen dieses Antrags dauerte ungefähr vier Monate und beinhaltete mehrere Kontaktaufnahmen mit Vertreterinnen und Vertretern der unterschiedlichen Sektionen des BFS. Begleitend stellte das BFS eine Kontaktperson für das Projekt. Ein Hauptteil der Arbeit war, die Liste der Variablen aus den Registern zu erstellen, welche verknüpft werden sollen. Eine Herausforderung dabei bestand darin, dass die Dokumentationen der Registerdaten beim BFS für Aussenstehende zum Teil schlecht zugänglich oder teilweise unverständlich waren. Erst durch persönliche Gespräche mit den entsprechenden Kontaktpersonen konnte geklärt werden, welche Informationen mit welchen praktischen Limitationen beim BFS verfügbar sind. Nach Eingabe des Antrags beim BFS wurde dieser durch die juristischen Dienste sowie den involvierten Sektionen geprüft und in die Unterschriftenrunde gegeben. Die Datenlieferung erfolgte nach abgeschlossener Unterschriftenrunde im Juni 2021.

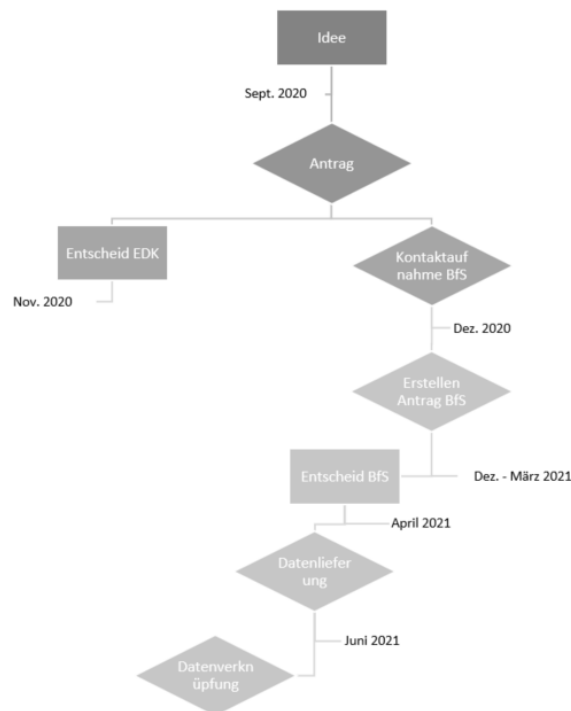


Abbildung 1 Verknüpfung des SUF der ÜGK-Erhebungen 2016 und 2017

Trotz intensiver Vorabklärungen wurden bei der ersten Datenbestellung zwei für die Analyse relevante Informationen nicht bestellt, was erst bei der ersten Prüfung der Daten bemerkt wurde. Hier

war eine Nachbestellung jedoch einfach, da lediglich eine informelle Zustimmung der beteiligten Personen benötigt wurde, weil die Variablen aus der gleichen Population, denselben Jahren und derselben Datengrundlage stammten, auf welche sich bereits die Erstbestellung bezog. Konkret war der komplette Datensatz im September 2021 beim ICER.

Zentrale Erkenntnisse der Verknüpfung ÜGK 2016/ÜGK 2017 mit Registerdaten (Projekt 1)

1. Voraussetzungen für Verknüpfungen: Projekt 1 zeigt, dass folgende Punkte für eine erfolgreiche Verknüpfung des ÜGK-SUF mit Registerdaten notwendig sind: **Bewilligung durch den Kosta Har-moS; Verknüpfungsantrags zu Händen des BFS; Ansprechperson für ÜGK-Verknüpfungen.**

2. Im vorliegenden Fall verlief die im Rahmen der ÜGK standardmässige Beantragung der Verknüpfung nicht einwandfrei: In Bezug auf die Hinterlegung des Verknüpfungsschlüssels beim BFS durch die PL ÜGK sowie in Bezug auf die Absprache zwischen der EDK und der datensichernden Institution hinsichtlich der Aufbereitung und Bereitstellung der Daten (insbesondere des zweiten Verknüpfungsschlüssels) scheint das Vorgehen zu wenig klar gewesen zu sein. Da die datensichernde Institution den zweiten Verknüpfungsschlüssel dem BFS nicht ordnungsgemäss zur Verfügung gestellt hatte, war die Verknüpfung der ÜGK-Daten 2016 und 2017 anfänglich nicht möglich. Es musste in diesem konkreten Fall auf archivierte Dokumente des Konsortium ÜGK zurückgegriffen werden, welche in diesem Prozess offiziell nicht relevant sein sollten und im Normalfall bereits vernichtet gewesen wären. **Es ist daher unabdingbar, dass dieser Prozess zwischen der EDK als Dateninhaber-in und der datensichernden Institution rückwirkend für die bereits veröffentlichten ÜGK-SUFs, aber auch für die Veröffentlichung zukünftiger ÜGK-SUFs geklärt wird.** Anhand dieses Projekts lässt sich konstatieren, dass die Möglichkeit der Nutzung von Registerdaten für den ÜGK-Erstberichterstattungsprozess oder für wissenschaftliche Forschungsvorhaben auf dem durch das Datennutzungskonzept (EDK 2021) vorgesehen Weg derzeit nicht gegeben ist.

3. **Aktuell fehlt eine aussagekräftige und nutzbare Dokumentation zu Registerdaten, um Verknüpfungprojekte effizient zu planen.** Die Datenbeschreibungen des BFS sind für Aussenstehende schwer auffindbar und zum Teil nicht klar nachvollziehbar, sodass direkt mit den entsprechenden Sektionen des BFS-Kontakt aufgenommen werden muss, wenn neue oder andere Daten verknüpft werden wollen.

4. **Der effektive Verknüpfungsprozess dauert zu lang,** um das Ergebnis der Verknüpfungen in das ÜGK-SUF zu integrieren und für den Erstberichterstattungsprozesses zu nutzen.

Vor diesem Hintergrund prüfte das ICER den Ablauf des Verknüpfungsprozesses auf Optimierung und lancierte dafür ein Anschlussprojekt. In Projekt 2 strebte das ICER an, den Verknüpfungsprozess zu beschleunigen und für einen Einsatz im Rahmen des Erstberichterstattungsprozesses nutzbar zu machen.

3.2. Projekt 2: Verknüpfung der Daten aus der ÜGK H4 Präpilotierung

Projekt 1 enthüllte einige Mängel und Herausforderungen im Verknüpfungsprozess, welche die Nutzung von Registerdaten für eine Erstberichterstattung erschweren. Ein zentraler Punkt dabei stellt die Dauer des Verknüpfungsprozesses dar, die es nicht erlaubt, die verknüpften Registerdaten zeitnah in den Datensatz für die Erstberichterstattung zu integrieren und dafür zu nutzen. Deshalb testete das ICER im Rahmen der geplanten ÜGK H4 Präpilotierung, ob die frühzeitige Integration des Datenverknüpfungsprozesses in die Erhebungsplanung und -vorbereitung diesbezüglich Optimierungsmöglichkeiten bietet. So wurde aufgrund der Erkenntnis, dass verschiedene Punkte für eine optimierte Verknüpfung erfüllt sein müssen (Vorliegen der Bewilligung durch den Kosta HarmoS; Vorliegen des Antrags zuhanden des BFS; Vorliegen der Stichprobe mit entsprechenden Merkmalen; Ansprechperson für ÜGK-Verknüpfungen), beschlossen, *den Verknüpfungsprozess frühzeitig – also bereits während der Planung und Vorbereitung der Erhebung – aufzugleisen und denselben Vertrag mit dem BFS als Grundlage zu nehmen*. Nach einem Austausch mit zwischen dem BFS und dem ICER wurde für dieses Projekt von Seiten des BFS von Beginn an eine Ansprechperson definiert, welche für die ÜGK-Verknüpfungen zuständig ist.

In Projekt 2 wurde der positive Beschluss des Kosta HarmoS weit vor dem Start der Datenerhebung im Rahmen der ÜGK H4 Präpilotierung eingeholt. Zudem ging das ICER davon aus, dass der erste Antrag an das BFS als Grundlage der Verknüpfung für Projekt 2 dient, um entsprechend die Datenbasis der Registerdaten der Verknüpfung der ÜGK-Daten 2016 und 2017 zu replizieren ³.

Dies bedeutet eine Umkehrung der zeitlichen Abfolge der relevanten Voraussetzungen: Da der positive Beschluss des Kosta HarmoS weit vor dem Start der Datenerhebung im Rahmen der ÜGK H4 Präpilotierung vorlag, war die erste Voraussetzung für die Datenverknüpfung (Zuordnung der ÜGK-Identifikationsnummern zu AHV-Nummern, welche die PL ÜGK jeweils nach der effektiven Erhebung mit dem teilgenommenen Sample veranlasst) noch nicht gegeben. In der Folge lieferte das ICER dem BFS bereits als die Stichprobe vorbereitet war die identifizierenden Merkmale der SuS in einer

³ Grundsätzlich könnte aus Sicht des ICER zukünftig im Rahmen des ÜGK-Erstberichterstattungsprozesses durchaus eine entsprechende inhaltliche Konsistenz des Vertrags über die Erhebungen hinweg angedacht werden.

vom BFS vorgegebenen Liste. Das ermöglichte dem BFS die Vorbereitung der Verknüpfung auf der Grundlage des geplanten Samples, so dass nach der Erhebung nur noch die Abgleichung auf das effektiv an der Erhebung teilgenommene Sample vorgenommen werden musste. Das führte zu substantziellen Zeitersparnissen. Gemäss Projekt 2 haben 429 SuS an der ÜGK H4 Präpilotierung teilgenommen. Davon konnten 97 % (417 SuS) eindeutig eine AHVN13 zugeordnet werden.

Durch ein Missverständnis zwischen dem ICER und dem BFS umfasste die Datenlieferung des BFS im November 2021 jedoch lediglich Informationen der LABB (Längsschnitts Analyse im Bildungsbe- reich) und nicht, wie vom ICER angedacht die mit dem ersten Vertrag konsistent gehaltenen Daten- quellen und Merkmale wie in Projekt 1. Entsprechend fehlten alle Informationen zu den Eltern, wel- che die Nutzung von Registerdaten erst motivieren. Unsere Ansprechperson beim BFS wies auf Anfrage aus, dass es ihr auf Basis der LABB nicht möglich ist, eine Verknüpfung entsprechend dem ersten Antrag umzusetzen. Dazu benötige sie ebenfalls die Eltern der Kinder, welche in den Regis- terdaten identifiziert werden müssen. Folglich wurde dem ICER erneut die Ansprechperson aus dem ersten Projekt zugewiesen, welche mit STATPOP Daten dies ermöglichen konnte. Entsprechend ist bei Verknüpfungsprojekten, welche beabsichtigen, Informationen zu den Eltern zu erhalten, darauf zu achten, dass eine Ansprechperson gestellt wird, welche Zugriff auf die STATPOP Daten hat.

Mit dieser Ansprechperson wurde vereinbart, dass die Verknüpfung der ÜGK H4 Präpilotierung im Rahmen eines Vertragszusatzes behandelt werden sollte. Ein solches Vorgehen hätte Teile des ad- ministrativen Prozesses (insbesondere die neue Beurteilung des Antrags durch das BFS) obsolet gemacht, da ein Vertragszusatz nur die Zustimmung aller am Vertrag beteiligten Parteien voraus- setzt. Ein entsprechendes Rundschreiben wurde durch die Ansprechperson beim BFS versandt, um das informelle Einverständnis der Beteiligten zum Vorgehen einzuholen. Auf Seiten der EDK wurde dieses Vorgehen jedoch nicht gutgeheissen, was einen neuen Antrag notwendig machte.

Um die Verknüpfung wie geplant umzusetzen, musste daher ein neuer Antrag zuhanden des BFS erstellt werden, der diesem Ende Januar 2022 zur Beurteilung übergeben wurde. Der Beurteilungs- prozess wurde nach vier Monaten abgeschlossen und dem Antrag wurde stattgegeben. Der Vertrag wurde jedoch erst im Mai 2022 in die Unterschriftenrunde zu allen Beteiligten versandt. Nach Ab- schluss der Unterschriftsrunde erhielt das ICER die finalen Daten im Juli 2022.

Bei der Überprüfung der gelieferten Daten zeigte sich, dass nur 391 statt der 417 vom BFS angege- benen SuS verknüpft werden können. 20 der insgesamt 38 Fälle können auf eine fehlende Anwe- senheitsliste zurückgeführt werden. Die restlichen 18 Fälle beziehen sich auf SuS, welche während der Erhebung ein Ersatzlogin und nicht ihr vorgesehenes Login verwendet haben, was eine rückwir- kende Identifikation verunmöglichte.

Dieser Umstand verdeutlicht, dass auch während der Erhebung Vorkehrungen getroffen werden müssen, damit später eine Verknüpfung mit administrativen Daten sichergestellt werden kann.

Zentrale Erkenntnisse aus der Verknüpfung ÜGK H4 Präpilotierung mit Registerdaten (Projekt 2)

Hauptgrund für die Lancierung von Projekt 2 war der Versuch, die zeitliche Dauer des Verknüpfungsprozesses zu beschleunigen, um die verknüpften Registerdaten zeitnah in den Datensatz für die Erstberichterstattung aufzunehmen und dafür zu nutzen. Dafür wurden eine frühzeitige Integration des Datenverknüpfungsprozesses in die Erhebungsplanung und -vorbereitung, die Nutzung eines bereits bestehenden Vertrags mit dem BFS sowie eine frühzeitige Kopplung der Identifikationsmerkmale als Optimierungsmöglichkeiten geprüft. Die daraus gewonnenen Erkenntnisse waren die folgenden:

- 1. Die frühzeitige Integration des Datenverknüpfungsprozesses in die Erhebungsplanung und -vorbereitung beschleunigt den Verknüpfungsprozess**, da die Zuordnung von AHV-Nummern zu den SuS wie auch die Datenlieferung durch das BFS zeitnah erfolgen kann.
2. Projekt 2 betonte zudem die **Bedeutung der vorgängigen Klärung des rechtlichen Rahmens eines Verknüpfungsprojekts**: Der Rückgriff auf einen bereits bestehenden Vertrag mit dem BFS ist im Einzelfall vorgängig zu klären, damit sich die Prozesse nicht unerwartet in die Länge ziehen.
3. Weiter verdeutlichte Projekt 2 zudem das **Optimierungspotenzial bei der Schulung und der Supervision der Testadministrierenden**. Im vorliegenden Fall wurde während den Erhebungen die Vergabe von Ersatzlogins auf der AL zu wenig deutlich dokumentiert, was den Zuordnungsprozess von SuS zu Logins erschwerte und in wenigen Einzelfällen verunmöglichte. Es wird deshalb auf Durchführungsseite eine ungenügende Instruktion bzw. Kontrolle der Testadministrierenden attestiert, weshalb Optimierungen empfohlen werden.

Aufgrund der Erfahrungen in Projekt 2 wurde beschlossen, den Verknüpfungsprozess während der Erhebungsplanung und -vorbereitung im Rahmen eines weiteren Projekts mit Daten der ÜGK H4 Piloterhebung auf zusätzliches Optimierungspotential zu testen.

3.3. Projekt 3: Verknüpfung der Daten aus der ÜGK H4 Piloterhebung

In Projekt 3 wurde darauf geachtet, die bei Projekt 1 und 2 aufgetauchten Probleme zu umgehen, indem frühzeitig die Bewilligung des Kosta HarmoS eingeholt wurde (Mai 2022) und zugleich bei der EDK abgeklärt wurde, ob die Daten aus der ÜGK H4 Piloterhebung im Rahmen eines Vertragszusatzes zur Verknüpfung mit den Daten ÜGK H4 Präpilotierung verknüpft werden dürfen. Dabei wurde festgestellt, dass in diesem Fall ein Vertragszusatz ausreicht, da es sich um eine Erhebung auf derselben Schulstufe (H4) handelt. Mit Unterstützung durch das BFS wurde der bestehende Vertrag angepasst und erfolgreich abgeschlossen (individuelle Abwesenheiten während den Sommer- und den Herbstferien führten dazu, dass dies erst im Oktober möglich war).

Parallel zur Unterschriftenrunde liefert das ICER die Informationen der teilnehmenden SuS an das BFS, welches in der Lage war, mehr als 99% (4880 von 4887) der SuS, die an der ÜGK H4 Piloterhebung teilgenommen haben, zu identifizieren und den Schlüssel 1 zu erstellen. Sobald die Identifikation erfolgt war, konnte das BFS das Zusammentragen der bestellten Daten vorbereiten. Die Verknüpfung mit den Registerdaten erfolgte jedoch erst nach Abschluss der Datenaufbereitungen am ICER Ende November 2022. Der Abschluss der Datenaufbereitungen war in diesem Fall erforderlich, da der ursprüngliche Identifikator der ÜGK durch einen Identifikator des BFS ersetzt wurde, was ein nachträgliches Hinzufügen von nicht im Vertrag mit dem BFS definierten Daten verunmöglicht.

Zentrale Erkenntnisse aus der Verknüpfung ÜGK H4 Piloterhebung mit Registerdaten (Projekt 3)

1. Projekt 3 zeigt, dass eine Verknüpfung zwischen ÜGK- und Registerdaten zeitnah (Juli bis November 2022) erfolgen kann. Eine zentrale Voraussetzung hierbei ist, dass die **Verträge zwischen den beteiligten Institutionen frühzeitig abgeschlossen werden**.
2. Zusätzlich lässt sich Zeit einsparen, indem **die ÜGK-seitige Datenaufbereitung parallel zur Datenzusammenstellung beim BFS vorgenommen** wird.
3. Weiter zeigte sich, dass das **mittlerweile aufeinander abgestimmte Zusammenspiel der Beteiligten am ICER und beim BFS den Prozess begünstigte**.

3.4. Fazit der Dokumentation und Evaluation des Verknüpfungsprozesses mit einem Fokus auf die Nutzung von verknüpften Daten für den Erstberichterstattungsprozess

Die verschiedenen Projekte und die damit gesammelten Erfahrungen des Verknüpfungsprozesses von ÜGK-Daten mit Registerdaten des BFS ermöglichen dem ICER, einen Idealprozess für die Vorbereitungen und Nutzung von Verknüpfungen zu skizzieren, der einen maximalen Zugewinn für die Erstberichterstattung verspricht. Dabei wird deutlich, dass der Verknüpfungsprozess durch eine gute Planung dahingehend optimiert werden kann, dass die verknüpften Daten wenige Monate nach den Erhebungen vorliegen und im Rahmen des Datenaufbereitungsprozesses und des Erstberichterstattungsprozesses genutzt werden können.

Neben dem Erfordernis der zeitnahen Verknüpfung drängt sich die Frage auf, ob und in welcher Form die Datengrundlage des nationalen Erstberichtes nach dessen Publikation im Rahmen eines SUF der Forschungsgemeinschaft zur Verfügung gestellt werden kann, um Replizierbarkeit zu gewährleisten. Während dies in einem folgenden Bericht ausführlicher dargelegt wird, lässt hier bereits festhalten, dass Originaldaten des BFS nie weitergegeben werden dürfen. Werden sie allerdings in Skalierungsmodellen oder für die Imputation verwendet und werden lediglich diese skalierten oder imputierten Variablen veröffentlicht, die keinen Rückschluss auf originale Registerdaten erlauben, ist deren Publikation in einem SUF zulässig und ermöglicht die Reproduzierbarkeit von in der Berichterstattung präsentierten Ergebnissen.

Mittel- und langfristig lässt sich das Aufwands-Ertrags-Verhältnis durch eine Institutionalisierung der Datenverknüpfung optimieren. Insbesondere für eine wiederkehrende Erhebung wie die ÜGK wird dieses Vorgehen als praktikabel eingeschätzt. Dies ist jedoch auch davon abhängig, ob die verknüpften Daten überhaupt eine entsprechend hohe Qualität aufweisen, dass sie im Datenaufbereitungs- und den Erstberichterstattungsprozesses einen Gewinn darstellen. Entsprechend wurde als Forschungsvorhaben im Rahmen von Projekt 1 (Verknüpfung des SUF aus den ÜGK-Erhebungen 2016 und 2017) eine Validierung von Angaben der SuS (Kapitel 4) angestrebt. Dabei stand die Frage im Zentrum, ob Angaben von SuS zum sozialen Kontext fehlerbehaftet sind, was zu systematischen Verzerrungen der geschätzten Effekte führen könnte und dem Argument bei künftigen Erhebungen, andere Datenquellen (Registerdaten oder Elternfragebogen) miteinzubeziehen, um valide(re) Ergebnisse zu erhalten, Vorschub leisten würde. Diese Frage stellte sich insbesondere im Rahmen der Vorbereitung der ÜGK 2024, im Rahmen derer die schulischen Leistungen von SuS auf Stufe H4 analysiert werden.

4. Validierung von Konstrukten der ÜGK durch Registerdaten

Nach dem in den vorangegangenen Kapiteln die einzelnen Verknüpfungsprojekte beschrieben wurden, werden in diesem Kapitel erste Resultate der „Validierung von Konstrukten der ÜGK durch Registerdaten“ präsentiert. Des Weiteren liefert dieses Kapitel erste Erkenntnisse zur Abdeckung bestimmter Merkmale – etwa der Bildung der Eltern – in den Registerdaten und damit zum Nutzen der Verknüpfung im Hinblick auf die Umsetzung der „once-only“-Strategie sowie auf potenzielle Kosteneinsparungen für die ÜGK.

Das vorliegende Forschungsprojekt (Ackermann, 2023) untersucht Messfehler sowie fehlende Angaben bei den Antworten der SuS. Hierfür werden die Variablen zum Migrationsstatus, der elterlichen Bildung und des elterlichen Berufs aus den mit Registerdaten verknüpften Daten des SUF der ÜGK-Erhebungen 2016 und 2017 untersucht. Diese Variablen wurden ausgewählt, weil sie für die ÜGK (Pham et al., 2016, 2017) und andere Large-Scale Assessments äusserst relevante Kontextualisierungsmerkmale darstellen (OECD, 2019). Leider ist der Vergleich zwischen Registerdaten und Aussagen der SuS zur Anzahl Bücher im Elternhaushalt und der Zuhause gesprochenen Sprache nicht möglich, da dazu keine Registerdaten vorliegen. Mit der Variable zur Anzahl der Bücher hätten alle Variablen, die zur Bildung einer-Variable zum sozio-ökonomischen Status (SES) verwendet werden, verglichen werden können.

Die präsentierten Resultate beziehen sich alle auf eigene Berechnungen mit den ÜGK Daten (Nidegger, 2019, 2021) sowie Daten des Bundesamts für Statistik (BFS, 2021 a-m).

4.1. Messfehler und fehlende Angaben bei Antworten von Schülerinnen und Schülern

Fehlende Angaben rühren daher, dass nicht alle SuS alle Fragen bei Leistungstests oder in einem Fragebogen beantworten. Einige werden unabsichtlich übersprungen oder bewusst nicht beantwortet. Fehlende Angaben sind für LSA wie die ÜGK problematisch, da auf der Individualebene zentrale Informationen fehlen, die zur inhaltlichen Kontextualisierung oder auch zur Erstellung von Plausible-Values herbeigezogen werden. Zudem schränken fehlende Angaben das Sample ein oder führen zu Verzerrungen, wenn die fehlenden Antworten nicht zufällig zustande kommen.

Die inhaltliche Kontextualisierung und die Plausible-Values sind jedoch ebenfalls von Messfehlern betroffen. Konkret bedeutet das, dass die Angaben von SuS fehlerhaft sein können, da die SuS willentlich falsch antworten oder die korrekte Antwort nicht wissen. In diesem Kontext bedeutet Messfehler also, dass die SuS keine validen Aussagen zu den gestellten Fragen machen. Es wird vermutet, dass dies insbesondere dann häufig vorkommt, wenn die Fragen nicht die SuS selbst, sondern deren Eltern betreffen, wie das im Fall der hier im Zentrum stehenden Angaben (Bildung und Beruf

der Eltern) zutrifft. Während zufällig verteilte Messfehler zu einer generellen Unterschätzung tatsächlicher Zusammenhänge führen, können systematische Messfehler die Zusammenhänge in Analysen verzerren oder sogar umdrehen.

Eine zentrale Erkenntnis des Forschungsprojekts ist, dass eine gute schulische Leistung die Wahrscheinlichkeit von Messfehlern und fehlenden Angaben der SuS reduziert (Ackermann, 2023). Dieser Befund bestätigt die Resultate aus früheren Studien (Ensminger et al., 2000; Lien et al., 2001; Kreuter et al., 2010; Ridolfo & Maitland, 2011; Jerrim & Micklewright, 2014; Engzell & Jonsson, 2015; Wittrock et al., 2017; Hovestadt & Schneider, 2021). Zum einen bedeutet dies, dass Aussagen leistungsstärkerer SuS besser mit den Angaben in den Registerdaten übereinstimmen. Zum anderen ist davon auszugehen, dass leistungsschwächere SuS vermehrt fehlende Angaben machen. Es kann also nicht ausgeschlossen werden, dass Analysen, welche sich auf ein Sample beschränken für das vollständige Information auf allen Variablen vorliegen, sich auf ein verzerrtes Sample beziehen, oder dass z.B. der Effekt der elterlichen Bildung auf die schulische Leistung (im Vergleich zu Registerdaten) und damit auch auf das Erreichen der Grundkompetenzen in der Erstberichtserstattung unterschätzt wird.

Neben dem Zusammenhang mit der schulischen Leistung untersucht das Forschungsprojekt weitere Einflussfaktoren auf Messfehler und das Auftreten fehlende Angaben. Einer davon ist das Alter der SuS zum Befragungszeitpunkt: Jüngere Kindern machen signifikant häufiger fehlerbehaftete oder fehlende Angaben (Ackermann, 2023; siehe u.a. auch Ensminger et al., 2000; Kreuter et al., 2010; Ridolfo & Maitland, 2011; Wittrock et al., 2017). Ein weiterer Einflussfaktor betrifft die gestellten Fragen an sich: Die Resultate zeigen, dass während die Antworten zum Migrationsstatus problemlos gegeben werden können, die Fragen zur elterlichen Bildung bereits viel schwieriger zu beantworten sind für SuS. Die meisten Messfehler und fehlenden Antworten bezogen sich auf die Frage zum elterlichen Beruf.

Ausserdem gibt das Forschungsprojekt Hinweise darauf, dass auch die Lebensumstände der Kinder eine Rolle spielen. So steigt die Wahrscheinlichkeit für fehlende Angaben und Messfehler, wenn der Elternteil, auf den sich die Frage bezieht, nicht im selben Haushalt wie das Kind lebt (Ackermann 2023; siehe u.a. auch Ensminger et al., 2000; West et al., 2001). Gleichzeitig sind Messfehler weniger häufig, wenn das Kind dasselbe Geschlecht wie das Elternteil hat, auf welches sich eine Frage bezieht. Das Auftreten von fehlenden Antworten ist jedoch nicht von einem solchen Rollenmodell-Effekt betroffen (Ackermann, 2023).

Die Resultate offenbaren weiter, dass bei SuS mit Migrationshintergrund und SuS aus Haushalten, in denen die Testsprache nicht gesprochen wird, häufiger fehlende Werte auftreten. Messfehler treten jedoch nur häufiger auf, wenn die Testsprache im Haushalt nicht gesprochen wird (Ackermann, 2023; ähnliche Resultate bei Nordahl et al., 2011). Hinsichtlich des Geschlechts findet das Forschungsprojekt keinen Unterschied hinsichtlich der Messfehler zwischen Jungen und Mädchen. Allerdings ist die Wahrscheinlichkeit für fehlende Antworten bei Jungen höher (Ackermann, 2023; in der Literatur finden sich unterschiedliche Ergebnisse zum Einfluss von Geschlecht: z.B. (Lien et al., 2001; West et al., 2001).

In einem zusätzlichen Schritt wurde der Einfluss von Merkmalen der Eltern der SuS auf Messfehler und fehlende Antworten untersucht. Während die höchste abgeschlossene Bildung der Eltern zu weniger fehlenden Angaben führt, ist dieser Effekt bei Messfehlern nicht eindeutig. Wiederum besitzt der höchste soziale Status des Berufs der Eltern keinen Einfluss auf fehlende Antworten jedoch korreliert dieser mit weniger Messfehlern. Die Resultate in diesem letzten Schritt sind allerdings nur bedingt aussagekräftig, da die beiden Variablen durch viele fehlende Werte das Sample deutlich einschränken (Ackermann, 2023).

Was für Folgerungen können aus den Resultaten für die ÜGK abgeleitet werden? Einerseits ist die Korrelation zwischen schulischer Leistung und fehlenden Angaben sowie Messfehlern für die inhaltliche Kontextualisierung problematisch, da offenbar eine Systematik besteht. Wird dem nicht Rechnung getragen, reproduzieren Imputationen andererseits die verzerrten Zusammenhänge bei Kindern mit fehlenden Angaben (Keogh & Bartlett, 2019). Das bedeutet, dass sich verzerrte Zusammenhänge bereits bei der Skalierung der Items, der Ziehung der Plausibel Values und schliesslich bei der Schwellenwertsetzung einschleichen.

Zudem offenbart sich, dass Angaben, die für die Bildung eines sozioökonomischen Status (SES) wichtig sind (wie die elterliche Bildung oder der Beruf der Eltern), besonders bei jungen SuS fehlerhaft oder gar fehlend sind. Folglich ist für eine Befragung von jüngeren Kindern zu überdenken, wie entsprechende Informationen erfasst werden können, um zu garantieren, dass diese valide sind. Ebenfalls ist denkbar, mittels einer Validationsstudie zu versuchen, den Messfehler zu korrigieren oder zumindest quantifizieren zu können (Keogh und Bartlett 2019).

4.2. Abdeckung der Registerdaten

Die Nutzung der Registerdaten als Ersatz zu der Befragung der SuS stellt eine Möglichkeit dar, valide Informationen⁴ zu erhalten. Das vorliegende Forschungsprojekt zeigt allerdings, dass die geringe Abdeckung zentraler Merkmale in Registerdaten diese Möglichkeit enorm einschränkt (siehe Abbildung 2: Abdeckung nach Datenquelle, Jahr und Familienmitglied).

Zu beachten ist, dass sich die verwendeten Merkmale (Migrationsstatus, elterliche Bildung und Beruf der Eltern) auf die Eltern der Kinder beziehen. Entsprechend müssen diese in den Registerdaten identifiziert werden können. Für die Samples aus dem Jahr 2016 und dem Jahr 2017 kann für jeweils mehr als 10% der Kinder kein Elternteil in den Registerdaten identifiziert werden (Vergleich Abdeckung „Kind STATPOP“ 97.39 Prozent bzw. 97.09 Prozent zu Abdeckung „Eltern STATPOP“ 82.23 Prozent bzw. 85.95 Prozent). Selbst wenn für diese Eltern potenziell Informationen vorhanden wären, können diese nicht zum Kind verlinkt werden (Daten: Nidegger, 2019, 2021; BFS 2021 a-m).

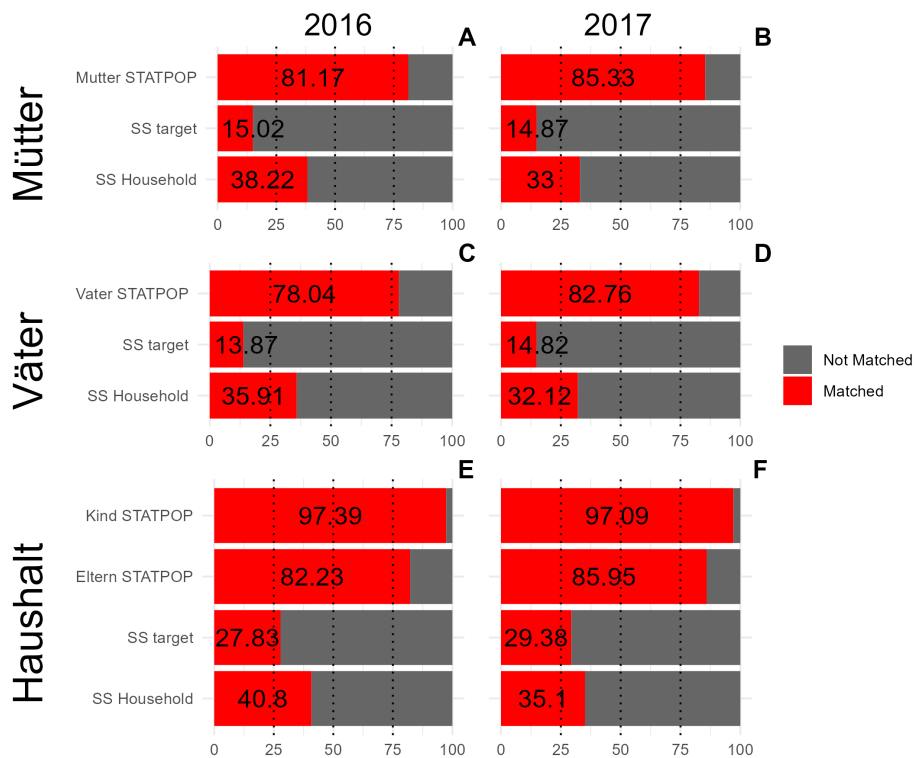


Abbildung 2: Abdeckung nach Datenquelle, Jahr und Familienmitglied

⁴ Es kann davon ausgegangen werden, dass Registerdaten durchaus eine höhere Validität besitzen, da die Angaben durch die Eltern selbst getätigt wurden.

Zudem sind die für die Bildung des SES wichtigen Merkmale zur elterlichen Bildung und dem elterlichen Beruf (OECD, 2019; Pham et al., 2016, 2017) unzureichend abgedeckt. Damit könnte eines der wichtigsten Merkmale zur sozialen Kontextualisierung der schulischen Leistung, der SES, nur bedingt gebildet werden. Dies liegt daran, dass die entsprechenden Merkmale bloss im Rahmen der Strukturhebung erhoben werden. Die Strukturhebung ist eine jährliche Befragung einer Stichprobe der erwachsenen Bevölkerung in der Schweiz. Daher kann für viele der Kinder, respektive deren Eltern, keine Information zu diesen Merkmalen aus den Registerdaten gewonnen werden. Im Fall dieses Forschungsprojekts wurden jeweils fünf Jahre der Strukturhebung zusammengefasst, um die Abdeckung und damit die Chance auf gültige Informationen zu erhöhen. Gleichwohl ist der Anteil an Eltern, die an der Strukturhebung teilgenommen und damit Informationen zu den entsprechenden Merkmalen haben, gering.

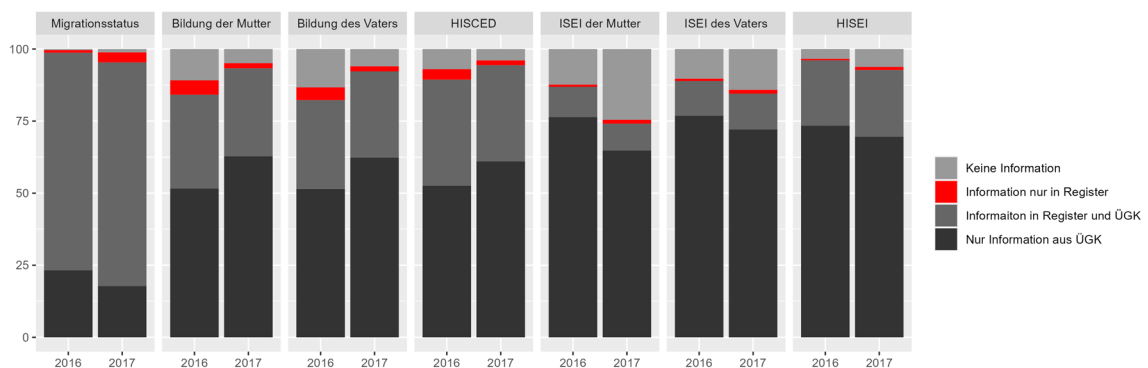


Abbildung 3 Abdeckung in den Daten nach Variable und Jahr

Abbildung 3 stellt für die ausgewählten Variablen dar, wo eine Substitution der ÜGK-Information durch Registerdaten überhaupt möglich ist („Information nur in Register“ plus „Information in Register und ÜGK“). Rot dargestellt ist der Anteil (zwischen 0.5% und 4.9%) an Informationsgewinn durch Registerdaten, in Fällen, welche fehlende Werte bei der jeweiligen Frage im SFB aufweisen (Daten: Nidegger, 2019, 2021; BFS 2021 a-m).

Künftige Registerdaten besitzen allerdings das Potential, diese Einschränkungen zu reduzieren. So kann die höchste elterliche Bildung künftig aus den LABB-Daten aggregiert werden. Zudem könnte es in Zukunft auch möglich sein, Angaben zum Beruf der Eltern aus Steuerdaten zu gewinnen. Wie oben angedeutet, können Registerdaten allerdings bereits heute im Sinne einer Validationsstudie verwendet werden, um Messfehler zumindest zu quantifizieren und im Optimalfall dazu verwendet werden, Messfehler zu „korrigieren“. Die Untersuchung im Rahmen des ersten Verknüpfungsprojekts

verdeutlicht jedoch, dass derzeit die geringe Abdeckung der Registerdaten eine ersatzlose Streichung eines SFB nicht erlaubt.

5. Zusammenfassung und Bewertung des Nutzens einer Verknüpfung für die ÜGK

In diesem Kapitel wird auf der Grundlage der Erfahrungen im Rahmen der Verknüpfungsprojekte ein idealtypischer Verknüpfungsprozess empfohlen sowie eine Empfehlung, den Pseudonymisierungsprozess des ÜGK-SUF retrospektiv und prospektiv zu überprüfen, ausgesprochen. Weiter werden die ersten Resultate aus der Verknüpfung des SUF der ÜGK-Erhebungen 2016 und 2017 mit Registerdaten zusammenfassend festgehalten, wobei das Augenmerk auf den Nutzen einer Verknüpfung für die ÜGK gelegt wird. Die Erkenntnisse münden in einer Empfehlung an die PL ÜGK.

Empfehlung: Idealtypischer Verknüpfungsprozess für eine Erstberichterstattung

Die Ziele der realisierten Verknüpfungsprojekte waren, den Prozess kennenzulernen und zu prüfen, ob sich Verknüpfungen für künftige Erhebungen bewähren und allenfalls für weitere Prozesse der ÜGK (wie z.B. den Erstberichterstattungsprozess, Skalierung der Items) genutzt werden können. Die drei Projekte zeigten, dass die administrativen Prozesse für die Nutzung von Registerdaten in Rahmen der ÜGK für eine Erstberichterstattung eine grosse Hürde darstellen. Für kommende Projekte wird aufgrund der gemachten Erfahrungen folgender idealtypische Verknüpfungsprozess empfohlen, der zu keinen vermeidbaren Verzögerungen in der Datenaufbereitung für die Erstberichterstattung führen sollte:

1. Antrag und Bewilligung durch Kosta HarmoS
 2. Antrag und Bewilligung durch BFS
- ERHEBUNG**
3. Lieferung der identifizierenden Merkmale der Stichprobe an das BFS
 4. Zuordnung ÜGK-Identifikationsnummer zur AHV-Nummer durch das BFS (Schlüssel 1)
 5. Datenzusammenstellung durch das BFS/Datenaufbereitung durch das ICER
 6. Datenlieferung durch das ICER
 7. Verknüpfung der Daten

Dabei ist es besonders wichtig, Schritt 1 und 2 weit vor (ca. 6 Monate) einer ÜGK-Erhebung abgeschlossen zu haben. In der Folge sind die weiteren Schritte zeitnah zur Erhebung umsetzbar. Im Anschluss an die Erhebung können zudem die Datenaufbereitung und die Datenzusammenstellung parallel geschehen (Punkt 5). Zu betonen ist auch, dass ab Schritt 2 eine Kontaktperson vom BFS gestellt werden sollte. Dabei hat sich bewährt, mit bereits bekannten Personen zusammenzuarbeiten, die bereits mit den Grundzügen des Projekts vertraut sind.

Eine solche Herangehensweise stellt sicher, dass eine Verknüpfung mit Registerdaten zu keiner Verzögerung bei der Erstberichterstattung führt und dass die Informationen aus den Registern bereits für die Datenaufbereitung bereitstehen. Auch hat eine Abklärung beim BFS gezeigt, dass Registerdaten in aggregierter Form in einem SUF veröffentlicht werden dürfen, beispielsweise wenn diese in einem Imputationsmodell verwendet wurden. Originaldaten dürfen allerdings nie veröffentlicht werden. Wird dies in der Erstberichterstattung berücksichtigt, bleibt auch die Reproduzierbarkeit gewährleistet.

Empfehlung: Klärung der Pseudonymisierung des ÜGK-SUF vor deren Veröffentlichung

Als äusserst vordringlich stellt sich die Klärung der Pseudonymisierung des ÜGK-SUF vor deren Veröffentlichung dar. Im „Konzept Datennutzung im Rahmen der Überprüfung des Erreichens der Grundkompetenzen (ÜGK)“ (EDK 2021) wird Folgendes bestimmt:

„Zum Abschluss der Anonymisierung wird der bearbeitete Datensatz an die datensichernde Institution übermittelt. Diese ersetzt darin die bisher verwendete ÜGK-Identifikatorvariable durch eine Pseudo-ID. Sie verwendet dazu ein in Zusammenarbeit mit dem wissenschaftlichen Konsortium entwickeltes Verfahren und stellt sicher, dass die ursprüngliche ID nur mit dem bei ihr aufbewahrten Schlüssel gelesen und nicht anderweitig rekonstruiert werden kann. Bei der Erstellung des Schlüssels wird darauf geachtet, dass eine Verknüpfung mit dem Personenidentifikator AHVN13 des BFS möglich bleibt (siehe 5.2.2).“

Die Erfahrung aus den Projekten und mit Blick auf die zukünftige Nutzung von Registerdaten wird empfohlen, diesen Prozess auf Übereinstimmung mit der aktuellen Situation bei der Verknüpfung der ÜGK-Daten aus 2016 und 2017 zu prüfen. Es zeigten sich diesbezüglich diverse Probleme aufseiten der datensichernden Institution: Aus den Erfahrungen mit Projekt 1 scheint die datensichernde Institution aktuell über keinen korrekten Schlüssel 2 zu verfügen, was zur Folge hat, dass die Verknüpfung über den Schlüssel 1 zu der AHVN13 nicht genutzt werden kann. Wobei der Umstand, dass keine der involvierten Stellen allein eine Verknüpfung vornehmen kann, als sinnvoll erachtet wird.

Die Klärung, was beim Prozess betreffend der Daten aus 2016 und 2017 zu einem inkorrekten Schlüssel 2 geführt hat, stellt auch im Hinblick auf zukünftigen SUF und deren Sicherung und Bereitstellung als äusserst relevant dar, und es lassen sich handlungsleitende Anweisungen für alle beteiligten Institutionen ableiten.

Empfehlung: Kein Ersatz des SFB durch Registerdaten

Das erste Forschungsprojekt zeigt auf, dass die Registerdaten sehr wohl das Potenzial besitzen, um Konstrukte aus der Befragung zu validieren. Eine solche Validierung bringt wichtige Erkenntnisse über die Qualität der Angaben der Kinder und deren Auswirkungen auf die Ergebnisse von Analysen mit diesen Daten.

Weiter lieferte dieses Forschungsprojekt Erkenntnisse zu möglichen Kosteneinsparungen bei der ÜGK durch den Einsatz von Verknüpfungen mit Registerdaten sowie zur Verwendung von Verknüpfungen zur Einlösung der „once-only“-Strategie des Bundes. Dazu muss gesagt werden, dass dieses Potenzial zurzeit stark limitiert ist durch die mangelhafte Abdeckung essenzieller Merkmale der für die ÜGK relevanten Population in den Registerdaten. Einerseits können bei einem substanziellen Teil der Kinder keine Eltern identifiziert werden, andererseits ist der Informationsgehalt selbst bei einer Identifikation oftmals gering, da zentrale Merkmale (z.B. die elterliche Bildung oder der Beruf der Eltern) nicht flächendeckend erhoben werden. Folglich birgt eine Verknüpfung aus der Perspektive der „once-only“ Strategie des Bundes und mit Sicht auf Kosteneinsparungen für die ÜGK (noch) wenig Potenzial. Die ÜGK bleibt weiterhin auf die Angaben der Kinder oder der Eltern angewiesen.

Zusammenfassung der Empfehlung an die PL ÜGK

Insgesamt verspricht die Verknüpfung von ÜGK-Daten mit Registerdaten einige Vorteile (u.a. die Validierung von Konstrukten). Die verknüpften Daten bergen Potenzial, das bei der Datenaufbereitung⁵ unterstützend sein kann. Darüber hinaus bieten verknüpfbare Datensätze die Möglichkeit für vielfältige Forschungsvorhaben mit den ÜGK-Daten. Daher wird der Nutzen der Verknüpfung mit Registerdaten grundsätzlich als positiv beurteilt und es wird empfohlen, diese weiterhin anzustreben

⁵ Wie dies bei Keogh und Bartlett (2019) vorgeschlagen wird, könnten Registerdaten gemeinsam mit den Angaben der SuS imputiert werden. Dadurch kann der Messfehler potenziell verringert werden, würde allerdings die Datenaufbereitung sowie die Datenanalyse komplexer machen.

und deren Vorteile zu nutzen. Dabei sollte entlang des vorgeschlagenen idealtypischen Verknüpfungsprozesses vorgegangen werden. Ebenso sollte retrospektiv die Pseudonymisierung der Daten aus 2016 und 2017 überprüft werden um diesen – essenziellen – Prozess auf der Handlungsebene zu standardisieren.

Zudem wird empfohlen, zu einem späteren Zeitpunkt noch einmal das Potenzial von Verknüpfungen mit Sicht auf Kosteneinsparungen für die ÜGK und aus der Perspektive der „once-only“ Strategie des Bundes zu untersuchen.

6. Fazit und Ausblick

Mit diesem Bericht können drei der fünf eingangs aufgezählten Punkte beantwortet werden:

- Dokumentation und Evaluation des Verknüpfungsprozesses als Information für die Projektleitung ÜGK (PL ÜGK): Mit der Projektbeschreibungen der einzelnen Verknüpfungsprojekte (Kapitel 1) sowie der Zusammenfassung der zentralsten Erkenntnisse (Kapitel 5) wurde ein Idealprozess für Verknüpfungen entwickelt, der den Nutzen von Registerdaten in Bezug auf die Erstberichterstattung gewährleisten kann.
- Nutzung von Registerdaten für die Validierung von Informationen aus dem Schülerfragebogen, die zur Generierung von projekt- und forschungsrelevanten Konstrukten (wie z.B. dem sozioökonomischen Status) genutzt werden: Das Forschungsprojekt zeigt auf, dass Registerdaten durchaus genutzt werden können, um Kinderaussagen zu validieren. Die Resultate des Forschungsprojekts bilden wichtige Erkenntnisse für künftige ÜGK-Erhebungen. Unter anderem zeigen sie, dass Kinderangaben bei relevanten Fragen zu den Eltern fehlerbehaftet sind und dass dies noch deutlicher ausgeprägt ist, je jünger die SuS sind.
- Beitrag zur „once-only“-Strategie des Bundes und Prüfung potenzieller Kosteneinsparungen für die ÜGK: Das Forschungsprojekt zeigt, dass die Abdeckung und damit der Informationsgehalt von Registerdaten generell und insbesondere bei SES-relevanten Merkmalen zu gering ist, als das eine Verknüpfung mit Registerdaten einen SFB oder einen EFB ersetzen könnte. Eine Verknüpfung kann aktuell demnach noch keinen Beitrag zur potenziellen Kosteneinsparungen für die ÜGK leisten und leistet demnach für die ÜGK noch keinen Beitrag im Sinne der „once-only“-Strategie des Bundes.

Offen bleibt in diesem ersten Bericht, inwiefern sich Registerdaten hinsichtlich der Datenaufbereitung und –verwertung sowie zur erweiterten Kontextualisierung der Erhebungsdaten (zusätzliche Informationen, Informationen aus anderen Quellen) verwenden lassen. Dazu wird in einem folgenden Bericht mehr Auskunft geben, der darlegen wird, inwiefern Registerdaten als zusätzliche Informationen zum Schülerfragebogen genutzt werden, um in Imputationsmodellen den Messfehler zu verringern. Durch den Elternfragebogen, der bei der ÜGK H4 Piloterhebung eingesetzt wurde, können die Ergebnisse mit einer aus naheliegenden Gründen als plausibel eingeschätzten Datenquelle zur sozialen Herkunft von SuS abgeglichen werden. Weiter wird dieser Folgebericht eine Empfehlung zur Datenaufbewahrung respektive zur Datennutzung für externe Forschende im Sinn der Reproduzierbarkeit beinhalten. Dabei wird beispielhaft dargestellt, welche Variablen in welcher Form in ein SUF einfließen können. Damit wird ein zusätzlicher Beitrag geleistet, den Nutzen einer Verknüpfung zwischen Register- und ÜGK-Daten aufzuzeigen.

7. Literatur

- Ackermann, Tobias (2023). Asking Students about their Parents: How Item Non-response and Measurement Error Depend on Construct Salience and Students' Cognitive Abilities. (Workingpaper – Zugang auf Anfrage)
- EDK (2021). Konzept Datennutzung im Rahmen der Überprüfung des Erreichens der Grundkompetenzen (ÜGK). Bern 07.05.2021
- Engzell, P., & Jonsson, J. O. (2015). Estimating Social and Ethnic Inequality in School Surveys: Biases from Child Misreporting and Parent Nonresponse. *European Sociological Review*, 31(3), 312–325. <https://doi.org/10.1093/esr/jcv005>
- Ensminger, M. E., Forrest, C. B., Riley, A. W., Kang, M., Green, B. F., Starfield, B., & Ryan, S. A. (2000). The Validity of Measures of Socioeconomic Status of Adolescents. *Journal of Adolescent Research*, 15(3), 392–419. <https://doi.org/10.1177/0743558400153005>
- Hovestadt, T., & Schneider, T. (2021). Liefern Jugendliche valide Informationen zum Bildungsstand ihrer Eltern in standardisierten Erhebungen? Befunde zu Schülerinnen und Schülern der 9. Jahrgangsstufe in Deutschland. *Zeitschrift für Erziehungswissenschaft*, 24(3), 715–742. <https://doi.org/10.1007/s11618-021-01016-5>
- Jerrim, J., & Micklewright, J. (2014). Socio-economic Gradients in Children's Cognitive Skills: Are Cross-Country Comparisons Robust to Who Reports Family Background? *European Sociological Review*, 30(6), 766–781. <https://doi.org/10.1093/esr/jcu072>
- Keogh, R. H., & Bartlett, J. W. (2019). Measurement error as a missing data problem. *arXiv*. <https://doi.org/10.48550/arXiv.1910.06443>
- Kreuter, F., Eckman, S., Maaz, K., & Watermann, R. (2010). Children's Reports of Parents' Education Level: Does it Matter Whom You Ask and What You Ask About? *Survey Research Methods*, 4(3), 127–138.

- Lien, N., Friestad, C., & Klepp, K.-I. (2001). Adolescents' proxy reports of parents' socioeconomic status: How valid are they? *Journal of Epidemiology & Community Health*, 55(10), 731–737. <https://doi.org/10.1136/jech.55.10.731>
- Nordahl, H., Krølner, R., Páll, G., Currie, C., & Andersen, A. (2011). Measurement of Ethnic Background in Cross-national School Surveys: Agreement Between Students' and Parents' Responses. *Journal of Adolescent Health*, 49(3), 272–277. <https://doi.org/10.1016/j.jadohealth.2010.12.013>
- OECD. (2019). *PISA 2018 Results (Volume II): Where All Students Can Succeed*. OECD. <https://doi.org/10.1787/b5fd1b8f-en>
- Pham, G., Helbling, L., Verner, M., & Ambrosetti, A. (2017). *ÜGK – COFO – VeCoF 2017 results: Technical appendices*. Pädagogische Hochschule St.Gallen (PHSG) & Service de la recherche en éducation (SRED).
- Pham, G., Helbling, L., Verner, M., Petrucci, F., Angelone, D., & Ambrosetti, A. (2016). *ÜGK – COFO – VECOF 2016 results: Technical appendices*. Pädagogische Hochschule St.Gallen (PHSG) & Service de la recherche en éducation (SRED)
- Ridolfo, H., & Maitland, A. (2011). Factors that influence the accuracy of adolescent proxy reporting of parental characteristics: A research note. *Journal of Adolescence*, 34(1), 95–103. <https://doi.org/10.1016/j.adolescence.2010.01.008>
- West, P., Sweeting, H., & Speed, E. (2001). We Really Do Know What You Do: A Comparison of Reports from 11 Year Olds and their Parents in Respect of Parental Economic Activity and Occupation. *Sociology*, 35(2), 539–559. <https://doi.org/10.1177/S0038038501000268>
- Wittrock, J., Kimmel, L., Hunscher, B., & Le, K. T. (2017). Proxy reporting in education surveys: Factors influencing accurate reporting in the 2012 Qatar Education Study. *International Journal of Social Research Methodology*, 20(6), 737–748. <https://doi.org/10.1080/13645579.2017.1301078>

8. Datensätze

Nidegger, Christian (2019). ÜGK / COFO / VECOF 2016: Competencies of Swiss pupils in mathematics (1.0.0) [Dataset]. FORS data service. <https://doi.org/10.23662/FORS-DS-1004-1>

Nidegger, Christian (2021). ÜGK / COFO / VECOF 2017: Competencies of Swiss pupils in languages (1.0.1) [Dataset]. FORS data service. <https://doi.org/10.23662/FORS-DS-876-2>

Bundesamt für Statistik (2021a). STATPOP – Stand: 31.12.2011.

Bundesamt für Statistik (2021b). STATPOP - Stand: 31.12.2012.

Bundesamt für Statistik (2021c). STATPOP - Stand: 31.12.2013.

Bundesamt für Statistik (2021d). STATPOP - Stand: 31.12.2014.

Bundesamt für Statistik (2021e). STATPOP - Stand: 31.12.2015.

Bundesamt für Statistik (2021f). STATPOP - Stand: 31.12.2016.

Bundesamt für Statistik (2021g). STATPOP - Stand: 31.12.2017.

Bundesamt für Statistik (2021h). Strukturhebung- Stand: 31.12.2012.

Bundesamt für Statistik (2021i). Strukturhebung- Stand: 31.12.2013.

Bundesamt für Statistik (2021j). Strukturhebung- Stand: 31.12.2014.

Bundesamt für Statistik (2021k). Strukturhebung- Stand: 31.12.2015.

Bundesamt für Statistik (2021l). Strukturhebung- Stand: 31.12.2016.

Bundesamt für Statistik (2021m). Strukturhebung- Stand: 31.12.2017.