

Bellmann, Lisa et al.

Research Report

Machbarkeitsstudie Weiterentwicklung Betriebsbefragungen:
Teilprojekt des BMAS-Projekts zur "Förderung innovativer Ansätze zur
Stärkung von Dateninfrastruktur und Methoden"

IAB-Forschungsbericht, No. 2/2025

Provided in Cooperation with:

Institute for Employment Research (IAB)

Suggested Citation: Bellmann, Lisa et al. (2025) : Machbarkeitsstudie Weiterentwicklung Betriebsbefragungen: Teilprojekt des BMAS-Projekts zur "Förderung innovativer Ansätze zur Stärkung von Dateninfrastruktur und Methoden", IAB-Forschungsbericht, No. 2/2025, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg, <https://doi.org/10.48720/IAB.FB.2502>

This Version is available at:

<https://hdl.handle.net/10419/313031>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-sa/4.0/deed.de>



INSTITUT FÜR ARBEITSMARKT- UND
BERUFSFORSCHUNG
Die Forschungseinrichtung der Bundesagentur für Arbeit

IAB-FORSCHUNGSBERICHT

Aktuelle Ergebnisse aus der Projektarbeit des Instituts für Arbeitsmarkt- und Berufsforschung

2|2025 Machbarkeitsstudie Weiterentwicklung Betriebsbefragungen

Teilprojekt des BMAS-Projekts zur „Förderung innovativer Ansätze zur Stärkung von Dateninfrastruktur und Methoden“

Lisa Bellmann, Nicole Gürtzgen, Sophie Hensgen, Susanne Kohaut, Alexander Kubis, Michael Oberfichtner,
André Pirralha



Finanziert von der
Europäischen Union
NextGenerationEU

ISSN 2195-2655

Machbarkeitsstudie Weiterentwicklung Betriebsbefragungen

Lisa Bellmann (IAB Nürnberg)

Nicole Gürtzgen (IAB Nürnberg, Universität Regensburg),

Sophie Hensgen (IAB Nürnberg),

Susanne Kohaut(IAB Nürnberg),

Alexander Kubis (IAB Nürnberg),

Michael Oberfichtner (IAB Nürnberg, Friedrich-Alexander-Universität Erlangen-Nürnberg),

André Pirralha (IAB Nürnberg)

In aller Kürze

- Betriebsbefragungen sind eine wertvolle Datengrundlage für Forschung, Politikberatung und die amtliche Statistik, wobei sie zugleich eine Reihe unterschiedlicher Anforderungen erfüllen sollen. Um beiden Zielen gerecht zu werden, führt das IAB seit über 30 Jahren mit dem IAB-Betriebspanel und der IAB-Stellenerhebung regelmäßig zwei große Betriebsbefragungen durch.
- Im Teilprojekt „Machbarkeitsstudie zur Weiterentwicklung der Betriebsbefragungen am IAB“ des vom BMAS finanzierten Projekts „Förderung innovativer Ansätze zur Stärkung von Dateninfrastruktur und Methoden“ wird einerseits die Möglichkeit untersucht, die bestehenden Befragungen um unterjährige Elemente zu erweitern. Andererseits soll geprüft werden, inwieweit sich beide Befragungen zu einer einzigen zusammenführen lassen.
- Dieser Bericht stellt erste Ergebnisse dar, liefert jedoch noch keine abschließenden Befunde, da das Projekt noch fortgeführt wird.
- Erste Ergebnisse weisen darauf hin, dass beide bestehenden Befragungen um hochfrequente Elemente erweitert werden können. Allerdings stehen in beiden Befragungen die finalen Ergebnisse zu den Effekten einer solchen Erweiterung auf den erwarteten Rücklauf noch aus.
- Mit Blick auf die Integration der beiden Befragungen wurde zunächst theoretisch die Machbarkeit eines rollierenden Panels untersucht. Hierbei wird eine jährliche Panelstichprobe von Betrieben in vier Substichproben aufgeteilt, welche quartalsweise als Panelerhebung befragt werden. Grundsätzlich scheint eine Integration der beiden Befragungen auf diesem Weg möglich, im Vergleich zum bisherigen IAB-Betriebspanel würde ein rollierendes Panel jedoch substantielle zusätzliche Sach- und Personalmittel erfordern. Eine Anknüpfung an die bestehenden Befragungen dürfte allerdings nicht ohne Brüche in den Zeitreihen realisierbar sein.

Inhalt

In aller Kürze	3
Inhalt.....	4
Zusammenfassung	6
Summary.....	8
Danksagung.....	9
1 Einleitung	10
2 Status Quo der Betriebsbefragungen am IAB	11
2.1 IAB-Betriebspanel	11
2.1.1 Stichprobe	12
2.1.2 Datenerhebung	14
2.1.3 Fragebogen und Inhalte der Befragung	14
2.1.4 Hochrechnung.....	15
2.1.5 Datenqualität	15
2.1.6 Datenprodukte auf Basis des IAB-Betriebspanels	16
2.2 IAB-Stellenerhebung.....	16
2.2.1 Stichprobe	17
2.2.2 Datenerhebung	18
2.2.3 Fragebogen und Inhalte der Befragung	18
2.2.4 Hochrechnung.....	20
2.2.5 Datenqualität	20
2.2.6 Datenlieferung an Eurostat und die IAB-Arbeitszeitrechnung	24
3 Implementierung hochfrequenter Elemente.....	26
3.1 IAB-Betriebspanel	26
3.1.1 Aufbau des Experiments	27
3.1.2 Fragebogen.....	29
3.1.3 Stichprobe und Fallzahlen.....	30
3.1.4 Teilnahmebereitschaft an der unterjährigen Befragung und Selektivität.....	32
3.1.5 Auswertungen auf Basis der unterjährigen Erhebung.....	35
3.1.6 Fazit auf Basis der bisherigen Ergebnisse	38
3.2 IAB-Stellenerhebung.....	38
3.2.1 Aufbau des Experiments	39
3.2.2 Treatment: 2-seitiger Fragebogen und 4-seitiger Fragebogen.....	40
3.2.3 Ergebnisse zum prozeduralen Verfahren: <i>Call for Questions</i> Rücklauf.....	41
3.2.4 Pretest	43
3.2.5 Ergebnisse des Längen-Experiments.....	44
3.2.6 Fazit auf Basis der bisherigen Ergebnisse	81

4	Konzeptionelle Überlegungen zur Integration der Befragungen	82
4.1	Rollierendes Panel	82
4.1.1	Überblick	82
4.1.2	Beschreibung der Variante.....	83
4.1.3	Skizze eines Erhebungszyklus	89
4.1.4	Bewertung hinsichtlich der Projektziele	92
4.1.5	Fazit	93
4.2	Integration ausgehend von IAB-Stellenerhebung	94
4.2.1	Offene noch zu diskutierende Punkte	96
5	Fazit auf Basis der bisherigen Ergebnisse und Ausblick.....	100
	Literaturverzeichnis	102
	Anhang 1: Pretest Ergebnisse.....	107
	Anhang 2: Teilnahmewahrscheinlichkeit multivariaten Modell.....	108
	Abbildungsverzeichnis.....	121
	Tabellenverzeichnis.....	122
	Impressum	124

Zusammenfassung

Betriebsbefragungen sind eine wertvolle Datengrundlage für Forschung, Politikberatung und die amtliche Statistik. Die Anforderungen an die resultierenden Datensätze sind seit vielen Jahren vielfältig und reichen von einem Paneldatensatz als Grundlage für die internationale Forschung bis zu vierteljährlichen Datenlieferungen an Eurostat innerhalb weniger Wochen nach dem Ende eines Quartals. Aufgrund der Krisen der letzten Jahre gewinnt zudem die zeitnahe Verfügbarkeit von krisenrelevanten Informationen über die Betriebe an Bedeutung. Um diese Vielzahl an Anforderungen zu erfüllen, führt das IAB seit über 30 Jahren mit dem IAB-Betriebspanel und der IAB-Stellenerhebung regelmäßig zwei große, aufwändige Betriebsbefragungen durch.

Als ein Teilprojekt des vom BMAS finanzierten Projekts „Förderung innovativer Ansätze zur Stärkung von Dateninfrastruktur und Methoden“ führt das IAB die „Machbarkeitsstudie zur Weiterentwicklung der Betriebsbefragungen am IAB“ durch. Der hier für das BMAS vorgelegte Endbericht enthält erste Ergebnisse zu folgenden Kernfragestellungen des Projekts: Ist es möglich, die beiden bestehenden Betriebsbefragungen (IAB-Stellenerhebung und IAB-Betriebspanel) des IAB um hochfrequente Elemente zu erweitern? Welche Möglichkeiten bestehen, die beiden Befragungen in eine Befragung zu integrieren? Die Ergebnisse in diesem Bericht stellen sind noch nicht abschließend, da das Projekt über die vom BMAS finanzierte Periode hinaus am IAB fortgeführt werden wird.

Zur zweiten Fragestellung, der Integration der beiden Betriebsbefragungen, wurden und werden im weiteren Projektverlauf die Machbarkeit sowie die Vor- und Nachteile unterschiedlicher Integrationsvarianten untersucht. Dabei sollen in jedem Fall die folgenden Anforderungen erfüllt werden: Es erfolgt weiterhin eine Panelbefragung von Betrieben. Die Lieferverpflichtungen an Eurostat werden eingehalten. Es werden weiterhin Informationen zu den Einstellungsprozessen erhoben. Beratungs- und forschungsrelevante Informationen stehen öfter als nur einmal jährlich zur Verfügung.

Zur ersten Fragestellung, der Integration hochfrequenter Elemente in die bestehenden Befragungen, wurden in beiden Befragungen Experimente durchgeführt, um die Effekte hochfrequenter Elemente vor allem auf den erwarteten Rücklauf zu untersuchen. Im Rahmen des IAB-Betriebspanels wurde im Jahr 2024 eine unterjährige telefonische Befragung zwischen den Haupterhebungen der Jahre 2023 und 2024 durchgeführt. Auf dieser Basis lässt sich die Teilnahmebereitschaft von Betrieben an einer zusätzlichen Erhebung sowie deren Umsetzbarkeit beurteilen. Um auch den Effekt der unterjährigen Zusatzerhebung auf die Teilnahmebereitschaft an künftigen Erhebungen untersuchen zu können, wurden dafür nur zufällig ausgewählte erst- und wiederholtbefragte Betriebe eingeladen. Die bisherigen Analysen zeigen, dass eine unterjährige Befragung in diesem Umfang prinzipiell umsetzbar ist. Allerdings stimmt nur ein relativ geringer Teil der befragten Betriebe einer Kontaktaufnahme für eine zusätzliche unterjährige Befragung zu. Bei den kontaktierten Betrieben bewegt sich die tatsächliche Teilnahmebereitschaft

jedoch im erwarteten Rahmen. Der Einfluss der unterjährigen Befragung auf die Teilnahmebereitschaft an künftigen Erhebungen wird im weiteren Projektverlauf untersucht.

In der IAB-Stellenerhebung werden derzeit bereits quartalsweise Nachbefragungen durchgeführt. Im Rahmen dieser Nachbefragungen wurde ein Survey-Experiment implementiert, um die Einführung weiterer hochfrequenter Fragen und die Auswirkungen verlängerter Fragebögen vor allem auf die Rücklaufquote zu untersuchen. Nach Einteilung einer Experimentalstichprobe in zwei Gruppen, von der jeweils eine in den Quartals-Nachbefragungen einen 2-seitigen Fragebogen und die andere einen 4-seitigen Fragebogen erhielt, zeigen sich keine signifikant unterschiedlichen Rücklaufquoten zwischen den beiden Gruppen. Auch im Vergleich zum Status-Quo, in dem die Betriebe in der Nachbefragung einen 1-seitigen Fragebogen erhalten, fallen die Unterschiede gering aus. Bezüglich der Response Burden (Teilnahmebelastung) deuten die Ergebnisse jedoch auf leichte Vorteile des 2-seitigen im Vergleich zu einem 4-seitigen Fragebogen hin. Dies deutet darauf hin, dass die kürzere Variante etwas besser geeignet sein könnte, die Teilnahmebereitschaft langfristig aufrechtzuerhalten, ohne die Belastung der Befragten zu erhöhen. Die Experimental-Effekte auf den langfristigen Rücklauf in allen Nachbefragungen werden im weiteren Projektverlauf untersucht.

Zur Frage, ob die beiden bestehenden Betriebsbefragungen des IAB um hochfrequente Elemente erweitert werden können, zeigen die ersten Ergebnisse somit insgesamt, dass eine solche Erweiterung grundsätzlich durchführbar wäre. Dies steht allerdings unter dem Vorbehalt, dass in beiden Befragungen die finalen Ergebnisse zu den Rücklaufeffekten noch ausstehen.

Mit Blick auf die Integration der beiden Befragungen wurde zunächst ein rollierendes Panel untersucht. Hier wird eine jährliche Panelstichprobe von Betrieben in vier Substichproben aufgeteilt, welche quartalsweise als Panelerhebung befragt werden. Dadurch können Inhalte häufiger als einmal jährlich abgefragt und ein für die Forschung attraktiver Paneldatensatz bereitgestellt werden. Die Lieferverpflichtungen an Eurostat können mit einem rollierenden Panel grundsätzlich erfüllt werden, allerdings wäre ein speziell darauf angepasstes Hochrechnungsverfahren nötig. Eine Integration der Fragebögen der beiden Erhebungen erscheint grundsätzlich möglich, würde es aber erfordern, Inhalte, die bisher erhoben werden, ganz wegzulassen oder nicht mehr in jeder Befragung zu erheben. Im Vergleich zum bisherigen IAB-Betriebspanel würde ein rollierendes Panel zudem substanzielle zusätzliche Sach- und Personalmittel erfordern. Eine Anknüpfung an die bestehenden Befragungen dürfte allerdings nicht ohne Brüche in den Zeitreihen zu bewerkstelligen sein.

Summary

Establishment surveys provide valuable data for research, policy advice, and official statistics. Their scope ranges from providing panel data for international research to providing quarterly data to Eurostat within weeks after the end of each quarter. Recent crises increased the importance of promptly gathering and making available relevant information from an establishment perspective. To meet these diverse demands, the IAB conducts two long-standing, large-scale establishment surveys – the IAB Establishment Panel and the IAB Job Vacancy Survey.

As part of the BMAS-funded project “Promoting Innovative Approaches to Strengthen Data Infrastructure and Methods”, the IAB is conducting the “Feasibility Study for the Advancement of Establishment Surveys at the IAB”. This report, prepared for the BMAS, presents preliminary results addressing the project's two key questions. First, can the existing surveys (the IAB Job Vacancy Survey and the IAB Establishment Panel) be extended with high-frequency elements? Second, is it feasible to integrate the two surveys into a single survey? The findings in this report are not yet conclusive, as the project will continue beyond the BMAS-funded phase.

Regarding the integration of both surveys into one survey, different approaches were and will be examined as a part of the project. Key conditions for integration include maintaining a panel survey of establishments, fulfilling data delivery obligations to Eurostat, collecting information on hiring processes, and providing relevant data at a higher frequency than annually.

Regarding the integration of high-frequency elements, experiments were conducted in both surveys to assess the impact of high-frequency elements, particularly on expected response rates. The IAB Establishment Panel implemented an additional mid-year telephone survey in 2024 between the main surveys of 2023 and 2024. This addition tests establishments’ willingness to participate in additional surveys and the feasibility of such an approach. Only random subsets of first-time and repeat participants were invited to uncover the mid-year survey’s impact on future response rates. The initial results indicate that a mid-year survey of this scale is feasible. However, only a relatively small proportion of establishments agreed to be contacted for additional surveys. Among the contacted establishments, the actual response rate met expectations. Further analysis will examine the effect of the mid-year survey on future participation.

The IAB Job Vacancy Survey regularly conducts quarterly follow-up surveys. Within these follow-up surveys, a survey experiment was conducted to test the introduction of additional high-frequency questions and the impact of longer questionnaires on response rates. The experiment split participants into two groups, one receiving a 2-page questionnaire and the other a 4-page version. The results showed no significant differences in response rates between the groups or compared to the status quo (a 1-page questionnaire). However, regarding response burden, the 2-page version seemed slightly better at maintaining long-term participation without increasing respondent strain. Long-term effects on response rates will be examined further.

The preliminary results from both experiments jointly suggest that extending both establishment surveys with high-frequency elements is feasible. However, final results on the impact on response rates are still pending.

Regarding the integration of both surveys into one survey, a rotating panel approach was evaluated, dividing the annual panel sample into four subsamples surveyed quarterly. This approach could enable more frequent data collection while providing an attractive panel dataset for research. Data delivery obligations to Eurostat can be met with this approach, though doing so would require specially tailored extrapolation methods. While integrating the questionnaires from the two surveys is generally feasible, some previously collected content would have to be omitted or collected less frequently. A rotating panel would also demand significantly more resources, both financial and human, compared to the current IAB Establishment Panel. Moving from the current surveys to a rotating panel does not seem feasible without creating breaks in existing time-series.

Danksagung

Diese Studie wurde von der Europäischen Union – NextGenerationEU finanziert.

1 Einleitung

Betriebsbefragungen sind eine wertvolle Datengrundlage für Forschung, Politikberatung und die amtliche Statistik. Die Anforderungen an die daraus resultierenden Datensätze sind seit vielen Jahren vielfältig. Sie reichen von einem Paneldatensatz, der Längsschnittauswertungen auf Betriebsebene ermöglicht, als Grundlage für internationale Forschung bis zu vierteljährlichen Datenlieferung an Eurostat innerhalb weniger Wochen nach dem Ende eines Quartals. Aufgrund der Krisen der letzten Jahre gewinnt zudem die zeitnahe Verfügbarkeit von in Krisen relevanten Informationen aus betrieblicher Perspektive an Bedeutung. Um diese Vielzahl an Anforderungen zu erfüllen führt das IAB seit über 30 Jahren mit dem IAB-Betriebspanel und der IAB-Stellenerhebung regelmäßig zwei große, aufwändige Betriebsbefragungen durch, die jeweils spezifische wissenschaftliche Fragestellungen und Beratungsbedarfe adressieren und dafür eigene Erhebungsinstrumente verwenden.

Als ein Teilprojekt des vom BMAS finanzierten Projekts „Förderung innovativer Ansätze zur Stärkung von Dateninfrastruktur und Methoden“ führt das IAB die „Machbarkeitsstudie zur Weiterentwicklung der Betriebsbefragungen am IAB“. Diese Machbarkeitsstudie untersucht zwei Kernfragestellungen: die Integration hochfrequenter Elemente in die bestehenden Betriebsbefragungen und die Machbarkeit der Integration der beiden Befragungen in eine Befragung.

Mit hochfrequenten Elementen sollen beratungs- und forschungsrelevante Informationen in höherer Frequenz als einmal jährlich zur Verfügung stehen. Diese Anforderung ergibt sich aus dem Bedarf nach kurzfristig verfügbaren Ergebnissen, der in den letzten Jahren infolge verschiedener Krisen an Bedeutung gewonnen hat. Daher wird in der Machbarkeitsstudie in einem ersten Schritt untersucht, inwieweit sich hochfrequente Elemente unter der Restriktion bestehender extramuraler Vergaben in die bestehenden Betriebsbefragungen am IAB integrieren lassen. Hierzu sollen auch experimentelle Ansätze im Rahmen der laufenden Befragungen genutzt werden.

Um die Machbarkeit der Integration der beiden Befragungen zu untersuchen, werden u.a. auf Basis dieser Ergebnisse mögliche übergeordnete Konzepte entwickelt, wie sich die verschiedenen Anforderungen an die Betriebsbefragungen des IAB durch eine einzelne Befragung erfüllen ließen. Neben der Berücksichtigung von hochfrequenten Elementen sollen dabei die zentralen Merkmale der bereits bestehenden Umfragen integriert werden.

Für die Weiterentwicklung der Betriebsbefragungen am IAB sind dabei folgende Nebenbedingungen zu beachten:

1. Für wissenschaftliche und beratungsrelevante Untersuchungen ist es aus inhaltlicher und methodischer Perspektive weiterhin erforderlich, eine Panelbefragung von Betrieben durchzuführen.
2. Die amtliche Verpflichtung an Eurostat zu der Gesamtzahl der offenen Stellen und Beschäftigten zu berichten muss auch in Zukunft unter Beachtung der hierfür erforderlichen Qualitäts- und

Zeitrestriktionen (45 Tage nach Quartalsende) gemäß den in Abschnitt 2.2.6 dargestellten Anforderungen erfüllt werden.

3. Als Komplementärstatistik zu den amtlichen Daten der gemeldeten offenen Stellen der Bundesagentur für Arbeit (BA) sollen weiterhin Informationen zu den Einstellungsprozessen erhoben werden.
4. Vor dem Hintergrund des erhöhten Bedarfs nach kurzfristig verfügbaren Ergebnissen sollen beratungs- und forschungsrelevante Informationen in höherer Frequenz als einmal jährlich zur Verfügung stehen.

Dieser Forschungsbericht präsentiert die bisherigen Ergebnisse des Teilprojekts „Machbarkeitsstudie zur Weiterentwicklung der Betriebsbefragungen am IAB“, liefert jedoch noch keine abschließenden Befunde, da das Projekt noch bis Ende 2025 weiterläuft. Zunächst gibt Kapitel 2 einen kurzen Überblick über das IAB-Betriebspanel und die IAB-Stellenerhebung als die beiden großen Betriebsbefragungen des IAB. Kapitel 3 beschreibt die Experimente zur Implementierung hochfrequenter Elemente in beiden Befragungen und präsentiert, soweit bereits verfügbar, die Ergebnisse dieser Experimente. Kapitel 4 legt den aktuellen Stand der konzeptionellen Überlegungen zur Integration der Befragungen dar. Schließlich zieht Kapitel 5 ein kurzes Fazit der bisherigen Ergebnisse.

2 Status Quo der Betriebsbefragungen am IAB

2.1 IAB-Betriebspanel

Das IAB-Betriebspanel (IAB-BP) ist eine jährliche repräsentative Arbeitgeberbefragung zu betrieblichen Determinanten der Beschäftigung und stellt eine zentrale Quelle für Analysen zur Arbeitskräftenachfrage auf dem Arbeitsmarkt in Deutschland dar. Die Erhebung wird seit 1993 in Westdeutschland und seit 1996 auch in Ostdeutschland durchgeführt.

Jährlich werden von Ende Juni bis etwa Ende Oktober bundesweit rund 15.000 Betriebe aller Wirtschaftszweige und Größenklassen befragt. Die Befragung wird überwiegend in persönlich-mündlichen Interviews von Verian im Auftrag des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) durchgeführt. Das IAB-Betriebspanel repräsentiert die Gesamtheit aller Betriebe mit mindestens einem sozialversicherungspflichtig Beschäftigten in Deutschland. Das IAB hat den Bundesländern die Möglichkeit eröffnet, die Stichprobe aus Landesmitteln so aufzustocken, dass hinreichende Nettofallzahlen auch für landesspezifische Auswertungen zur Verfügung stehen. Diese Möglichkeit nutzen derzeit elf Bundesländer. Zusätzlich finanziert das IWH Halle eine Aufstockungsstichprobe für das verarbeitende Gewerbe in Ostdeutschland. Das IAB finanziert rund 9.250 der etwa 15.000 Interviews. Das IAB-Betriebspanel ist als Längsschnitterhebung konzipiert, in der möglichst jedes Jahr dieselben Betriebe befragt werden. Die Befragung umfasst ein breites Fragespektrum zu einer Vielzahl beschäftigungspolitischer Themen. Wiederkehrende Fragen nehmen etwa drei Viertel des Fragebogens

ein und ein Viertel steht für zusätzliche Fragen zur Verfügung. Auf diese Weise bietet der Datensatz nicht nur interessante Auswertungsmöglichkeiten für die Politikberatung, sondern auch vielfältige Analysemöglichkeiten für die Wissenschaft. Die Daten werden der wissenschaftlichen Öffentlichkeit über das Forschungsdatenzentrum (FDZ) zur Verfügung gestellt. Das IAB-Betriebspanel ist darüber hinaus Basis weiterer Datenprodukte des FDZ wie der verschiedenen LIAB-Varianten, in denen Betriebs- und Beschäftigtendaten zusammengefügt werden.

2.1.1 Stichprobe

Die Grundgesamtheit des IAB-Betriebspanels umfasst alle Betriebsstätten mit mindestens einer oder einem sozialversicherungspflichtig Beschäftigten zum Stichtag 30. Juni. Basis der Stichprobenziehung ist die Betriebsdatei der Bundesagentur für Arbeit (BA). Sie enthält alle Betriebe, die im Rahmen des Meldeverfahrens zur Sozialversicherung ihre sozialversicherungspflichtig Beschäftigten an die Sozialversicherungsträger melden und eine Betriebsnummer erhalten.

Die Stichprobe ist nach Branche (20 Kategorien), Betriebsgröße (10 Kategorien) und Bundesland (16 Kategorien) disproportional geschichtet. Für die Branchengliederung wird seit der Erhebungswelle 2009 die Wirtschaftszweigklassifikation WZ 2008 herangezogen. Die Stichprobe wird nach dem Prinzip der optimalen Schichtung beschäftigtenproportional gezogen. Große Betriebe sind aus diesem Grund in der Stichprobe überproportional vertreten. Diese Disproportionalitäten werden im Rahmen der Gewichtung wieder ausgeglichen.

Die jährlichen Bruttostichproben des IAB-Betriebspanels setzen sich aus drei Teilstichproben zusammen:

- **Wiederholerstichprobe:** Sie umfasst alle Betriebe mit einem gültigen Interview aus dem Vorjahr. In der Erhebung 2023 waren das rund 14.500 Betriebe (brutto). Die wiederholt befragten Betriebe bilden den Kern des IAB-Betriebspanel, weil sie über einen längeren Zeitraum hinweg beobachtet werden können. Sie ist notwendig, damit man über reine Zeitreihenanalysen hinausgehend Panelauswertungen durchführen kann.
- **Nachbearbeitungsstichprobe:** Sie umfasst alle (teilnahmebereiten) Betriebe (1.200 Betriebe (brutto) in der Erhebung 2023) mit einem gültigen Interview aus dem Vorvorjahr. Diese Teilstichprobe erhöht die im Querschnitt auswertbare Anzahl von Fällen.
- **Ergänzungsstichprobe:** Betriebe, die neu aus der Betriebsdatei der Bundesagentur für Arbeit gezogen wurden. Diese Betriebe werden zum Ausgleich der Panelmortalität und zur laufenden Anpassung der Stichprobe an den wirtschaftlichen Strukturwandel (neue Betriebsnummern) ergänzt. In der Ergänzungsstichprobe wurden 2023 rund 41.000 Adressen (brutto) eingesetzt.

Tabelle 1 gibt einen Überblick über die Entwicklung der Nettofallzahlen in der Stichprobe seit 1993. Diese Fälle stehen jährlich für Querschnittsanalysen zur Verfügung. Die Tabelle zeigt, dass bis 1996 nur gut 4.000 Fälle verfügbar waren, da nur westdeutsche Betriebe befragt wurden. Durch die Einbeziehung Ostdeutschlands in die Stichprobe 1996 und die finanzielle Beteiligung der ostdeutschen Bundesländer

stieg die Anzahl auswertbarer Interviews auf über 9.000. Ab 2000 kam die Finanzierung durch westdeutsche Bundesländer hinzu, was zu einem zusätzlichen Anstieg des Stichprobenumfangs führte.

Tabelle 1: Auswertbare Fälle des IAB-Betriebspanels

1993 bis 2023

Leere Zelle	Westdeutschland	Ostdeutschland	Gesamt
1993	4.265	0	4.265
1994	4.138	0	4.138
1995	4.096	0	4.096
1996	4.029	4.313	8.342
1997	4.102	4.748	8.850
1998	4.289	4.905	9.194
1999	4.427	5.335	9.762
2000	8.416	5.515	13.931
2001	9.849	5.688	15.537
2002	10.105	5.303	15.408
2003	10.090	5.767	15.857
2004	10.104	5.585	15.689
2005	10.103	5.718	15.821
2006	9.856	5.593	15.449
2007	9.630	6.014	15.644
2008	9.361	6.095	15.456
2009	9.525	5.998	15.523
2011	9.596	6.019	15.615
2012	9.418	5.865	15.283
2013	9.560	5.996	15.556
2014	9.589	6.136	15.725
2015	9.474	6.103	15.577
2016	9.410	6.090	15.500
2017	9.416	5.925	15.341
2018	9.491	5.930	15.421
2019	9.149	6.114	15.263
2020	9.398	6.041	15.439
2021	10.481	6.205	16.686
2022	9.128	6.093	15.217
2023	8.133	6.442	14.575

Quelle: IAB-Betriebspanel; eigene Darstellung. © IAB

2.1.2 Datenerhebung

Die Befragung wird den Betrieben in der Stichprobe in einem gemeinsamen Brief der Vorsitzenden des Vorstands der BA und des Direktors des IAB angekündigt. Dem Anschreiben liegt außerdem ein Unterstützungsschreiben des Präsidenten der Bundesvereinigung der Deutschen Arbeitgeberverbände (BDA) bei, das die besondere Bedeutung der Teilnahme an der Befragung betont.

Die Daten des IAB-Betriebspanels werden in einem sogenannten Mixed-Mode Ansatz erhoben. Das bedeutet, dass das computergestützte Erhebungsinstrument so aufgebaut ist, dass es sowohl in persönlich-mündlichen Interviews eingesetzt werden kann als auch von der befragten Person selbst online ausgefüllt werden kann. Rund 75 Prozent der Interviews werden persönlich-mündlich mit Interviewenden im Betrieb oder am Telefon durchgeführt. Die Interviewenden sind speziell für Betriebsbefragungen geschult.

Die durchschnittliche Interviewdauer betrug 2023 rund 40 Minuten. Die Dauer der Interviews ist von der Größe des Betriebs abhängig. So dauert ein Interview bei Betrieben mit bis zu 10 Beschäftigten nur etwa eine halbe Stunde, während es bei großen Betrieben über eine Stunde dauern kann. Das liegt vor allem daran, dass kleinere Betriebe ad hoc einen besseren Überblick über das betriebliche Geschehen haben als größere Betriebe. Außerdem treffen für kleinere Betriebe bestimmte Fragen seltener zu, so dass sie deutlich weniger Fragen beantworten müssen.

2.1.3 Fragebogen und Inhalte der Befragung

Der Fragebogen des IAB-Betriebspanels umfasst ein großes Spektrum an Themen, wie beispielsweise Fragen zur Aus- und Weiterbildung, zur Lohnfindung, zu betriebliche Kenngrößen wie Umsatz und Vorleistungen und zur Beschäftigtenstruktur. Da das IAB-Betriebspanel als Längsschnitterhebung (Panel) konzipiert ist, ist es wichtig die relevanten Fragen jährlich zu wiederholen, um langfristige Entwicklungen beobachten und analysieren zu können. Aus diesem Grund ist der Fragebogen modular aufgebaut. Zu jedem Themenblock (Modul) gibt es ein definiertes Basisprogramm an Fragen. Diese sind in jeder Welle fester Bestandteil des Fragebogens. Darüber hinaus gibt es für verschiedene Themenblöcke zusätzliche vertiefende Fragen, die in einem zweijährigen Turnus abgefragt werden (Weiterbildung, betriebliche Innovationen oder Arbeitszeit). Fragen zur Altersstruktur der Beschäftigten und zu Frauen in Führungspositionen werden abwechselnd alle zwei Jahre gestellt. Mit diesem Modulsystem wird gewährleistet, dass wesentliche Informationen in jedem Jahr vorhanden sind. Außerdem wird über die (in zweijährigem Turnus gestellten) vertiefenden Fragen das Analysepotential des IAB-Betriebspanels erheblich erhöht.

Darüber hinaus sind in jedem Jahr für aktuellen Schwerpunktthemen etwa vier der 20 inhaltlichen Seiten des Fragebogens reserviert. Die wichtigsten Schwerpunkte der letzten Jahre waren das Thema Fachkräftebedarf und Fachkräftesicherung und ergänzende Fragen zum Ausbildungsmodul. Während der Corona-Pandemie wurde der Fragebogen um Fragen zu den Folgen der Pandemie und dem betrieblichen Umgang mit der Krise ergänzt. Ähnliche Fragen standen auch zu Beginn des russischen Angriffskriegs gegen die Ukraine im Zentrum des Interesses.

2.1.4 Hochrechnung

Eine Gewichtung der Daten des IAB-Betriebspanels ist aufgrund des disproportionalen Stichprobenansatzes sowie zum Ausgleich möglicher Differenzen zwischen Ist- und Sollstärke der einzelnen Schichtungszellen nötig. Die Gewichtung des IAB-Betriebspanels erfolgt grundsätzlich in Form einer Hochrechnung auf die Grundgesamtheit. Dabei wird die Gewichtung sowohl auf die Anzahl der Betriebe in der Grundgesamtheit insgesamt, als auch in Ost- und Westdeutschland und in den einzelnen Bundesländern sowie auf die Verteilung der Betriebe des Verarbeitenden Gewerbes in Ostdeutschland angepasst. Für die Gewichtung kommt ein generalisierter Regressions-Schätzer (GREG) zur Anwendung. Die Soll-Strukturen ergeben sich aus der Betriebsdatei der Bundesagentur für Arbeit. Maßgeblich ist die Struktur (Verteilung der Betriebe) zum Zeitpunkt der Stichprobenziehung für die jeweilige Welle (also zum 30. Juni des jeweiligen Vorjahres). Durch eine solche Hochrechnung werden sowohl die Disproportionalitäten als auch die unterschiedlichen Ausschöpfungsraten entlang der Schichtungszellen in einem Schritt ausgeglichen.

Im Rahmen der Querschnittgewichtung wird darauf geachtet, dass die aus der gewichteten Stichprobe hochgerechneten Beschäftigtenzahlen (sozialversicherungspflichtig Beschäftigte zum 30. Juni des Vorjahres) auf der Ebene der Bundesländer den Soll-Vorgaben aus der Beschäftigtenstatistik der Bundesagentur für Arbeit entsprechen.

Ergebnis der Gewichtung ist ein integrierter betriebs- und beschäftigtenproportionaler Gewichtungsfaktor, so dass mit den gewichteten Daten nicht nur betriebsproportionale, sondern auch beschäftigtenproportionale Auswertungen möglich sind.

2.1.5 Datenqualität

Die Sicherung einer gleichbleibend hohen Datenqualität hat bei der Erhebung des IAB-Betriebspanels große Bedeutung. So werden schon bei Erstellung des Fragebogens hohe Standards angelegt. Zur Qualitätssicherung werden in einem kognitiven Pretest neue Fragen daraufhin getestet, ob sie sich für eine Betriebsbefragung eignen und ob die Betriebe voraussichtlich über die abgefragten Sachverhalte auskunftsfähig sind, also die gewünschten Informationen auf Betriebsebene vorliegen. Dazu werden in rund 90 Betrieben aus unterschiedlichen Branchen und Betriebsgrößenklassen Interviews durchgeführt. Das Pretest-Interview besteht aus zwei Teilen. Die Testperson wird zunächst gebeten, den Fragebogen zu beantworten. Der zweite Schritt beinhaltet dann den eigentlichen kognitiven Test. Die Betriebe werden gebeten, jede Frage zu kommentieren, und gefragt, ob sie bei der Beantwortung Schwierigkeiten hatten. Von besonderem Interesse sind dabei die Auskunftsfähigkeit der Betriebe sowie Verständnisschwierigkeiten und Interpretationsprobleme. Die Ergebnisse des kognitiven Pretests fließen in die weitere Fragenentwicklung ein. Der Pretest wird von speziell geschulten Projektmitarbeiterinnen und Projektmitarbeitern des IAB durchgeführt.

Für die Befragung des IAB-Betriebspanel werden die Interviewenden jährlich ausführlich geschult. Während der Erhebung wird die Arbeit der Interviewenden laufend kontrolliert. Diese Kontrolle soll sicherstellen, dass die Interviews korrekt durchgeführt werden und es zu keinen Verfälschungen der

Untersuchungsergebnisse kommt. Darüber hinaus kommt nach Abschluss der Erhebung ein aufwändiges Datenprüfverfahren zur Anwendung. Dabei werden die Daten durch ein umfangreiches Programm nach über 200 Kriterien überprüft. Bei Verletzung eines der Prüfkriterien wird ein Fehlerprotokoll erstellt. Falls es zur Aufklärung von Prüf Fehlern erforderlich ist, wird der befragte Betrieb telefonisch kontaktiert.

2.1.6 Datenprodukte auf Basis des IAB-Betriebspanels

Das IAB-Betriebspanel ist die Basis für drei Standarddatenprodukte, die über das Forschungsdatenzentrum der BA (FDZ) auch externen Forschenden zugänglich sind:

- **Betriebspanel:**
Die Stichprobe aus der jährlichen Erhebung wird der wissenschaftlichen Öffentlichkeit auf Antrag zugänglich gemacht
- **LIAB:**
Das FDZ bietet mehrere Datensätze an, in denen die Daten des IAB-Betriebspanels mit den Beschäftigteninformationen aus den administrativen Daten der BA verknüpft werden. Diese Datensätze werden speziell für wissenschaftliche Analysen zur Verfügung gestellt. Zurzeit nutzen etwa 420 Forschende das IAB-Betriebspanel und den LIAB.
- **LPP (einschließlich LPP-ADIAB):**
Seit 2012 wird das IAB-Betriebspanel jedes zweite Jahr durch eine vertiefenden Betriebsbefragung für das Projekt „Arbeitsqualität und wirtschaftlicher Erfolg“ (LPP) ergänzt (Ausnahme: Die Erhebung für 2022 wurde auf 2023 verschoben). In diesem Projekt stehen personalökonomische Fragestellungen im Mittelpunkt. Die Erhebung ist als Zusatzbefragung im Rahmen des IAB-Betriebspanels konzipiert. 2023 wurden erstmals auch kleine und mittelgroße Betriebe mit unter 50 sozialversicherungspflichtig Beschäftigten in diesem Rahmen befragt. Zusätzlich werden im Rahmen dieses Projekts auch Beschäftigte der interviewten Betriebe befragt, was interessante Auswertungsperspektiven für die personalökonomische Forschung in Deutschland eröffnet. Die simultane Betrachtung von Arbeitgeber- und Beschäftigtenperspektive ermöglicht sowohl Quer- als auch Längsschnittanalysen von Wirkungszusammenhängen zwischen Personalmanagement, individueller Arbeitsqualität und betrieblichem Erfolg.
Die Verknüpfung der Befragungen des LPP mit dem IAB-Betriebspanel eröffnet weitere Analysepotentiale. Darüber hinaus werden die Befragungsdaten des LPP auch mit administrativen Daten verknüpft (LPP-ADIAB).

2.2 IAB-Stellenerhebung

Ein wichtiger Grund dafür, warum die Nachfrageseite des Arbeitsmarktes noch unzureichend erforscht ist, ist der Mangel an geeigneten Mikrodatsätzen. Vorhandene Daten, die z.B. die realisierte Beschäftigung messen, liefern nur ein unvollständiges Bild, da sie die ungedeckte Nachfrage in Form unbesetzter Stellen nicht erfassen. Zudem beschränken sich prozessproduzierte Daten zu offenen

Stellen auf dem deutschen Arbeitsmarkt auf bei der Bundesagentur für Arbeit mit Vermittlungsauftrag gemeldete offene Stellen und bilden somit nur einen Teil aller offenen Stellen ab (Bossler et al., 2020).

Die IAB-Stellenerhebung ist für den deutschen Arbeitsmarkt der einzige Datensatz, der Informationen über die Struktur und das Niveau sowohl der gedeckten als auch der ungedeckten Arbeitsnachfrage bereitstellt (Bossler et al., 2020). Hierbei verfolgt die IAB-Stellenerhebung, die seit 1989 vom IAB durchgeführt wird, zwei wesentliche Ziele:

Das erste Ziel ist es, umfassende Informationen über die Struktur und Anzahl gemeldeter und nicht gemeldeter offener Stellen sowie der Gesamtbeschäftigung auf dem deutschen Arbeitsmarkt bereit zu stellen. Als solches ist die Umfrage auch amtlicher Datenlieferant für die europäische Statistik zu offenen Stellen, die von Eurostat verwaltet wird und ein vollständiges Bild der offenen Stellen sowie (Gesamt-)Beschäftigung in Deutschland liefert (Eurostat, 2024b). In Deutschland sind, ähnlich wie in vielen anderen Ländern, Betriebe nicht gesetzlich verpflichtet, ihre offenen Stellen zu melden. Daher enthält die amtliche Statistik der Bundesagentur für Arbeit (BA) nur einen Teil der gesamten offenen Stellen, d.h. die Anzahl der gemeldeten offenen Stellen, für die die Betriebe einen Vermittlungsauftrag erteilt haben. Aufgrund der Schwankungen des Anteils der gemeldeten offenen Stellen an allen offenen Stellen über die Zeit, kann die gesamte Arbeitsnachfrage nicht zuverlässig aus den bei der Bundesagentur für Arbeit gemeldeten Stellen abgeleitet werden. Gemeldete offene Stellen repräsentieren somit weder in ihrer Struktur noch in ihrem Volumen die gesamten offenen Stellen (Bossler et al., 2020). Die Meldequote schwankte in den vergangenen 10 Jahren zwischen 58 und 39 Prozent, sie lag im vierten Quartal 2023 bei 41 Prozent. Auch liegen der BA in Bezug auf die Gesamtbeschäftigung zwar (mit einem 6-monatigen Zeitverzug) Informationen über die sozialversicherungspflichtige Beschäftigung sowie die (ausschließlich) geringfügig Beschäftigten vor, ihr fehlen jedoch notwendige Informationen zu Zahl der Beamtinnen und Beamten, Richterinnen und Richter sowie tätigen Inhaberinnen und Inhabern sowie mithelfenden Familienangehörigen.

Das zweite Ziel der IAB-Stellenerhebung ist es, detaillierte Informationen über die Such- und Rekrutierungsprozesse der Arbeitgeber zu erheben. Damit zielt die Erhebung darauf ab, eine größere inhaltliche Lücke in der Verfügbarkeit von Mikrodaten zu schließen, da es (auch international) bislang sehr wenige Datensätze gibt, die das Rekrutierungsverhalten von Arbeitgebern auf vergleichbare Weise abbilden können. Konkret werden im Rahmen der IAB-Stellenerhebung Personalmanager oder Führungskräfte mit Personalverantwortung befragt. Die Befragten werden gebeten, u.a. Informationen über die Anzahl und Struktur offener Stellen, Rekrutierungsprozesse und Entwicklungen, die für die Arbeitsmarktpolitik relevant sind, zu liefern.

2.2.1 Stichprobe

Die IAB-Stellenerhebung wird als wiederholte Querschnittsbefragung deutscher Betriebe durchgeführt. Die Grundgesamtheit der Hauptumfrage im vierten Quartal jedes Jahres besteht aus allen Betrieben in Deutschland mit mindestens einer oder einem sozialversicherungspflichtigen Beschäftigten. Da der Anteil der Betriebe, die über offene Stellen verfügen, tendenziell gering ist, erfordert die Umfrage eine große Stichprobe. Aktuell wird für die IAB-Stellenerhebung jedes Jahr erneut eine geschichtete

Zufallsstichprobe von maximal 140.000 Betrieben aus der Grundgesamtheit von rund 2.200.000 Betrieben gezogen. Dies bedeutet, dass jedes Jahr etwa sechs Prozent aller Betriebe zur Teilnahme eingeladen werden. Seit dem vierten Quartal 2015 wird die Stichprobe nach Region (West- und Ostdeutschland), sechs Betriebsgrößenklassen und 24 Branchen gemäß der deutschen Branchenklassifikation 2008 (NACE Rev. 2) in einer dreidimensionalen Stichprobenmatrix stratifiziert. Die Kontaktwahrscheinlichkeit variiert stark innerhalb jeder Stratifikationszelle.

2.2.2 Datenerhebung

Die IAB-Stellenerhebung wird viermal jährlich, jeweils in jedem Quartal, mit einem Push-to-Web-Modus zunächst mit einer Online-Antwortoption (Computer-Assisted Web Interviewing, CAWI) und der anschließenden Möglichkeit, mittels eines schriftlichen Fragebogens (Paper-and-Pencil Interviewing, PAPI) zu antworten, durchgeführt. Die Hauptbefragung startet in jeder Welle jeweils im vierten Quartal eines Kalenderjahres gefolgt von drei Nachbefragungen in den drei Quartalen des Folgejahres. In den Nachbefragungen wird eine zufällig ausgewählte Teilgruppe der Teilnehmer aus der Nettostichprobe des vierten Quartals gebeten, einige der (für die hochfrequente Berichterstattung an Eurostat benötigten) Kerninformationen der Befragung zu aktualisieren. In den drei Nachbefragungen werden somit nur kürzere Fragebögen (im PAPI-Modus ein 1-Seiter) verwendet, während im vierten Quartal ein ausführlicherer Fragebogen zum Einsatz kommt (im PAPI-Modus ein 8-Seiter).

Im Rahmen des für die IAB-Stellenerhebung gewählten Push-to-Web-Modus erfolgt, wie von Dillman (2017) definiert, der Erstkontakt der Teilnehmer auf postalischem Wege, mit der Einladung, an der Befragung online teilzunehmen. Alternative Beantwortungsmöglichkeiten werden erst später standardmäßig zur Verfügung gestellt. Im Falle der IAB-Stellenerhebung wird dies so umgesetzt, indem den Teilnehmern in jedem Quartal der schriftliche Fragebogen erst mit dem Erinnerungsschreiben zugesandt wird. Bei der Wahl des Erhebungsmodus wurde darauf geachtet, dass das Design der Umfrage in allen Erhebungsquartalen gleichbleibt und damit mögliche Modus-Effekte vermieden werden.¹ Die IAB-Stellenerhebung wird derzeit von dem Erhebungsinstitut „Economics and Data ED23 GmbH“ (früher: „Economix Research and Consulting“) durchgeführt, das vom IAB mit der Erhebung beauftragt wird. Die Hotline wird durch den Unterauftragnehmer T.I.P. Biehl und Partner betreut. Dienstleistungspartner für den Druck und Versand ist derzeit die Marketing Service Magdeburg KG.

2.2.3 Fragebogen und Inhalte der Befragung

Wie oben dargestellt, erfolgt im vierten Quartal die ausführliche Hauptbefragung der Betriebe mit einem längeren Befragungsinstrument, während sich die Nachbefragungen in den drei Folgequartalen auf die Abfrage einiger Kerninformationen zu den offenen Stellen und zur Beschäftigung sowie der Arbeitszeitrechnung konzentrieren (siehe hierzu auch Abschnitt 2.2.6).

¹ Derartige Modus-Effekte konnten beim bis zur Erhebungswelle 2022/2023 geltenden Erhebungsmodus nicht ausgeschlossen werden. Dieser sah einen PAPI-Modus mit einer zusätzlichen Online-Option in der Hauptbefragung im vierten Quartal vor, während die Nachbefragungen in den Folgequartalen hingegen mittels telefonischer Interviews (CATI) durchgeführt wurden.

Der im vierten Quartal verwendete Fragebogen besteht aus drei Modulen. *Das erste Modul* erfasst Informationen über die Anzahl der Beschäftigten, die Zusammensetzung der Belegschaft und die Anzahl und Struktur offener Stellen auf Betriebsstättenebene. Darüber hinaus werden die Betriebe nach ihrer aktuellen wirtschaftlichen Situation, Einstellungen und Kündigungen in den letzten 12 Monaten und ihren Erwartungen hinsichtlich der zukünftigen Arbeitsnachfrage gefragt. Fragen zu industriellen Beziehungen vervollständigen das erste Modul. *Das zweite Modul* des Fragebogens erfasst Informationen zu Themen, die von Erhebungswelle zu Erhebungswelle variieren und gesondert über SGB II Mittel finanziert sind. Im Allgemeinen sammelt dieses Modul Informationen über die Bewertung und Nutzung spezifischer arbeitsmarktpolitischer Instrumente durch die Arbeitgeber. Darüber hinaus enthält dieser Teil auch Fragen zur Bereitschaft der Arbeitgeber, Personen mit spezifischen soziodemografischen Merkmalen einzustellen, wie z.B. langzeitarbeitslose Personen oder Migranten. *Das dritte Modul* des Fragebogens konzentriert sich auf konkrete Einzelfälle, den letzten erfolgreichen Einstellungsprozess und den letzten abgebrochenen Suchprozess. Nicht nur die Frage nach Fachkräftengpässen zeigt, dass die Betriebsebene oftmals die tatsächliche Situation am Arbeitsmarkt in einem Berufsfeld verdeckt. Dieser Teil erhebt deshalb detaillierte Informationen zu den konkret eingestellten Personen, zu prädeternierten Merkmalen der offenen Stelle sowie zum Suchprozess. Beispielsweise werden Arbeitgeber in diesem Teil der Umfrage gebeten, genaue Informationen über die Dauer der Rekrutierung, d.h. die Dauer vom Beginn der Suche bis zur Entscheidung für eine Bewerberin oder einen Bewerber, bereitzustellen. Weitere Fragen beziehen sich auf Einstellungskonditionen oder die Such- und Besetzungskanäle, die Arbeitgeber im Zusammenhang mit dem berichteten Einstellungsprozess nutzen, wie beispielsweise soziale Netzwerke, Zeitungsanzeigen, private und öffentliche Arbeitsvermittlungen oder die Nutzung von Unternehmenswebsites und Online-Jobbörsen. Die Angaben sind damit anschlussfähig zur Engpassanalyse der BA. Sowohl das erste als auch das dritte Modul zu Einstellungsprozessen und Rekrutierungsfehlschlägen sind, abgesehen von geringfügigen Änderungen, über die Zeit gleichgeblieben und ermöglichen daher auch die Untersuchung der Entwicklung der erhobenen Items auf betrieblicher sowie auf individueller Ebene über die Zeit. Der Anteil der Suchabbrüche an allen Personalsuchen lag im Jahr 2023 bei 23 Prozent. Dies zeigt zusätzlich die Notwendigkeit, neben erfolgreichen Einstellungsprozessen auch mögliche Suchabbrüche im konkreten Fall umfassend zu erheben. Zusammenfassend bietet die Stellenerhebung detaillierte und umfangreiche Informationen zu den folgenden Themen:

- Anzahl und Struktur der Beschäftigung sowie der betrieblichen Arbeitsnachfrage
- Wirtschaftliche Lage in den befragten Betrieben
- Informationen zu Rekrutierungsprozessen und den dafür verwendeten Instrumenten
- Nutzung und Bewertung arbeitsmarktpolitischer Programme durch Arbeitgeber
- Informationen zu aktuellen arbeitsmarktpolitischen Themen
- Umfang, Ursachen und betriebliche Auswirkungen von abgebrochenen Suchprozessen (Bossler et al., 2020)

2.2.4 Hochrechnung

Da das Stichprobendesign der Stellenerhebung auf einem disproportional geschichteten Stichprobenverfahren basiert, gibt es Strata, die in den Rohdaten über- oder unterrepräsentiert sind. Da dies zu einer potenziellen Antwortverzerrung führt, müssen die Antworten der Betriebe hochgerechnet werden, um Statistiken berechnen zu können, die repräsentativ für die Grundgesamtheit, d.h. Betriebe mit mindestens einer oder einem sozialversicherungspflichtigen Beschäftigten, sind. Konkret werden die Antworten der teilnehmenden Betriebe unter Verwendung eines generalisierten Regressions-Schätzers (GREG) hochgerechnet. Das Verfahren beinhaltet die Ermittlung sowohl von Design- als auch *Nonresponse*-Gewichten, um Aussagen über die Entwicklung der Arbeitsnachfrage und der Rekrutierungsprozesse ermöglichen, die repräsentativ für die jeweiligen Branchen, Regionen und Größenklassen sind (Brenzel et al., 2016). Die Betriebsgewichte werden mit Hilfe der folgenden Schritte ermittelt:

1. Die Designgewichte werden für jede Zelle der Stichprobenmatrix als die Inverse des Verhältnisses der Nettostichprobe in einer Schicht zu der jeweiligen Schichtgröße berechnet.
2. Die Designgewichte werden mit einem Gewicht multipliziert, welches mit Hilfe eines *Nonresponse*-Modells ermittelt wird. Letzteres basiert auf einem logistischen Regressionsmodell, das die Betriebsgrößenkategorie, die Branche, den durchschnittlichen Tageslohn und das durchschnittliche Mitarbeiteralter aus den administrativen Daten der Bundesagentur für Arbeit als Hilfsvariablen verwendet. Die *Nonresponse*-Gewichte werden als Inverse der geschätzten Teilnahmewahrscheinlichkeit ermittelt.
3. Die Gewichte aus Schritt 2 werden als Startgewichte verwendet, um die Gewichte basierend auf einem GREG-Schätzer so zu kalibrieren, dass Eckwerte der Ankervariablen „Anzahl der Betriebe“ und „Anzahl der sozialversicherungspflichtigen Beschäftigten“ in allen Zellen der Stichprobenmatrix genau erreicht werden. Da die amtlichen Quartalsinformationen zu den beiden Ankervariablen erst mit einer Verzögerung von 6 Monaten verfügbar sind, müssen diese Variablen für die zeitnahe Hochrechnung am aktuellen Rand zunächst geschätzt werden (Hutter, 2020).

Gemäß den Eurostat-Vorgaben wird das Kalibrierungsverfahren jeweils im August für das Vor-Vor-Jahr auf Basis der endgültigen Informationen zu den Ankervariablen wiederholt und überarbeitete Hochrechnungsgewichte zur Verfügung gestellt.

2.2.5 Datenqualität

Die Notwendigkeit zur Qualitätssicherung der erhobenen Daten ergibt sich neben den hohen IAB-internen Qualitätsansprüchen sowohl aus den Erfordernissen des EU- Qualitätsberichts im Rahmen der Lieferung der offenen Stellen und Beschäftigtenzahl an Eurostat (siehe hierzu auch Abschnitt 2.2.6) als auch aus dem Beratungsauftrag gegenüber der BA, zum Beispiel im Rahmen der BA-Engpassanalyse.

Ein erstes Kriterium für die Qualität der erhobenen Daten ist die Präzision der Schätzung. Auf Basis des aktuellen EU-Durchführungsrechtsakts zur Statistik der offenen Stellen werden derzeit für die Gesamtheit aller Betriebe, aber auch für die Betriebsgrößenklasse mit mindestens 10 Beschäftigten, die zu den offenen Stellen der verschiedenen Wirtschaftsabschnitte gehören Variationskoeffizienten

einmal jährlich im Rahmen eines Qualitätsreports an Eurostat berichtet (Tabelle 2). Nach derzeitigem Stand ist im Zuge der anstehenden Novellierung künftig mit Präzisionszielen für die nicht-saisonbereinigte Gesamtzahl der offenen Stellen (Variationskoeffizient unter 5%) und die nicht-saisonbereinigte Gesamtzahl der Beschäftigten (Variationskoeffizient unter 1%) zu rechnen, die in den jeweiligen Quartalen nicht überschritten werden dürfen.

Tabelle 2: Geplante EU-Präzisionsziele am Beispiel der Erhebungswelle 2023/2024

2022, Quartale, Variationskoeffizienten

Leere Zelle	Art. 5: CoV Precision targets	Q1/2022	Q2/2022	Q3/2022	Q4/2022
Beschäftigung insgesamt	< 0,01	N/V	N/V	N/V	N/V
darunter:		Leere Zelle			
(B) Bergbau und Gewinnung von Steinen und Erden	≤ 0,25	N/V	N/V	N/V	N/V
...	...	Leere Zelle	Leere Zelle	Leere Zelle	Leere Zelle
(S) Erbringung von sonstigen Dienstleistungen	≤ 0,25	N/V	N/V	N/V	N/V
Offene Stellen insgesamt	< 0,05	0,03	0,03	0,03	0,02
darunter:		Leere Zelle			
(A) Land- und Forstwirtschaft, Fischerei	-	0,22	0,22	0,22	0,13
(B) Bergbau und Gewinnung von Steinen und Erden	≤ 0,25	0,13	0,66	0,16	0,15
(C) Verarbeitendes Gewerbe	≤ 0,25	0,04	0,05	0,05	0,03
(D) Energieversorgung	≤ 0,25	0,10	0,13	0,14	0,10
(E) Wasser/Abfall/Umweltverschmutzungen	≤ 0,25	0,08	0,10	0,09	0,07
(F) Baugewerbe	≤ 0,25	0,12	0,11	0,12	0,07
(G) Handel; Instandh. und Reparatur von Fahrzeugen	≤ 0,25	0,11	0,13	0,10	0,07
(H) Verkehr und Lagerei	≤ 0,25	0,12	0,10	0,09	0,06
(I) Gastgewerbe	≤ 0,25	0,10	0,12	0,12	0,09
(J) Information und Kommunikation	≤ 0,25	0,15	0,10	0,13	0,08
(K) Finanz- und Versicherungsdienstleistungen	≤ 0,25	0,25	0,22	0,15	0,18
(L) Grundstücks- und Wohnungswesen	≤ 0,25	0,23	0,22	0,21	0,15
(M) Freiberufl., wissenschaftl. u. techn. Dienstleistungen	≤ 0,25	0,09	0,12	0,09	0,11
(N) Sonstigen wirtschaftlichen Dienstleistungen	≤ 0,25	0,09	0,09	0,11	0,07
(O) Öffentliche Verwaltung/Verteidigung/ Sozialvers.	≤ 0,25	0,08	0,10	0,13	0,08
(P) Erziehung und Unterricht	≤ 0,25	0,08	0,14	0,10	0,09
(Q) Gesundheits- und Sozialwesen	≤ 0,25	0,09	0,09	0,09	0,13
(R) Kunst, Unterhaltung und Erholung	≤ 0,25	0,08	0,08	0,08	0,05
(S) Erbringung von sonstigen Dienstleistungen	≤ 0,25	0,18	0,16	0,19	0,10

Anmerkung Derzeit besteht keine Lieferverpflichtung des Variationskoeffizienten für die Beschäftigung.

Quelle: IAB-Stellenerhebung, eigene Berechnungen für EU-Qualitätsbericht 2022. Präzisionsziele: Durchführungsrechtsakt zur vierteljährlichen Statistik der offenen Stellen (Entwurf) © IAB

Zum anderen sollen künftig diese Werte sowie die Werte für die Wirtschaftsabschnitte B bis S gemäß Wirtschaftszweigklassifikation WZ08 (bzw. Abschnitte B bis T gem. WZ25) bei einem zugrundeliegenden Variationskoeffizienten von über 25 Prozent als „unzuverlässig“ gekennzeichnet werden. Schon heute werden seitens des IAB als unsicher identifizierte Werte entsprechend geflaggt (d.h. gekennzeichnet). In wenigen Einzelfällen betraf dies in der jüngeren Vergangenheit (Corona) atypische Entwicklungsmuster der Beschäftigung bei oftmals sehr schwach besetzten Wirtschaftsabschnitten. Die Ursachen werden eingehend untersucht und spätestens im Rahmen der planmäßigen Revision entsprechend bei der Meldung berücksichtigt, so dass der dann gemeldete Wert in der Regel nicht mehr geflaggt ist.

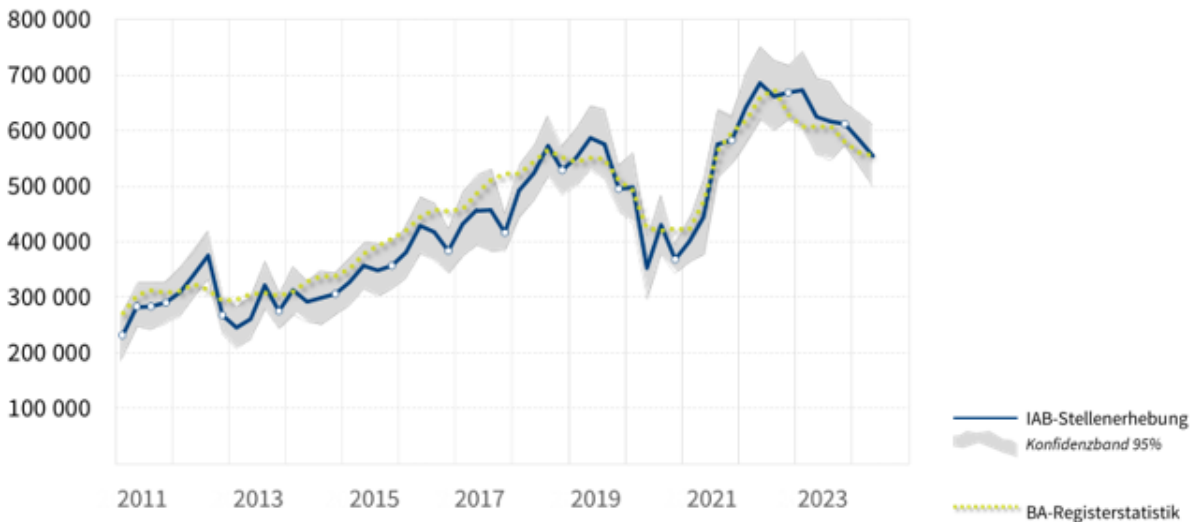
Ein zweites Kriterium für die Qualität der erhobenen Daten ist die Frage, wie gut die wahre Anzahl der offenen Stellen tatsächlich durch die IAB-Stellenerhebung getroffen wird. Zur Validierung der Umfragedaten – etwa durch einen Vergleich mit einer alternativen Datenquelle – kämen prinzipiell externe, z.B. prozessproduzierte Daten in Frage. Diese stehen allerdings für die Gesamtzahl der offenen Stellen nicht zur Verfügung. Verfügbar sind jedoch administrative Daten zu den der Bundesagentur für Arbeit gemeldeten offenen Stellen. Ein Abgleich der auf Basis der IAB-Stellenerhebung ermittelten Anzahl der gemeldeten offenen Stellen mit den administrativen Daten der Statistik der Bundesagentur für Arbeit bietet sich insbesondere vor dem Hintergrund an, da der GREG-Schätzer des aktuell verwendeten Hochrechnungsverfahrens die Hochrechnungsgewichte nicht (mehr) so anpasst, dass die hochgerechnete Zahl der gemeldeten offenen Stellen in einzelnen Stratifizierungszellen der Zahl der gemeldeten offenen Stellen auf Basis der BA Daten entspricht.² Dies ist relevant, da sich der Vergleich der Umfrage- mit den administrativen Daten für eine externe Validierung der Daten auf Basis der IAB-Stellenerhebung nur dann eignet, wenn eine Übereinstimmung nicht „per constructionem“ durch das Hochrechnungsverfahren herbeigeführt wird.

Aus diesem Grund wird neben der Zahl der offenen Stellen auch die Anzahl der von den Betrieben an die BA mit Vermittlungsauftrag gemeldeten Stellen erhoben und getrennt vom Erfassungssystem der BA hochgerechnet. Neben der externen Validierung der Daten wird hiermit auch die Anschlussfähigkeit an die administrativen Daten der BA dokumentiert. Konkret zeigt der Vergleich der auf Basis der IAB-Stellenerhebung erhobenen Daten zu den gemeldeten offenen Stellen mit denen der BA-Statistik, dass sich die im System der BA registrierte Zahl der mit Vermittlungsauftrag gemeldeten Stellen (ohne Stellen der Arbeitnehmerüberlassung, WZ08: 782 und 783) in der Regel im 95 Prozent-Konfidenzband der im Rahmen der IAB-Stellenerhebung erhobenen Zahl befindet (Abbildung 1).

² Stattdessen werden die Gewichte derzeit so kalibriert, dass die Anzahl der sozialversicherungspflichtig Beschäftigten sowie die Anzahl der Betriebe reproduziert werden können (vgl. Abschnitt 2.2.4).

Abbildung 1: Anzahl gemeldete offene Stellen ohne Arbeitnehmerüberlassung* im Vergleich zur BA-Registerstatistik

Erstes Quartal 2011 zweites Quartal 2024



Anmerkung: * Arbeitnehmerüberlassung gem. WZ08: 782 und 783. Konfidenzintervalle wurden auf dem 95 Prozent-Niveau berechnet.
Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

Grundsätzlich ergeben sich Abweichungen zwischen den Umfrage- und administrativen Daten aus methodischen Unterschieden in den beiden Erhebungen und insbesondere durch Besonderheiten des Meldeverhaltens im Wirtschaftsbereich „Arbeitnehmerüberlassung“.³ In allen Wirtschaftsbereichen mit Ausnahme der Arbeitnehmerüberlassung fallen die Abweichungen i.d.R. gering aus.

Bezieht man die Arbeitnehmerüberlassung mit ein, liegt die hochgerechnete Zahl der als gemeldet angegebenen Stellen aus der IAB-Stellenerhebung unter der Zahl der gemeldeten Stellen aus der BA-Registerstatistik (hier nicht abgebildet). Diese Abweichungen beruhen auf dem besonderem Rekrutierungsverhalten dieser Arbeitnehmerüberlassung.⁴ Stellenmeldungen aus der Arbeitnehmerüberlassung richten sich stärker auf erwartete Aufträge in der Zukunft. Den Arbeitsagenturen oder Jobcentern werden somit auch dann Vakanzen gemeldet, wenn dahinter keine aktuell zu besetzenden offenen Stellen stehen. Solche potenziellen Besetzungsbedarfe (für die Adresskartei des Betriebs aus der im Bedarfsfall schnell rekrutiert wird) werden statistisch ordnungsgemäß als Aufträge zur Arbeitsvermittlung registriert, decken sich aber nicht mit den

³ Weitere Erklärungen für potenzielle Abweichungen sind unterschiedliche Stichtagskonzepte, der aus der Befragung resultierende Stichprobenfehler und eine mögliche Zeitverzögerung bei der Abmeldung von Stellen aus dem Register der Bundesagentur für Arbeit.

⁴ Verschiedene gemeinsam mit der BA durchgeführte Analysen weisen auf diese Unterschiede hin (siehe unter anderem Qualitätsbericht der BA zur Statistik der gemeldeten Arbeitsstellen von 2018). Die Anschlussfähigkeit ist jedoch insbesondere zur Engpassanalyse der BA gewährleistet, da diese auf Basis der registerbasierten BA-Daten ohne die dort gemeldeten Stellen der Arbeitnehmerüberlassung erstellt wird.

Befragungsergebnissen aus der IAB-Stellenerhebung und entsprechen auch nicht der von Eurostat vorgegeben Definition einer offenen Stelle.⁵

Der Niveauabgleich der gemeldeten offenen Stellen (ohne Arbeitnehmerüberlassung) auf Basis der IAB-Stellenerhebung und der BA-Statistik wird seit einigen Jahren regelmäßig mit der Quartalspublikation der offenen Stellen von Seiten des IAB an die Arbeitsmarktberichterstattung der BA kommuniziert. Ebenso erfolgt sowohl für die Teilgruppe der sozialversicherungspflichtigen Beschäftigung und geringfügigen Beschäftigung sowie der gemeldeten Stellen ein interner Datenabgleich auf der Ebene der publizierten Wirtschaftsabschnitte und Betriebsgrößenklassen. Insgesamt ist auf Basis dieser externen Validierung eine hohe Passungsgüte der erhobenen Daten zu beobachten.

2.2.6 Datenlieferung an Eurostat und die IAB-Arbeitszeitrechnung

Das IAB hat die Europäische Kommission bei der Einführung einer europaweiten Statistik offener Stellen umfänglich beraten. Im Jahr 2009 ist eine europäische Verordnung in Kraft getreten, die alle Länder verpflichtet, ab 2010 quartalsweise Informationen über die Zahl der offenen Stellen und die Zahl der Beschäftigten getrennt nach Wirtschaftssektoren und Betriebsgrößen zu ermitteln und die Daten zu festgesetzten Terminen an die Europäische Kommission/EUROSTAT zu liefern. In Deutschland erfüllt das IAB als ‚Other National Authorities‘ (ONA’s) im Europäischen Statistischen System (ESS) mit der IAB-Stellenerhebung diese Aufgabe im Auftrag des BMAS bzw. der Bundesregierung. Sie ist derzeit die einzige Erhebung in Deutschland, die dies ermöglicht.

Auf die vierteljährlichen Daten über offene Stellen und die Gesamtbeschäftigung sind zusätzlich Saisonbereinigungsverfahren anzuwenden, welche im Artikel 5 und 7 des Beschlusses 1999/468/EG unter Beachtung von dessen Artikel 8 festgelegt sind. Das derzeit gültige Saisonbereinigungsverfahren ist JDemetra+, entsprechend der von Eurostat veröffentlichten ESS guidelines on seasonal adjustment und Eurostat (Eurostat, 2024a) i.V.m. dem Statistical requirements compendium (Eurostat, 2024c). Jede Saisonbereinigung erfordert eine ausreichend lange valide Zeitreihe der benötigten Information.

Die Berichtspflichten zu den verschiedenen Zeitreihen sowie der zugehörigen Saisonbereinigung leitet sich aus den EU-Verordnungen (EG) 453/2008, 1062/2008 und 19/2009 unmittelbar ab. In der Verordnung 19/2009 der Kommission zur Durchführung der Verordnung 453/2008 ist in Artikel 3, Absatz 1 geregelt, dass Deutschland seine Gesamtzahl der offenen und besetzten Stellen [Gesamtbeschäftigung] sowie die entsprechenden Metadaten innerhalb von 45 Tagen nach Ablauf des Referenzquartals an Eurostat

⁵ Eine „offene Stelle“ ist gemäß VERORDNUNG (EG) Nr. 453/2008 (Artikel 2 Nr. 1) eine neu geschaffene, nicht besetzte oder demnächst freiwerdende bezahlte Stelle, a) zu deren Besetzung der Betrieb aktive Schritte unternimmt, um eine geeignete Bewerberin bzw. einen geeigneten Bewerber außerhalb des betreffenden Unternehmens zu finden, und bereit ist, weitere Schritte zu unternehmen, und b) die der Betrieb sofort oder innerhalb eines bestimmten Zeitraums zu besetzen beabsichtigt. Inhaltlich ist der Punkt b) insbesondere bei Betrieben der Arbeitnehmerüberlassung mit ihrem eigenen Rekrutierungsverhalten kritisch zu betrachten.

Eine ausführliche Beschreibung der Unterschiede zwischen der Statistik der gemeldeten Stellen und den auf Basis der IAB-Stellenerhebung berichteten gemeldeten Stellen findet sich im Qualitätsbericht der BA zur Statistik der gemeldeten Arbeitsstellen (muk) sowie im Methodenbericht Umstellung der Statistik der gemeldeten Arbeitsstellen von (Hartmann und Reimer, 2010). Weil die Zahl der als gemeldet angegebenen Stellen aus der IAB-Erhebung von der Zahl der gemeldeten Stellen aus der BA-Registerstatistik abweicht, können die gemeldeten Stellen aus der BA-Registerstatistik nicht mit der inversen Meldequote zum gesamt-gesamtwirtschaftlichen Stellenangebot hochgerechnet werden.

übermittelt. Aufgrund dieser Berichtspflichten des IAB darf die Gefahr einer Nichterfüllung der Qualitätsansprüche an die Durchführung der Erhebung und an ihre Ergebnisse nicht bestehen

Um diese Verpflichtung zu erfüllen, muss das mit der Erhebung beauftragte Institut die benötigten Kern-Informationen in aggregierter und hochgerechneter Form dem IAB spätestens 36 Tage nach Quartalsende zur Weiterleitung zur Verfügung stellen. Der mit dem IAB erforderliche Abstimmungsprozess zur vorläufigen Abnahme der vierteljährlichen Hochrechnung des neuen Quartals (auf welcher die verschiedenen Datenprodukte für Eurostat aufbauen) findet innerhalb dieser 36 Tage statt. Die aggregierten Quartalsdaten sind dann sowohl als saisonbereinigte als auch als nicht-saisonbereinigte Zeitreihen spätestens 45 Tage nach Quartalsende durch das IAB über die EDAMIS-Schnittstelle an Eurostat zu liefern.

Eine weitere Lieferverpflichtung ergibt sich aus dem Qualitätsbericht, welcher der EU zum 30. Juni jeden Jahres zu der Erhebung des vorangegangenen Jahres vorgelegt werden muss. Die Einzelheiten der Anforderungen dieses Qualitätsberichts sind der EU-Verordnung 1062/2008 zu entnehmen. Unter anderem erfolgt hier derzeit eine Bewertung der systematischen Fehler der absoluten Zahl der offenen Stellen für die Gesamtzahl der offenen Stellen und nach Möglichkeit auf der Aggregationsebene der NACE Rev. 2 sowie nach Größenklassen (siehe Abschnitt 6.2 im jährlich aktualisierten Qualitätsberichts, zu finden unter folgendem Link: https://ec.europa.eu/eurostat/cache/metadata/EN/jvs_esqrs_de.htm).

Im Zuge von (derzeit laufenden) Novellierungen der EU-Vorordnungen sind zusätzliche oder geänderte Lieferverpflichtungen beziehungsweise Qualitätsstandards möglich. So wurden zunächst in der Version vom Oktober 2023 zur Neufassung des Durchführungsrechtsaktes zur Statistik der offenen Stellen (Item 5) erstmals explizite Präzisionsziele genannt. Derzeit befindet sich die Fassung in der Trilog-Abstimmung zwischen Rat, Kommission und Parlament. Der Vorschlag des Rates sieht eine Streichung der Präzisionsziele vor, während Kommission und Parlament weiterhin an einer Einführung der Ziele festhalten wollen. Konkret ist vorgesehen, dass mit dem jährlichen Qualitätsbericht die gemeldeten Werte für die Gesamtzahl an Beschäftigten beziehungsweise offenen Stellen einen zugrundeliegenden Variationskoeffizienten von 1 Prozent bzw. 5 Prozent nicht überschreiten. Für die NACE Rev. 2 Abschnitt B bis S (bzw. NACE Rev. 2.1 Abschnitt B bis T) wäre bei einem Variationskoeffizienten von über 25 Prozent der entsprechende Wert bei der Meldung mit *bzw. NACE Rev. 2.1 Abschnitt B bis T* als unzuverlässig zu markieren.

Weitere (institutsinterne) Anforderungen an die Datenlieferung ergeben sich im Status-Quo aufgrund von IAB-internen Lieferungen an die vierteljährliche Arbeitszeitrechnung. Die IAB-Arbeitszeitrechnung (AZR) ist das Schlüsselprodukt zu den geleisteten Arbeitsstunden in Deutschland und liegt den amtlichen Statistiken zum Arbeitseinsatz in den Volkswirtschaftlichen Gesamtrechnungen zugrunde. Um die Lieferverpflichtung des IAB zur Arbeitszeit gegenüber DESTATIS in hoher Qualität zu erfüllen, erhebt die IAB-Stellenerhebung seit dem zweiten Quartal 2013 in identischer Form verschiedene Quartalsinformationen zur Arbeitszeit (rund 5 Fragen). Die mittlerweile ausreichend lange Zeitreihe ermöglicht es der IAB-Arbeitszeitrechnung auch diese wichtigen Informationen für die Berechnungen zur Arbeitszeit in Deutschland zu nutzen.

3 Implementierung hochfrequenter Elemente

3.1 IAB-Betriebspanel

Das IAB-Betriebspanel kann als jährliche Befragung nur verzögert auf aktuelle Entwicklungen (z.B. die COVID-19-Pandemie oder den russischen Angriffskrieg gegen die Ukraine) reagieren. Um auch zeitnah aktuelle Entwicklungen abbilden zu können, müssten zwischen den jährlichen Erhebungen zusätzlich kurze Befragungen der Betriebe durchgeführt werden. In diesem Teilprojekt soll die Machbarkeit zusätzlicher Kurzbefragungen unter den Betrieben, die am IAB-Betriebspanel teilnehmen, untersucht werden. Solche Zusatzerhebungen ermöglichen es, aktuelle Themen schneller abzufragen und gleichzeitig die bereits aus der Haupterhebung vorhandenen Betriebsinformationen zu nutzen. Allerdings ist die Teilnahmebereitschaft der Betriebe an Zusatzerhebungen unklar. Des Weiteren erhöhen zusätzliche Erhebungen die Belastung der teilnehmenden Betriebe, was zu vermehrten Panelausstiegen führen könnte. Um diese Aspekte genauer zu erforschen, wurde in einem Experiment eine unterjährige Zusatzbefragung für eine zufällige Stichprobe von Betrieben des Betriebspanels durchgeführt. Hierbei müssen Betriebe, die erstmals 2023 am IAB-Betriebspanel teilnahmen von jenen unterschieden werden, die bereits mehrfach an der Erhebung teilgenommen haben. Aus datenschutzrechtlichen Gründen musste den wiederholt befragten Betrieben am Ende der Haupterhebung 2023 die Frage gestellt werden, ob sie bereit wären an einer Zusatzerhebung teilzunehmen. Nur die Betriebe welche einer Zusatzerhebung zugestimmt hatten wurde in das Experiment mit aufgenommen. Die erstbefragten Betriebe erhielten diese Frage nicht.

Das Ziel des Experimentes ist es daher, folgende Fragen zu beantworten:

1. Wie groß ist zum Zeitpunkt der Haupterhebung 2023 die Bereitschaft der wiederholt befragten Betriebe an der unterjährigen Zusatzbefragung teilzunehmen?
2. Wie hoch ist die tatsächliche Teilnahme der erst- und der wiederholt befragten Betriebe an der unterjährigen Zusatzbefragung?
3. Unterscheiden sich die teilnehmenden Betriebe der Zusatzerhebung systematisch von den Betrieben, die an der Haupterhebung 2023 teilnahmen?
4. Wie beeinflusst die zusätzliche unterjährige Kontaktaufnahme die Rücklaufquote der kontaktierten Betriebe (Erstbefragte und Wiederholer) in der Haupterhebung 2024?

Zum Zeitpunkt der Erstellung dieses Berichts können nur Antworten zu den Fragen eins bis drei gegeben werden, da die Haupterhebung 2024 noch nicht abgeschlossen ist. Ergebnisse zur vierten Frage werden im Laufe des Jahres 2025 vorliegen.

Auf Basis der Daten der unterjährigen Erhebung ist ein Forumsbeitrag zur Rekrutierung von Fachkräften aus dem Ausland als Veröffentlichungen geplant. Nach Rücksprache mit dem FDZ soll der Datensatz aus der einmaligen Erhebung wegen des vergleichsweise großen Aufwands nicht veröffentlicht werden.

3.1.1 Aufbau des Experiments

Um die grundsätzliche Machbarkeit unterjähriger Zusatzerhebungen zu testen, wurde in dem Experiment eine unterjährige Kurzbefragung zwischen den Haupterhebungen 2023 und 2024 durchgeführt (Abbildung 2). Die Beschränkung auf nur eine unterjährige Befragung erfolgte, um die Panelbetriebe nicht unnötig zu belasten.

Abbildung 2: Aufbau des Experiments

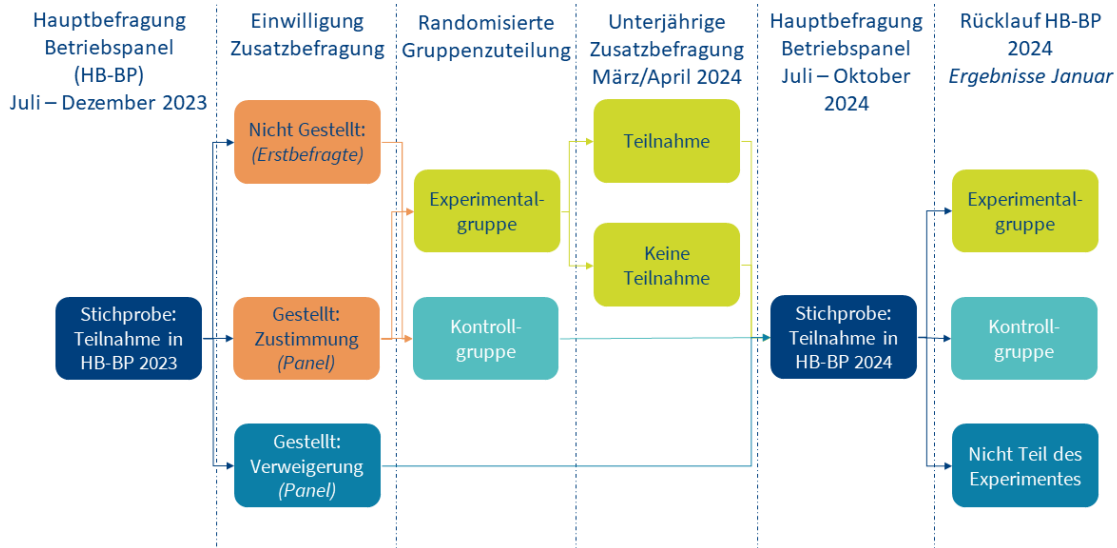


Quelle: IAB-Betriebspanel; eigene Darstellung. © IAB

Das Experiment basiert auf erst- und wiederholt befragten Betrieben, die an der Haupterhebung 2023 teilgenommen haben und mindestens eine sozialversicherungspflichtige Person beschäftigen. Zusätzlich waren wir aufgrund datenschutzrechtlicher Vorgaben verpflichtet, für die wiederholt befragten Betriebe eine Einwilligung für die unterjährige Befragung einzuholen. Daher beschränkt sich die Stichprobe der wiederholt befragten Betriebe auf jene, die ihre Einwilligung am Ende des Fragebogens der Haupterhebung 2023 gegeben haben (Abbildung 3).⁶ Der Aufbau des Experiments ist ansonsten für beide Teilgruppen identisch. Aus der Gruppe der erstbefragten Betriebe und der wiederholt Befragten mit Zustimmung wurden zufällig eine Experimental- und eine Kontrollgruppe gebildet (ausführliche Erläuterung der Stichprobenziehung in Abschnitt 1.2.2). Die unterjährige Zusatzbefragung erstreckte sich über den Zeitraum von zwei Monaten (März und April 2024). Für die Haupterhebung 2024 (Welle 32) wurden alle Betriebe kontaktiert, die an der Haupterhebung 2023 teilgenommen hatten, unabhängig von ihrer Teilnahme an der unterjährigen Befragung.

⁶ Diese Filterfrage kann zu Selektionseffekten in den Analysen, für die Teilgruppe der wiederholt befragten Betriebe, führen.

Abbildung 3: Detaillierter Ablauf des Experiments



Quelle: IAB-Betriebspanel; eigene Darstellung. © IAB

Erhebungsmodus

Die Entscheidung für ein Erhebungsdesign beruht auf der Abwägung von verschiedenen Faktoren (Rücklaufquote, Kosten, zukünftige Umsetzung und Kompatibilität mit dem Experiment), die im Folgenden näher erläutert werden. Für die unterjährige Befragung kommen drei computergestützte Erhebungsdesigns infrage:

1. ein monomodales Erhebungsdesign mit persönlich-mündlichen (CAPI) Interviews oder Interviews per Telefon (CATI)
2. ein monomodales Erhebungsdesign mit selbstadministrierten Interviews durch einen Online-Fragebogen oder
3. ein sequenzielles Erhebungsdesign, das beide Erhebungsformen kombiniert und zunächst eine Beantwortung des Fragebogens online ermöglicht und erst, wenn der Fragebogen nach einer gewissen Zeit nicht selbstständig ausfüllt wurde, zu einem Interviewer gestützten Modus wechselt.

Interviews, die persönlich durch geschulte Interviewerinnen und Interviewer durchgeführt werden (1) sind relativ teuer, dafür ist erfahrungsgemäß die Antwortbereitschaft der Betriebe relativ hoch (Gleiser et al, 2022). Neben den Fixkosten für die Programmierung des Fragebogens müssen die interviewenden Personen bezahlt und die Fahrtkosten zu den Betrieben erstattet werden. Bei Interviews durch Interviewende im Telefonstudio entfallen die Fahrtkosten. Außerdem können in relativ kurzer Zeit viele Betriebe befragt werden, weil die Fahrtzeiten zu den Betrieben entfallen.

Bei selbstadministrierten Interviews (2) fallen neben den Fixkosten nur relativ niedrige Zusatzkosten an. Dieser Erhebungsmodus ist somit kostengünstiger als andere Erhebungsarten (Gleiser et al, 2022). Allerdings muss den befragten Personen mehr Zeit für die Beantwortung der Fragen eingeräumt und

mehrfach an die Befragung erinnert werden, um nicht zu hohe Ausfälle zu generieren. Zudem sind die erwarteten Rückläufe von 10 bis 15 Prozent bei Online-Befragungen zu gering, um eine für dieses Experiment ausreichende Anzahl an Interviews zu bekommen.

Sequenzielle Erhebungsdesigns (3) sind insgesamt kostengünstiger, da der Fragebogen online ohne Unterstützung ausgefüllt werden kann. Nur die Betriebe, die nicht selbständig antworten, werden nach einer gewissen Zeit von Interviewenden kontaktiert, um den Rücklauf zu erhöhen.

Für das Experiment ist es erforderlich, alle Betriebe im gleichen Erhebungsmodus zu befragen, um nicht unkontrollierte Einflüsse auf die Teilnahmebereitschaft der Betriebe zu bekommen. Dies schließt ein sequenzielles Erhebungsdesign aus. Zudem sollte die unterjährige Erhebung innerhalb eines relativ kurzen Zeitraums von 2 Monaten durchgeführt und eine möglichst hohe Rücklaufquote erreicht werden, was mit einer Online-Befragung nicht erreicht werden kann. Aus den genannten Gründen kommen nur telefonische Interviews (CATI) durch geschulte Interviewende in einem Telefonstudio in Frage.

3.1.2 Fragebogen

Der Fragebogen einer Zusatzbefragung, die aktuelle Entwicklungen aufgreifen möchte, sollte kurz sein und ohne große betriebsinterne Recherche oder einen Wechsel der Kontaktperson am Telefon beantwortet werden können. Darüber hinaus muss der Fragebogen die Fragen zu den offenen Stellen der IAB-Stellenerhebung enthalten, da es Teil des Experiments ist, zu überprüfen, ob sie in einer unterjährigen Zusatzerhebung abgefragt werden können.

Um eine realistische und gleichzeitig intern valide Experimentalsituation zu schaffen, sollte der Fragebogen folgende Bedingungen erfüllen:

Erstens soll die Befragung durchschnittlich ca. 10 Minuten dauern und dabei keine große Varianz der Befragungsdauer für Untergruppen von Betrieben aufweisen damit der Einfluss der Kurzbefragung auf alle Untergruppen möglichst gleich ist. Für die Gestaltung des Fragebogens bedeutet das, dass möglichst keine Fragen gestellt werden sollten, die für einen größeren Teil der Betriebe nicht zutreffen.

Zweitens sollte die Abfrage von konkreten Zahlenwerten (z.B. Anzahl der Beschäftigten) auf ein Minimum begrenzt werden, ausgenommen die Abfrage der offenen Stellen. Dies hat den Zweck, zum einen die Belastung der Teilnehmenden zu minimieren und zum anderen eine realistische Durchführung der Befragung zu gewährleisten, die einen Fokus auf die Einschätzungen der Betriebe zu aktuellen Geschehnissen vorsieht.

Drittens müssen die Fragen einfach und kurz gehalten werden, damit sie am Telefon beantwortet werden können. Damit soll die Belastung für die Teilnehmenden geringgehalten werden.

Der finale Fragebogen enthält 16 inhaltliche Fragen und die datenschutzrechtlich notwendige Frage, ob der Betrieb mit einer Verknüpfung der Daten der unterjährigen Erhebung mit Daten, die im IAB vorliegen, einverstanden ist. Darüber hinaus werden einige Informationen zur befragten Person erhoben. Es werden vier Fragen zu den offenen Stellen im Betrieb, neun Fragen zu der (Nicht-) Beschäftigung von ausländischen Arbeitskräften und drei Fragen zum Mindestlohn gestellt. Die Fragen

den offenen Stellen orientieren sich am Fragebogen der IAB-Stellenerhebung, während die restlichen Fragen neu entwickelt wurden bzw. aus früheren Erhebungen des IAB-Betriebspanels stammen.

Die Befragung dauerte im Durchschnitt 5,5 Minuten und war damit deutlich kürzer als die vom Umfrageinstitut anhand von Pretests und Berechnungen angenommene Zeit. Die Standardabweichung von 1,9 Minuten zeigt, dass die Dauer der Interviews für alle Betriebe relativ gleich lang war. Die mittlere Interviewdauer unterscheidet sich nicht zwischen wiederholt befragten und erstbefragten Betrieben.

3.1.3 Stichprobe und Fallzahlen

Berechnung der Stichprobengröße

Die Berechnung der Mindestgröße der Stichprobe erfolgt mit Hilfe sogenannter Power-Analysen. Um sicherzustellen, dass der Einfluss der unterjährigen Kontaktaufnahme auf die Rücklaufquote in der Haupterhebung 2024 (HE24) mit einer ausreichenden Sicherheit gemessen werden kann und die Daten der unterjährigen Befragung statistisch belastbar sind, muss die Stichprobe pro Gruppe eine bestimmte Mindestgröße erreichen. Zu diesem Zweck wurde eine Power-Analyse durchgeführt, die die Mindestgröße der Stichprobe bestimmt. Um aussagekräftige Analysen zu ermöglichen, mussten vor der Berechnung der Mindeststichprobengröße verschiedene Parameter festgelegt werden:

1. **Kleinster identifizierbarer Effekt:** Ziel der Forschungsfrage 4 ist es, Unterschiede in den Rücklaufquoten zwischen der Kontroll- und Experimentalgruppe festzustellen. Der kleinste identifizierbare Effekt gibt an, welche Differenz der Rücklaufquoten als bedeutsamer Unterschied interpretiert werden kann. Je kleiner der Effekt (die Differenz der Rücklaufquoten), desto größer muss die Stichprobengröße sein. Berechnet wird diese als die Differenz der (angenommenen) Rücklaufquoten, geteilt durch die gepoolte Standardabweichung.
2. **Power Level:** Die Wahrscheinlichkeit, die Nullhypothese abzulehnen, wenn die Alternativhypothese zutrifft, muss festgelegt werden. Üblicherweise wird ein Wert von 0,8 oder 0,9 angesetzt.
3. **Signifikanz Level:** Für die Wahrscheinlichkeit, die Nullhypothese abzulehnen, obwohl sie wahr ist, wird üblicherweise ein Wert von 0,05 verwendet.

Aus der Forschungsfrage 4 ergeben sich folgende Hypothesen:

Ho: Rücklaufquoten der Kontrollgruppe in HE 24 = Rücklaufquoten der Experimentalgruppe (Kontaktversuch) in HE 24

H1: Rücklaufquoten der Kontrollgruppe in HE 24 $\neq$ Rücklaufquoten der Experimentalgruppe (Kontaktversuch) in HE 24

Um die erforderliche Fallzahl zu schätzen, müssen Annahmen über die Rücklaufquoten der Kontroll- und Experimentalgruppe getroffen werden. Diese Annahmen basieren auf den Rücklaufquoten der BeCovid-Studie, einer monatlichen Betriebsbefragung, die ausschließlich telefonisch durchgeführt wurde (Backhaus et al, 2022).

Für die wiederholt befragten Betriebe gehen wir von einer Differenz zwischen den Rücklaufquoten der Experimental- und Kontrollgruppe von fünf Prozentpunkten aus. Genauer, wir vermuten, dass die

zusätzliche Kontaktierung die Rücklaufquote um fünf Prozentpunkte verringern wird. Zudem wird angenommen, dass die Rücklaufquote der Haupterhebung 2024 für die Kontrollgruppe bei 75 Prozent und für die Experimentalgruppe bei 70 Prozent liegt. Dies ergibt einen kleinsten identifizierbaren Effekt von 0,112. Das Power-Level wurde auf 0,8 und das Signifikanzniveau auf 0,05 festgelegt, woraus sich eine erforderliche Stichprobengröße von 1.250 Betrieben pro Gruppe für die unterjährige Befragung ergibt.

Auch bei den Erstbefragten wird von einer Differenz von 5 Prozentpunkten ausgegangen. Bei den Erstbefragten wird jedoch von einer Rücklaufquote von 60 Prozent für die Kontrollgruppe und 55 Prozent für die Experimentalgruppe ausgegangen, da diese Betriebe noch keine starke Bindung zum Panel haben. Bei Annahme des gleichen Power-Levels und Signifikanzniveaus wie bei den wiederholt befragten Betrieben ergibt die Power-Analyse eine Stichprobengröße von mindestens 1.533 Betrieben pro Gruppe für die unterjährigen Befragung.

Stichprobenziehung

Die Experimental- und die Kontrollgruppe wurden mit einem sogenannten systematischen Random-Sampling-Verfahren gezogen. Dafür wurde ein Datensatz aus den erst- und wiederholt befragten Betrieben des IAB-Betriebspanels 2023 erstellt, die für das Experiment angesprochen werden dürfen. Zunächst wurde jedem Betrieb eine Zufallszahl zugeordnet und gespeichert. Die Zufallsvariable dient dazu, Betriebe, die sich in allen anderen Variablen gleichen, zufällig anzuordnen, um eine zufällige Auswahl der Gruppen zu gewährleisten. Als nächstes wurden die Betriebe nach dem Erhebungsmodus bei der Hauptbefragung, dem Erhebungspfad (d. h. der Abfolge der angebotenen Erhebungsmodi), der Branche, der Größenklasse, dem Bundesland und der Zufallsvariable geordnet. Um die berechnete Stichprobengröße sicher zu erreichen, müssen die Mindeststichprobenumfänge von Kontroll- und Experimentalgruppen gleich groß sein. Jeder zweite Betrieb im Datensatz wurde hierfür der Experimentalgruppe zugewiesen, was zu 2.061 Fällen bei den wiederholt befragten Betrieben und 1.902 Fällen bei den Erstbefragten führte.

Ausfälle und finale Fallzahlen

Eine CATI-Befragung setzt voraus, dass für alle zu kontaktierenden Betriebe eine Telefonnummer bekannt ist. Da in der Haupterhebung des IAB-Betriebspanels verschiedene Erhebungsmodi zum Einsatz kommen, liegen nicht für alle Betriebe aktuelle Telefonnummern vor. So hatten 63 Betriebe in der gezogenen Stichprobe keine Telefonnummer (Tabelle 3). Daher wurde im Vorfeld eine Telefonnummernrecherche für diese Betriebe durchgeführt. Trotz der Recherche konnten für 7 Betriebe keine aktuelle Telefonnummer gefunden werden. Diese Fälle mussten aus der Stichprobe entfernt werden, wobei ein Fall auf die wiederholt befragten Betriebe (nun 2.060) und sechs Fälle auf die Erstbefragten (nun 1.896) entfielen. Zudem mussten 4 Fälle aus der Stichprobe genommen werden, weil die Telefonnummern doppelt in der Stichprobe vorkamen, sodass die Stichprobe der wiederholt befragten Betriebe sich auf 2.056 reduzierte.

Weitere Betriebe mussten aus der Stichprobe ausgeschlossen werden, weil sie auf der ADM-Telefonsperrliste standen. Dies ist eine Liste, auf die sich Betriebe setzen lassen können, wenn sie

grundsätzlich an keiner Telefonbefragung teilnehmen möchten. Da die unterjährige Befragung ausschließlich telefonisch durchgeführt wurde, durften Betriebe, die auf dieser Liste stehen, nicht kontaktiert werden. Dies führte zur Sperrung von 58 Fällen in der Erstbefragten-Stichprobe (nun 1.838). Die wiederholt befragten Betriebe, die der Zusatzbefragung zugestimmt und als bevorzugten Erhebungsmodus entweder Telefon oder keine Präferenz gewählt haben, wurden trotz Sperrliste angeschrieben und telefonisch kontaktiert, da sie explizit der telefonischen Befragung zugestimmt hatten. Deshalb verringerte sich die Stichprobe der wiederholt befragten Betriebe nur um 21 Fälle auf 2.039.

Tabelle 3: Aufschlüsselung der Fallzahlen für die Experimentalgruppe

Leere Zelle	Wiederholt befragte Betriebe	Erstbefragte Betriebe	Gesamt
Haupterhebung	10.887	3.955	14.842
Ungültige Fälle	247	152	399
Betriebe ohne SVB	425	0	425
Keine Zustimmung	6.093	0	6.093
Verbleibende Betriebe	4.122	3.803	7.925
Gezogene Stichprobe	2.061	1.902	3.963
Kein Ergebnis bei Telefonnummernrecherche	1	6	7
ADM-Sperrliste	21	58	79
Dublette bei Telefonnummern	4	0	4
Anzurufende Stichprobe	2.035	1.838	3.873

Quelle: IAB-Betriebspanel; eigene Berechnungen. © IAB

3.1.4 Teilnahmebereitschaft an der unterjährigen Befragung und Selektivität

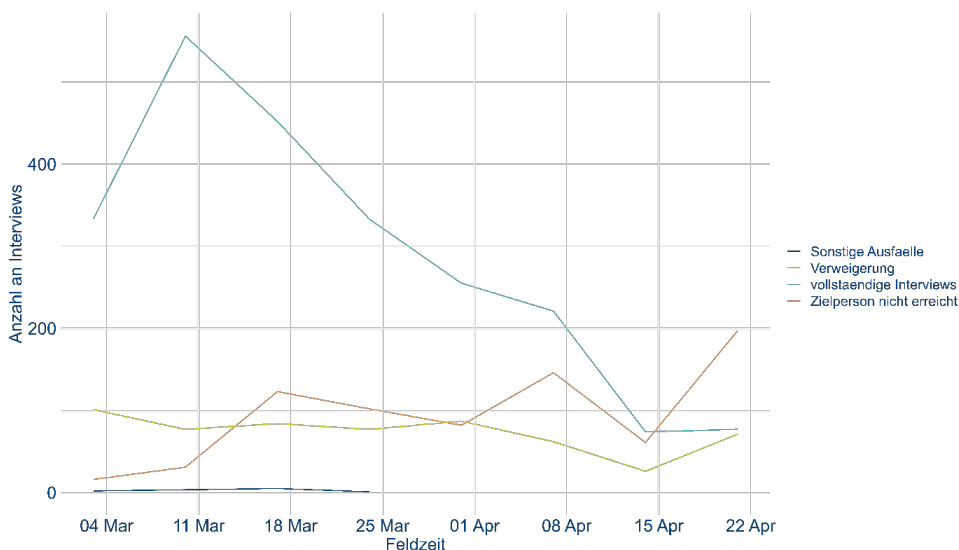
Teilnahmebereitschaft und präferierter Erhebungsmodus

Die wiederholt befragten Betriebe wurden – wie bereits erwähnt - gefragt, ob sie bereit wären, an einer unterjährigen Befragung teilzunehmen und welchen Erhebungsmodus sie bevorzugen würden. Von den 10.887 wiederholt befragten Betrieben haben 39,34 Prozent einer unterjährigen Befragung zugestimmt. Von diesen Befragten entschieden sich 53,33 Prozent für eine Online-Befragung, 33,57 Prozent für eine Telefonbefragung, und 11,07 Prozent hatten keine Präferenz für einen bestimmten Erhebungsmodus.

Rücklaufquoten

Die Feldphase begann am 06.03.2024 mit einem Anschreiben, das an alle Betriebe in der Experimentalgruppe versendet wurde. Abbildung 4 zeigt den Verlauf der Feldphase sowie die Anzahl der Verweigerungen und sonstigen Ausfälle. In den ersten beiden Wochen konnten viele vollständige Interviews geführt werden. In telefonisch gut zu erreichenden und kooperativen Betrieben können relativ schnell Interviews geführt werden. Bei den telefonisch schlecht zu erreichenden Betrieben dauert es länger bis ein Interview zu Stande kommt, was an der fallenden Linie der vollständigen Interviews erkennbar ist. Auch die zunehmende Anzahl nicht erreichbarer Betriebe am Ende der Feldzeit erklärt den Rückgang der Anzahl der vollständigen Interviews. Die Anzahl der Verweigerungen bleibt relativ konstant.

Abbildung 4: Rücklauf vollständiger Interviews und Ausfälle während der Feldzeit



Quelle: IAB-Betriebspanel; eigene Darstellung. © IAB

Tabelle 4 enthält eine detaillierte Aufschlüsselung der unterschiedlichen Ausfälle von Betrieben. Sogenannte neutrale Ausfälle werden dabei separat aufgeführt – sie beschreiben Ausfälle, die durch Umstände außerhalb des Betriebs verursacht sind (z. B. keine oder falsche Telefonnummer etc.). Diese Fälle werden nicht in die Berechnung der Rücklaufquote einbezogen.

Als nicht neutrale Ausfälle werden hier die „Zielperson ist nicht erreichbar“, die „Verweigerung der Teilnahme am Interview“ und „sonstige Ausfallgründe“ unterschieden. Diese Ausfälle werden in der Berechnung der Rücklaufquote berücksichtigt. Zudem sind die Rücklaufquoten der gesamten Stichprobe sowie die der einzelnen Teilstichproben angegeben. Für die gesamte Stichprobe beträgt die Rücklaufquote 63 Prozent, wobei die wiederholt befragten Betriebe eine Rücklaufquote von 68 Prozent und die Erstbefragten eine Rücklaufquote von 57 Prozent erzielten.

Tabelle 4: Ausfälle in der unterjährigen Befragung

	Wiederholt befragte Betriebe		Erstbefragte Betriebe		Gesamt	
A Gezogene Stichprobe	2.061	100 %	1.902	100 %	3.963	100 %
B Neutrale Ausfälle	118	6 %	185	10 %	303	8 %
Keine Telefonnummer	1	0 %	6	0 %	7	0 %
Sperrliste	22	1 %	58	3 %	80	2 %
Dublette bei Telefonnummer	4	0 %	0	0 %	4	0 %
Falsche Telefonnummer	74	4 %	102	5 %	176	4 %
Betrieb erloschen	17	1 %	19	1 %	36	1 %
C Bereinigtes Brutto	1.943	100 %	1.717	100 %	3.660	100 %
Zielperson nicht erreicht	340	17 %	418	24 %	758	21 %
Verweigerung	276	14 %	317	18 %	593	16 %
Sonstige Ausfälle	2	0 %	6	0 %	8	0 %
D Ausfälle gesamt (bezogen auf das bereinigte Brutto)	618	32 %	741	43 %	1.359	37 %
E Interviews insgesamt (bezogen auf das bereinigte Brutto)	1.325	68 %	976	57 %	2.301	63 %

Quelle: IAB-Betriebspanel; eigene Berechnungen. © IAB

Analyse der Ausfälle

Die Teilnahme an der unterjährigen Befragung ist freiwillig, sodass Betriebe selbst entscheiden können, ob sie an der Zusatzbefragung teilnehmen möchten oder nicht. Falls die Ausfälle in der Befragung systematisch sind, d.h. sich die teilnehmenden Betriebe von den nicht-teilnehmenden signifikant unterscheiden, entsteht eine Selektion, die zu einer Verzerrung der Stichprobe der unterjährigen Befragung führen kann. Im nächsten Schritt untersuchen wir daher Selektionseffekte mithilfe einer Nonresponse-Bias-Analyse (Gleiser et al. 2022). Ziel dieser Analyse ist es, festzustellen, ob sich die teilnehmenden Betriebe in den jeweiligen Stichprobengruppen stark von den ursprünglich gezogenen Betrieben unterscheiden. Hierfür wird der "Average Absolute Relative Nonresponse Bias" (AARNB) verwendet. Der AARNB beschreibt die durchschnittliche relative Differenz der gezogenen Stichprobe und der teilnehmenden Stichprobe in der zu betrachteten Variable. Hierfür wird zunächst der Anteil der gezogenen Stichprobe ($\bar{Y}_{c,n}$) und der teilnehmenden Stichprobe ($\bar{Y}_{c,r}$) pro Ausprägung (c) berechnet. Die relative Differenz pro Ausprägung wird berechnet, indem der Anteil der teilnehmenden Stichprobe von dem Anteil der gezogenen Stichprobe subtrahiert wird und dieser Wert durch den Anteil der gezogenen Stichprobe geteilt wird. Die Summe aller relativen Differenzen einer Variablen werden summiert und durch die Anzahl der Ausprägungen geteilt.

$$AARNB = \frac{\sum_{c=1}^C \left| \frac{\bar{Y}_{c,r} - \bar{Y}_{c,n}}{\bar{Y}_{c,n}} \right|}{C}$$

Tabelle 5 zeigt die AARNB-Werte für die wiederholt Befragten und die Erstbefragten in Bezug auf drei Variablen: Größenklasse, Branche und Region. Es ist festzustellen, dass die Gruppe der teilnehmenden wiederholt befragten Betriebe im Vergleich zur gezogenen Stichprobe nur geringe bis moderate Unterschiede hinsichtlich der betrachteten Variablen aufweist. Als gering bis moderat gelten dabei Differenzen, die unter 15 Prozent liegen. Der größte durchschnittliche Unterschied ist bei der Branchenvariable zu erkennen.⁷ Die Gruppe der teilnehmenden Erstbefragten zeigt dagegen einen moderaten bis starken Unterschied (über 15 Prozent). Auch hier ist der größte Unterschied in der Branchenvariable zu beobachten.

Tabelle 5: Average Absolute Relative Nonresponse Bias

Variable	Wiederholt befragte Betriebe	Erstbefragte Betriebe
Größenklasse	3,57 %	11,83 %
Branche	7,66 %	16,86 %
Westdeutschland/Ostdeutschland	0,38 %	4,39 %

Quelle: IAB-Betriebspanel; eigene Berechnungen. © IAB

3.1.5 Auswertungen auf Basis der unterjährigen Erhebung

Insgesamt liegen 2.300 gültige Interviews für inhaltliche Analysen vor. Die Daten der unterjährigen Erhebung sind seit Juli verfügbar und können mit den mitgelieferten Hochrechnungsfaktoren auch betriebs- und beschäftigungsproportional hochgerechnet werden. Der Hochrechnungsfaktor korrigiert die disproportionale Stichprobenziehung, die dazu führt, dass Großbetriebe überrepräsentiert sind.

Bei der Prüfung des Datensatzes ist aufgefallen, dass die hochgerechnete Gesamtzahl der Beschäftigten der unterjährigen Erhebung mit rund 48 Mio. stark von der Gesamtzahl der Beschäftigten in der BA-Betriebsdatei von knapp 43 Mio. abweicht. Zwar handelt es sich bei der Beschäftigungszahl der Haupterhebung um den Vorjahreswert, allerdings ist die Abweichung zwischen den beiden Angaben zu hoch, als dass sie sich durch eine Veränderung zwischen zwei Erhebungszeiten erklären ließe.

Im Unterschied zur Haupterhebung wurde im Interview der unterjährigen Erhebung die Angabe der befragten Person zur Beschäftigtenzahl ohne weitere Prüfungen oder Nachfrage übernommen. Da es sich um Panelbetriebe handelt, die bereits im Vorjahr interviewt wurden, war die Annahme, dass die Angaben weitgehend übereinstimmen sollten. Auch die Anzahl der sozialversicherungspflichtigen Beschäftigten wurde nicht erneut erfragt. Der Abgleich der Angaben zur Anzahl der Sozialversicherungspflichtigen mit denen der BA-Betriebsdatei ist wichtig, da die BA-Betriebsdatei wiederum Basis der Stichprobenziehung und Hochrechnung ist. In der Haupterhebung wird deshalb durch die Anzahl der sozialversicherungspflichtigen Beschäftigten die Betriebseinheit oder

⁷ Welche Kategorien hierbei den größten Unterschied aufweisen, kann durch den berechneten Durchschnitt nicht erschlossen werden. Hierzu dient die absolute Differenz der Anteile an den Kategorien zwischen den teilnehmenden Betrieben und der gezogenen Stichprobe.

Betriebsstätte identifiziert, die für die entsprechende Betriebsnummer in der BA-Betriebsdatei erfasst ist.

Um besser zu verstehen, wie diese Abweichung zustande kommt und welche Faktoren dahinterliegen, wurden die Informationen aus der Haupterhebung und die Angaben aus der unterjährigen Erhebung auf Betriebsebene zusammengespielt und die Differenz zwischen der Angaben zur Beschäftigtenzahl in der unterjährigen Erhebung und den Angaben in der Haupterhebung berechnet. Bei der Auswertung wurde zwischen Betrieben bis zu 20 Beschäftigten (1.260 Betriebe) und Betrieben mehr als 20 Beschäftigten (1.040 Betriebe) unterschieden. Bei den kleineren Betrieben wurden die absoluten Differenzen berechnet, bei den Betrieben mit mehr als 20 Beschäftigten prozentuale Anteile. Bei kleinen Betrieben würden bereits geringe absolute Abweichung zu hohen prozentualen Veränderungen führen. Umgekehrt sind absolute Differenzen bei großen Betrieben schwierig zu interpretieren, wenn man die Betriebsgröße nicht mit betrachtet.

Tabelle 6 zeigt die absoluten Abweichungen zwischen den Angaben der unterjährigen Erhebung und den Angaben im IAB-Betriebspanel 2023 für die kleinen Betriebe nach Perzentilen. Dargestellt sind die Beträge der Abweichung, d.h. es wird nicht zwischen positiven und negativen Abweichungen differenziert. Man sieht, dass 50 Prozent der Betriebe eine Abweichung von einem Beschäftigten oder weniger im Vergleich zum Vorjahreswert haben. Kleinere Abweichungen von Jahr zu Jahr sind plausibel, da Betriebe wachsen und schrumpfen können. Lediglich 10 Prozent der Betriebe weisen eine Abweichung von 4 Beschäftigten oder mehr auf.

Tabelle 6: Abweichungen bei der Anzahl Beschäftigten für Betriebe mit ≤ 20 Beschäftigten in 2023 (N=1.260). Absolute Werte

Quantil	Abweichung
5 %	0
25 %	0
50 %	1
75 %	2
90 %	4
95 %	8

Quelle: IAB-Betriebspanel; eigene Berechnungen. © IAB

In Tabelle 7 sind die prozentualen Abweichungen für Betriebe mit mehr als 20 Beschäftigten dargestellt. Die Angaben hier beziehen sich auf die Abweichung zum Vorjahreswert in Prozent. Die prozentualen Veränderungen sind hier aussagekräftiger, da beispielsweise bei einem Betrieb mit 2.000 Beschäftigten eine Abweichung von 10 Beschäftigten weniger relevant wäre als die gleiche Veränderung bei einem Betrieb mit 200 Beschäftigten. Hier weisen mehr als 50 Prozent der Betriebe eine Abweichung zum Vorjahreswert von 7,1 Prozent oder mehr auf, 25 Prozent der Betriebe eine Abweichung von 15,9 Prozent oder mehr und 5 Prozent sogar Abweichungen von über 100 Prozent.

Zum Vergleich: Bei der Haupterhebung des IAB-Betriebspanels liegt der Grenzwert für Abweichungen bei der Anzahl der Beschäftigten bei Betrieben mit mehr als 20 Beschäftigten, der ohne Nachfrage beim Betrieb akzeptiert wird, bei +/-5 Prozent. Sind die Abweichungen größer wird bei dem Betrieb nachgefragt und versucht eine Klärung herbeizuführen. Dies bedeutet, von den 1.036 Betrieben mit mehr als 20 Beschäftigten in der unterjährigen Erhebung (Tabelle 2) wären in der Haupterhebung etwa die Hälfte der Fälle im Rahmen der Datenprüfung überprüft worden.

Tabelle 7: Abweichungen bei den Beschäftigten für Betriebe mit >20 Beschäftigten in 2023 (N=1.040). Werte in Prozent

Quantil	Abweichung
5 %	0
25 %	2,9
50 %	7,1
75 %	15,9
95 %	104,9

Quelle: IAB-Betriebspanel; eigene Berechnungen. © IAB

Die Abweichungen zwischen den Angaben zu der Gesamtzahl der Beschäftigten 2023 in der Haupterhebung und den Angaben hierzu in der unterjährigen Erhebung sind so erheblich, dass sie nicht den üblichen Veränderungen (Wachstum und Schrumpfung von Betrieben) zwischen zwei Zeitpunkten entsprechen. Im Folgenden versuchen wir möglichen Erklärungen nachzugehen. Ein Ansatz wäre, dass die befragte Person am Telefon nur mehr oder weniger genaue bzw. gerundete Schätzungen für die Betriebsgröße angibt. Ein Indikator hierfür ist, dass sie Angaben zu den Beschäftigtenzahlen gemacht haben, die durch 5 oder 10 teilbar sind.

Insgesamt sind bei 447 Betrieben, also 19,5 Prozent, die Angaben zur Anzahl der Beschäftigten durch 10 teilbar und bei 753 Betrieben bzw. 32,9 Prozent sind die Angaben durch 5 teilbar. Zum Vergleich sind beim Vorjahreswert, der aus der Haupterhebung stammt, nur bei 197 Betrieben bzw. 8,6 Prozent die Angaben durch 10 teilbar und bei 429 Betrieben bzw. 18,7 Prozent durch 5 teilbar. Dies legt die Vermutung nahe, dass Betriebe ungenaue oder gerundete Angaben machen, wenn sie am Telefon ohne Hinweis auf die Angaben in der BA-Statistik befragt werden.

Aus der BA-Statistik werden die Angaben zur Anzahl der Beschäftigten für eine Betriebsnummer aggregiert. Dabei handelt es sich um eine regionale Betriebsstätte, die beispielsweise eine Filiale oder Niederlassung sein kann. Eine mögliche Erklärung, warum die am Telefon berichtete Gesamtzahl der Beschäftigten bei der unterjährigen Erhebung nicht mit der aus dem Vorjahr übereinstimmt, ist, dass die Betriebe für eine andere Betriebseinheit also z. B. für das ganze Unternehmen berichten. Das würde einen Hinweis darauf liefern, warum die Gesamtzahl der Beschäftigten deutlich über der der Haupterhebung 2023 liegt. Um das zu überprüfen, haben wir die Anzahl der Betriebe, die letztes Jahr angaben, eine Filiale oder Niederlassung zu sein, ausgewertet. Es fällt auf, dass es sich bei Betrieben, die

in der unterjährigen Erhebung größere (>5 %) prozentuale Abweichungen haben, deutlich häufiger (+8%-Punkte) um Niederlassung lt. Haupterhebung handelt als im Durchschnitt der Stichprobe, was für diese Vermutung spricht.

3.1.6 Fazit auf Basis der bisherigen Ergebnisse

Mit dem hier beschriebenen Experiment sollte getestet werden, ob die Betriebe des IAB-Betriebspanels 2023 (erst und wiederholt befragte Betriebe) bereit sind an einer zusätzlichen, kurzen telefonischen Erhebung zwischen den Haupterhebungen teilzunehmen. Für die wiederholt befragten Betriebe aus 2023 wurde bereits in der Haupterhebung 2023 ihre Zustimmung zu einer zusätzlichen Befragung eingeholt. Die Bereitschaft zur Teilnahme an der unterjährigen Erhebung war mit etwa 40 Prozent vergleichsweise gering.

Aus der Gruppe der Betriebe, die ihre Zustimmung gegeben hatten, sowie aus der Gruppe der erstbefragten Betriebe wurde eine zufällige Stichprobe für die unterjährige Erhebung gezogen. Insgesamt wurden knapp 4.000 Betriebe für die unterjährige Erhebung kontaktiert. Die Teilnahmebereitschaft der kontaktierten Betriebe lag mit einer Rücklaufquote von durchschnittlich 63 Prozent im erwarteten Bereich, welcher sich aus Rücklaufquoten des BeCovid Panels ergibt. Der Rücklauf reichte aus, um die Zielgröße von 2.300 auswertbaren Interviews zu erreichen.

Die Auswertung der Daten zeigt, dass die hochgerechnete Anzahl der Beschäftigten stark von dem Vorjahreswert aus der Haupterhebung nach oben abweicht. Offensichtlich berichteten einige der befragten Personen wohl über das gesamte Unternehmen und nicht über die lokale betriebliche Einheit wie in der Haupterhebung. Im Unterschied zur Haupterhebung wurde im Interview der unterjährigen Erhebung die Angabe zur Beschäftigtenzahl ohne weitere Prüfungen oder Nachfrage übernommen, in der Annahme, dass Panelbetriebe vergleichbare Zahlen wie im Vorjahr angeben. Diese Annahme hat sich als nicht richtig erwiesen. Zukünftig müssten die Angaben zu den Beschäftigten, wie in der Haupterhebung, unter Zuhilfenahme der Werte aus dem Vorjahr abgefragt werden und gegebenenfalls durch das Erhebungsinstitut überprüft werden.

Zusammenfassend ist festzustellen, dass eine unterjährige Befragung in der Größe gut umsetzbar war. Die Folgen der unterjährigen Erhebung auf die Teilnahmebereitschaft der Betriebe in der Haupterhebung des IAB-Betriebspanels 2024 können zum jetzigen Zeitpunkt noch nicht beurteilt werden.

3.2 IAB-Stellenerhebung

Im Rahmen der Stellenerhebung wurde ein Survey-Experiment mit dem Ziel durchgeführt, die Implementierung neuer und rotierender Fragen in jedem Quartal sowie deren Verfahrensschritte, wie den Aufruf zur Fragenstellung, den Pretest, das Fragebogendesign und die Feldarbeit zu testen (hier nachfolgend „Prozedurales Verfahren“). Zusätzlich sollte untersucht werden, welche Auswirkungen die Hinzufügung weiterer Fragen auf die Antwortrate, die Belastung der Teilnehmer (*Response Burden*) und weitere Ergebnisse hat (hier nachfolgend als „Längen-Experiment“ bezeichnet).

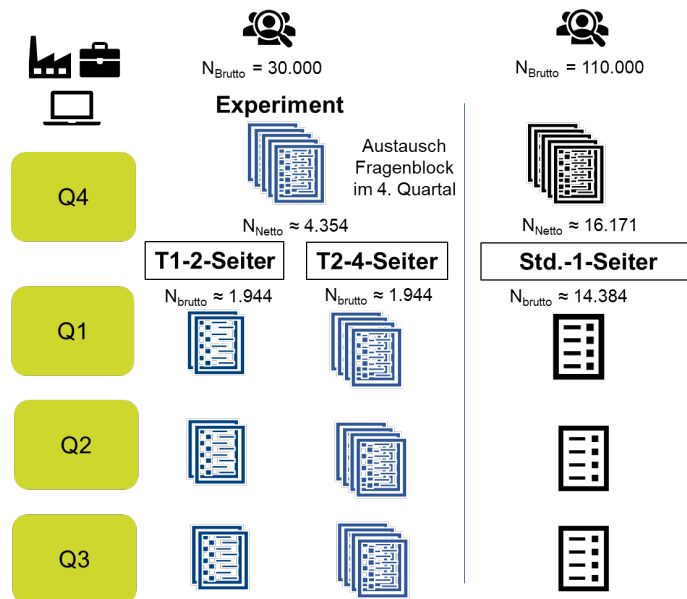
3.2.1 Aufbau des Experiments

Zur Umsetzung dieser Ziele wurde das Experiment wie folgt konzipiert: Zusätzlich zur geplanten Bruttostichprobe von 110.000 Betrieben wurde eine weitere Experimentalstichprobe von 30.000 Betrieben gezogen. Im vierten Quartal 2023 wurden für diese Experimentalgruppe (Exp.) innerhalb des 8-seitigen Hauptfragebogens bestehende Fragen durch 7 hochfrequente Fragen ausgetauscht und für diese Hochfrequenzfragen das prozedurale Verfahren getestet. Mit der Implementierung von Hochfrequenzfragen bereits im vierten Quartal 2023 wurde insbesondere ermöglicht, einen Teil der Hochfrequenzfragen im gesamten unterjährigen Panel abzufragen.

Da der Hauptfragebogen aus Belastungsgründen für die Betriebe auf maximal 8 Seiten mit derzeit 77 Fragen begrenzt ist, ist das Längen-Experiment auf die Nachbefragungen im ersten, zweiten und dritten Quartal 2024 beschränkt. Die kurzen 1-seitigen Nachbefragungen (im Folgenden Std.-1-Seiter genannt) mit rund 10 Fragen zu den Kerninformationen wurden in der Experimentalgruppe in jedem Quartal um hochfrequente Frageblöcke erweitert. Hierzu wurde die Experimentalstichprobe in zwei Gruppen unterteilt: eine Gruppe erhielt einen 4-seitigen Fragebogen mit bis zu 40 zusätzlichen Fragen, während die andere einen kürzeren, 2-seitigen Fragebogen mit bis zu 11 zusätzlichen Fragen erhalten hat. Die Seitenangaben beziehen sich hier auf den PAPI-Modus. Auch hier erfolgt neben dem sogenannten Längen-Experiment in jedem Quartal die Prüfung des prozeduralen Verfahrens.

Aus den 30.000 zusätzlichen Fällen der Experimentalstichprobe (Brutto) resultierte aufgrund eines unerwartet hohen Rücklaufs eine Nettostichprobe von 4.290 Betrieben, die im vierten Quartal 2023 an der Befragung teilgenommen haben. Die Bruttostichprobe der Experimentalstichprobe ist in den drei Nachbefragungen aus vertraglichen Gründen jeweils auf 3.900 Fälle begrenzt. Bereinigt um 12 Testadressen zur Überprüfung von Versandzeitpunkt und -qualität, wurden aus der Stichprobe von 4.290 Betrieben jeweils 1.944 Betriebe dem Zufallsprinzip folgend (Randomisierung) entweder der kurze Fragebogen (T1-2-Seiter -Treatment 1) oder der lange Fragebogen (T2-4-Seiter - Treatment 2) zugewiesen (Abbildung 5). Die Randomisierung folgte einer Block-Randomisierung, um eine ausgewogene Verteilung der Schichtungsvariablen (Region, Betriebsgrößenkategorie und Wirtschaftszweige) über die Experimentalgruppen zu gewährleisten. Mit diesem Vorgehen wurde angestrebt, die Repräsentativität der Stichproben zu sichern, sodass sich mit Hilfe der Ergebnisse Rückschlüsse auf die Grundgesamtheit ziehen lassen.

Abbildung 5: Experimentelles Design der Hochfrequenz im Rahmen der IAB-Stellenerhebung
 Viertes Quartal 2023 bis drittes Quartal 2024



Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

Das Experiment zielt darauf ab, die Machbarkeit und Auswirkungen der Einführung neuer Fragen zu bewerten, einschließlich der Prozessschritte für die Integration neuer Fragen und ihrer Auswirkungen auf die Rücklaufquote und die Response Burden.

3.2.2 Treatment: 2-seitiger Fragebogen und 4-seitiger Fragebogen

Der Fragebogen für die Nachbefragungen enthält Kerninformationen zu offenen Stellen, Beschäftigung und Arbeitszeit, um den Datenverpflichtungen gegenüber Eurostat sowie dem Statistischen Bundesamt nachzukommen. Während die Hauptbefragung, die in jedem vierten Quartal eines Kalenderjahres durchgeführt wird, auch mehrere andere Themen umfasst, enthalten die Nachbefragungen in den verbleibenden Folgequartalen im Wesentlichen ein Update der Informationen zu den offenen Stellen, zur Beschäftigung, zu den Beschäftigungserwartungen sowie zu betrieblichen Arbeitszeitkonten. Da diese Kerninformationen jedes Jahr identisch abgefragt werden, unterscheidet sich das hierzu erforderliche prozedurale Verfahren grundlegend von dem Verfahren für regelmäßig wechselnde und zeitnahe auswertbare Hochfrequenzfragen. Dies gilt insbesondere für den Pretest sowie die einzuhaltenden Fristen.

Um den Fragebogen der Nachbefragungen im Rahmen des Längen-Experiments um zusätzliche Inhalte zu erweitern, wurden zwei Ansätze verfolgt: Zum einen die systematische Integration eines erweiterten festen Satzes von Fragen, der spezifische Themenbereiche abdeckt – bezeichnet als "Panel-Fragen" – und zum anderen die Durchführung eines regelmäßigen internen Aufrufs für Fragen (Call for Questions), um neue Themenbereiche zu identifizieren und zu ergänzen. In Tabelle 8 ist die Anzahl der Fragen nach

thematischem Block für jedes Treatment vom ersten bis zum dritten Quartal dargestellt. Die Fragen lassen sich in die Standardfragen der Nachbefragungen, zusätzliche Panel-Fragen, Ergebnisse des internen Call for Questions, sowie in zusätzliche Fragen zur Response Burden einteilen. Die Tabelle zeigt für jedes Quartal die Gesamtzahl der Fragen, differenziert nach der Seitenanzahl des Fragebogens (2 Seiten oder 4 Seiten). Diese Aufstellung bietet einen detaillierten Einblick in die Struktur und den Umfang des Fragebogens über die Quartale hinweg.

Tabelle 8: Anzahl der treatmentspezifischen Fragen in den drei Nachbefragungen

Erstes Quartal bis drittes Quartal 2024

Leere Zelle	Q1/2024		Q2/2024		Q3/2024	
	2 Seiten	4 Seiten	2 Seiten	4 Seiten	2 Seiten	4 Seiten
Standardfragen	10	10	11	11	10	10
Zusätzliche Panel-Fragen	6	7	5	8	6	9
Zusätzliche Fragen aus Call for Questions	3	18	3	13	3	19
Zusätzliche Response Burden-Fragen	2	2	2	2	2	2
Gesamtzahl Fragen	21	37	21	34	21	40

Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

3.2.3 Ergebnisse zum prozeduralen Verfahren: *Call for Questions* Rücklauf

Wie zuvor erwähnt, ist neben den Ergebnissen des Längen-Experiments ebenso die Bewertung des prozeduralen Verfahrens der Implementierung hochfrequenter Fragen von zentralem Interesse. Um die Prozesse der Befragung weiter verbessern zu können, soll daher neben der Ermittlung der Auswirkungen auf die Rücklaufquote und die *Response Burden* auch der prozedurale Ablauf des *Call for Questions* und des Pretests im Rahmen des Längen-Experiments untersucht werden.

Mit dem Ziel, neue hochfrequente Fragen in die Befragungen der IAB-Stellenerhebung zu implementieren, ergibt sich grundsätzlich die Gelegenheit, die Befragung weiter zu öffnen und in Bezug auf das Forschungs- und Beratungsziel des IAB besser zu positionieren. Zu diesem Zwecke wurde im Rahmen des Projektes ein regelmäßiger IAB-interner *Call for Questions* durchgeführt, um Forschenden am IAB und ihren Kooperationspartnern und -partnerinnen zu ermöglichen, Fragen für ihre eigenen Forschungsinteressen zu entwickeln und in die IAB-Stellenerhebung einzubringen. Es wurden vier separate *Call for Questions* veröffentlicht, jeweils einer pro Quartal, die in jedem Quartal im IAB-Intranet⁸ bekannt gemacht wurden. Unter anderem auch aufgrund vertraglicher Erfordernisse werden (zwei bis) drei Monate vor Beginn der jeweiligen Feldphasen als spätest möglicher Start der Fragengenesse betrachtet. Nach der Veröffentlichung der *Call for Questions* wurde eine Frist von zwei Wochen eingeräumt, innerhalb derer sich alle interessierten Forschenden bewerben konnten. Bei einer zu

⁸ Ein Beispiel für den *Call for Questions* ist hier zu finden: <https://www.baintranet.de/info/iab/Daten/IAB-Stellenerhebung/Seiten/IAB-Stellenerhebung---Call-for-Questions-Q1-2024.aspx>.

geringen Anzahl an Bewerbungen war eine Fristverlängerung um eine Woche möglich (wie beispielsweise im ersten Quartal 2024). Die Anzahl der Bewerbungen sowie deren Themen und das Quartal der Einreichung finden sich in Tabelle 9.

Tabelle 9: Anzahl der Neueinreichungen aus dem IAB aufgrund des Call for Questions und ihre umgesetzten Themen

Viertes Quartal 2023 bis drittes Quartal 2024

Leere Zelle	IAB-Neueinreichungen	Themen
Q4/2023	4	Neu: Betrieblicher Urlaubsanspruch; Sonderfragen Energie; Response Burden, Die letzten 3 Monate
Q1/2024	2	Neu: Mindestlohn; Rente und Arbeitsmarkt Panel: Sonderfragen Energie; Response Burden, Die letzten 3 Monate
Q2/2024	2	Neu: Deutsche Sprachkenntnisse; Menschen mit Schwerbehinderung Panel: Sonderfragen Energie; Response Burden; Die letzten 3 Monate
Q3/2024	2	Neu: Ausbildung; Quereinsteiger/-innen Panel: Sonderfragen Energie; Response Burden; Die letzten 3 Monate

Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

Wegen des Längen-Experiments sollte jede Hochfrequenz-Einreichung in den drei Nachbefragungen im ersten, zweiten und dritten Quartal 2024 Fragenvarianten sowohl für die Lang- als auch für die Kurzversion des Fragebogens umfassen, und zwar im Schnitt drei bis vier Fragen für den langen sowie eine dieser Fragen für den kurzen Fragebogen. Sogenannte Filterfragen, d.h. Fragen bei denen eine bestimmte Antwort zur Überspringung von Folgefragen führt, sollten dabei vermieden werden. Hiermit sollte gewährleistet werden, dass sich beide Treatmentgruppen hinsichtlich der Länge des zu beantwortenden Fragebogens tatsächlich unterscheiden. Zudem sollten alle Fragen innerhalb des eingereichten Moduls in engem inhaltlichem Zusammenhang stehen, um ein unbeabsichtigtes Treatment durch verschiedene Themen zu vermeiden.

Wie die Anzahl der Bewerbungen zeigt, fiel die Resonanz auf den *Call for Questions* in allen vier Quartalen eher gering aus, obgleich in jedem Quartal Einreichungen zu verzeichnen waren. Derzeit ist noch unklar, ob die geringe Anzahl der Einreichungen eher auf ein geringes Interesse oder die Komplexität des *Call for Questions* mit der Anforderung, jeweils eine Kurz- und Langversion der Fragen einzureichen, zurückzuführen ist.

Auf Basis der Daten des ersten Quartals wurde ein IAB-Forumsbeitrag zur betrieblichen Betroffenheit einer Mindestloohnerhöhung auf 14 Euro publiziert (Börschlein und Diegmann 2024). Die zusätzlich erhobenen Daten werden außerdem über das FDZ als zusätzliches Datenmodul der IAB-Stellenerhebung auch externen Forschenden zur Verfügung gestellt werden.

3.2.4 Pretest

Ein weiterer wichtiger prozeduraler Arbeitsschritt bei der Implementierung hochfrequenter Fragen ist der sogenannte Pretest. Der Pretest ist ein Instrument, mit dem Erhebungsfragen getestet werden, indem sie einer kleinen Gruppe von Teilnehmenden (in unserem Fall also einigen Betrieben) gestellt werden. Im Wesentlichen liefert das Pretesting eine Möglichkeit, im Voraus darüber Kenntnis zu erlangen, ob eine konkrete Fragestellung Probleme für Interviewer/-innen oder Teilnehmende verursacht (Presser et al., 2004).

In der Forschungsliteratur werden verschiedene Methoden zum Pretesting von Erhebungen diskutiert, wie z.B. kognitive Interviews, Verhaltenscodierung und ein sogenanntes *Respondent-Debriefing* (Presser et al., 2004). Alle Pretests der Umfragen beim IAB werden aktuell von ProIAB⁹ unter Verwendung der Methode *Respondent Debriefing Pretest* durchgeführt. Die Methode des *Respondent-Debriefings* ermöglicht direktes Feedback der Teilnehmenden nach Abschluss der Umfrage. Dies kann wertvolle Einblicke in mögliche Missverständnisse oder Schwierigkeiten liefern, die beim Beantworten der Fragen aufgetreten sind.

In jedem Quartal wurden alle neuen Fragen, die in den Fragebogen aufgenommen werden sollten, zum Pretest an die ProIAB gesendet. Zusammen mit den Fragen wurde ein Interviewskript mit etwa fünf Fragen vorbereitet, um detailliert zu erfragen, ob die Betriebe die Erhebungsfragen gut verstanden haben oder ob es Schwierigkeiten bei der Beantwortung gab. Zudem wurden zu einigen Erhebungsfragen detaillierte Fragen zu den eingeführten Konzepten oder zu den Antwortskalen gestellt. Die ProIAB leitet die neuen Fragen im späteren Layout bereits im Vorfeld der Befragung an etwa 10 Betriebe weiter. Nach dem (gemeinsamen) Ausfüllen des Fragebogens können so direkt etwaige Probleme bei der Beantwortung identifiziert und entsprechende Änderungen aufgenommen werden, um die Erhebungsfragen zu verbessern.

Um den Nutzen des Pretesting-Verfahrens zu verdeutlichen, wird im Folgenden ein konkretes Beispiel beschrieben. Im ersten Quartal 2024 wurden Fragen zum Halten rentenberechtigter Beschäftigter gestellt. Nach Erhalt der ProIAB-Pretestergebnisse zeigte sich, dass der Begriff „Rentenberechtigung“ für die Betriebe schwer nachvollziehbar war, da sie nicht genau verstanden, worauf sich der Begriff bezog. Deshalb wurde dem Fragewortlaut eine von verschiedenen Betrieben vorgeschlagene Formulierung hinzugefügt, die den Begriff „Rentenberechtigt“ präziser definiert. Konkret enthielt die endgültige Frageformulierung den Erklärungszusatz: „die das reguläre Rentenalter erreicht haben und somit rentenberechtigt waren oder geworden sind“. Diese Anpassung führte zu einer deutlichen Verbesserung des Verständnisses bei den Teilnehmern.

Insgesamt lieferte die Pretesting-Phase nach jedem Call for Questions Rückmeldungen, die signifikante Verbesserungen der Fragen und Antwortskalen ermöglichten¹⁰. Jede Pretesting-Phase machte Probleme

⁹ Die ProIAB bestehen aus Projektmitarbeitern, die über die Agenturen für Arbeit im gesamten Bundesgebiet verteilt sind und die IAB-Wissenschaftler bei ihrer Feldforschung unterstützen. Für weitere Informationen über das ProIAB siehe Winters, J., und Kargus, A. (2012). Die ProIAB: Stets zur Stelle im Dienste der Wissenschaft. Ein Gespräch mit Christine Hense und Christian Sprenger. *IAB-Forum*, (H. 2), 34-39. <https://doi.org/10.3278/IFO1202W034>.

¹⁰ Ein Beispiel für die Ergebnisse des Pretests befindet sich in Anhang 1.

bei der Beantwortung der neuen Fragen deutlich, die während der Entwicklung nicht sofort erkennbar waren—sowohl in Bezug auf die Formulierung der Fragen als auch auf die Struktur der bereitgestellten Antwortoptionen.

Zusammenfassend waren die Ergebnisse aus dem Pretest von großem Nutzen, insbesondere um die Klarheit und Verständlichkeit der zusätzlichen Fragen zu verbessern. Das erhaltene Feedback wurde sorgfältig analysiert und genutzt, um entsprechende Anpassungen vorzunehmen. Dieser Prozess konzentrierte sich nicht nur auf die Präzisierung der Fragen, sondern auch darauf, den ursprünglich gewählten Ansatz zur Messung der Antworten neu zu bewerten, um sicherzustellen, dass die Antwortskalen mit den Perspektiven und Erfahrungen der Teilnehmer übereinstimmen. Der Pretest kam somit auch den Kolleginnen und Kollegen des IAB, die die Fragen entworfen hatten, zugute, da er ihnen ermöglichte, ihre Fragen zu präzisieren und ihre Forschungsziele umfassender zu reflektieren. Dieser iterative Prozess aus Testen und Feedback bot die Möglichkeit, potenzielle Fehler zu identifizieren und zu vermeiden, die die Datenqualität der Umfrage hätten beeinträchtigen können.

In Zukunft wird das ProIAB möglicherweise nicht mehr zur Verfügung stehen, um die Pretests der hochfrequenten Fragen durchzuführen. Für den Fall, dass hochfrequente Fragen auch in Zukunft dauerhaft in die IAB-Stellenerhebung eingebracht werden sollen, ist es daher erforderlich, alternative Möglichkeiten von Pretests auszuloten. Eine Möglichkeit besteht darin, den Pretest extramural an das mit der Befragung beauftragte Erhebungsinstitut (resp. den Unterauftragnehmer T.I.P. Biehl und Partner) zu vergeben. Auch hier würde im Vorfeld eine formatierte Version der Erhebungsfragen an eine kleinere Stichprobe geliefert und telefonisch das Feedback von den Pretest-Teilnehmenden eingeholt.

3.2.5 Ergebnisse des Längen-Experiments

Die *Response Burden* (Teilnahmebelastung) ist seit langem ein wichtiges Konzept in der Umfrageforschung, da sie insbesondere die Datenqualität und die Rücklaufquote beeinflusst (Yan und Williams, 2022). Das Konzept der *Response Burden* ist besonders bei Betriebsumfragen relevant, da die Anforderungen zu Beantwortung der Umfrage für Betriebe erheblichen Zeit- und Kostenaufwand bedeuten können. Dies kann sowohl die Teilnahmebereitschaft in Bezug auf die gesamte Befragung (*Unit-Response*) als auch auf einzelne Fragen (*Item-Response*) beeinträchtigen sowie Messfehler (*Measurement Error*) und Umfragekosten erhöhen (Bavdaž, 2010; Erikson et al., 2023; Giesen, 2012). Aktuelle Studien betonen die Unterscheidung zwischen der tatsächlichen (*Actual*) und der wahrgenommenen (*Perceived*) *Response Burden* – wobei erstere sich auf die tatsächlich aufgewendeten Ressourcen für die Beendigung einer Umfrage und letztere auf die subjektive Wahrnehmung des damit verbundenen Aufwands bezieht (Erikson et al., 2023; Haraldsen et al., 2013).

Die Länge des Fragebogens wird als eine zentrale Determinante der *Response Burden* angesehen. So gibt es Evidenz dazu, dass längere Fragebögen die Antwortquoten bei PAPI-Befragungen (d.h. mit schriftlichen Fragebögen) negativ beeinflussen (Dillman et al., 1993; Sandelin, 2022). Weitere Forschungsergebnisse zeigen, dass Fragen, die gegen Ende des Fragebogens platziert wurden, größere Messfehler aufweisen, was das Risiko für eine schlechtere Datenqualität und einen erhöhten *Nonresponse-Bias* erhöht (Krosnick et al., 2002; Olson und Peytchev, 2007). Die Auswirkungen der

Fragebogenlänge auf CAWI-Umfragen sind jedoch weniger eindeutig. Einige Studien finden keinen Einfluss der Fragebogenlänge auf die Rücklaufquote (Blumenberg et al., 2019; Koitsalu et al., 2018). Andere Studien legen nahe, dass eine längere erwartete Antwortzeit die Antwortquoten verringerte und dass Web-Umfragen maximal 20 Minuten dauern sollten und idealerweise nur 10 Minuten, um die Zufriedenheit der Teilnehmenden zu maximieren (Galesic und Bosnjak, 2009; Revilla und Ochoa, 2017). Wird ein Fragebogen über diese Länge hinaus erweitert, besteht das Risiko, dass die Teilnehmenden gegen Ende des Interviews infolge von Ermüdung und mangelnder Motivation ein unerwünschtes Antwortverhalten aufweisen (Cernat et al., 2022).

In der IAB-Stellenerhebung kann eine Implementierung von hochfrequenten Fragen im Rahmen der drei Nachbefragungen nur als Ergänzung zu den zur Erfüllung der Lieferverpflichtungen zu erfassenden Kerninformationen erfolgen. Aus diesem Grund ist es erforderlich, die damit verbundene *Response Burden* abzuschätzen. Das Längen-Experiment zielt darauf ab, die Auswirkungen der tatsächlichen Belastung auf die Rücklaufquote im Kontext der IAB-Stellenerhebung zu untersuchen. Durch die experimentelle Variation der Fragebogenlänge können die direkten Auswirkungen der tatsächlichen Belastung auf die Teilnahmebereitschaft untersucht werden. Gleichzeitig wurden Fragen zur Messung der wahrgenommenen Belastung integriert, um Einblicke darüber zu gewinnen, wie die subjektiven Erfahrungen der Betriebe ihre Bereitschaft zur Teilnahme und die Genauigkeit ihrer Antworten beeinflussen können. Dieser Ansatz kann dazu beitragen, die *Response Burden* zu verstehen und das Umfragedesign sowie die Teilnahmebereitschaft zu verbessern.

Die nachfolgenden Unterabschnitte enthalten eine detaillierte Beschreibung der Ergebnisse des Längen-Experiments zur Untersuchung der Auswirkungen der Fragebogenlänge auf die Rücklaufquote und die *Response Burden*. Hierzu werden im Folgenden im Wesentlichen die Unterschiede in verschiedenen Ergebnisvariablen zwischen den beiden Treatmentgruppen (2-Seiter vs. 4-Seiter) ermittelt. Um zusätzlich die Standardgruppe mit dem Std.-1-Seiter Fragebogen in die Untersuchung einzubeziehen, sollen ebenfalls Vergleiche zwischen dieser Gruppe und den beiden Treatment-Gruppen angestellt werden.

Zu beachten ist jedoch, dass es erhebliche Größenunterschiede zwischen den Gruppen gibt. So enthalten die Treatmentgruppen jeweils 1.944 Fälle, während die Standardgruppe (Std.) aus 14.384 Fällen besteht. Zusätzlich zeigt Abbildung 1, dass es zwischen den Treatmentgruppen und der Standardgruppe auch geringfügige Unterschiede in den Fragen des Hauptfragebogens im vierten Quartal gab. Konkret enthielten die jeweiligen Fragebögen für die Standardgruppe und die Experimentalgruppe im vierten Quartal 2023 drei unterschiedliche Fragen. Trotz dieser Unterschiede, die unerwünschte Effekte auf die Rücklaufquote haben könnten, können durch den Vergleich der Treatmentgruppen mit der Standardgruppe möglicherweise wichtige Erkenntnisse gewonnen werden. Da die Voraussetzung hierfür ist, dass die Experimentalgruppe und die Standardgruppe im vierten Quartal 2023 keine signifikanten Diskrepanzen im Teilnahmeverhalten aufweisen, wird für dieses Quartal im Folgenden auch der Unterschied in der Rücklaufquote zwischen der Experimentalgruppe und der Standardgruppe untersucht.

Größere Gruppenunterschiede stellen eine Herausforderung für die Randomisierung und Vergleichbarkeit dar. Aufgrund der begrenzten Anzahl an Fällen in den Experimentalgruppen war es nicht möglich, drei Treatmentgruppen zu randomisieren, ohne die statistische Power zu beeinträchtigen. Nach Durchführung einer Poweranalyse wurde daher entschieden, nur zwei Treatmentgruppen zu bilden, um sicherzustellen, dass das Experiment eine ausreichende statistische Power aufweist, um die erwarteten Effekte zu detektieren. Dennoch besteht weiterhin die Möglichkeit, die Standardgruppe mit den Treatmentgruppen zu vergleichen.

Allerdings stellt die unterschiedliche Gruppengröße zwischen der Std.-1-Seiter-Gruppe und den Treatmentgruppen ein potenzielles Problem für die Vergleichbarkeit dar. Um dieser Problematik zu begegnen, werden neben den Vergleichsanalysen der Rücklaufquoten auch die Ergebnisse einer Sensitivitätsanalyse, die mit der G*Power-Software (Faul et al., 2007) durchgeführt wurde, berichtet. Diese Sensitivitätsanalyse ermöglicht es uns, die minimal nachweisbaren Effektgrößen für das gewählte Studiendesign zu bestimmen und die Robustheit der Ergebnisse angesichts der unterschiedlichen Gruppengrößen zu bewerten. Konkret zeigt die Analyse, welche Effektgrößen mit ausreichender statistischer Power detektiert werden können und inwieweit die ungleichen Gruppengrößen die Genauigkeit und Verlässlichkeit der geschätzten Effekte beeinflussen. Dadurch wird sichergestellt, dass die Vergleiche zwischen der Std.-1-Seiter Gruppe und den Treatmentgruppen valide und aussagekräftig sind, selbst unter den gegebenen strukturellen Einschränkungen des Experiments.

Zur Überprüfung signifikanter Unterschiede in den Rücklaufquoten wird die Z-Teststatistik angewendet.¹¹ Zudem werden Effektgrößen berichtet. Cohen's h eignet sich dazu, über die statistische Signifikanz hinaus die Stärke des Effekts zu bewerten, wobei ein Wert von 0,2 auf einen kleinen, 0,5 auf einen mittleren und 0,8 auf einen großen Effekt hinweist. In unserem Fall bezieht sich dieser Unterschied auf den Einfluss der Fragebogenlänge auf die Rücklaufquote der Teilnehmer.

Quartal Rücklaufquote

Als erste wichtige Ergebnisvariable wird die Rücklaufquote ermittelt. Zur Ermittlung der Rücklaufquote wird die AAPOR RR1 Definition der Rücklaufquote verwendet (American Association for Public Opinion Research, 2023). Die Formel für die Rücklaufquote 1 (RR1) lautet:

$$RR1 = \frac{\text{Anzahl der gültigen Teilnehmer (Eurostat Fragen beantwortet)}}{\text{Gesamtzahl der Fälle in der Stichprobe}}$$

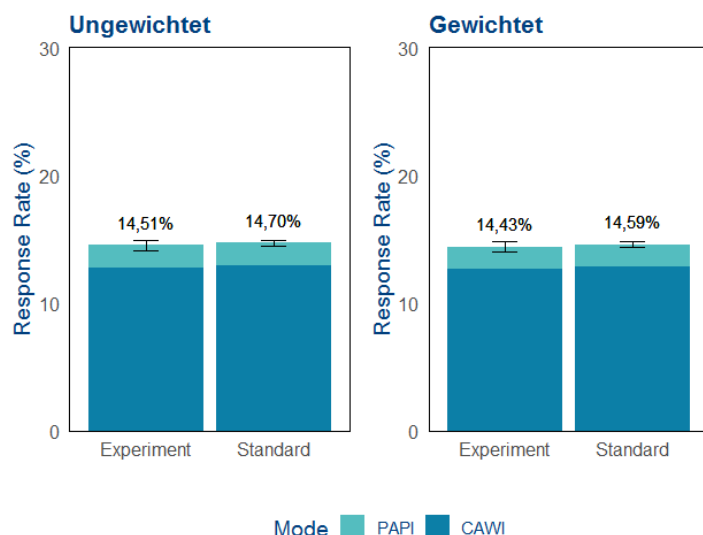
¹¹ Die Z-Teststatistik ist entscheidend, um die statistische Signifikanz zwischen den *Rücklaufquoten* der Gruppen zu bestimmen, da sie die unterschiedlichen Gruppengrößen durch die Einbeziehung der Stichprobengrößen in die Berechnung des Standardfehlers der Rücklaufquote berücksichtigt. Ein hoher absoluter Wert des Z-Teststatistik deutet darauf hin, dass der Unterschied zwischen den Gruppen statistisch signifikant ist, was bedeutet, dass der Unterschied nicht zufällig ist, sondern mit hoher Wahrscheinlichkeit auf einen tatsächlichen Effekt zurückzuführen ist. Dies gewährleistet eine korrekte Bewertung der Signifikanzunterschiede zwischen den Gruppen, unabhängig von den unterschiedlichen Gruppengrößen, während Effektgrößen wie Cohen's h die Größenordnung der beobachteten Unterschiede quantifizieren.

Als gültige Teilnehmer werden in der IAB-Stellenerhebung alle Betriebe betrachtet, die mindestens zwei offizielle Eurostat-Fragen - konkret zur Anzahl der Mitarbeitenden und den offenen Stellen - beantworten, entweder durch Zusendung des schriftlichen Fragebogens oder durch Klicken auf den *Submit-Button* in der Webversion. Diese Fragen sind immer direkt am Anfang des Fragebogens platziert. In der obigen Gleichung entspricht somit die Anzahl der gültigen Teilnehmer im Zähler der Anzahl der Teilnehmer, die die Eurostat-Fragen beantwortet haben, während die Gesamtzahl der Fälle in der Stichprobe im Nenner die gültigen Teilnehmer, Nicht-Interviews (wie Ablehnungen und keine Kontaktaufnahmen) sowie andere Fälle mit unbekannter Teilnahmeberechtigung umfasst.

Im Folgenden werden sowohl ungewichtete als gewichtete Rücklaufquoten ausgewiesen. Die gewichteten Rücklaufquoten gewährleisten, dass die Ergebnisse repräsentativ sind und die Zielpopulation genau widerspiegeln. Gleichzeitig ermöglichen sie eine transparente Bewertung der Repräsentativität der Umfrage und potenzieller Verzerrungen. Die Gewichtung erfolgt durch Inklusionswahrscheinlichkeiten, die als Inverse der Wahrscheinlichkeit berechnet werden, dass ein Betrieb basierend auf den drei spezifischen Schichtungsvariablen – Betriebsgröße, Wirtschaftszweig und Region (Ost/West Gebiet) – in die Stichprobe gelangt.

Abbildung 6: Rücklaufquoten nach Gruppe und Erhebungsmodus in der Hauptbefragung

Viertes Quartal 2023, Anteile in Prozent



Anmerkung: Konfidenzintervalle wurden auf dem 95 Prozent-Niveau berechnet.

Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

Abbildung 6 zeigt die gewichteten und ungewichteten Rücklaufquoten des vierten Quartals sowohl für die Experimentalgruppe als auch für die Standardgruppe, jeweils getrennt nach Erhebungsmodus. Tabelle 10 weist die Ergebnisse des Z-Tests der Unterschiede der gewichteten als auch der

ungewichteten Rücklaufquoten zwischen der Standard- und der Experimentalgruppe sowie die Effektstärken (Cohen's h) aus. Es zeigen sich keine statistisch signifikanten Unterschiede zwischen den Rücklaufquoten im vierten Quartal. Gemäß den herkömmlichen Effektstärkenschwellen (kleiner Effekt: 0,20; mittlerer Effekt: 0,50; großer Effekt: 0,80) fallen auch die Effektstärken sehr gering aus. Abbildung 6 zeigt außerdem, dass die Moduswahl in beiden Gruppen ähnlich verläuft, wobei über als 85 Prozent der Stichprobe den CAWI-Modus wählen.

Tabelle 10: Vergleich der Rücklaufquoten zwischen Standardgruppe und Experimentalgruppe in der Hauptbefragung

Viertes Quartal 2023

	Z-Wert	P-Wert	95% Konfidenzintervall	Cohen's h
Ungewichtet	0,818	0,412	[-0,002; 0,006]	0,005
Gewichtet	0,819	0,412	[-0,002; 0,005]	0,004

Anmerkungen: Keine statistisch signifikanten Unterschiede (Z-test) in der Rücklaufquote zwischen der Standard- und der Experimentalgruppe zum 5 Prozent-Niveau. P-Wert inkl. Bonferroni-Holm Korrektur. Die Holm-Bonferroni-Korrektur verringert das Risiko von Fehlern erster Art (falsch positive Ergebnisse), das mit der Anzahl gleichzeitiger Tests zunimmt.

Quelle: IAB-Stellenerhebung; eigene Berechnungen. © IAB

Da die Sensitivitätsanalyse basierend auf den tatsächlichen Fallzahlen in den Gruppen durchgeführt wird, erfolgt diese Analyse nur für die ungewichteten Rücklaufquoten. Die Sensitivitätsanalyse wurde mit einem zweiseitigen Test durchgeführt, da keine gerichtete Hypothese vorlag. Es wurde somit keine spezifische Annahme getroffen, dass die eine Rücklaufquote höher oder niedriger als die jeweils andere ausfällt.

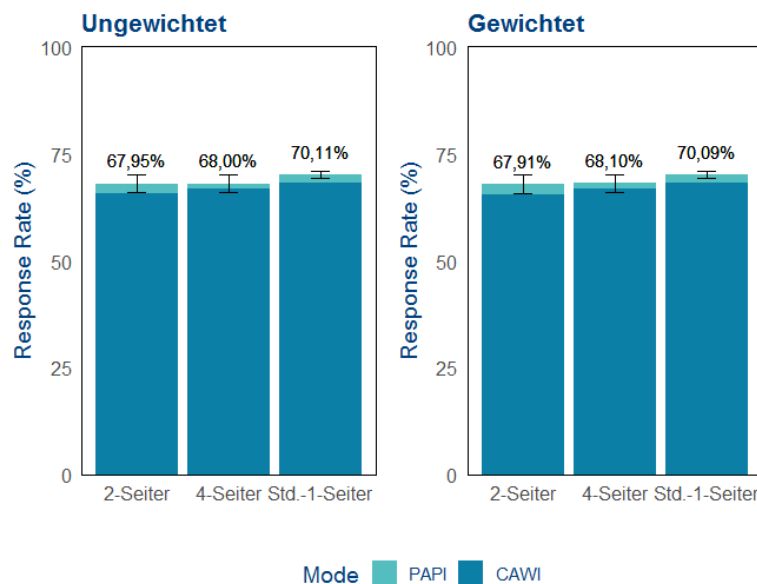
Für die Rücklaufquote der Experimentalgruppe ($p_1 = 0,1451$) und die berechnete minimale Rücklaufquote der Standardgruppe ($p_2 = 0,1516$) ergab sich, dass bei den vorliegenden Stichprobengrößen von 29.998 für die Experimentalgruppe und 109.989 für die Standardgruppe eine minimale Differenz der Rücklaufquoten von 0,0065 mit ausreichender statistischer Power (80%) erkannt werden kann. Der kritische Z-Wert betrug 1,960. Vergleicht man diese minimale detektierbare Differenz mit der tatsächlichen Rücklaufquote der Standardgruppe von 0,1470, ergibt sich eine Differenz von 0,0019 zwischen den Gruppen ($p_2 = 0,1470$ vs. $p_1 = 0,1451$). Diese Differenz liegt deutlich unter der minimalen detektierbaren Differenz von 0,0065. Dies bedeutet, dass mit dem vorliegenden Design so kleine Unterschiede, wie sie zwischen der Experimentalgruppe und der tatsächlichen Rücklaufquote der Standardgruppe existieren, nicht mit ausreichender statistischer Power entdeckt werden können. Zusätzlich weisen die Cohen's h-Werte in Tabelle 3 (0,005 für ungewichtete und 0,004 für gewichtete Rücklaufquoten) auf extrem kleine Effektstärken hin. Da Cohen's h eine standardisierte Effektgröße zwischen zwei Rücklaufquoten angibt, zeigt dieser Wert, dass der beobachtete Unterschied statistisch und praktisch vernachlässigbar ist. Die Ergebnisse der Sensitivitätsanalyse untermauern dies: Der gemessene Unterschied liegt unter der minimalen detektierbaren Differenz, was bestätigt, dass keine

bedeutenden Effekte zwischen den Gruppen existieren, die das Studiendesign erfassen könnte. Da die beobachteten Unterschiede in den Rücklaufquoten somit unterhalb der Schwelle für eine statistisch signifikante Entdeckung liegen und auch Cohen's h auf praktisch irrelevante Unterschiede hindeutet, bestätigen die Ergebnisse der Sensitivitätsanalyse, dass die vorhandenen Unterschiede in den Rücklaufquoten, wenn überhaupt, minimal sind.

Abbildung 7 zeigt die Rücklaufquoten des ersten Quartals 2024. Da in den Nachbefragungen die Experimentalgruppe in zwei Treatmentgruppen eingeteilt wurde, wird bei der Experimentalgruppe im Folgenden zwischen der 2- und 4-Seiter-Gruppe unterschieden.

Abbildung 7: Rücklaufquoten nach Gruppe und Erhebungsmodus in der ersten Nachbefragung

Erstes Quartal 2024, Anteile in Prozent



Anmerkung: Konfidenzintervalle wurden auf dem 95 Prozent-Niveau berechnet.

Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

Gemäß Abbildung 7 sind die Unterschiede in der ungewichteten Rücklaufquote zwischen der 2- und 4-Seiter-Gruppe sehr gering (0,05 Prozentpunkte). Der Unterschied in der gewichteten Rücklaufquote fällt etwas größer aus (0,19 Prozentpunkte). Im Vergleich zur Std.-1-Seiter-Gruppe liegt die ungewichtete Rücklaufquote in beiden Treatmentgruppen um ca. 2 Prozentpunkte höher. Tabelle 11 zeigt die Z-Werte, p-Werte (sowohl unkorrigiert als auch mittels der Holm-Bonferroni-Methode korrigiert), Konfidenzintervalle und Cohen's h für die Vergleiche zwischen den verschiedenen Gruppen im ersten Quartal. Die Korrektur der p-Werte mit der Holm-Bonferroni-Methode wurde durchgeführt, um das Risiko eines fälschlichen Zurückweisens der Nullhypothese (Typ 1 -Fehler) bei multiplen Tests zu reduzieren. Zwischen den beiden Treatmentgruppen zeigen sich auch im ersten Quartal keine

signifikanten Unterschiede in den Rücklaufquoten, und auch das Effektstärkenmaß fällt nur gering aus. Es zeigt sich jedoch ein statistisch signifikanter Unterschied zwischen den Rücklaufquoten der Std.-1-Seiter und der 4-Seiter-Gruppe. Diese Signifikanz bleibt jedoch nach Anwendung der Holm-Bonferroni-Korrektur nicht bestehen.

Im Hinblick auf den Erhebungsmodus verzeichnen im ersten Quartal 2024 alle Gruppen eine noch stärkere Tendenz als im vierten Quartal 2023, die Umfrage online zu beantworten (mehr als 95 Prozent).¹²

Tabelle 11: Vergleich der Rücklaufquoten zwischen Standardgruppe und Experimentalgruppen in der ersten Nachbefragung

Erstes Quartal 2024

Leere Zelle		Z-Wert	p-Wert	p-Wert (Holm-Bonferroni- Korrektur)	95% Konfidenzintervall	Cohen's h
2-Seiter vs. 4-Seiter	Ungewichtet	-0,034	0,972	0,972	[-0,000; 0,043]	-0,001
	Gewichtet	-0,135	0,892	0,892	[-0,027; 0,024]	-0,003
Std.-1-Seiter vs. 2-Seiter	Ungewichtet	1,918	0,055	0,165	[-0,000; 0,043]	0,046
	Gewichtet	2,169	0,030*	0,090	[0,002; 0,041]	0,047
Std.-1-Seiter vs. 4-Seiter	Ungewichtet	1,873	0,061	0,165	[-0,000; 0,043]	0,045
	Gewichtet	1,991	0,046*	0,092	[0,000; 0,039]	0,043

Anmerkungen: Die Sterne zeigen einen statistisch signifikanten Unterschied (Z-test) in der Rücklaufquote zwischen der Std.-1-Seiter, 2-Seiter und 4-Seiter-Gruppen an: *p < 0,05 ** p < 0,01 *** p < 0,001. Die Holm-Bonferroni-Korrektur verringert das Risiko von Fehlern erster Art (falsch positive Ergebnisse), das mit der Anzahl gleichzeitiger Tests zunimmt.

Quelle: IAB-Stellenerhebung; eigene Berechnungen. © IAB

Die Sensitivitätsanalyse wurde für den Vergleich der Treatmentgruppen mit der Std.-1-Seiter-Gruppe durchgeführt, um die minimal detektierbare Differenz der Rücklaufquoten zu bestimmen. Da die Hypothese besteht, dass die Rücklaufquote der Gruppen mit mehr Fragen (2- und 4-Seiter) kleiner ist, wurde ein einseitiger Test verwendet.

Für den Vergleich zwischen der 2-Seiter-Gruppe ($p_1 = 0,6795$) und der Std.-1-Seiter-Gruppe ($p_2 = 0,7071$) wurde mit den vorliegenden Stichprobengrößen von 1.944 für die 2-Seiter-Gruppe und 14.380 für die Std.-1-Seiter-Gruppe eine minimale Differenz von 0,0276 als detektierbar bestimmt. Der kritische Z-Wert betrug 1,645. Der tatsächlich beobachtete Unterschied zwischen den Gruppen (0,6795 vs. 0,7011) beträgt jedoch nur 0,0216, was unter der Schwelle der detektierbaren Differenz liegt. Dies bedeutet,

¹² Während der Durchführung des Umfrageprotokolls enthielten die an die Teilnehmer gesendeten Anschreiben versehentlich nicht die vorfrankierten Rückumschläge, die für die Rücksendung der schriftlichen Fragebögen (PAPI) vorgesehen waren. Dies stellt eine Abweichung vom geplanten Protokoll dar, welches den Versand vorfrankierter Umschläge vorsah, um die Rücksendung der Antworten zu erleichtern und die potenzielle Belastung der Teilnehmer zu minimieren. Der fehlende Versand der vorfrankierten Rückumschläge könnte somit die Rücklaufquote beeinträchtigt haben, da die Teilnehmer unerwartet die Portokosten für die Rücksendung ihrer Fragebögen tragen mussten. Nichtsdestotrotz wurden schriftliche Fragebögen seitens der Teilnehmer zurückgesendet und in die Rücklaufquote einbezogen. In den folgenden Quartalen erfolgte wieder der Versand der vorfrankierten Umschläge an die Teilnehmer.

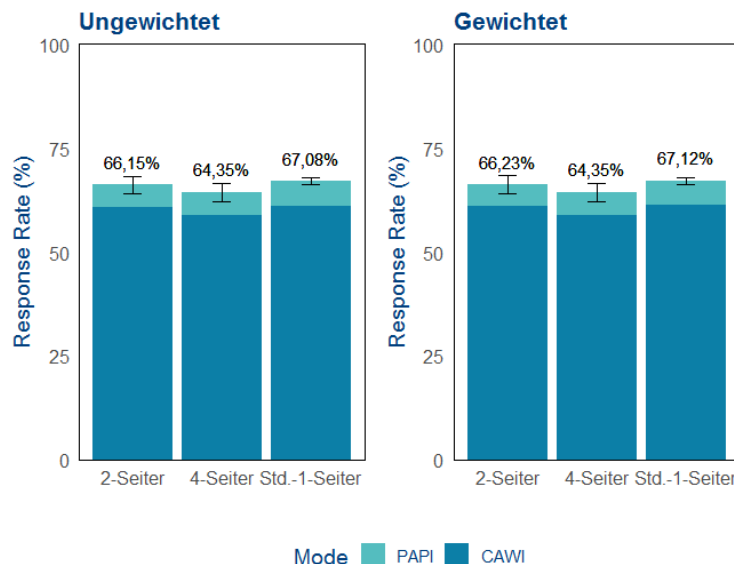
dass die beobachteten Unterschiede nicht groß genug sind, um mit der gegebenen Power entdeckt zu werden.

Für den Vergleich zwischen der 4-Seiter-Gruppe ($p_1 = 0,6800$) und der Std.-1-Seiter-Gruppe ($p_2 = 0,7076$) ergab die Sensitivitätsanalyse eine minimale detektierbare Differenz von 0,0276. Auch hier betrug der kritische Z-Wert 1,645. Der tatsächliche Unterschied von 0,0211 liegt ebenfalls unter dieser Schwelle, was darauf hinweist, dass der beobachtete Unterschied mit der aktuellen Stichprobengröße nicht zuverlässig entdeckt werden konnte.

Zusätzlich zeigen die Cohen's h-Werte in Tabelle 4 (0,046 bzw. 0,045 für den Vergleich der 2-Seiter- bzw. 4-Seiter-Gruppe mit der Std.-1-Seiter-Gruppe) kleine Effektstärken, was darauf hinweist, dass die Größenordnungen der Unterschiede zwischen den Gruppen gering sind. Da die Cohen's h-Werte nahe bei 0 liegen, wird bestätigt, dass selbst wenn Unterschiede vorhanden wären, diese minimal und statistisch schwer detektierbar sind. Dies wird durch die Ergebnisse der Sensitivitätsanalyse weiter untermauert, die zeigen, dass die beobachteten Unterschiede unterhalb der detektierbaren Schwelle liegen.

Abbildung 8: Rücklaufquoten nach Gruppe und Erhebungsmodus in der zweiten Nachbefragung

Zweites Quartal 2024, Anteile in Prozent



Anmerkung: Konfidenzintervalle wurden auf dem 95 Prozent-Niveau berechnet.

Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

Abbildung 8 stellt die Rücklaufquoten des zweiten Quartals, getrennt nach Gruppe und Erhebungsmodus dar. Für dieses Quartal werden die Rücklaufquoten für die Std.-1-Seiter-Gruppe sowie die beiden Treatmentgruppen (2-Seiter und 4-Seiter) ausgewiesen. Die Abbildung zeigt, dass die

Rücklaufquote der 2-Seiter-Gruppe im Vergleich zum vorherigen Quartal mit einem Rückgang um 1,68 Prozentpunkte (gewichtet) am geringsten gesunken ist. Sowohl die Std.-1-Seiter-Gruppe als auch die 4-Seiter-Gruppe verzeichnen (gewichtet) gegenüber dem Vorquartal größere Rückgänge, nämlich 2,97 Prozentpunkte bzw. 3,75 Prozentpunkte. Hinsichtlich des Befragungsmodus zeigt sich in allen Gruppen ein sehr ähnlicher Trend. So ist ein leichter Rückgang der Präferenz für den CAWI-Modus im Vergleich zu den 95 Prozent des ersten Quartals zu beobachten, da im zweiten Quartal über 91 Prozent der Teilnehmer den CAWI-Modus wählten. Insgesamt zeigt sich, dass sich die Wahl des Modus zwischen den Gruppen wenig unterscheidet, mit einer weiterhin vorherrschenden Präferenz für den web-basierten Erhebungsmodus.

Tabelle 12 enthält die Z-Werte, p-Werte, die mit der Holm-Bonferroni-Methode korrigierten p-Werte, Konfidenzintervalle und Cohen's h für die Vergleiche zwischen den verschiedenen Gruppen im zweiten Quartal. Der Vergleich zwischen der 2-Seiter- und der 4-Seiter-Gruppe zeigt keine statistisch signifikanten Unterschiede in den Rücklaufquoten, mit p-Werten, die deutlich über der konventionellen Signifikanzschwelle ($p < 0,05$) liegen. Dies gilt auch nach Anwendung der Holm-Bonferroni-Korrektur. Die Effektstärken (Cohen's h) sind ebenfalls klein und weisen somit auf nur geringe Unterschiede zwischen diesen beiden Gruppen hin. Der Vergleich zwischen der Std.-1-Seiter- und der 2-Seiter-Gruppe kann ebenfalls keine signifikanten Unterschiede in den Rücklaufquoten nachweisen, mit kleinen Effektstärken sowohl für die ungewichteten als auch die gewichteten Werte. Die Sensitivitätsanalyse für den Vergleich zwischen der 2-Seiter-Gruppe ($p_1 = 0,6615$) und der Std.-1-Seiter-Gruppe ($p_2 = 0,6708$) ergab, dass bei den vorliegenden Stichprobengrößen von 1.944 für die 2-Seiter-Gruppe und 10.082 für die Std.-1-Seiter-Gruppe eine minimale Differenz von 0,0287 als detektierbar bestimmt wurde. Der kritische Z-Wert betrug 1,645. Der tatsächlich beobachtete Unterschied zwischen den Gruppen (0,6615 vs. 0,6708) beträgt jedoch nur 0,0093, was deutlich unter der Schwelle der detektierbaren Differenz liegt. Dies bedeutet, dass die beobachteten Unterschiede zu gering sind, um mit der gegebenen Power statistisch detektierbar zu sein.

Im Gegensatz dazu deutet der Vergleich zwischen der Std.-1-Seiter-Gruppe und der 4-Seiter-Gruppe auf statistisch signifikante Unterschiede hin. Der ungewichtete Vergleich ergibt einen Z-Wert von 2,361 und einen p-Wert von 0,018, der nach Korrektur jedoch zunimmt ($p = 0,054$). Auf Basis des gewichteten Vergleichs zeigt sich jedoch ein statistisch signifikanter Unterschied mit einem Z-Wert von 2,687 und einem korrigierten p-Wert von 0,021, insgesamt jedoch mit kleinen, aber nicht vernachlässigbaren Effektstärken (Cohen's h von 0,057 bzw. 0,058). Für den Vergleich zwischen der 4-Seiter-Gruppe ($p_1 = 0,6435$) und der Std.-1-Seiter-Gruppe ($p_2 = 0,6708$) ergab die Sensitivitätsanalyse eine minimale detektierbare Differenz von 0,0291. Der tatsächliche Unterschied von 0,0273 liegt knapp unter dieser Schwelle, was darauf hinweist, dass dieser Unterschied nur schwer detektierbar ist. Zusätzlich verdeutlichen die Cohen's h-Werte in Tabelle 5 (0,019 für den Vergleich der 2-Seiter-Gruppe mit der Std.-1-Seiter-Gruppe und 0,057 für den Vergleich der 4-Seiter-Gruppe mit der Std.-1-Seiter-Gruppe) kleine Effektstärken.

Tabelle 12: Vergleich der Rücklaufquoten zwischen Standardgruppe und Experimentalgruppe in der zweiten Nachbefragung

Zweites Quartal 2024

Leere Zelle		Z-Wert	p-Wert	p-Wert (Holm-Bonferroni-Korrektur)	95% Konfidenzintervall	Cohen's h
2-Seiter vs. 4-Seiter	Ungewichtet	1,179	0,238	0,476	[-0,011; 0,047]	0,037
	Gewichtet	1,376	0,168	0,337	[-0,007; 0,045]	0,039
Std.-1-Seiter vs. 2-Seiter	Ungewichtet	0,811	0,417	0,476	[-0,013; 0,031]	0,019
	Gewichtet	0,878	0,379	0,379	[-0,011; 0,028]	0,018
Std.-1-Seiter vs. 4-Seiter	Ungewichtet	2,361	0,018*	0,054	[0,004; -0,049]	0,057
	Gewichtet	2,687	0,007***	0,021*	[0,007; 0,047]	0,058

Anmerkungen: Die Sterne zeigen einen statistisch signifikanten Unterschied (Z-test) in der Rücklaufquote zwischen der Std.-1-Seiter, 2-Seiter und 4-Seiter-Gruppen an: * $p < 0,05$ ** $p < 0,01$ *** $p < 0,001$. Die Holm-Bonferroni-Korrektur verringert das Risiko von Fehlern erster Art (falsch positive Ergebnisse), das mit der Anzahl gleichzeitiger Tests zunimmt.

Quelle: IAB-Stellenerhebung; eigene Berechnungen. © IAB

Zusammenfassend zeigen die Ergebnisse des zweiten Quartals 2024, dass die 2-Seiter-Gruppe zwar den geringsten Rückgang in der Rücklaufquote verzeichnet, jedoch keine signifikanten Unterschiede zur Std.-1-Seiter-Gruppe oder zur 4-Seiter-Gruppe bestehen. Auch wenn der Vergleich zwischen der 4-Seiter- und der Std.-1-Seiter-Gruppe statistisch signifikante Unterschiede aufweist, bleibt die Größenordnung der Effekte gering, wie die Cohen's h-Werte verdeutlichen. Die Sensitivitätsanalyse bestätigt, dass die beobachteten Unterschiede größtenteils unterhalb der detektierbaren Schwelle liegen. Insgesamt deuten die Ergebnisse darauf hin, dass sich keine der Gruppen signifikant in ihrer Fähigkeit unterscheidet, stabile Teilnahmequoten aufrechtzuerhalten. Die minimalen Effektstärken und die nicht signifikanten Unterschiede weisen darauf hin, dass alle drei Formate in etwa vergleichbare Ergebnisse hinsichtlich der Rücklaufquote liefern.

Kumulative Rücklaufquote

Wie bei jeder Umfrage ist zu erwarten, dass die Rücklaufquote im Laufe der Zeit abnimmt. So ist die so genannte *Survey Attrition*, definiert als *Unit-Nonresponse* nach der ersten Erhebungswelle ein bekanntes Phänomen in der Umfrageforschung. Speziell im Fall der Stellenerhebung ist zu erwarten, dass die Rücklaufquote in allen Gruppen im Laufe der Quartalsnachbefragungen abnimmt.

Die Kumulative Rücklaufquote (KRR) ist ein wichtiger Indikator für die Teilnahmebereitschaft der Teilnehmer über mehrere Erhebungswellen hinweg und ist somit insbesondere für Langzeitstudien oder Untersuchungen mit mehreren Befragungszeitpunkten relevant. Die KRR wird im Folgenden gemäß der *Standarddefinition* der *American Association for Public Opinion Research* (AAPOR) berechnet (American Association for Public Opinion Research, 2023). Für *probability*-basierte Online-Panels sehen die AAPOR-Standards vor, dass die *Nonresponse* auf allen Stufen, einschließlich der Rekrutierungs- und Studienstufe, in die Berechnung der Rücklaufquote einfließen sollte (McPhee et al., 2022). Im konkreten Fall der Stellenerhebung wird die kumulative Rücklaufquote definiert durch die Rücklaufquote der

Rekrutierungswelle (viertes Quartal) multipliziert mit der Rücklaufquote jeder der drei Nachbefragungen (Q1-Q3).

Konkret lautet die Formel wie folgt:

$$Kumulative\ Rücklaufquote(KRR_i) = RECR(Q4) \times COMR(Q_{i=1,2,3})$$

Die kumulative Rücklaufquote ist entscheidend, um die langfristige Teilnahmebereitschaft der Teilnehmer zu messen. Ein hoher KRR-Wert ist ein Indikator dafür, dass die Teilnehmer über mehrere Erhebungswellen hinweg aktiv teilnehmen. Im Gegensatz dazu kann ein niedriger KRR-Wert auf Probleme bei der Teilnahmebereitschaft hinweisen, was zu Verzerrungen und einer verminderten Aussagekraft der Ergebnisse führen kann.

Tabelle 13: Kumulative Rücklaufquoten (KRR)

Viertes Quartal 2023 bis zweites Quartal 2024, Anzahl Bruttostichprobe sowie jeweilige Anteile in Prozent

Leere Zelle	Bruttostichprobe			KRR (ungewichtet)			KRR (gewichtet)		
	Standard	Treatmentgruppen		Standard	Treatmentgruppen		Standard	Treatmentgruppen	
	Std.-1-Seiter	2-Seiter	4-Seiter	Std.-1-Seiter	2-Seiter	4-Seiter	Std.-1-Seiter	2-Seiter	4-Seiter
Q4/2023	110.000	30.000		14,70	14,51		14,59	14,43	
Q1/2024	14.384	1.944	1.944	10,30	9,85	9,86	10,22	9,79	9,82
Q2/2024	14.384	1.944	1.944	9,86***	9,59	9,33*	9,79***	9,55	9,28**

Anmerkungen: Die Sterne zeigen einen statistisch signifikanten Unterschied (Z-test) in der Rücklaufquote zwischen den Quartalen an: *p < 0,05 ** p < 0,01 *** p < 0,001.

Quelle: IAB-Stellenerhebung; eigene Berechnungen. © IAB

Tabelle 13 zeigt die kumulativen Rücklaufquoten (ungewichtet und gewichtet) für das erste und zweite Quartal. Es zeigt sich, dass die Rücklaufquoten über die Quartale hinweg relativ stabil bleiben, obwohl sich im zweiten Quartal einige signifikante Rückgänge zeigen. Insbesondere die 4-Seiter-Gruppe weist in der ungewichteten (0,53 Prozentpunkte) und in der gewichteten (0,54 Prozentpunkte) KRR eine signifikante Abnahme gegenüber dem ersten Quartal auf. Die Std.-1-Seiter-Gruppe zeigt eine vergleichsweise höhere KRR als die anderen Gruppen, jedoch auch eine statistisch signifikante Abnahme im zweiten Quartal (gewichtet 0,43 Prozentpunkte) gegenüber dem ersten Quartal. Die geringste Abnahme in der gewichteten Rücklaufquote verzeichnet die 2-Seiter-Gruppe, mit einem Rückgang von 0,24 Prozentpunkten.

Insgesamt verdeutlichen die Ergebnisse, dass trotz der erwarteten *Survey Attrition* über die Quartale hinweg die 2-Seiter-Gruppe über die Zeit die stabilste Teilnahmequote aufweist, was darauf hindeutet, dass dieses Format zu einer höheren langfristigen Teilnahmebereitschaft beitragen könnte.

Repräsentativität: Rücklaufquoten von Untergruppen

Subgruppen- Rücklaufquoten weisen auf Unterschiede in der Teilnahmebereitschaft über verschiedene Kategorien vorhandener beobachtbarer Merkmale hin. In den drei vorliegenden Tabellen werden die Subgruppen- Rücklaufquoten nach den Schichtungsvariablen Wirtschaftszweig, Betriebsgrößenklassen und Region für das vierte Quartal sowie das erste und zweite Quartal dargestellt.

Tabelle 14: Vergleich ausgewählter Rücklaufquoten nach 24 Wirtschaftszweigen sowie Cohen's h

Viertes Quartal 2023 bis zweites Quartal 2024, Anteile in Prozent

Leere Zelle	Q4/2023 Vergleich Std. vs. Exp.			Q1/2024 Vgl. Treatmentgruppen			Q2/2024 Vgl. Treatmentgruppen		
	Std.	Exp.	Cohen's h	2-Seiter	4-Seiter	Cohen's h	2-Seiter	4-Seiter	Cohen's h
Land- und Forstwirtschaft (A)	20,3	20,0	0,008	69,0	75,0	0,133	63,4	62,5	0,018
Bergbau, Steine und Erden (B)	12,3	11,0	0,037	81,2	80,0	0,031	75,0	100,0*	1,047
Ernährung, Textil, Bekleidung, Möbel u.a. (C1)	15,9	15,9	0,001	71,8	64,4	0,160	64,1	67,3	0,068
Holz, Papier, Druck (C2)	17,5	18,6	0,028	68,8	76,8	0,181	72,5	73,2	0,015
Chemie, Kunststoff, Glas, Baustoffe (C3)	16,2	16,6	0,010	74,5	66,3	0,179	66,3	66,3	0,000
Metalle, Metallerzeugung (C4)	19,3	17,3	0,051	72,5	68,9	0,079	75,8	72,2	0,082
Maschinen, Elektrotechnik, Fahrzeuge (C5)	16,6	15,8	0,019	71,0	68,0	0,064	66,0	66,0	0,000
Energieversorgung (D)	13,4	12,7	0,018	62,9	67,6	0,098	51,4	51,4	0,001
Wasserversorgung, Abfallentsorgung (E)	17,0	18,3	0,034	70,5	77,8	0,166	70,5	71,6	0,024
Baugewerbe (F)	19,3	19,1	0,005	59,8	69,9	0,212	63,4	67,5	0,085
Handel, Reparatur (G)	12,1	12,9	0,024	72,5	67,4	0,111	68,1	55,8	0,254
Verkehr, Lagerei (H)	14,1	13,8	0,008	61,4	63,6	0,045	63,9	63,6	0,004
Gastgewerbe (I)	12,3	11,0	0,040	64,4	52,0	0,251	58,9	60,0	0,022
Information und Kommunikation (J)	11,9	12,9	0,030	58,3	69,1	0,223	64,6	63,9	0,013
Finanzdienste, Versicherung (K)	9,3	9,6	0,010	64,9	69,9	0,106	75,7	65,8*	0,218
Grundstücks- und Wohnungswesen (L)	19,1	18,4	0,016	73,5	70,2	0,072	77,1	63,1*	0,308
Freiberufl. wissenschaftl. und techn. Dienste (M)	15,3	14,8	0,015	64,0	77,1	0,290	62,8	71,1	0,176
Sonst. wirtschaftl. Dienste ohne Arbeitnehmerüberlassung (N ohne 782 und 783)	12,8	12,4	0,012	76,9	67,0	0,221	57,1	69,1	0,249
Arbeitnehmerüberlassung (WZ 782 und 783)	9,1	8,9	0,009	63,2	65,2	0,039	60,3	59,1	0,024
Öffentliche Verwaltung (O)	17,2	15,5	0,044	67,9	61,0	0,144	69,1	57,3	0,245
Erziehung und Unterricht (P)	15,5	15,8	0,009	60,2	66,0	0,119	60,2	59,8	0,008
Gesundheits- und Sozialwesen (Q)	14,6	14,9	0,007	72,0	68,7	0,073	65,3	57,4	0,161
Kunst, Unterhaltung, Erholung (R)	14,0	11,5*	0,075	56,1	56,1	0,000	59,6	47,4	0,246
Sonstige Dienstleistungen (S)	17,6	18,9	0,032	72,5	70,2	0,051	73,6	69,1	0,099

Anmerkungen: Die Sterne zeigen einen statistisch signifikanten Unterschied (Z-test) in der Rücklaufquote zwischen den betrachteten Teilgruppen an: *p < 0,05 ** p < 0,01 *** p < 0,001. Die Brancheneinteilung erfolgt gemäß der 24 Wirtschaftszweigaggregate aus der Stratifizierung.

Quelle: IAB-Stellenerhebung; eigene Berechnungen. © IAB

Bezogen auf die Wirtschaftszweige zeigt Tabelle 14, dass im vierten Quartal 2023 nur im Bereich „Kunst, Unterhaltung, Erholung“ ein statistisch signifikanter Unterschied zwischen der Standard- und der Experimentalgruppe nachgewiesen werden kann, wobei die Standardgruppe eine höhere Rücklaufquote aufweist. Der Cohen's h-Wert von 0,075 deutet jedoch auf einen vernachlässigbaren Effekt hin, da er deutlich unter der Schwelle von 0,2 für einen kleinen Effekt liegt.¹³

Im ersten Quartal 2024 sind keine statistisch signifikanten Unterschiede zwischen der 2-Seiter- und der 4-Seiter-Gruppe zu beobachten. Allerdings zeigen einige Wirtschaftszweige, wie die Erbringung von „Freiberufliche wissenschaftl. und techn. Dienste“ ($h=0,290$) und „Sonstige wirtschaftliche Dienste ohne Arbeitnehmerüberlassung“ ($h=0,221$) kleine Effekte.

Im zweiten Quartal 2024 können signifikante Unterschiede in den Wirtschaftszweigen Bergbau, Finanz- und Versicherungswesen, im „Grundstücks- und Wohnungswesen“ sowie im „Bergbau Steine und Erden“ nachgewiesen werden.¹⁴

Im „Finanzdienste- und Versicherungen“ und im „Grundstücks- und Wohnungswesen“ weist die 2-Seiter-Gruppe jeweils eine signifikant höhere Rücklaufquote auf ($h = 0,218$, $h = 0,308$).

Insgesamt zeigen die Ergebnisse, dass die Unterschiede in den Rücklaufquoten zwischen den Gruppen über die Quartale und Wirtschaftszweige hinweg variieren. Während in einigen Fällen statistisch signifikante Unterschiede auftreten, sind diese nicht konsistent über alle Zeiträume und Sektoren hinweg. Gemäß den Cohen's h-Werten liegen die meisten beobachteten Effekte im Bereich kleiner Größenordnungen ($h < 0,5$), mit Ausnahme des Bergbausektors im zweiten Quartal 2024.

Interessanterweise zeigen im zweiten Quartal sowohl der Finanz- als auch der Immobiliensektor eine Präferenz für den kürzeren Fragebogen, was auf mögliche sektorspezifische Unterschiede in der Reaktion auf Fragebogenlänge hindeutet.

Bei den Betriebsgrößenklassen (Tabelle 15) zeigen sich im vierten Quartal 2023 keine signifikanten Unterschiede zwischen der Standard- und der Experimentalgruppe mit Ausnahme der Klasse " ≥ 1000 ", wo ein statistisch signifikanter, aber sehr kleiner Effekt ($h = 0,008$) zu beobachten ist. Die Cohen's h-Werte deuten durchweg auf vernachlässigbare Effekte hin.

¹³ Ein Wert von 0,2 weist auf einen kleinen, 0,5 auf einen mittleren und 0,8 auf einen großen Effekt hin.

¹⁴ Im Bergbau ist die Rücklaufquote in der 4-Seiter-Gruppe deutlich höher, mit einem sehr großen Effekt ($h = 1,047$). Dieser Wert überschreitet die Schwelle von 0,8 für einen großen Effekt deutlich. Dieses Ergebnis sollte aber aufgrund der möglicherweise kleinen Fallzahlen mit Vorsicht interpretiert werden.

Tabelle 15: Vergleich ausgewählter Rücklaufquoten nach 7 Betriebsgrößenklassen sowie Cohen's h
Viertes Quartal 2023 bis zweites Quartal 2024, Anteile in Prozent

Leere Zelle	Q4/2023 Vergleich Std. vs. Exp.			Q1/2024 Vgl. Treatmentgruppen			Q2/2024 Vgl. Treatmentgruppen		
	Std.	Exp.	Cohen's h	2-Seiter	4-Seiter	Cohen's h	2-Seiter	4-Seiter	Cohen's h
< 10	16,6	16,0	0,017	64,6	65,0	0,007	61,0	60,6	0,009
10-49	16,5	16,9	0,010	66,5	69,6	0,066	69,0	65,4	0,077
20-49	15,8	15,5	0,006	72,6	73,1	0,011	67,8	68,5	0,016
450-249	12,9	12,7	0,007	71,2	68,3	0,062	71,2	64,8	0,137
250-499	7,9	7,5	0,011	58,2	51,8	0,128	56,4	64,3	0,162
500-999	6,8	6,6	0,008	75,0	42,3*	0,678	50,0	38,5	0,232
>= 1000	6,2	6,0*	0,008	38,6	50,0	0,232	38,5	60,0	0,434

Anmerkungen: Die Sterne zeigen einen statistisch signifikanten Unterschied (Z-test) in der Rücklaufquote zwischen den betrachteten Teilgruppen an: *p < 0,05 ** p < 0,01 *** p < 0,001. Die Einteilung erfolgt gemäß der 7 Betriebsgrößenklassen aus der Stratifizierung.

Quelle: IAB-Stellenerhebung; eigene Berechnungen. © IAB

Im ersten Quartal 2024 ist ein statistisch signifikanter Unterschied in der Klasse „500-999“ Beschäftigte zu beobachten. Hier weist die 2-Seiter-Gruppe eine deutlich höhere Rücklaufquote auf, mit einem mittleren bis großen Effekt (h = 0,678). Dies könnte darauf hindeuten, dass größere Betriebe in diesem Quartal möglicherweise besser auf den kürzeren Fragebogen reagiert haben.

Im zweiten Quartal 2024 sind keine signifikanten Unterschiede erkennbar, wobei die Cohen's h-Werte größtenteils auf kleine Effekte hindeuten. Auffällig ist jedoch der moderate Effekt (h = 0,434) in der Größenklasse ">= 1000", auch wenn dieser nicht statistisch signifikant ist.

In Bezug auf die regionale Verteilung (Tabelle 16) zeigen sich über alle drei Quartale hinweg keine signifikanten Unterschiede zwischen der Standard- und der Experimentalgruppe bzw. der 2-Seiter- und der 4-Seiter-Gruppe. Die Cohen's h-Werte weisen durchgehend auf vernachlässigbare bis sehr kleine Effekte hin.

Tabelle 16: Vergleich ausgewählter Rücklaufquoten in Ost- und Westdeutschland sowie Cohen's h
Viertes Quartal 2023 bis zweites Quartal 2024, Anteile in Prozent

Leere Zelle	Q4/2023 Vergleich Std. vs. Exp.			Q1/2024 Vgl. Treatmentgruppen			Q2/2024 Vgl. Treatmentgruppen		
	Std.	Exp.	Cohen's h	2-Seiter	4-Seiter	Cohen's h	2-Seiter	4-Seiter	Cohen's h
Westdeutschland	15,2	14,9	0,008	68,1	68,3	0,002	65,9	64,6	0,026
Ostdeutschland	13,7	13,7	0,001	67,4	67,4	0,001	67,0	63,8	0,066

Anmerkungen: Die Sterne zeigen einen statistisch signifikanten Unterschied (Z-test) in der Rücklaufquote zwischen den betrachteten Teilgruppen an: *p < 0,05 ** p < 0,01 *** p < 0,001. Die Einteilung erfolgt gemäß der 2 Regionalklassen aus der Stratifizierung.

Quelle: IAB-Stellenerhebung; eigene Berechnungen. © IAB

Insgesamt legen diese Befunde nahe, dass die Unterschiede in den Subgruppen- Rücklaufquoten zwischen den Gruppen minimal sind und die Repräsentativität der Ergebnisse nicht beeinträchtigen. Die Fragebogenlänge scheint unterschiedliche Auswirkungen auf verschiedene Subgruppen zu haben, wobei der Wirtschaftszweig den stärksten Einfluss zu haben scheint. Die Effekte sind jedoch meist klein bis moderat und variieren über den Quartalen.

Teilnahmewahrscheinlichkeit im multivariaten Modell

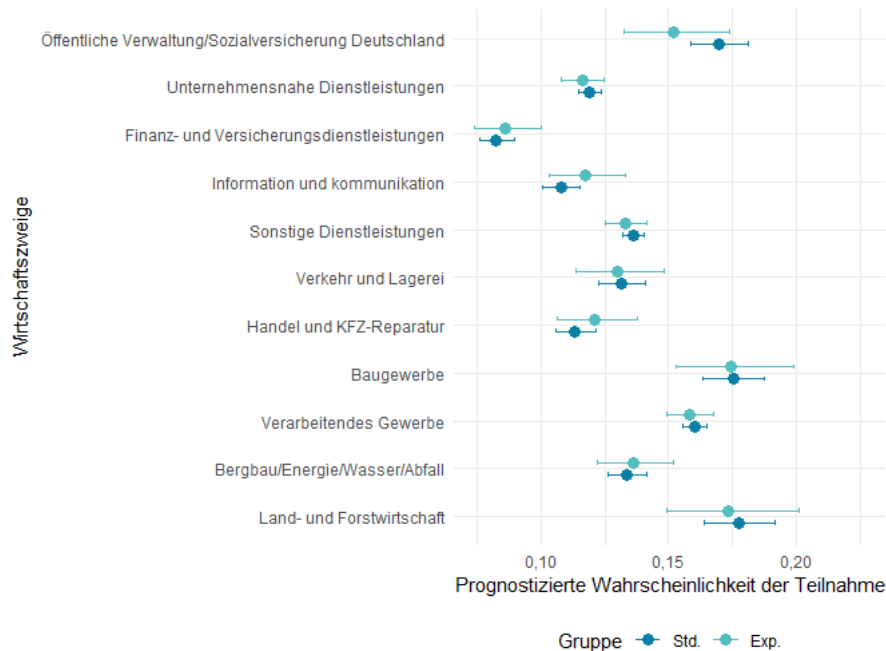
Im vorliegenden Abschnitt werden die Faktoren, die die Teilnahmequoten an der Erhebung in Abhängigkeit von der Länge des Fragebogens beeinflussen, in einem multivariaten Modell untersucht, welches mögliche Korrelationen zwischen verschiedenen erklärenden Faktoren berücksichtigt. Das Hauptziel ist es zu bewerten, ob die Länge des Fragebogens die Teilnahmequoten signifikant beeinflusst und ob sich dieser Effekt zwischen den verschiedenen Untergruppen (Std.-1-Seiter-, 2-Seiter- und 4-Seiter-Fragebögen) unterscheidet. Zu diesem Zweck wird ein logistisches Regressionsmodell geschätzt, bei dem die Teilnahme an der Umfrage die abhängige Variable ist. Die Teilnahme wird auf die gleiche Weise wie bei der Berechnung der Rücklaufquote definiert. Die Treatmentgruppen sind die primären unabhängigen Variablen. Zudem wurden Indikatoren für die Wirtschaftszweige und Betriebsgrößenklassen als Kontrollvariablen einbezogen, um mögliche Korrelationen zwischen den Stratifizierungsvariablen und Treatmentgruppen berücksichtigen zu können¹⁵.

Um die mögliche Heterogenität in der Wirkung der Fragebogenlänge zu untersuchen, wurden Interaktionseffekte zwischen der Fragebogensgruppe und den Stratifizierungsvariablen in das Modell aufgenommen. Dadurch kann überprüft werden, ob sich der Einfluss der Fragebogenlänge auf die Teilnahmequoten in verschiedenen Wirtschaftszweigen oder Betriebsgrößen unterscheidet. Die vollständigen Ergebnisse der logistischen Regressionen sind in Anhang 2 dargestellt.

Im Anschluss an die logistische Regression wurden mithilfe des *emmeans* Paket (Lenth, 2024) marginale Mittelwerte ermittelt, um die vorhergesagten Teilnahmewahrscheinlichkeiten für spezifische Kombinationen von Treatmentgruppe, Wirtschaftszweig und Betriebsgröße zu berechnen. Dies ermöglichte es, die Unterschiede zwischen den Gruppen innerhalb jeder Stratifizierungsvariable als Kontraste statistisch zu validieren und grafisch darzustellen.

¹⁵ In dieser Analyse konzentrieren wir uns auf die Betriebsgrößenklassen und die Wirtschaftszweige als zentrale erklärende Faktoren, die die Teilnahmequoten an der Umfrage beeinflussen. Obwohl die Region ein weiterer potenzieller Stratifizierungsfaktor sein könnte, wurde sie aufgrund theoretischer Überlegungen ausgeschlossen, die die Betriebsmerkmale und sektoralen Unterschiede in den Vordergrund stellen. Die Einbeziehung der Region hat keinen zusätzlichen Erklärungswert in Bezug auf die untersuchten Treatmenteffekte geliefert und hätte unnötige Komplexität in das Modell eingeführt, ohne eine ausreichende theoretische Grundlage zu haben.

Abbildung 9: Teilnahmewahrscheinlichkeit nach Wirtschaftszweigen (11 Kategorien) im vierten Quartal 2023
 Viertes Quartal 2023, Anteile in Prozent



Anmerkungen: Die vollständigen Ergebnisse sind in Anhang 2 dargestellt. Die Ergebnisse wurden mit Inklusionswahrscheinlichkeitsgewichte gewichtet. Konfidenzintervalle wurden auf dem 95 Prozent-Niveau berechnet. Signifikante Gruppenunterschiede (Kontrasts) sind durch eckige Klammern dargestellt. * $p < 0,05$ ** $p < 0,01$ *** $p < 0,001$. Die Brancheneinteilung erfolgt gemäß der 11 hochaggregierten Wirtschaftszweigaggregate auf der IAB-Homepage.

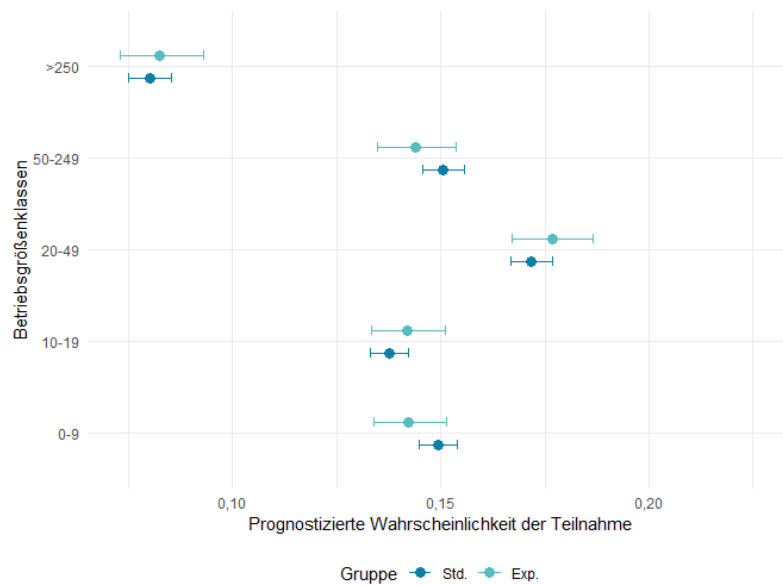
Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

Abbildung 9 zeigt die vorhergesagten Teilnahmewahrscheinlichkeiten im vierten Quartal 2023 nach Wirtschaftszweigen, welche der einfacheren Darstellung halber auf elf Kategorien reduziert wurden. Insgesamt deuten die Ergebnisse darauf hin, dass es wenig signifikante Unterschiede in der Teilnahmewahrscheinlichkeit zwischen der Standard- und der Experimentalgruppe in den einzelnen Wirtschaftszweigen gibt. Eine Ausnahme bildet die Öffentliche Verwaltung, deren Betriebe in der Experimentalgruppe eine geringere Wahrscheinlichkeit als in der Standardgruppe aufweisen, an der Umfrage teilzunehmen. Dennoch ist dieser Unterschied nicht signifikant.

Die Ergebnisse in Abbildung 10, in der die Teilnahmewahrscheinlichkeit nach Betriebsgrößenklassen im vierten Quartal dargestellt werden, deuten auf ein ähnliches Bild hin. Die Betriebsgrößenklassen wurden hier der besseren Darstellung halber auf fünf Kategorien reduziert. Auch hier können keine signifikanten Unterschiede zwischen der Experimental- und der Standardgruppe in den einzelnen Betriebsgrößenklassen nachgewiesen werden. Die Abbildung unterstreicht außerdem den bekannten Trend in der IAB-Stellenerhebung, dass größere Betriebe eine geringere Wahrscheinlichkeit als kleinere Betriebe haben, an der Umfrage teilzunehmen.

Abbildung 10: Teilnahmewahrscheinlichkeit nach 5 Betriebsgrößenklassen im vierten Quartal 2023

Viertes Quartal 2023, Anteile in Prozent

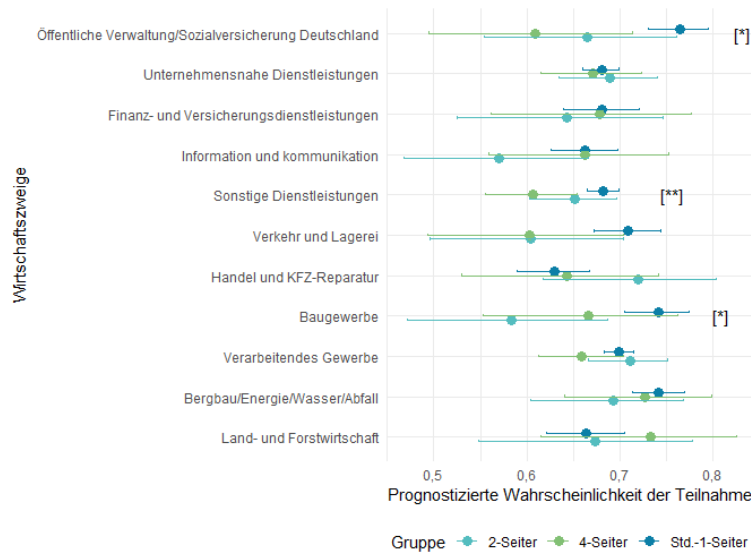


Anmerkungen: Die vollständigen Ergebnisse sind in Anhang 2 dargestellt. Die Ergebnisse wurden mit Inklusionswahrscheinlichkeitsgewichte gewichtet. Konfidenzintervalle wurden auf dem 95%-Niveau berechnet. Signifikante Gruppenunterschiede (Kontrasts) sind durch eckige Klammern dargestellt * $p < 0,05$ ** $p < 0,01$ *** $p < 0,001$.

Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

Abbildung 11 und Abbildung 12 zeigen die vorhergesagten Teilnahmewahrscheinlichkeiten für das erste Quartal, jeweils nach Wirtschaftszweigen und Betriebsgrößenklassen. Gemäß Abbildung 11 gibt es signifikante Unterschiede zwischen der Std.-1-Seiter-, der 2-Seiter- und der 4-Seiter-Gruppe. So zeigt sich ein signifikanter Unterschied zwischen der Std.-1-Seiter- und der 2-Seiter-Gruppe im Baugewerbe ($p = 0,011$), wobei die 2-Seiter-Gruppe verglichen mit der Std.-1-Seiter-Gruppe im Durchschnitt eine geringere Wahrscheinlichkeit aufweist an der Umfrage teilzunehmen. Ein weiterer signifikanter Unterschied besteht zwischen der Std.-1-Seiter- und der 4-Seiter-Gruppe in den Sonstigen Dienstleistungen ($p = 0,009$). Schließlich zeigt sich auch ein signifikanter Unterschied zwischen der Std.-1-Seiter- und der 4-Seiter-Gruppe in der Öffentlichen Verwaltung/Sozialversicherung Deutschland ($p = 0,012$), wobei die Std.-1-Seiter-Gruppe eine höhere Wahrscheinlichkeit als die 4-Seiter-Gruppe aufwies, an der Umfrage teilzunehmen.

Abbildung 11: Teilnahmewahrscheinlichkeit nach Wirtschaftszweigen (11 Kategorien) im ersten Quartal 2024
Erstes Quartal 2024, Anteile in Prozent



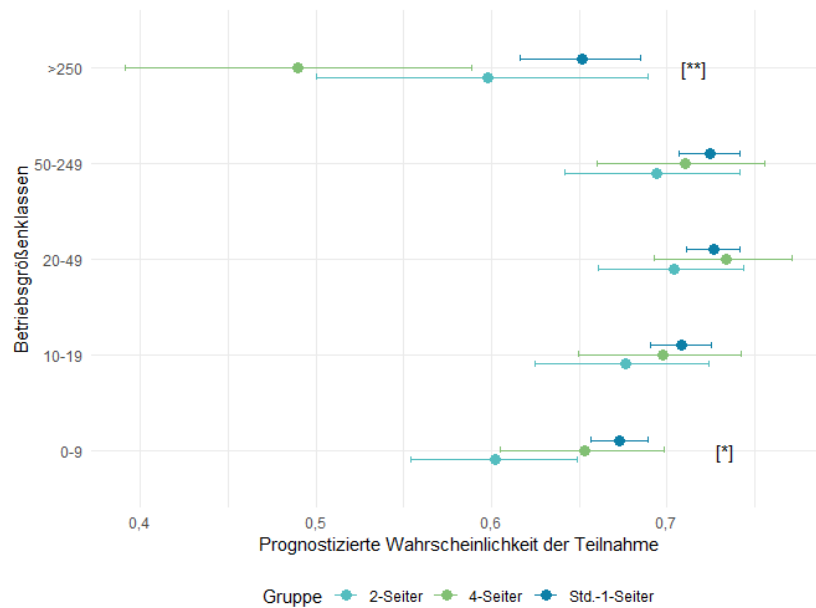
Anmerkungen: Die vollständigen Ergebnisse sind in Anhang 2 dargestellt. Die Ergebnisse wurden mit Inklusionswahrscheinlichkeitsgewichte gewichtet. Konfidenzintervalle wurden auf dem 95 Prozent-Niveau berechnet. Signifikante Gruppenunterschiede (Kontrasts) sind durch eckige Klammern dargestellt. * $p < 0,05$ ** $p < 0,01$ *** $p < 0,001$. Die Brancheneinteilung erfolgt gemäß der 11 hochaggregierten Wirtschaftszweigaggregate auf der IAB-Homepage.

Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

Abbildung 12, die die Teilnahmewahrscheinlichkeit nach Betriebsgrößenklassen darstellt, zeigt einen signifikanten Unterschied ($p=0,006$) zwischen der Std.-1-Seiter- und der 4-Seiter-Gruppe in der Betriebsgrößenklassen-Kategorie mit mehr als 250 Mitarbeitenden. Obwohl die Konfidenzintervalle sehr groß ausfallen, zeigt sich, dass unter den Betrieben mit mehr als 250 Mitarbeitenden die Betriebe der 4-Seiter-Gruppe im Durchschnitt eine geringere Wahrscheinlichkeit aufweisen zu antworten als die der Std.-1-Seiter-Gruppe. Ein weiterer signifikanter Unterschied ($p = 0,013$) besteht zwischen der Std.-1-Seiter- und der 2-Seiter-Gruppe in der Betriebsgrößenklasse „1–9 Mitarbeitende“. In dieser Kategorie weist die 2-Seiter-Gruppe im Durchschnitt eine geringere Wahrscheinlichkeit auf an der Befragung teilzunehmen, verglichen mit der Std.-1-Seiter-Gruppe.

Abbildung 12: Teilnahmewahrscheinlichkeit nach 5 Betriebsgrößenklassen im ersten Quartal 2024

Erstes Quartal 2024, Anteile in Prozent

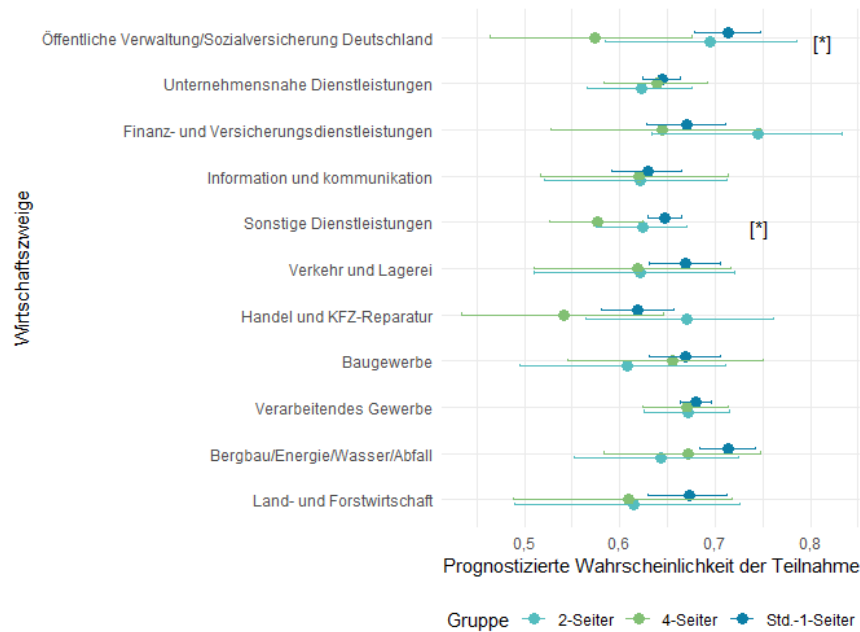


Anmerkungen: Die vollständigen Ergebnisse sind in Anhang 2 dargestellt. Die Ergebnisse wurden mit Inklusionswahrscheinlichkeitsgewichte gewichtet. Konfidenzintervalle wurden auf dem 95 Prozent-Niveau berechnet. Signifikante Gruppenunterschiede (Kontrasts) sind durch eckige Klammern dargestellt * $p < 0,05$ ** $p < 0,01$ *** $p < 0,001$.

Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

Abbildung 13 zeigt die Teilnahmewahrscheinlichkeit nach Wirtschaftszweigen für das zweite Quartal 2024. Auch hier gibt es zwei signifikante Unterschiede: in den Kategorien Öffentliche Verwaltung/Sozialversicherung Deutschland und in den Sonstigen Dienstleistungen. Im ersten Fall zeigt sich das Muster, das bereits im ersten Quartal beobachtet wurde, auch im zweiten Quartal, wobei die Std.-1-Seiter-Gruppe im Durchschnitt signifikant wahrscheinlicher ($p = 0,031$) an der Umfrage teilnimmt als die 4-Seiter-Gruppe. Dasselbe gilt auch für den Wirtschaftszweig Sonstige Dienstleistungen, in dem wie im ersten Quartal die 4-Seiter-Gruppe eine geringere Wahrscheinlichkeit aufweist teilzunehmen ($p = 0,019$).

Abbildung 13: Teilnahmewahrscheinlichkeit nach Wirtschaftszweigen (11 Kategorien) im zweiten Quartal 2024
Zweites Quartal 2024, Anteile in Prozent

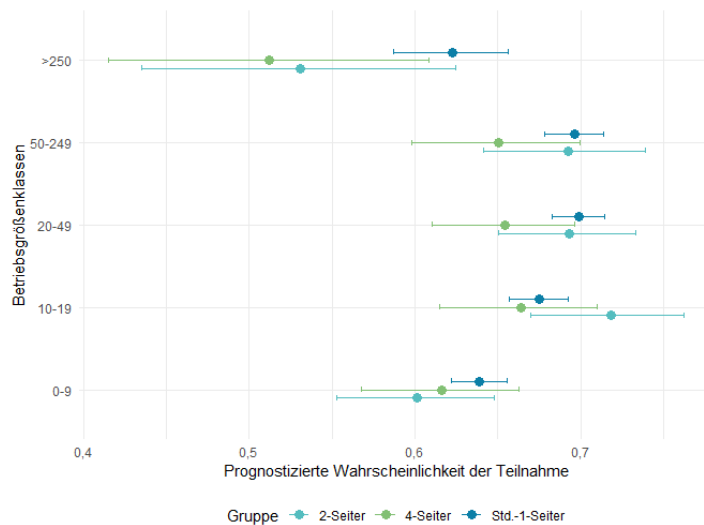


Anmerkungen: Die vollständigen Ergebnisse sind in Anhang 2 dargestellt. Die Ergebnisse wurden mit Inklusionswahrscheinlichkeitsgewichte gewichtet. Konfidenzintervalle wurden auf dem 95 Prozent-Niveau berechnet. Signifikante Gruppenunterschiede (Kontrasts) sind durch eckige Klammern dargestellt. *p < 0,05 ** p < 0,01 *** p < 0,001. Die Brancheneinteilung erfolgt gemäß der 11 hochaggregierten Wirtschaftszweigaggregate auf der IAB-Homepage.

Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

Abbildung 14 zeigt die Ergebnisse der Teilnahmewahrscheinlichkeit nach Betriebsgrößenklassen im zweiten Quartal 2024. Die signifikanten Unterschiede, die im ersten Quartal in den Größenkategorien 1-9 und über 250 beobachtet wurden, sind im zweiten Quartal nicht mehr festzustellen.

Abbildung 14: Teilnahmewahrscheinlichkeit nach 5 Betriebsgrößenklassen im zweiten Quartal 2024
Zweites Quartal 2024, Anteile in Prozent



Anmerkungen: Die vollständigen Ergebnisse sind in Anhang 2 dargestellt. Die Ergebnisse wurden mit Inklusionswahrscheinlichkeitsgewichte gewichtet. Konfidenzintervalle wurden auf dem 95 Prozent-Niveau berechnet. Signifikante Gruppenunterschiede (Kontrasts) sind durch eckige Klammern dargestellt * $p < 0,05$ ** $p < 0,01$ *** $p < 0,001$.

Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

Insgesamt zeigen die Ergebnisse, dass es nur wenig konsistente signifikante Unterschiede in der Teilnahmewahrscheinlichkeit zwischen den verschiedenen Gruppen gibt. So weist die 2-Seiter-Gruppe in den kleineren Betriebsgrößenklassen (1–9 Mitarbeitende) eine geringere Wahrscheinlichkeit auf, an der Umfrage teilzunehmen, verglichen mit der Std.-1-Seiter-Gruppe. Ebenso zeigt sich ein signifikanter Unterschied zwischen der Std.-1-Seiter- und der 4-Seiter-Gruppe in der Betriebsgrößenklassen-Kategorie mit mehr als 250 Mitarbeitenden sowie in bestimmten Wirtschaftszweigen, wie dem Baugewerbe und den Sonstigen Dienstleistungen. Diese Befunde deuten darauf hin, dass die Länge des Fragebogens in spezifischen Kontexten, abhängig von der Betriebsgröße und dem Wirtschaftszweig, einen Einfluss auf die Teilnahmebereitschaft haben kann.

Nonresponse-Bias

Ein *Nonresponse-Bias* tritt auf, wenn systematische Unterschiede zwischen den Teilnehmern und Nicht-Teilnehmern der Umfrage bestehen, was potenziell zu Ergebnissen führen kann, die die Zielpopulation nicht vollständig repräsentieren (Dillman et al., 2014). Selbst bei sorgfältiger Stratifizierung und Untersuchung der Rücklaufquoten in den Untergruppen kann die Möglichkeit eines *Nonresponse-Bias* nicht vollständig ausgeschlossen werden. Wie Groves and Peytcheva (2008) betonen, können trotz ähnlicher Rücklaufquoten in verschiedenen Untergruppen wesentliche Unterschiede zwischen Teilnehmern und Nicht-Teilnehmern innerhalb dieser Untergruppen bestehen. Daher ist es notwendig, eine gezielte Analyse und Quantifizierung des *Nonresponse-Bias* durchzuführen, um ein klareres Bild von

der Repräsentativität unserer Umfrageantworten zu erhalten. Das Verständnis des Ausmaßes und der Richtung des *Nonresponse-Bias* ist entscheidend, um die Umfrageergebnisse korrekt zu interpretieren und fundierte Entscheidungen auf Basis der Daten treffen zu können (Wagner, 2012).

Der vorliegende Abschnitt enthält eine detaillierte Untersuchung des potenziellen *Nonresponse-Bias* in Rahmen der IAB-Stellenerhebung. Der *Nonresponse-Bias* ergibt sich als Differenz zwischen der Schätzung der interessierenden Größe (z.B. dem Mittelwert einer Variablen) welche basierend auf den Antworten der Teilnehmenden ermittelt wird und dem entsprechenden Wert welcher auf Basis der Bruttostichprobe ermittelt wird (Küfner et al., 2022). Der *Nonresponse-Bias* wird wie folgt berechnet:

$$\text{Nonresponse Bias}_i = \hat{Y}_{i,r} - \hat{Y}_{i,n}$$

Dabei ist $\hat{Y}_{i,r}$ der Schätzwert für die Variablenkategorie i , basierend auf den Antworten der Teilnehmer, und $\hat{Y}_{i,n}$ ist der entsprechende Schätzwert basierend auf der gesamten Bruttostichprobe (Teilnehmer und Nicht-Teilnehmer).

Der *Absolute Nonresponse Bias* ist als der Betrag dieser Differenz definiert:

$$\text{Absolute Nonresponse Bias}_i = |\text{Nonresponse Bias}_i|$$

Der durchschnittliche Absolute Nonresponse Bias einer Variablen ergibt sich als der Durchschnitt dieser absoluten Verzerrungen über alle Kategorien hinweg:

$$\text{Average Absolute Nonresponse Bias}_i = \frac{\sum_{i=1}^K \text{Absolute Nonresponse Bias}_i}{K}$$

Hierbei bezeichnet K die Anzahl der Kategorien einer bestimmten Variablen (z. B. Betriebsgrößenklassen).

Der *Average Absolute Nonresponse-Bias* (AANB) wird im Folgenden separat für die Stratifizierungsvariablen Wirtschaftszweige, Betriebsgrößenklassen und Regionen ermittelt, um zu untersuchen, welche dieser Variablen am stärksten von einem möglichen *Nonresponse-Bias* betroffen ist. Da alle Werte auf Anteilen basieren, können sie auf derselben Skala verglichen werden (Küfner et al., 2022). Tabelle 10 enthält die Ergebnisse für das vierte Quartal 2023 sowie für das erste und das zweite Quartal 2024. Für das vierte Quartal 2023 fallen die Werte des AANB für die Betriebsgrößenklassen in der Standardgruppe, die einen Bias von 2,11 Prozent aufweist, und in der Experimentalgruppe, bei der der Bias leicht höher bei 2,15 Prozent liegt, ähnlich aus. Bei den Wirtschaftszweigen liegen die Werte in beiden Gruppen ebenfalls nahe beieinander, mit 0,72 Prozent in der Standard-Gruppe und 0,75 Prozent in der Experimentalgruppe. Bei der Betrachtung der regionalen Klassifikationen (West/Ost) zeigt sich

jedoch ein deutlicher Unterschied: Die Standardgruppe weist einen *Bias* von 2,32 Prozent auf, während die Experimentalgruppe einen niedrigeren *Bias* von 1,79 Prozent zeigt.

Tabelle 17 enthält für das erste und zweite Quartal 2024 detailliertere Vergleiche zwischen der Std.-1-Seiter-Gruppe und den beiden Treatmentgruppen. Für die Betriebsgrößenklassen zeigt sich im ersten Quartal, dass die Std.-1-Seiter-Gruppe einen *Bias* von 0,50 Prozent aufweist und die 2-Seiter-Gruppe und 4-Seiter-Gruppe einen höheren *Bias* von jeweils 0,74 Prozent und 0,69 Prozent verzeichnen. Im zweiten Quartal steigt der *Bias* für die Std.-1-Seiter leicht auf 0,58 Prozent und für die 2-Seiter-Gruppe auf 0,86 Prozent an, während die 4-Seiter-Gruppe einen etwas niedrigeren Wert von 0,60 Prozent verzeichnet. Wie schon im ersten Quartal weist die 2-Seiter-Gruppe den höchsten *Bias*-Prozentsatz für Betriebsgrößen auf.

Tabelle 17: Average Absolute Nonresponse Bias im ersten und zweiten Quartal 2024

Erstes und zweites Quartal 2024, Anteile in Prozent

Leere Zelle	Q4/2023		Q1/2024			Q2/2024		
	Std.	Exp.	Std.-1-Seiter	2-Seiter	4-Seiter	Std.-1-Seiter	2-Seiter	4-Seiter
Betriebsgrößenklassen	2,11	2,15	0,50	0,74	0,69	0,58	0,86	0,60
Wirtschaftszweige	0,72	0,75	0,18	0,30	0,25	0,17	0,30	0,33
Region (West/Ost)	2,32	1,79	0,05	0,21	0,29	0,41	0,34	0,30

Anmerkung: Die Ergebnisse wurden mit Inklusionswahrscheinlichkeiten gewichtet.

Quelle: IAB-Stellenerhebung; eigene Berechnungen. © IAB

Für die Wirtschaftszweige fällt der *Bias* in allen Gruppen und in beiden Quartalen durchweg niedrig aus. Im ersten Quartal weist die Std.-1-Seiter-Gruppe einen *Bias* von 0,18 Prozent auf, die 2-Seiter-Gruppe verzeichnet einen leicht höheren *Bias* von 0,30 Prozent, und der Wert der 4-Seiter-Gruppe liegt bei 0,25 Prozent. Dieses Muster setzt sich im zweiten Quartal fort, wobei der *Bias* leicht ansteigt, aber insgesamt niedrig bleibt: 0,17 Prozent für Std.-1-Seiter-Gruppe, 0,30 Prozent für die 2-Seiter-Gruppe und 0,33 Prozent für die 4-Seiter-Gruppe. Diese Ergebnisse zeigen, dass die Repräsentation der Wirtschaftszweige generell nicht von der Fragebogenlänge beeinflusst zu sein scheint.

Schließlich zeigt die regionale Klassifizierung (West/Ost) im ersten Quartal einen minimalen *Bias*, welcher im zweiten Quartal in allen Gruppen leicht ansteigt: 0,41 Prozent für die Std.-1-Seiter-, 0,34 Prozent für die 2-Seiter- und 0,30 Prozent für die 4-Seiter-Gruppe. Interessanterweise blieb der *Bias* der 4-Seiter-Gruppe, die im Vergleich zu den anderen Gruppen im ersten Quartal einen höheren *Bias* aufweist, im zweiten Quartal stabil, während die Werte der anderen Gruppen Zuwächse verzeichnen.

Insgesamt deuten diese Ergebnisse darauf hin, dass der *Nonresponse Bias* insgesamt niedrig ausfällt und hinsichtlich der Fragebogenlänge keine der Varianten (Std.-1-Seiter, 2- oder 4-Seiter) durchgehend in allen Gruppen und Quartalen besser abschneidet als die anderen.

Response Burden

Nach der Analyse der Auswirkungen der Fragebogenlänge auf die Rücklaufquote und andere wichtige Responseindikatoren sollen im folgenden Abschnitt die Auswirkungen der Fragebogenlänge zunächst auf die sogenannte *Perceived Response Burden* untersucht werden. Wie zuvor erläutert, stellt die *Response Burden* einen wichtigen Indikator dar, über welchen mögliche Effekte der Fragebogenlänge auf die Rücklaufquote vermittelt werden können. Da es Evidenz dafür gibt, dass auch der Erhebungsmodus die *Response Burden* erklären kann (Haas et al., 2021), werden diese Fragen im Folgenden zusätzlich getrennt nach CAWI- und PAPI-Teilnehmern ausgewertet. Zur Messung der Response Burden wurden im Fragebogen zwei Fragen aufgenommen.

Die erste *Response Burden* Frage zur Messung der *Perceived Response Burden* erhebt, welche Aspekte das Ausfüllen des Fragebogens für die Betriebe erschwert haben. Konkret lautete die Frage mit den dazugehörigen Antwortoptionen, bei denen Mehrfachnennungen möglich waren:

Hat einer der folgenden Punkte das Ausfüllen des Fragebogens erschwert?

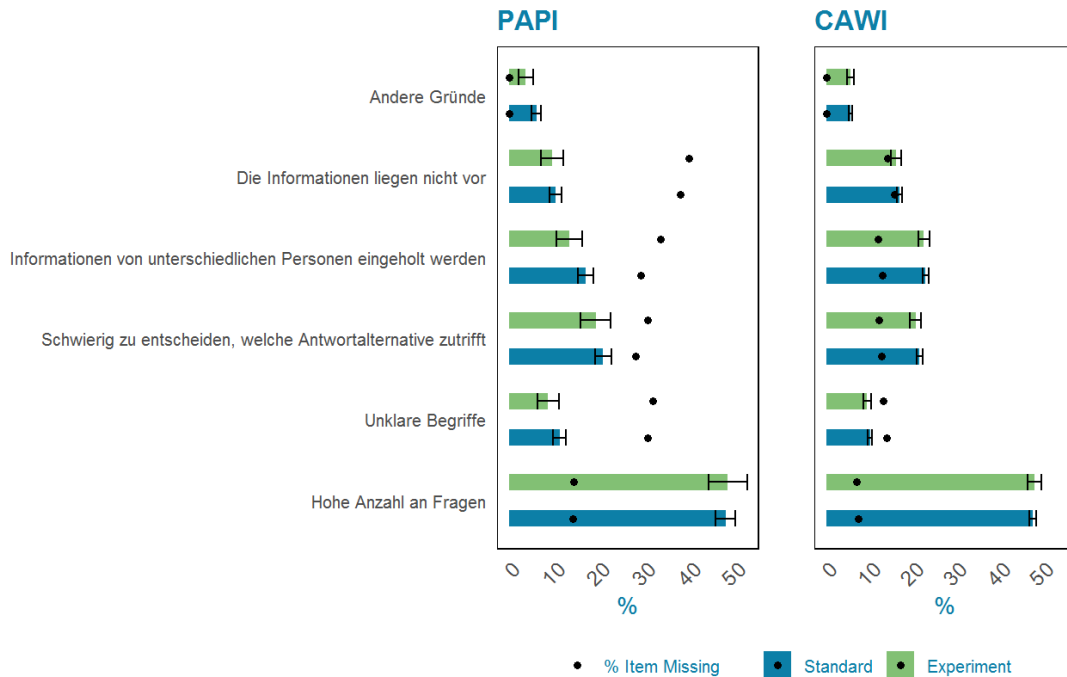
- Hohe Anzahl an Fragen
- Unklare Begriffe und Erklärungen
- Es war schwierig zu entscheiden, welche Antwortalternative zutrifft
- Informationen mussten von unterschiedlichen Personen eingeholt werden;
- Die abgefragten Informationen liegen nicht vor
- Andere Gründe.

Abbildung 15 weist die jeweiligen Anteile der Teilnehmer im vierten Quartal 2023 aus, die angegeben haben, dass die genannten Punkte das Beantworten der Umfrage erschwert haben. Links sind die Antworten der Teilnehmer im PAPI Modus und rechts die Antworten der CAWI-Teilnehmer dargestellt. Die Balken unterscheiden jeweils zwischen der Standardgruppe (grün) und der Experimentalgruppe (blau). Insgesamt ist ersichtlich, dass für beide Erhebungsmodi und Gruppen der Hauptgrund, der das Ausfüllen des Fragebogens erschwert hat, die hohe Anzahl der Fragen war. So haben in beiden Gruppen und beiden Erhebungsmodi über 48 Prozent der Teilnehmer diesen Grund als erschwerenden Faktor genannt. Weitere erschwerende Faktoren waren die Umstände, dass „Informationen von unterschiedlichen Personen eingeholt werden“ und „es schwierig zu entscheiden war, welche Antwortalternative zutrifft“. Generell zeigt sich, dass die CAWI-Teilnehmer die Umfrage generell als weniger belastend wahrnehmen als die PAPI-Teilnehmer.

Neben den Anteilen der Teilnehmer, die angegeben haben, dass die genannten Punkte das Beantworten der Umfrage erschwert haben, sind in Abbildung 15 ebenfalls die Anteile der Teilnehmer ausgewiesen, die die Fragen zur Response Burden nicht beantwortet haben. Diese Anteile, die in Abbildung 11 mit Hilfe der Punkte dargestellt sind, fallen generell für den PAPI-Modus höher aus als für den CAWI-Modus.

Abbildung 15: Response Burden im vierten Quartal 2023

Viertes Quartal 2023, Anteile in Prozent



Anmerkung: Konfidenzintervalle wurden auf dem 95 Prozent-Niveau berechnet.

Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

Tabelle 18 weist die Ergebnisse der Differenzen (Z-Test) zwischen der Standard- und der Experimentalgruppe in den jeweiligen Anteilen der Teilnehmer, die angegeben haben, dass die genannten Punkte das Beantworten der Umfrage erschwert haben, aus. Insgesamt gab es nur zwei signifikante Unterschiede, beide im PAPI-Modus. Erstens gab es einen signifikanten Unterschied beim Punkt „Informationen von unterschiedlichen Personen eingeholt werden“, mit einem p-Wert von 0,048 und einer sehr kleinen Cohen's h Effektgröße von 0,099 (unter 0,2). Zweitens zeigte sich bei dem Punkt „Andere Gründe“ ein signifikanter Unterschied mit einem p-Wert von 0,036 und einer ebenfalls sehr kleinen Effektgröße von 0,109. Obwohl die Unterschiede in diesen Fällen signifikant sind, ist die Effektgröße sehr klein. Im vierten Quartal 2023 wurde diese letzte Frage offen gestellt. Dies bedeutet, dass die Teilnehmer im Gegensatz zu der anderen Response Burden-Kategorien die Frage nicht mit „Ja“ oder „Nein“ antworten mussten, sondern andere Gründe angeben konnten, die für sie die Teilnahme erschwert haben.

Tabelle 18: Ergebnisse von Standardgruppe und Experimentalgruppe zu den Perceived Response Burden Items im vierten Quartal 2023, differenziert nach dem Erhebungsmodus

Viertes Quartal 2023, Anteile in Prozent

Item:	Erhebungsmodus							
	PAPI				CAWI			
	Std.	Exp.	P-Wert	Cohen's h	Std.	Exp.	P-Wert	Cohen's h
Hohe Anzahl an Fragen	47,92	48,40	0,843	-0,009	47,64	48,03	0,670	-0,007
Unklare Begriffe und Erklärungen	11,22	8,66	0,090	0,085	10,06	9,39	0,220	0,022
Es war schwierig zu entscheiden, welche Antwortalternative zutrifft	20,79	19,21	0,422	0,039	21,55	20,61	0,209	0,023
Informationen von unterschiedlichen Personen eingeholt werden	16,93	13,37	0,048*	0,099	22,88	22,52	0,635	0,008
Die abgefragten Informationen liegen nicht vor	10,34	9,60	0,617	0,024	16,82	16,14	0,316	0,018
Andere Gründe	6,12	3,77	0,036*	0,109	5,62	5,57	0,902	0,002

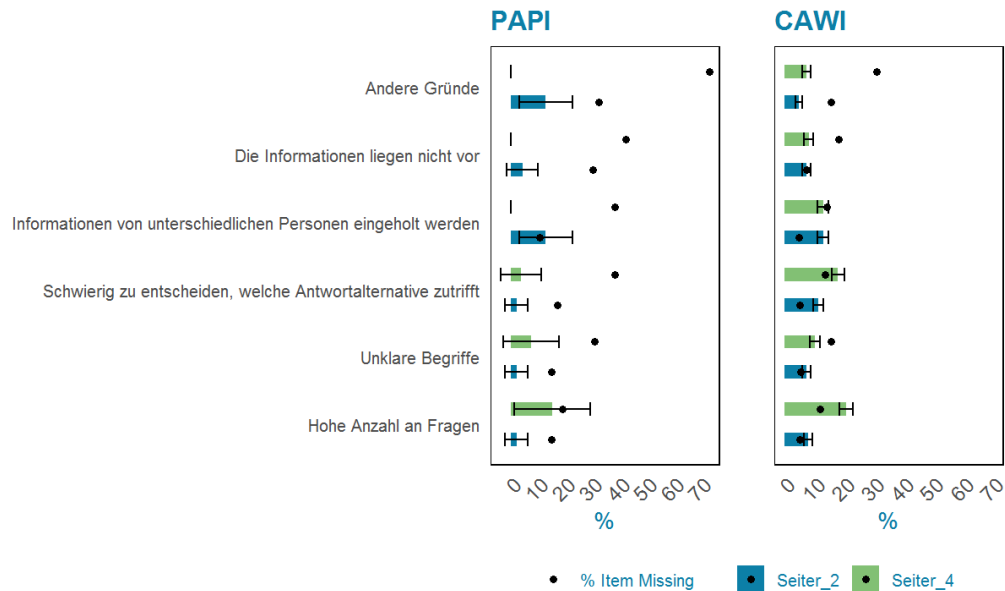
Anmerkung: p-Werte basieren auf Z-Tests. Signifikanzniveau: *p < 0,05 ** p < 0,01 *** p < 0,001.

Quelle: IAB-Stellenerhebung; eigene Berechnungen. © IAB

Abbildung 16 zeigt die jeweiligen Anteile der Teilnehmer im ersten Quartal 2024, in welchem die Experimentalgruppe in die beiden Treatmentgruppen (2-Seiter und 4-Seiter) aufgeteilt wurden. Da die Frage zur Perceived Response Burden im ersten Quartal nur den Experimentalgruppen gestellt wurde, können in dieser Abbildung anders als in Abbildung 11 die Ergebnisse für die Std.-1-Seiter-Gruppe nicht ausgewiesen werden. Links sind die Ergebnisse für den PAPI-Modus und rechts für den CAWI-Modus jeweils getrennt für die beiden Treatmentgruppen dargestellt. Aufgrund der geringen Anzahl eingegangener PAPI-Fragebögen fallen die Konfidenzintervalle für diesen Erhebungsmodus jedoch sehr groß aus.

Abbildung 16: Response Burden im ersten Quartal 2024

Erstes Quartal 2024, Anteile in Prozent



Anmerkung: Konfidenzintervalle wurden auf dem 95 Prozent-Niveau berechnet.

Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

Tabelle 19 weist ähnlich zu Tabelle 18 wieder die Ergebnisse der Vergleiche der Rücklaufquoten zwischen beiden Treatmentgruppen aus. Im PAPI- Modus gaben 15,38 Prozent der Teilnehmer in der 4-Seiter-Gruppe Schwierigkeiten aufgrund der „Hohen Anzahl an Fragen“ an. Dieser Anteil liegt deutlich höher ist als der entsprechende Anteil von 2,17 Prozent der Teilnehmer in der 2-Seiter-Gruppe. Dieser Unterschied ist signifikant ($p = 0,034$) und die Effektgröße ist ebenfalls mittelgroß ($h = -0,510$). Im CAWI-Modus können für mehrere Items signifikante Unterschiede zwischen den Treatmentgruppen nachgewiesen werden. So gaben 20,76 Prozent der 4-Seiter-Gruppe an, Schwierigkeiten mit der „Hohen Anzahl an Fragen“ gehabt zu haben, verglichen mit 7,92 Prozent in der 2-Seiter-Gruppe ($p < 0,000$). Die Effektgröße von $h = -0,375$ deutet jedoch auf einen kleinen bis mittleren Effekt hin. Zusätzlich fanden 10,19 Prozent der Teilnehmer in der 4-Seiter-Gruppe „Unklare Begriffe und Erklärungen“ problematisch, verglichen mit 7,45 Prozent in der 2-Seiter-Gruppe ($p = 0,014$). Die Effektgröße ist jedoch sehr klein ($h = 0,096$). Schließlich hatten 17,98 Prozent der 4-Seiter-Gruppe Schwierigkeiten bei der „Entscheidung, welche Antwortalternative zutrifft.“ verglichen mit 11,29 Prozent in der 2-Seiter-Gruppe ($p < 0,001$). Zudem gab die 4-Seiter Gruppe signifikant häufiger die Option „Andere Gründe“ an, obwohl die Effektgröße auch hier sehr klein ist ($h = 0,103$). Bei den Antwortkategorien, dass „Informationen von unterschiedlichen Personen eingeholt werden mussten“ und „Die abgefragten Informationen liegen nicht vor“ fallen die Unterschiede zwischen den Gruppen hingegen sehr gering aus.

Tabelle 19: Ergebnisse der Treatmentgruppen zu den Perceived Response Burden Items im ersten Quartal 2024, differenziert nach dem Erhebungsmodus

Erstes Quartal 2024, Anteile in Prozent

Item:	Erhebungsmodus							
	PAPI				CAWI			
	2-Seiter	4-Seiter	p-Wert	Cohen's h	2-Seiter	4-Seiter	p-Wert	Cohen's h
Hohe Anzahl an Fragen	2,17	15,38	0,034*	-0,510	7,92	20,76	0,000***	-0,375
Unklare Begriffe und Erklärungen	2,17	7,69	0,260	-0,266	7,45	10,19	0,014*	-0,096
Es war schwierig zu entscheiden, welche Antwortalternative zutrifft	2,17	3,85	0,678	-0,098	11,29	17,98	0,000***	-0,190
Informationen von unterschiedlichen Personen eingeholt werden	13,04	0,00	0,054	0,739	13,02	12,96	0,965	0,001
Die abgefragten Informationen liegen nicht vor	4,35	0,00	0,280	0,420	7,53	8,18	0,540	-0,024
Andere Gründe	13,04	0,00	0,054	0,739	4,86	7,33	0,009**	-0,103

Anmerkung: p-Werte basieren auf Z-Tests. Signifikanzniveau: *p < 0,05 ** p < 0,01 *** p < 0,001.

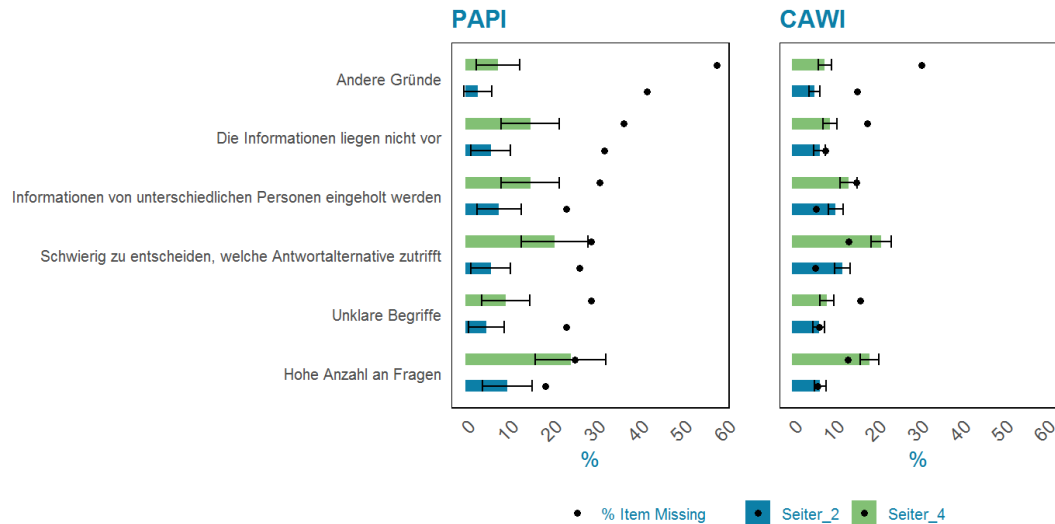
Quelle: IAB-Stellenerhebung; eigene Berechnungen. © IAB

Diese Ergebnisse deuten darauf hin, dass der 4-Seiter Fragebogen im ersten Quartal insgesamt eine größere Response Burden für die Teilnehmer mit sich brachte. Die hier dargestellten signifikanten Unterschiede basieren jedoch hauptsächlich auf dem CAWI-Modus, da im ersten Quartal 2024 insgesamt nur eine geringe Anzahl von Fragebögen mit dem PAPI Modus beantwortet wurden.

Die Ergebnisse der Analyse des zweiten Quartals, dargestellt in Abbildung 17 und Tabelle 20 bestätigen die Ergebnisse des ersten Quartals. Im PAPI-Modus berichtete die 4-Seiter-Gruppe weiterhin über signifikant höhere Schwierigkeiten als die 2-Seiter-Gruppe insbesondere in den Kategorien „Hohe Anzahl an Fragen“ (24,30 Prozent vs. 9,71 Prozent, p = 0,005, h = -0,397), Kategorien „Schwierigkeit bei der Entscheidung, welche Antwortalternative zutrifft“ (20,56 Prozent vs. 5,83 Prozent, p = 0,003, h = 0,453), und Kategorien „Die abgefragten Informationen liegen nicht vor“ (14,95 Prozent vs. 5,83 Prozent, p = 0,030, h = -0,306). Die Unterschiede sind signifikant, und die Effektgrößen liegen alle im Bereich zwischen klein und mittel.

Abbildung 17: Response Burden im zweiten Quartal 2024

Zweites Quartal 2024, Anteile in Prozent



Anmerkung: Konfidenzintervalle wurden auf dem 95 Prozent-Niveau berechnet.

Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

Im CAWI-Modus zeigen sich qualitativ ähnliche Unterschiede zwischen der 2- und 4-Seiter-Gruppe, mit signifikant höheren Anteilen in allen Antwortkategorien der 4-Seiter-Gruppe. Während die Anteile der sechs *Perceived Response Burden*-Kategorien in der 4-Seiter-Gruppe signifikant höher ausfallen, überschreitet die Effektgröße die Grenze eines kleinen Effekts nur bei den Antwortoptionen „Hohe Anzahl an Fragen“ (18,36 Prozent vs. 6,59 Prozent, $p < 0,001$, $h = -0,366$) und „Schwierigkeit bei der Entscheidung, welche Antwortalternative zutrifft“ (21,07 Prozent vs. 11,92 Prozent, $p < 0,001$, $h = -0,248$).

Tabelle 20: Ergebnisse der Treatmentgruppen zu den Perceived Response Burden Items im zweiten Quartal 2024, differenziert nach dem Erhebungsmodus

Zweites Quartal 2024, Anteile in Prozent

Item:	Erhebungsmodus							
	PAPI				CAWI			
	2-Seiter	4-Seiter	p-Wert	Cohen's h	2-Seiter	4-Seiter	p-Wert	Cohen's h
Hohe Anzahl an Fragen	9,71	24,30	0,005 **	-0,397	6,59	18,36	0,000***	-0,366
Unklare Begriffe und Erklärungen	4,85	9,35	0,206	-0,177	6,34	8,22	0,081	-0,072
Es war schwierig zu entscheiden, welche Antwortalternative zutrifft	5,83	20,56	0,001 **	-0,453	11,92	21,07	0,000***	-0,248
Informationen von unterschiedlichen Personen eingeholt werden	7,77	14,95	0,101	-0,229	10,31	13,37	0,022*	-0,095
Die abgefragten Informationen liegen nicht vor	5,83	14,95	0,030*	-0,306	6,51	9,00	0,024*	-0,093
Andere Gründe	2,91	7,48	0,137	-0,210	5,33	7,69	0,020*	-0,096

Anmerkung: p-Werte basieren auf Z-Tests. Signifikanzniveau: *p < 0,05 ** p < 0,01 *** p < 0,001.

Quelle: IAB-Stellenerhebung; eigene Berechnungen. © IAB

Insgesamt lässt sich aus der Auswertung der *Perceived Response Burden* Fragen schließen, dass der 4-Seiter-Fragebogen im ersten und zweiten Quartal durchweg als belastender wahrgenommen wird, wobei diese Unterschiede jedoch i.W. nur auf Basis der CAWI-Teilnehmer präzise ermittelt werden können.

Item Nonresponse

Item-Nonresponse stellt eine häufige Herausforderung in der Umfrageforschung dar und bedeutet, dass Teilnehmer bestimmte Fragen nicht beantworten und sich dadurch fehlende Informationen zu bestimmten abgefragten Items ergeben. Das Phänomen der Item-Nonresponse ist grundsätzlich von der *Unit-Nonresponse* zu unterscheiden, welche widerspiegelt, dass Teilnehmer überhaupt nicht an der Befragung teilnehmen. *Item-Nonresponse* kann aus verschiedenen Gründen auftreten, etwa infolge der mangelnden Bereitschaft der Teilnehmer sensible Fragen zu beantworten, mangelnder Informationen zu einem bestimmten Thema, einer Ermüdung durch den Fragebogen, aber auch infolge eines versehentlichen Überspringens von Fragen.

In diesem Abschnitt wird die Häufigkeit der *Item-Nonresponse* in Abhängigkeit von der Länge des Fragebogens untersucht. Die Analyse umfasst alle Fragen des Fragebogens mit Ausnahme der Eurostat-Fragen, die beantwortet werden müssen, damit die befragten Betriebe als gültige Teilnehmer berücksichtigt werden. Da die Teilnehmer zufällig in Gruppen (2-Seiter- vs. 4-Seiter-Gruppe) eingeteilt wurden, variiert die Anzahl der gestellten Fragen entsprechend (siehe Tabelle 8).

Aufgrund der Filterfragen in der IAB-Stellenerhebung wurde die Anzahl der für jeden Betrieb relevanten Fragen individuell angepasst. Bei der Berechnung der *Item-Nonresponse* wurde zwischen nicht anwendbaren und anwendbaren Fragen unterschieden. Ein Beispiel für eine Filterfrage aus der

Standardbefragung ist Frage 6: „Gibt es in Ihrem Betrieb/Ihrer Verwaltungsstelle Arbeitszeitkonten?“ mit den Antwortoptionen „Ja“ oder „Nein“. Bei einer „Nein“-Antwort werden die Teilnehmer gebeten direkt mit Frage 11 fortzufahren. Fehlende Antworten auf die Fragen 7 bis 10 gelten in diesem Fall nicht als *Item-Nonresponse*, da diese Fragen für den Betrieb nicht zutrafen. Diese Regel wurde für allen Gruppen angewendet: Nur fehlende Antworten auf anwendbare Fragen wurden als *Item-Nonresponse* gezählt.

Tabelle 21 weist die *Item-Nonresponse*-Raten für alle Gruppen —die Std.-1-Seiter, 2-Seiter- und 4-Seiter-Gruppe — im ersten Quartal aus. Die *Item-Nonresponse*-Rate ist als der Anteil der anwendbaren Fragen, die von den Teilnehmern nicht beantwortet wurden, an der Gesamtzahl der gestellten Fragen definiert. Wie oben dargelegt, wurde die Gesamtzahl der Fragen für jeden Teilnehmer angepasst, um die Anzahl der anwendbaren Fragen zu berücksichtigen.

Tabelle 21: Item-Nonresponse-Rate im ersten Quartal 2024 nach Gruppe und Erhebungsmodus, differenziert nach dem Erhebungsmodus

Erstes Quartal 2024, Anteile in Prozent

Erhebungsmodus	Gruppe	Item-Nonresponse-Rate	Gruppenvergleich	p-Wert	Cohen's h
Gesamt	Std.-1-Seiter	8,08	Std.-1-Seiter vs. 2-Seiter	0,008**	0,092
	2-Seiter	5,74	Std.-1-Seiter vs. 4-Seiter	0,021*	0,078
	4-Seiter	6,07	2-Seiter vs. 4-Seiter	0,718	-0,014
PAPI	Std.-1-Seiter	11,50	Std.-1-Seiter vs. 2-Seiter	1,000	0,058
	2-Seiter	9,69	Std.-1-Seiter vs. 4-Seiter	1,000	0,207
	4-Seiter	5,75	2-Seiter vs. 4-Seiter	1,000	0,148
CAWI	Std.-1-Seiter	7,98	Std.-1-Seiter vs. 2-Seiter	0,019*	0,094
	2-Seiter	5,60	Std.-1-Seiter vs. 4-Seiter	0,094	0,074
	4-Seiter	6,08	2-Seiter vs. 4-Seiter	1,000	-0,020

Anmerkungen: Die Sterne zeigen einen statistisch signifikanten Unterschied (Z-test) zwischen den Gruppen an. Für den Vergleich zwischen der Std.-1-Seiter, 2-Seiter- und 4-Seiter-Gruppe wurden die p-Werte mittels der Holm-Bonferroni-Korrektur angepasst um mehrfache Vergleiche zu berücksichtigen. Diese Anpassung verringert das Risiko von Fehlern erster Art (falsch positive Ergebnisse), das mit der Anzahl gleichzeitiger Tests zunimmt. * $p < 0,05$ ** $p < 0,01$ *** $p < 0,001$

Quelle: IAB-Stellenerhebung; eigene Berechnungen. © IAB

Die Std.-1-Seiter-Gruppe weist mit 8,08 Prozent im Vergleich zur 2-Seiter-Gruppe (5,74 Prozent) und zur 4-Seiter-Gruppe (6,07 Prozent) insgesamt eine höhere *Item-Nonresponse-Rate* auf. Über beide Erhebungsmodi hinweg zeigt sich ein statistisch signifikanter Unterschied zwischen der Std.-1-Seiter-Gruppe und der 2-Seiter-Gruppe ($p = 0,008$, $h = 0,092$), was auf einen kleinen Effekt hindeutet. Ebenso besteht ein signifikanter Unterschied zwischen der Std.-1-Seiter-Gruppe und der 4-Seiter-Gruppe ($p = 0,021$, $h = 0,078$), ebenfalls mit einem kleinen Effekt.

Im PAPI-Modus beträgt die *Item-Nonresponse-Rate* der Std.-1-Seiter-Gruppe 11,50 Prozent. Dieser Anteil fällt höher aus als in der 2-Seiter-Gruppe mit 9,69 Prozent, der Unterschied ist jedoch nicht signifikant ($p = 1,000$, $h = 0,058$). Der Unterschied zur 4-Seiter-Gruppe (5,75 Prozent) im PAPI-Modus ist ebenfalls

nicht signifikant, jedoch liegt der Effektgrößenwert leicht über der Schwelle für einen kleinen Effekt ($p = 1,000$, $h = 0,207$).

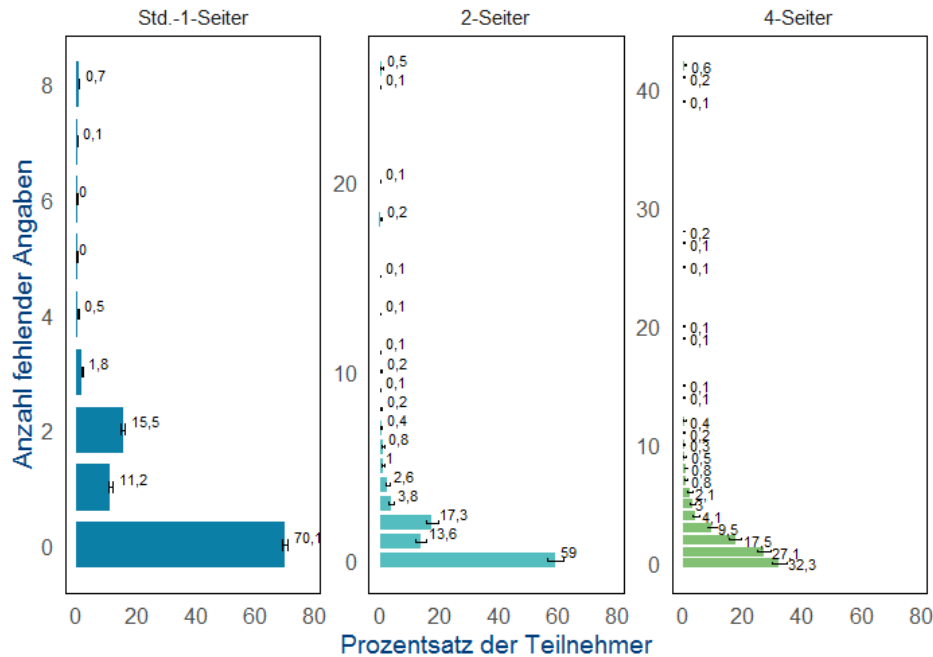
Im CAWI-Modus zeigt sich ein ähnliches Bild wie in der Gesamt- und PAPI-Betrachtung: Die Std.-1-Seiter-Gruppe weist die höchste *Item-Nonresponse-Rate* auf (7,98 Prozent). Dennoch ist keiner der Unterschiede zwischen der Std.-1-Seiter-Gruppe und den Treatmentgruppen signifikant, und die Effektgrößen sind durchweg sehr klein ($p = 1,000$, $h < 0,2$ in allen Fällen).

Zusammenfassend zeigt sich, dass das etwas unerwartete Ergebnis¹⁶, dass die Std.-1-Seiter-Gruppe eine höhere *Item-Nonresponse-Rate* aufweist, insbesondere durch den PAPI-Modus getrieben ist, wobei jedoch keine signifikanten Unterschiede zu den anderen Gruppen nachgewiesen werden können. In der 2-Seiter- und 4-Seiter-Gruppe fällt die *Item-Nonresponse-Rate* sehr ähnlich aus, und es bestehen keine signifikanten Unterschiede zwischen diesen beiden Gruppen in den verschiedenen Erhebungsmodi.

Abbildung 18 veranschaulicht die Verteilung des Prozentsatzes der fehlenden Items pro Teilnehmer in den jeweiligen Gruppen. Diese Verteilung liefert Aufschluss darüber, wie viele Teilnehmer in jeder Gruppe alle Fragen vollständig beantwortet haben, im Vergleich zu denjenigen Teilnehmern, die wenige oder mehrere Fragen ausgelassen haben.

¹⁶ Die höhere *Item-Nonresponse-Rate* in der Std.-1-Seiter-Gruppe scheint durch den PAPI-Modus beeinflusst zu sein, der im Durchschnitt eine höhere Rate fehlender Angaben aufweist. Zusätzlich spielt die größere absolute Anzahl der Teilnehmer in der Std.-1-Seiter-Gruppe eine Rolle. Auch wenn der prozentuale Anteil der PAPI-Teilnehmer in den Gruppen ähnlich ist, führt die höhere absolute Anzahl in der Std.-1-Seiter-Gruppe dazu, dass der Moduseffekt hier stärker zum Tragen kommt. Es ist somit wahrscheinlich weniger die prozentuale Verteilung der PAPI-Teilnehmer, sondern die Kombination aus Moduseffekt und Gruppengröße, die die Unterschiede in den *Item-Nonresponse-Raten* antreibt.

Abbildung 18: Gruppenspezifischer Anteil Teilnehmer mit fehlenden Angaben im ersten Quartal 2024
Erstes Quartal 2024, Anteile in Prozent



Anmerkung: Konfidenzintervalle wurden auf dem 95 Prozent-Niveau berechnet. Die unterschiedliche Skala auf der y-Achse spiegelt die unterschiedliche Anzahl der Fragen in jeder Gruppe wider.

Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

Die Ergebnisse zeigen, dass in der Std.-1-Seiter-Gruppe über 70,1 Prozent der Teilnehmer keine fehlenden Angaben aufweisen. In der 2-Seiter-Gruppe liegt der entsprechende Anteil bei 59,0 Prozent, was darauf hinweist, dass die meisten Teilnehmer dazu neigen, alle Fragen zu beantworten. In der 4-Seiter-Gruppe liegt der entsprechende Anteil hingegen nur bei 32,3 Prozent. Eine vollständige *Item-Nonresponse* ist nur bei einem sehr geringen Anteil der Teilnehmer zu beobachten, nämlich bei 0,70 Prozent in der Std.1-Seiter, 0,50 Prozent in der 2-Seiter- und 0,60 Prozent in der 4-Seiter-Gruppe. Weiterhin fällt auf, dass in allen Gruppen der Großteil der Teilnehmer entweder keine oder nur eine sehr geringe Anzahl unbeantworteter Fragen aufweist.

Tabelle 22 enthält die *Item-Nonresponse*-Raten für alle Gruppen im zweiten Quartal 2024. Im Vergleich zur Analyse des ersten Quartals zeigt sich für das zweite Quartal ein leicht anderes Muster. Die 4-Seiter-Gruppe weist mit 7,93 Prozent die höchste *Item-Nonresponse*-Rate auf, während die Std.-1-Seiter-Gruppe mit 4,62 Prozent die niedrigste Rate aufweist. Der Unterschied zwischen der Std.-1-Seiter- und der 4-Seiter-Gruppe ist dabei statistisch signifikant ($p < 0,001$, $h = -0,137$), was auf einen kleinen negativen Effekt hindeutet. Auch der Unterschied zwischen der 2-Seiter- und der 4-Seiter-Gruppe zeigt einen signifikanten Unterschied ($p = 0,012$, $h = -0,107$), der ebenfalls auf einen kleinen Effekt hindeutet.

Tabelle 22: Item-Nonresponse-Rate im zweiten Quartal 2024 nach Gruppe und Erhebungsmodus, differenziert nach dem Erhebungsmodus

Zweites Quartal 2024, Anteile in Prozent

Erhebungsmodus	Gruppe	Item-Nonresponse-Rate	Gruppenvergleich	p-Wert	Cohen's h
Gesamt	Std.-1-Seiter	4,62	Std.-1-Seiter vs. 2-Seiter	0,295	-0,029
	2-Seiter	5,27	Std.-1-Seiter vs. 4-Seiter	0,000***	-0,137
	4-Seiter	7,93	2-Seiter vs. 4-Seiter	0,012*	-0,107
PAPI	Std.-1-Seiter	11,90	Std.-1-Seiter vs. 2-Seiter	0,083	-0,225
	2-Seiter	20,10	Std.-1-Seiter vs. 4-Seiter	0,012*	-0,293
	4-Seiter	22,90	2-Seiter vs. 4-Seiter	1,000	-0,068
CAWI	Std.-1-Seiter	3,92	Std.-1-Seiter vs. 2-Seiter	1,000	-0,003
	2-Seiter	3,98	Std.-1-Seiter vs. 4-Seiter	0,000***	-0,118
	4-Seiter	6,53	2-Seiter vs. 4-Seiter	0,030*	-0,115

Anmerkungen: Die Sterne zeigen einen statistisch signifikanten Unterschied (Z-test) zwischen den Gruppen an. Für den Vergleich zwischen der Std.-1-Seiter, 2-Seiter- und 4-Seiter-Gruppe wurden die p-Werte mittels der Holm-Bonferroni-Korrektur angepasst, um mehrfache Vergleiche zu berücksichtigen. Diese Anpassung verringert das Risiko von Fehlern erster Art (falsch positive Ergebnisse), das mit der Anzahl gleichzeitiger Tests zunimmt. * $p < 0,05$ ** $p < 0,01$ *** $p < 0,001$

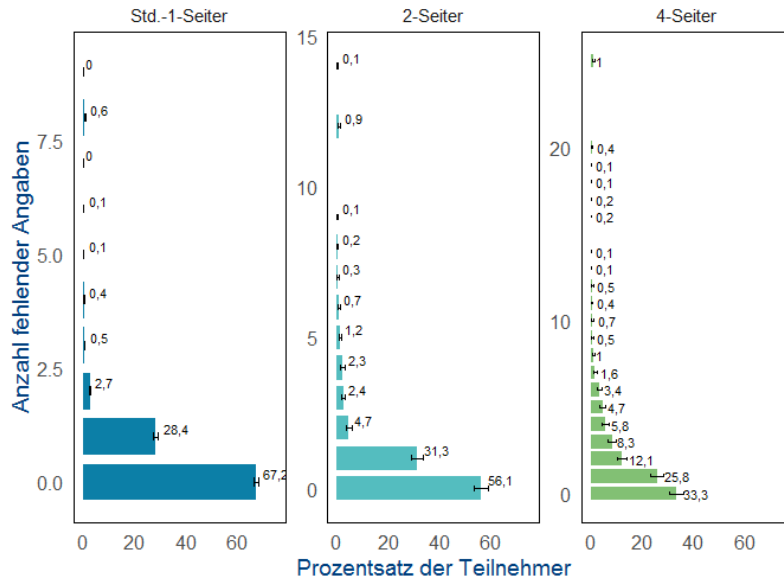
Quelle: IAB-Stellenerhebung; eigene Berechnungen. © IAB

Im PAPI-Modus zeigt sich eine höhere *Item-Nonresponse-Rate* in der 4-Seiter-Gruppe (22,90 Prozent) im Vergleich zur Std.-1-Seiter-Gruppe (11,90 Prozent). Der Unterschied ist statistisch signifikant ($p = 0,012$, $h = -0,293$) und deutet auf einen mittleren Effekt hin. Interessanterweise zeigt sich im Vergleich zwischen der Std.-1-Seiter- und der 2-Seiter-Gruppe im PAPI-Modus kein signifikanter Unterschied, allerdings fällt der Effektgrößenwert negativ aus ($h = -0,225$), was auf eine tendenzielle, jedoch nicht signifikante Verbesserung bei der kürzeren Version hindeutet.

Im CAWI-Modus beträgt die *Item-Nonresponse-Rate* der Std.-1-Seiter-Gruppe 3,92 Prozent, was im Vergleich zur 4-Seiter-Gruppe (6,53 Prozent) ebenfalls signifikant niedriger ist ($p < 0,001$, $h = -0,118$). Der Unterschied zwischen der 2-Seiter- und der 4-Seiter-Gruppe ist ebenfalls signifikant ($p = 0,030$, $h = 0,115$), mit einem kleinen Effekt zugunsten der kürzeren Fragebögen.

Abbildung 19 veranschaulicht die Verteilung der fehlenden Angaben pro Teilnehmer im zweiten Quartal 2024. Insgesamt zeigt sich auch hier, dass die Teilnehmer der Std.-1-Seiter-Gruppe weniger Fragen unbeantwortet lassen als die Teilnehmer der 4-Seiter-Gruppe.

Abbildung 19: Gruppenspezifischer Anteil Teilnehmer mit fehlenden Angaben im zweiten Quartal 2024
Zweites Quartal 2024, Anteile in Prozent



Anmerkung: Konfidenzintervalle wurden auf dem 95 Prozent-Niveau berechnet. Die unterschiedliche Skala auf der y-Achse spiegelt die unterschiedliche Anzahl der Fragen in jeder Gruppe wider.

Quelle: IAB-Stellenerhebung; eigene Darstellung. © IAB

Dieses Ergebnis ist ermutigend und deutet durchaus auf ein erhebliches Engagement der teilnehmenden Betriebe für die Erhebung hin – dies jedoch konditional darauf, dass die Betriebe überhaupt an der Befragung teilnehmen.

Zusätzliche Auswertung von Paradata

Der Fragebogen ist das primäre Instrument zur Datenerhebung in Umfragen und darauf ausgelegt, gültige und genaue Informationen zu sammeln. Dies zu erreichen erfordert eine sorgfältige Gestaltung des Fragebogens unter Berücksichtigung des jeweiligen Geschäftskontexts (Snijders et al., 2023; Willimack und Snijders, 2013). Selbst nach der Gestaltung und dem Pretest bleibt die grundlegende Frage, ob der Fragebogen wie beabsichtigt „funktioniert“ (Snijders et al., 2023). Online-Fragebögen vereinfachen die Beantwortung dieser Frage, indem sie eine weitreichendere Nutzung von sogenannten Paradata ermöglichen als schriftliche Fragebögen.

Bei Paradata handelt es sich um zusätzliche Daten, die den Umfrageprozess dokumentieren, wie etwa Ausfüllzeiten und Interaktionsmuster, also die Art und Weise, wie Teilnehmende mit dem Fragebogen interagieren, zum Beispiel durch ihr Navigationsverhalten oder Änderungen an Antworten während des Ausfüllens. Paradata können grundsätzlich dazu beitragen, die Qualität von Umfragedaten zu verbessern. Hinsichtlich der Befragungsqualität unterscheidet das sogenannte Total Survey Error Konzept (Groves et al., 2004) beim Umfrageprozess zwischen Repräsentativität und Messung, wobei jede

dieser Komponenten potenzielle Fehlerquellen mit sich bringen kann. Beispielsweise können bei der Repräsentation der Grundgesamtheit Abdeckungsfehler aus Ungenauigkeiten im Stichprobenrahmen resultieren, während Nichtantwortfehler auftreten können, wenn Teilnehmer die Teilnahme verweigern. Bei der Messung können Fehler auftreten, wenn Fragen die beabsichtigten Konzepte nicht genau erfassen oder es Probleme im Beantwortungsprozess gibt. Zudem können Fehler auch im Rahmen der Datenaufbereitung auftreten. Dieser Prozess umfasst die Bereinigung, Strukturierung und Anpassung der erhobenen Daten, um sie für die Analyse nutzbar zu machen. Paradata können helfen, diese Fehler während der Datenerhebung zu identifizieren und Informationen zur Modellierung der oben dargelegten Fehler bereitzustellen (Kreuter, 2013). Da der Erhebungsprozess sich vom ersten Kontakt mit den Teilnehmer bis zum Ende der Datenerhebung erstreckt, können Paradata grundsätzlich in vier Phasen erhoben werden: vor der Umfrage, während der Rekrutierung der Teilnehmer, dem Zugang und der Antwortphase (*Response-Phase*) (McClain et al., 2019).

Im Folgenden werden die im Rahmen der IAB-Stellenerhebung generierten Paradata des CAWI-Modus ausgewertet und somit im Wesentlichen die im Rahmen der Antwortphase generierten Paradata untersucht. Die Fokussierung auf die Antwortphase schließt explizit die Untersuchung der vorherigen Phasen, wie etwa der Stichprobenziehung, aus.

Die Analyse der Antwortphase bietet sich an, da es während dieser Phase möglich ist, äußerst umfangreiche Paradata zu erheben. In der detailliertesten Form können etwa Tastenanschläge, Mausbewegungen, Klicks oder Touch-Ereignisse, die auf Seite, Sitzung und Befragtebene aggregiert werden, erhoben werden. Häufig erhobene Paradata der Antwortphase umfassen zudem Antwortzeiten (Callegaro et al., 2009; Höhne und Schlosser, 2018; Yan und Tourangeau, 2008).

Die Antwortzeit ist ein besonders häufig untersuchter Indikator, da sie mögliche Ablenkungen, Sitzungszeitüberschreitungen und Multitasking aufdecken (Höhne und Schlosser, 2018; Yan und Tourangeau, 2008) und somit zur Beurteilung der Datenqualität beitragen kann. Paradata der Antwortphase können auch genutzt werden, um *Nonresponse-Fehler* zu untersuchen. Insbesondere können diese Paradata die Analyse einer spezifischen Form des *Nonresponse* Verhaltens, dem Abbruch der Befragung (des sogenannten *Breakoffs*), erleichtern und helfen zu verstehen, wie man diesen Fehler vorhersagen und verhindern kann (Mittereder und West, 2021).

Breakoff Rates

Im Folgenden werden *Breakoff Rates* getrennt nach den beiden Treatmentgruppen untersucht. Die *Breakoff Rate* (Abbruchrate) in Webumfragen (CAWI) misst den Anteil der Teilnehmer, die eine Online-Befragung beginnen, aber nicht bis zum Ende durchführen. Hohe *Breakoff Rates* können zu Verzerrungen führen und die Repräsentativität der Umfrageergebnisse verringern (Galesic und Bosnjak, 2009).

Bei dieser Analyse werden nur diejenigen Betriebe berücksichtigt, die mindestens die Eurostat-Fragen, welche zu Beginn des Fragebogens gestellt werden, beantwortet haben. In diesem Zusammenhang kann die Analyse der *Breakoff Rates* als Ergänzung zur Analyse der Rücklaufquote betrachtet werden. Die Rücklaufquote bezieht sich in diesem Fall auf den Anteil der Teilnehmer, die die Mindestanforderung

zur Teilnahme erfüllt haben, nämlich die Beantwortung der Eurostat-Fragen. Demgegenüber umfasst die *Breakoff Rate* jene Teilnehmer, die den Fragebogen zwar begonnen, aber nach der Beantwortung dieser ersten Fragen nicht bis zur letzten Seite des Webfragebogens navigiert haben und nicht auf „Absenden“ geklickt haben.

Die *Breakoff Rate* (BR) berechnet sich wie folgt:

$$BR = \frac{\text{Anzahl der CAWI Teilnehmer, die den Fragebogen nicht abgeschickt haben}}{\text{Gesamtzahl der gültigen CAWI Befragten (Eurostat Fragen beantwortet)}}$$

Im Hinblick auf die Unterscheidung zwischen *Breakoff* und *Item-Nonresponse* beschreibt ein *Breakoff* das vorzeitige Verlassen der gesamten Umfrage, während *Item-Nonresponse* das Auslassen einzelner Fragen bezeichnet. Alle Betriebe, die den Fragebogen abgebrochen haben, weisen zwangsläufig *Item-Nonresponse* auf, da sie mindestens eine Frage unbeantwortet lassen. Umgekehrt gilt jedoch nicht, dass alle Betriebe mit *Item-Nonresponse* den Fragebogen abgebrochen haben. So können Teilnehmer die Befragung insgesamt beendet und abgesendet haben, auch wenn sie einzelne Fragen ausgelassen haben.

Tabelle 23 weist die *Breakoff Rates* jeweils nach Gruppe und Quartal sowie die durchschnittliche Dauer bis zum *Breakoff* aus. Es wird deutlich, dass es einen deutlichen Unterschied in den *Breakoff Rates* zwischen dem vierten Quartal und dem ersten Quartal gibt. Dieser Unterschied kann auf den wesentlich längeren Fragebogen (8 Seiten) der Hauptbefragung im vierten Quartal als im ersten Quartal (1-, 2- oder 4-Seiter) zurückgeführt werden.

Tabelle 23: Breakoff Rates und Breakoff - Durchschnittliche Dauer nach Gruppe und Quartal

Viertes Quartal 2023 bis zweites Quartal 2024, Anteile in Prozent

Quartal	Gruppe	Breakoff Rate	Gruppenvergleich (p-value; Cohen's h)	Durchschnittliche Dauer (Min.)
4.Quartal	Std.	38,29	Standard vs. Experiment: p = 0,840; h = 0,003	7,47
	Exp.	38,10		7,35
1.Quartal	Std.-1-Seiter	4,02	Std.1-Seiter vs. 2-Seiter: p = 1,000; h = 0,018	1,94
	2-Seiter	4,39	Std.1-Seiter vs. 4-Seiter: p = 0,855; h = 0,032	3,03
	4-Seiter	4,68	2-Seiter vs. 4-Seiter: p = 1,000; h = 0,014	6,45
2.Quartal	Std.-1-Seiter	2,86	Std.1-Seiter vs. 2-Seiter: p = 0,302; h = 0,032	2,76
	2-Seiter	3,42	Std.1-Seiter vs. 4-Seiter: p = 0,000***; h = 0,119	2,68
	4-Seiter	5,19	2-Seiter vs. 4-Seiter: p = 0,070. h = 0,087	4,24

Anmerkungen: Die Sterne zeigen einen statistisch signifikanten Unterschied (Z-test) zwischen den Gruppen an. Für den Vergleich zwischen der Std.-1-Seiter, 2-Seiter- und 4-Seiter-Gruppe wurden die p-Werte mittels der Holm-Bonferroni-Korrektur angepasst, um mehrfache Vergleiche zu berücksichtigen. Diese Anpassung verringert das Risiko von Fehlern erster Art (falsch positive Ergebnisse), das mit der Anzahl gleichzeitiger Tests zunimmt. * p < 0,05 ** p < 0,01 *** p < 0,001

Quelle: IAB-Stellenerhebung; eigene Berechnungen. © IAB

Tabelle 23 zeigt zudem, dass im vierten Quartal nur ein sehr geringer Unterschied zwischen den Standard- und den Experimentalgruppen hinsichtlich der *Breakoff Rates* besteht ($p = 0,840$; $h = 0,003$). Auch der Unterschied in der durchschnittlichen Dauer der Teilnahme bis zum Abbruch ist sehr gering. Im ersten Quartal sind die Unterschiede jedoch deutlicher, wobei die Std.-1-Seiter-Gruppe die geringste *Breakoff Rate* (4,02 Prozent) aufweist. Der Unterschied in der *Breakoff Rate* zwischen der Std.-1-Seiter-Gruppe und der 2-Seiter-Gruppe beträgt 0,37 Prozentpunkte, ist allerdings statistisch nicht signifikant ($p = 1,000$; $h = 0,018$). Der Unterschied zwischen der 2-Seiter- und der 4-Seiter-Gruppe beträgt 0,29 Prozentpunkte und deutet auf eine sehr kleine Effektstärke ($h = 0,087$) hin. Der deutlichste Unterschied zwischen der 2-Seiter- und der 4-Seiter-Gruppe zeigt sich in der durchschnittlichen Dauer der Teilnahme bis zum *Breakoff*, die bei der 4-Seiter-Gruppe mehr als doppelt so lang ausfällt.

Im zweiten Quartal sinken sowohl die *Breakoff Rates* als auch die durchschnittliche Zeit bis zum *Breakoff* bei der Standard- und 2-Seiter-Gruppe. Auch wenn der Unterschied zwischen diesen beiden Gruppen signifikant ist, ist die Effektgröße sehr klein ($h = 0,119$). Bei der 4-Seiter-Gruppe gehen die Indikatoren in entgegengesetzte Richtungen: Während die *Breakoff Rate* im zweiten Quartal steigt, sinkt die durchschnittliche Zeit bis zum *Breakoff* bei der 4-Seiter-Gruppe. Die Ergebnisse in Tabelle 20 verdeutlichen die Tendenz, dass längere Fragebögen mit höheren *Breakoff Rates* einhergehen, und deuten zudem daraufhin, dass im Gegensatz zu der 2-Seiter-Gruppe die *Breakoff Rate* bei der 4-Seiter-Gruppe mit fortschreitenden Quartalen ansteigt.

3.2.6 Fazit auf Basis der bisherigen Ergebnisse

Im Rahmen der IAB-Stellenerhebung wurde ein Survey-Experiment durchgeführt, um die Einführung von hochfrequenten Fragen und die Auswirkungen verlängerter Fragebögen vor allem auf die Rücklaufquote und weitere Ergebnisvariablen zu untersuchen. Hierzu wurde eine Experimentalgruppe in zwei Gruppen unterteilt: eine Gruppe erhielt in den Quartalsnachbefragungen einen 2-seitigen Fragebogen, während die andere einen 4-seitigen Fragebogen erhielt. Der Fokus lag darauf, die Teilnahmebereitschaft der Betriebe und die *Response Burden* durch die erhöhte Befragungsdauer zu untersuchen.

Die Ergebnisse zeigen, dass es keine signifikanten Unterschiede zwischen den Gruppen gab, die den 2- und den 4-seitigen Fragebogen erhielten. Auch im Vergleich zur Standardgruppe, die im *Status-Quo* nur einen 1-seitigen Fragebogen erhielt, fallen die Unterschiede gering aus. Die Effektstärken waren für die meisten Vergleiche sehr klein, was darauf hinweist, dass die Auswirkungen der Fragebogenlänge auf die Rücklaufquote vernachlässigbar sind. Zudem zeigte sich, dass die Wahl des Befragungsmodus zwischen den Gruppen ähnlich verlief, wobei eine deutliche Präferenz für die Online-Antwortoption (CAWI) bestand.

Bezüglich der *Response Burden* deuten die Ergebnisse jedoch auf leichte Vorteile des 2-seitigen Fragebogens hin. Im Vergleich zur 4-Seiter-Gruppe berichteten die Betriebe, die den 2-seitigen Fragebogen erhielten, von einer geringeren wahrgenommenen Belastung. Dies deutet darauf hin, dass bei Verlängerung des Fragebogens im Vergleich zum *Status-Quo* die kürzere 2-Seiter-Variante etwas besser als die 4-Seiter-Variante geeignet sein könnte, die Teilnahmebereitschaft langfristig aufrechtzuerhalten, ohne die Belastung der Befragten zu erhöhen.

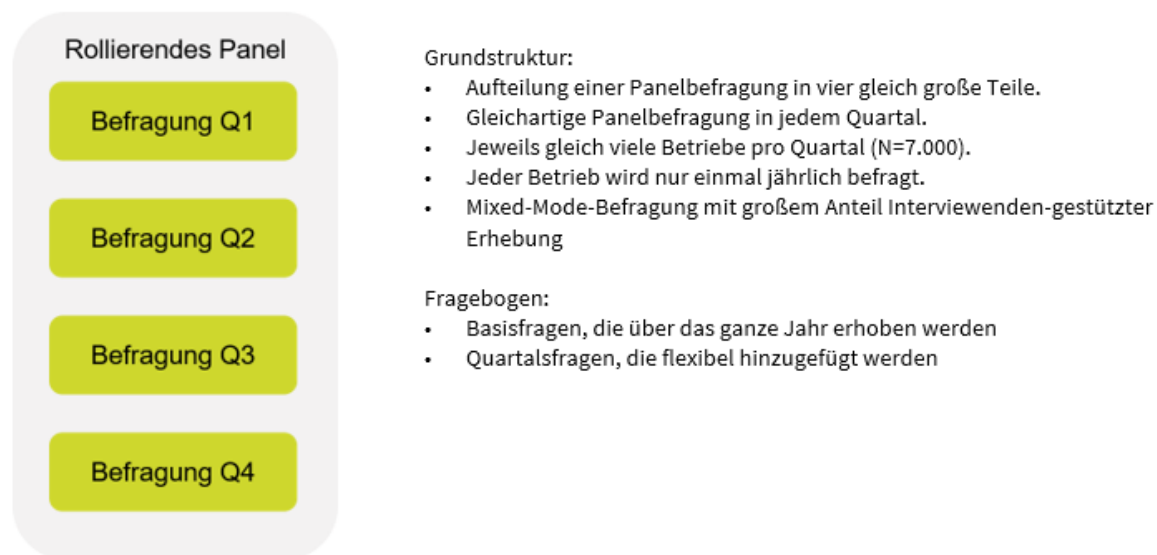
4 Konzeptionelle Überlegungen zur Integration der Befragungen

4.1 Rollierendes Panel

4.1.1 Überblick

Das rollierende Panel ist eine mögliche Variante zur Integration der beiden Befragungen, welche im Folgenden vorgestellt wird (Abbildung 20 zeigt einen Überblick). Dieses Kapitel gibt zunächst einen kurzen Überblick über die Variante, im Anschluss stellt es die zentralen Elemente des rollierenden Panels wie Fragebogen, Modus, Stichprobenziehung und Hochrechnung (und den Ressourcenbedarf) detaillierter vor. Abschließend wird ein kompletter Erhebungszyklus skizziert und die Variante hinsichtlich der Projektziele bewertet.

Abbildung 20: Rollierendes Panel – Überblick



Quelle: Eigene Darstellung. © IAB

Die Grundidee des rollierenden Panels ist, eine jährliche Panelstichprobe in vier Substichproben aufzuteilen, welche zu vier unterschiedlichen Zeitpunkten im Jahr (quartalsweise) als Panelerhebung befragt werden. Der Fragebogen hat dabei einen kleinen Anteil an Fragen zu aktuellen Inhalten, die von Quartal zu Quartal unterschiedlich sein können, wobei der übrige Teil des Fragebogens über längere Zeiträume oder dauerhaft gleich bleibt. Die höhere Befragungsfrequenz ermöglicht es, Fragen zu aktuellen Themen zeitnah zu erheben und unterjährige Auswertungen durchzuführen. Nach einem Erhebungszyklus, also einem Jahr, werden die Betriebe dann erneut befragt, um weiterhin eine jährliche

Panel-Erhebung fortzuführen. Jeder Betrieb kann nur Teil einer Quartalerhebung sein. Damit bietet dieses Design eine hochfrequente Erhebung ohne Mehrbelastung für die befragten Betriebe.

Neben einer Erhebung im Paneldesign und einer Erhebungsfrequenz, die häufiger ist als jährlich, muss ein rollierendes Panel auch die Berichtspflicht an Eurostat erfüllen können. Nach aktuellem Stand ist es für die sichere Erfüllung der Qualitätskriterien von Eurostat erforderlich, 7.000 Fälle pro Quartal bzw. 28.000 Fälle pro Jahr zu erheben.

Die Befragung soll als Modus Mix aus Selbstausfüller Interviews (CAWI) und persönlich mündlichen Interviews (CAPI) durchgeführt werden. Es ist geplant, dass die Betriebe zuerst in die Selbstausfüllerphase gehen und anschließend die Betriebe, von denen wir noch kein vollständig ausgefülltes Interview erhalten haben, von einem Interviewenden kontaktiert werden. Es ist somit auch ein Modus-Wechsel von CAWI zu CAPI innerhalb eines Interviews möglich. Informationen über einen Modus-Wechsel innerhalb eines Interviews liegen in den Paradata vor, wobei für jeden Screen nachvollzogen werden kann, in welchem Modus die Fragen beantwortet wurden. Hiermit ist es möglich, bei der Auswertung der Daten den verwendeten Modus zu berücksichtigen.

Die Länge der Befragung für einen einzelnen Betrieb orientiert sich an der bisherigen Interviewlänge im IAB-Betriebspanel, da sich eine längere Befragung negativ auf die Panelstabilität auswirken könnte. Diese Fragebogenlänge bei einer integrierten Befragung nicht zu überschreiten, erfordert deutliche inhaltliche Abstriche gegenüber dem bisherigen gemeinsamen Befragungsumfang von IAB-Betriebspanel und IAB-Stellenerhebung.

Die Stichprobenziehung erfolgt einmal jährlich im Vorjahr zur eigentlichen Erhebung, wobei sich die Stichprobe wie auch beim Betriebspanel aus wiederholt befragten und erstbefragten Betrieben zusammensetzt. Jede Quartalsbefragung ergibt einen eigenen, hochrechenbaren Datensatz. Nach einem Erhebungsjahr werden die vier Quartalsbefragungen zu einem Gesamtdatensatz zusammengefügt.

Wir gehen im Folgenden aus zwei Gründen davon aus, dass die Befragung komplett durch das IAB finanziert wird. Erstens ermöglicht die Fallzahl des Gesamtdatensatzes Auswertungen auf Bundeslandebene, womit ein wesentlicher Vorteil der Aufstockungsstichproben aus Sicht der Bundesländer entfällt. Zweitens wäre das IAB unabhängig von der Finanzierung verpflichtet, diese Fallzahl zu erreichen und könnte sie auch bei einem Ausstieg der Bundesländer nicht reduzieren. Beides spricht für eine konservative Kalkulation der notwendigen Ressourcen. Sollte es gelingen, die Bundesländer weiterhin von einer Kofinanzierung zu überzeugen, würde sich der Ressourcenbedarf entsprechend reduzieren.

4.1.2 Beschreibung der Variante

Struktur des Fragebogens

Der Fragebogen im rollierenden Panel besteht aus Basisfragen, die nicht variieren, und aus Quartalsfragen, die sich auf aktuelle Themen beziehen und zwischen den Quartalen variieren können. Die Basisfragen bleiben sowohl innerhalb eines Erhebungsjahres und größtenteils über mehrere Wellen

hinweg gleich, um eine Auswertung als Paneldatensatz zu ermöglichen. Die Basisfragen umfassen wie bisher im Betriebspanel jährlich oder zweijährlich gestellte Fragen sowie zusätzliche Schwerpunktfragen. Solche Schwerpunktfragen können jährlich aktuelle Themen aufgreifen, bei denen es inhaltlich oder aus Gründen der Fallzahlen geboten ist, diese nicht nur in einzelnen Quartalen zu erheben. Für die Basisfragen ist, analog zum jetzigen IAB-Betriebspanel, ein kognitiver Pretest durch die ProIAB vorgesehen, um neu entwickelte Fragen zu testen.

Die Quartalsfragen beziehen sich auf aktuelle Themen und können Themen abbilden, bei denen eine geringere Fallzahl sowie eine Erhebung zu einem Zeitpunkt ausreichen. Die Quartalsfragen sind so konzipiert, dass sie in erster Linie als Querschnitt (und nicht im Längsschnitt) ausgewertet werden sollen. Im Gegensatz zu einer reinen Querschnittsbefragung liegen jedoch umfassende Informationen zu den jeweiligen Betrieben auch zu anderen Zeitpunkten vor. Quartalsfragen sollten im Idealfall zwei Monate vor Feldstart eingebracht werden, es besteht aber auch die Möglichkeit, Notfallfragen noch vier bis sechs Wochen vor Start der Erhebung einzubringen. Aufgrund dieser Taktung und begrenzter zeitlicher Ressourcen der ProIAB können Quartalsfragen voraussichtlich nicht regelmäßig mit kognitiven Pretests überprüft werden. Für komplett neue Fragen ist jedoch trotzdem ein Pretest empfehlenswert.

Die Aufteilung der Fragebogenlänge auf Basis- und Quartalsfragen ermöglicht Schwerpunktsetzungen in Richtung Paneldimension bzw. Querschnittsdimension. Denkbar ist auch, bestimmte Fragen jedes Jahr nur in einem Quartal zu erheben, was zusätzliche Schwerpunktsetzungen zwischen einer großen Bandbreite an Informationen, dann allerdings für verschiedene Betriebe, und Auswertungstiefe mit Blick auf die Fallzahl erlaubt.

Inhalte aus den beiden bestehenden Befragungen

Für eine Integration der Fragebögen der beiden Betriebserhebungen ist es notwendig, die Inhalte der beiden Erhebungen zu konsolidieren und auszuwählen, welche Fragen aus dem bisherigen Frageprogramm fortgeführt werden sollen. Im Folgenden werden mögliche Kriterien für eine Konsolidierung aufgezeigt und der Umfang des Kürzungsbedarfs geschätzt. Dabei beziehen sich die Längenangaben auf Seiten in der derzeitigen Formatierung des Fragebogens für das IAB-Betriebspanel und als Maximum wird von den bisherigen ca. 20 inhaltlichen Fragebogenseiten bzw. etwa 90 Fragen ausgegangen.

Der reine IAB-Teil des Fragebogens beim IAB-Betriebspanel umfasst rund 18 Seiten. Um den Fragebogen des IAB-Betriebspanel zu evaluieren, wurde das Projekt „Fragebogen Redesign“ aufgesetzt, um zu untersuchen, welche Fragen wie häufig genutzt werden. Auf Basis des Themendossiers „IAB-Betriebspanel“ wurde für 192 Veröffentlichungen, die das IAB-Betriebspanel verwenden und im Zeitraum von Januar 2020 bis Juli 2023 erschienen sind, die Fragenutzung erfasst und ausgewertet. Am häufigsten verwendet wurden die Fragen zu Betriebsgröße, Branche und Tarifbindung. Die Informationen zur Tarifbindung sind als jährliche Angabe in keiner anderen Datenquelle in dieser Form für Deutschland verfügbar. Weitere häufig genutzte Fragen sind, ob ein Betrieb Weiterbildungsmaßnahmen für die Beschäftigten fördert sowie Fragen zur Beschäftigtenstruktur, z.B. die Qualifikation der Beschäftigten im Betrieb, Beschäftigung in Teilzeit, befristete Beschäftigung und

Personaleinstellungen, die Frage zur Höhe der Bruttolohn- und Gehaltssumme pro Monat, die Frage, ob es einen Betriebs- und Personalrat im Betrieb gibt, sowie die Fragen zur betrieblichen Berufsausbildung. Auswertungen auf Ebene der Themenmodule zeigen, dass Fragen aus dem Modul „Berufsausbildung und Ausbildungsstellen“ am häufigsten in Publikationen verwendet werden, gefolgt von Fragen aus „Personalbewegungen und Personalsuche“ sowie aus „allgemeine Angaben zum Betrieb/ zur Dienststelle“. Detaillierte Angaben zum methodischen Vorgehen und den Ergebnissen finden sich in Bellmann et al. (2024). Mit einer Fortführung und Weiterentwicklung dieses Projekts könnte eine Grundlage für eine evidenzbasierte Auswahl der weiterzuführenden Fragen geschaffen werden.

Die IAB-Stellenerhebung erfüllt derzeit die Berichtspflicht an Eurostat zu offenen Stellen und Beschäftigung. Diese sieht die quartalsweise Meldung der offenen (sofort und später zu besetzenden Stellen), die gemeldeten Stellen sowie Informationen zur Gesamtbeschäftigung im Betrieb (SVB, geringfügig Beschäftigte, Beamte, Richter, tätige Inhaber und mithelfende Familienangehörige) vor. Diese Informationen sind auch bei einer integrierten Betriebsbefragung weiterhin zu erheben. Im Betriebspanel-Fragebogen werden die besetzten Stellen sowie die Gesamtbeschäftigung (bis auf die Kurzarbeit) ebenfalls abgefragt. In diesem Kontext wird auch die Frage gestellt, ob es sich bei der Organisation um ein unabhängiges, eigenständiges Unternehmen beziehungsweise eine eigenständige Einrichtung ohne Niederlassung/Dienststellen an anderer Stelle handelt. Dies wird in ähnlicher Form im Betriebspanel-Fragebogen auch abgefragt. Im Vergleich zum bisherigen Fragebogen des IAB-Betriebspanel müssten in einen integrierten Fragebogen daher die Frage nach Leiharbeit und freien Mitarbeiter/-innen und die Frage, ob der Betrieb derzeit aktiv nach neuen Arbeitskräften (gegliedert nach Anforderungsniveau) sucht, aufgenommen werden. Hierfür ergäbe sich ein zusätzlicher Platzbedarf von etwa einer halben Seite.

Die IAB-Stellenerhebung erhebt zusätzlich quartalsweise Informationen für die Arbeitszeitrechnung. Hierfür werden fünf Fragen erhoben, die sich auf Arbeitszeitkonten beziehen. Beim Betriebspanel wird die Frage, ob es im Betrieb Arbeitszeitkonten gibt, ebenfalls abgefragt. Die Fragen, welche Ausgleichs- und Abrechnungszeiträume für die Arbeitszeitkonten gelten, wieviel Prozent der Beschäftigten ein Arbeitszeitkonto mit Ausgleichszeitraum von bis zu einem Jahr haben, wie viele Stunden die Beschäftigten im Durchschnitt auf ihren Konten haben und ob diese Zahl exakt oder geschätzt ist, sind nicht im Betriebspanel-Fragebogen enthalten. Sie entsprechen einem Platzbedarf von etwa einer drei Viertel Seite.

Des Weiteren erhebt die Stellenerhebung Informationen zu Rekrutierungsprozessen als Komplementärstatistik zur BA-Statistik der gemeldeten offenen Stellen. Dies umfasst einen Fragenblock zur letzten Neueinstellung einer Person in eine sozialversicherungspflichtige Beschäftigung in den letzten 12 Monaten und einen Fragenblock zum letzten Fall eines Suchabbruchs. Abgefragt werden im Fragenblock zum letzten Fall einer sozialversicherungspflichtigen Neueinstellung Merkmale zur Person, Fragen zum Einstellungsprozess sowie Merkmale der besetzten Stelle. Der Fragenblock zum letzten Fall des Suchabbruchs zielt auf Betriebe ab, die in den letzten 12 Monaten eine Personalsuche für eine sozialversicherungspflichtige Stelle abrechnen mussten, weil sie kein geeignetes Personal finden

konnten. Zusammen entsprechen die beiden Frageblöcke in ihrer jetzigen Form einem Umfang von etwa sechs Seiten. Diese Themen bildet das IAB-Betriebspanel bisher nicht ab. Die beiden Fragenblöcke werden von der Stellenerhebung nur einmal jährlich bei der Haupterhebung erhoben. Dabei ist jedoch zu beachten, dass nur rund 60 Prozent der Betriebe Angaben zum letzten Fall einer sozialversicherungspflichtigen Neueinstellung machen können und nur knapp 15 Prozent der Betriebe Informationen zum letzten Fall eines Suchabbruchs berichten können. Um im rollierenden Panel einen ausreichenden Rücklauf bei diesen Fragenblöcken erzielen zu können, müssen sie in zwei Quartalen erhoben werden.

Zusammenfassend bedeutet das, dass für die Integration der Inhalte aus der Stellenerhebung für zwei Quartale etwa sieben Seiten Platzbedarf kalkuliert werden müssen, für die anderen beiden Quartale jeweils etwas über eine Seite. Zusammen mit den derzeitigen Inhalten aus dem IAB-Betriebspanel würde der Fragebogen im Durchschnitt auf eine Länge von 25 Seiten kommen. Entsprechend wäre eine Reduktion um etwa ein Fünftel des momentanen Fragebogenumfangs nötig, beispielsweise in dem Fragen entsprechend der Struktur des Fragebogens nur einem Teil der Betriebe gestellt werden. Über diesen Kürzungsbedarf hinaus berücksichtigt dieses Vorgehen nur Inhalte der IAB-Stellenerhebung, die unmittelbar für die definierten Nebenbedingungen notwendig sind. Alle weiteren Inhalte der IAB-Stellenerhebung stünden ebenfalls nicht mehr zur Verfügung oder würden zu einem weiteren Kürzungsbedarf führen.

Zeitliche Bezugspunkte der Fragen

Einige Fragen des IAB-Betriebspanels und der IAB-Stellenerhebung beziehen sich auf definierte Zeitpunkte oder Zeiträume. Da sich bei einem rollierenden Panel die Erhebungszeitpunkte über ein Jahr verteilen, müssten die zeitpunktbezogenen Fragen daran angepasst werden. Das IAB-Betriebspanel enthält Fragen bezogen auf folgende Zeitpunkte beziehungsweise Zeiträume: letztes Geschäftsjahr, letztes Kalenderjahr, Fragen zum ersten Halbjahr des laufenden Jahres sowie Fragen zum Stichtag 30.06. des laufenden Jahres.

Fragen, die sich auf das letzte Geschäfts- oder Kalenderjahr beziehen, erscheinen auch beim rollierenden Panel als unproblematisch. Dies sind beispielsweise Fragen zu Investitionen und Innovationen („Wie hoch war die Investitionssumme aller Investitionen in Ihrem Betrieb im Vorjahr?“). Es muss allerdings bedacht werden, dass Informationen zum letzten Kalenderjahr bei Befragten des ersten Quartals eventuell noch nicht vorliegen und geschätzt werden müssen. Dies ist mit einem Pretest zu überprüfen.

Fragen, die sich auf das erste Halbjahr des laufenden Jahres oder auf den 30.06. beziehen, müssten im rollierenden Panel umstrukturiert werden. Fragen zum ersten Halbjahr sind zum Beispiel Fragen zur Weiterbildung („Fördert ihr Betrieb im 1. Halbjahr Weiterbildungsmaßnahmen?“) oder Fragen zur Personalbewegung („Haben Sie im ersten Halbjahr eingestellt?“). Die unserer Einschätzung nach beste Option¹⁷ wäre, im ersten und zweiten Quartal zum zweiten Halbjahr des Vorjahres zu fragen und

¹⁷ Diese Gestaltung bietet einerseits gute Anknüpfungsmöglichkeit an bisher vorliegende Daten und andererseits die Möglichkeit, Daten aus zwei Quartalen für Aushebungen zusammenzufassen.

Betriebe im dritten und vierten Quartal zum ersten Halbjahr des Befragungsjahres. Bei diesen Fragen ergäbe sich so ein Informationsgewinn, da zusätzlich Angaben zum zweiten Halbjahr vorlägen.

Fragen, die bisher den 30.06. als Referenzzeitpunkt haben, sind zum Beispiel Fragen zur Beschäftigten- und Qualifikationsstruktur („Wie verteilt sich die Gesamtzahl der Beschäftigten am 30.06. dieses Jahres auf die hier genannten Beschäftigtengruppen?“). Diese können im rollierenden Panel analog zu den Halbjahres-bezogenen Fragen so umgesetzt werden, dass im ersten und zweiten Quartal Angaben zum 31.12. des Vorjahres erhoben werden und im dritten und vierten Quartal zum 30.06. des Befragungsjahres. Bei dieser Aufteilung ist damit zu rechnen, dass die Angaben einer gewissen Saisonalität unterliegen, da beispielsweise die Beschäftigungsdaten Jahresend-Effekten unterliegen, weil dort häufiger als sonst Beschäftigungsverhältnisse enden.

Bei Fragen, die sich nicht auf einen konkreten Zeitpunkt oder Zeitraum beziehen, liegen zwischen den Antworten der Betriebe bis zu elf Monate. Wird ein Betrieb im Januar zu einem aktuellen Thema befragt (z.B. Krisenfragen) und ein anderer Betrieb erhält im Dezember dieselbe Frage, hat sich die politische Lage, auf die sich die Frage bezieht, unter Umständen sehr verändert. Zudem kann es zu sogenannten Recall-Effekten kommen, wenn die befragte Person sich nicht mehr korrekt erinnert. Abhängig von der Frage sind die Ergebnisse daher nur bedingt vergleichbar. Gleichzeitig ermöglicht das quartalsweise Design der Erhebung, die Entwicklung über das Jahr hinweg nachzuzeichnen. Um die korrekte Auswertung der unterschiedlichen Fragen zu ermöglichen, enthält der Gesamtdatensatz des rollierenden Panels einen Monatsindikator.

Die Erhebung der offenen und besetzten Stellen ist an die Vorgaben von Eurostat gebunden, siehe Abschnitte 2.2.5 und 2.2.6. Diese sehen vor, dass die Daten repräsentativ für das jeweilige Erhebungsquartal erhoben werden. Hierfür bietet Eurostat zwei Varianten an: Variante eins sieht vor, dass die Daten möglichst gleichmäßig über das gesamte Quartal erhoben werden. Bei Variante zwei werden die Daten zu einem bestimmten Referenzdatum in der Mitte des Quartals erhoben. Die IAB-Stellenerhebung setzt bisher Variante eins um. Bei einem rollierenden Panel sind beide Varianten möglich.

Modus, Stichprobenziehung, Datenbereitstellung

Bei der Auswahl des Erhebungsmodus sind zwei Aspekte von zentraler Bedeutung: hoher und schneller Rücklauf, um die Meldung an Eurostat zu gewährleisten, sowie Panelstabilität, um eine ausreichende Anzahl an Panelfällen sicher zu stellen.

Bislang werden den Betrieben des IAB-Betriebspanels verschiedene computergestützte Erhebungs-Modi angeboten. Auf Basis der Erfahrungen der letzten Wellen des IAB-Betriebspanels erscheint es realistisch, eine Kombination aus selbstadministrierten Interviews (CAWI) und durch Interviewende gestützte Interviews (CAPI) anzubieten. Ein Teil der Betriebe, nach den bisherigen Erfahrungen im IAB-Betriebspanel etwa 25 Prozent, würde dabei den Fragebogen selbstständig online ausfüllen. Die verbleibenden 75 Prozent der Betriebe werden von Interviewenden befragt, wobei sich dabei der Modus „CAPI am Telefon“ bewährt hat. Das bedeutet, dass speziell auf Betriebsbefragungen geschulte Interviewende den Betrieb telefonisch befragen, wobei die interviewte Person die Fragen online sehen

kann. Dies hat auch den Vorteil, dass Betriebe wiederholt vom selben Interviewer befragt werden können, was sich positiv auf die Panelstabilität auswirkt.¹⁸

Es wird einmal jährlich eine Ausgangsstichprobe gezogen. Dies bedeutet eine Ziehung und Adresslieferung von Erstbefragten, die in vier strukturgleiche Stichproben aufgeteilt werden. Eine Herausforderung ist dabei, dass die Adressen der Erstbefragten über ein Jahr hinweg veralten. Denkbar ist, bei der bereits gezogenen Stichprobe ein Update der Adressen zu machen. Eine häufigere Stichprobenziehung würde den Prozess hingegen komplizierter machen. Es erscheint sinnvoll, Betriebe, die in einem bestimmten Quartal nicht befragt werden können, in das nächste Quartal zu übernehmen. Allerdings sollte diese Möglichkeit zeitlich beschränkt werden, um zu verhindern, dass immer mehr Adressen in spätere Quartale geschoben werden und sich die Struktur der Stichprobe substantiell ändert. Insgesamt erscheint es daher sinnvoll, bei Panelbetrieben eine Verschiebung in das nächste Quartal zu akzeptieren, nicht jedoch bei erstbefragten Betrieben und auch keine Verschiebungen um mehrere Quartale zu erlauben.

Um eine Auswertung der Daten auch unterjährig zu ermöglichen, werden nach jedem Quartal Zwischendatensätze erstellt. Hierfür stellt das Erhebungsinstitut auch unterjährige Hochrechnungsfaktoren bereit und führt ein Editing der Daten durch. Die Daten für die Lieferung an Eurostat werden quartalsweise mit einem gesonderten Hochrechnungsverfahren berechnet. Der finale Quartalsdatensatz soll dem IAB zur Mitte des Folgequartals für Auswertungen zur Verfügung stehen. Zu Beginn des neuen Kalenderjahres wird ein integrierter Gesamtdatensatz erstellt, der ebenfalls Hochrechnungsfaktoren umfasst. Insgesamt müssen also vier Hochrechnungsfaktoren für die unterjährigen Erhebungen und ein Hochrechnungsfaktor für den Gesamtdatensatz erstellt werden. Eine ausführliche Ausreißersuche kann aus zeitlichen Gründen nur einmal im Jahr für den Gesamtdatensatz stattfinden.

Es werden jeweils vier kurze Methodenberichte für die Quartalerhebungen sowie ein ausführlicher Methodenbericht pro Jahr für den Gesamtdatensatz vom Erhebungsinstitut erstellt. Darin werden die wichtigsten Themen rund um den Datensatz wie beispielsweise Informationen zur Stichprobenziehung, Editing, Fragebogenerstellung, Feldarbeit oder Rücklauf dokumentiert. Daraus erstellt das IAB einmal pro Jahr ein FDZ-Datenreport (z.B. Bächmann et al., 2023), welcher über das FDZ den Datennutzenden zur Verfügung gestellt wird. Dieser dient als Grundlage für wissenschaftliche Auswertungen des Datensatzes und enthält Informationen zum Datensatz sowie Arbeitshilfen. Zudem wird dem FDZ ein integrierter Gesamtdatensatz¹⁹ im Folgejahr des Erhebungsjahres zur Verfügung gestellt. In diesem Zusammenhang erstellt das IAB, wie bisher für das IAB-Betriebspanel, eine Arbeitshilfe (z.B. ein Stata dofile), damit die einzelnen Wellen als Paneldatensatz zusammengespielt werden können.

¹⁸ Die Erfahrungen der Betriebspanel Welle 2024 zeigen, dass Interviews über ein Telefonstudio (CATI) für Betriebsbefragungen nicht gut funktionieren und der Rücklauf bei diesem Modus deutlich schlechter ausfällt als erwartet. Vor diesem Hintergrund sollte beim rollierenden Panel auf diesen Modus nur eine geringe Rolle spielen oder ganz darauf verzichtet werden.

¹⁹ Seitens des FDZ besteht der Wunsch, nur einmal im Jahr einen integrierten Datensatz zur Verfügung zu stellen.

4.1.3 Skizze eines Erhebungszyklus

Eine Herausforderung des rollierenden Panels ist der enge Zeitplan und die Parallelität mehrerer Wellen. Um zu veranschaulichen, wie ein Erhebungszyklus in der konkreten Umsetzung aussehen könnte, wird im Folgenden ein kompletter Erhebungszyklus der Welle Y (grau hinterlegt) skizziert. Tabelle 24 soll den Ablauf veranschaulichen und zeigen, welche Arbeitsschritte wann anfallen. Viele Arbeitsschritte, insbesondere die Planung und Erstellung des Fragebogens, finden bereits in Vorbereitung auf die eigentliche Erhebung statt.

Tabelle 24: Skizze eines Erhebungszyklus des rollierenden Panels

Leere Zelle	2026 (Welle X)				2027 (Welle Y)												2028 (Welle Z)			
	Q1	Q2	Q3	Q4	Q1			Q2			Q3			Q4			Q1	Q2	Q3	Q4
					J	F	M	A	M	J	J	A	S	O	N	D				
Vorbereitend																				
Stichprobenziehung				Y										Z						
Basis-Fragebogen																				
Einreichung und Auswahl der neuen Themen	Y				Z	Z														
Pretest (Fragebogen, Feldphase, Auswertung)		Y	Y					Z	Z	Z	Z									
Finalisierung Fragebogen u. Online Instrument			Y	Y									Z	Z						
Druck d. Anschreiben				Y										Z						
Fragebogen-Quartal																				
Aktuelle Ideenfindung/Call	X2	X3	X4	Y1		Y2			Y3			Y4			Z1		Z1	Z2	Z3	Z4
Abstimmung und Finalisierung	X2	X3	X4	Y1		Y2			Y3			Y4			Z1		Z1	Z2	Z3	Z4
Programmierung Instrument, Testung u. Interviewerschulung	X2	X3	X4	Y1			Y2			Y3			Y4			Z1	Z2	Z2	Z3	Z4
Feldphase																				
Versand d. Anschreiben	X2	X3	X4	Y1			Y2			Y3			Y4			Z1	Z1	Z2	Z3	Z4
Datenerhebung	X1	X2	X3	X4	Y1	Y1	Y1	Y2	Y2	Y2	Y3	Y3	Y3	Y4	Y4	Y4	Z1	Z2	Z3	Z4
Versand Reminder	X1	X2	X3	X4		Y1			Y2			Y3			Y4		Z1	Z2	Z3	Z4
Monitoring/Feldstatus	X1	X2	X3	X4	Y1	Y1	Y1	Y2	Y2	Y2	Y3	Y3	Y3	Y4	Y4	Y4	Z1	Z2	Z3	Z4
Nachbereitend																				
Datenprüfung/Editing/vorläufige Ausreißersuche/Gewichtung		X1	X2	X3	X4			Y1			Y2			Y3			Y4	Z1	Z2	Z3
Finaler Datensatz pro Quartal		X1	X2	X3	X4			Y1			Y2			Y3			Y4	Z1	Z2	Z3
Lieferung an Eurostat		X1	X2	X3		X4			Y1			Y2			Y3		Y4	Z1	Z2	Z3
Gesamtdatensatz erstellen					X	X											Y			
Datenreport					X	X											Y			
Daten-Paket für das FDZ erstellen						X	X	X									Y	Y		

Gelb: Y1 steht für die Quartalsbefragung im ersten Quartal 2027 der Welle Y; **Grün:** Y2 steht für die Befragung in Quartal 2 der Welle Y; **Blau:** Y3 steht für die Befragung in Quartal 3 der Welle Y; **Orange:** Y4 steht für die Befragung in Quartal 4 der Welle Y; **Lila:** Y steht für gleichbleibende, nicht rollierende Fragen der Welle Y

Quelle: Eigene Darstellung.

Aus dem Ablaufplan wird auch ersichtlich, dass die Vor- und Nachbereitung einer Welle immer jeweils parallel zur Feldphase der Folge-Welle stattfinden. In Tabelle 24 sind daher auch die Arbeitsschritte der Vorwelle X und der Nachwelle Z vermerkt.

Vor der Feldphase - Vorbereitend

Die vorbereitenden Arbeiten für Welle Y starten mit der Entwicklung des Basis-Fragebogens. Der Basis-Fragebogen wird einmal im Jahr erstellt und bleibt dann über alle vier Quartale gleich. Die Planung des Fragebogens startet mit einem internen Call for Questions für externe Forschende, die neue Themenvorschläge für die Schwerpunktfragen einbringen können. Es können sowohl Fragen für den Basisfragebogen und auch Fragen für die Quartalsfragebögen eingereicht werden. Nach einer Abstimmung der Themen wird der Pretest-Fragebogen erstellt und von den ProlAB in einem kognitiven Pretest getestet. Diese stehen auch weiterhin einmal im Jahr für den Pretest der Basisfragen zur Verfügung. Die Ergebnisse des Pretests fließen in die finale Abstimmung der neuen Fragen ein. Abschließend wird der Fragebogen finalisiert und an das Erhebungsinstitut für die Programmierung übergeben. Neben der Entwicklung des Fragebogens müssen vor der Erhebung zudem alle weiteren Dokumente aktualisiert werden. Hierzu zählen das Anschreiben und das Datenschutzblatt ebenso wie die Anleitungen für die interviewenden Personen. Zudem müssen die Interviewenden geschult werden.

Der Quartalsfragebogen wird immer im Vor-Quartal zur eigentlichen Erhebung, also für Quartal eins 2027 (Y1) im vierten Quartal 2026, vorbereitet. Hier können auch kurzfristig aktuelle Themen eingebracht werden, welche dann einmalig quartalsweise oder bei Bedarf über mehrere Quartale erhoben werden. Diese Fragen sollen auch in einem Pretest getestet werden, ein kognitiver Pretest ist aus zeitlichen Gründen jedoch nur realisierbar, wenn Fragen für den Quartalsfragebogen bereits mit dem allgemeinen Call for Questions zu Beginn des Vorjahres eingereicht werden. Nach dem Pretest werden die neuen Quartalsfragen abgestimmt und an das Erhebungsinstitut für die Programmierung des Online-Instruments übergeben.

Während der Feldphase

Die Feldphase startet jeweils zu Beginn eines Quartals und endet nach 12 Wochen mit der Zielgröße von 7.000 Interviews pro Quartal. Vier Wochen nach Feldstart werden die Erinnerungsschreiben an die Betriebe versendet, die bis dahin noch nicht teilgenommen haben.

Während der Feldphase findet ein wöchentliches Monitoring des Feldstatus statt, bei dem die Soll- und Ist-Zahlen der Rückläufe von Interviews kontrolliert werden. Aufgrund der Lieferverpflichtung an Eurostat ist die Feldzeit zeitlich begrenzt und kann nicht ins nachfolgende Quartal ausgedehnt werden, weshalb die Einhaltung der angestrebten Rücklaufquoten umso wichtiger ist. Die Quartale drei und vier könnten wegen der Urlaubszeiten in den Sommer- und Weihnachtsferien problematisch sein. Die Betriebe sind in dieser Zeit schwieriger zu erreichen und es stehen auch weniger Interviewende zur Verfügung. Das Monitoring des Feldverlaufs dient dazu, rechtzeitig Anpassung (wie beispielsweise Aufstockungstichproben oder zusätzliche Erinnerungsschreiben) vornehmen zu können, sollte der Feldverlauf nicht planmäßig laufen. Hierzu ist auch der regelmäßige Austausch mit dem Erhebungsinstitut unerlässlich.

Nach der Feldphase – Nachbereitend

Nach der Feldphase einer Quartalerhebung (Y1) finden die Validierung der erhobenen Daten, das Editing und die Erstellung der Hochrechnungsfaktoren durch das Erhebungsinstitut statt. Für die unterjährigen Wellen werden nur kurze Methodenbeschreibungen (max. 10 Seiten) vom Erhebungsinstitut angefertigt.

Die Lieferung der finalen Quartalsdaten erfolgt dann zu Beginn des Folgequartals (die Daten für das erste Quartal (Y1) werden zu Beginn des zweiten Quartals geliefert). Nun erfolgt die Auswertung der Daten durch das IAB und die Lieferung an Eurostat. Die Lieferverpflichtung an Eurostat sieht vor, dass die Daten 45 Tage nach Quartalsende vorliegen.

Nach einem kompletten Erhebungszyklus der Welle Y wird ein Gesamtdatensatz aus den Daten der vier Quartale (Y1- Y4) erstellt. An dieser Stelle werden auch Hochrechnungsfaktoren für die Auswertung des Gesamtdatensatzes erstellt. Am Ende eines Zyklus, also einmal pro Jahr, werden auch die vollständigen Bruttodaten und die Paradata geliefert. Es wird ein Datenreport für den Gesamtdatensatz erstellt (ähnlich dem Datenreport für das Betriebspanel) und das Datenpaket aufbereitet, das über das FDZ der Fachöffentlichkeit zur Verfügung gestellt wird.

4.1.4 Bewertung hinsichtlich der Projektziele

Abgleich mit den Nebenbedingungen

Im Projektantrag werden vier Nebenbedingungen für eine integrierte Erhebung genannt. Im Folgenden wird zu jeder Nebenbedingung erläutert, inwieweit ein rollierendes Panel sie erfüllt und welche Einschränkungen zu erwarten sind.

Eine Panelbefragung von Betrieben kann mit einem rollierenden Panel weiterhin durchgeführt werden. Eine Anschlussfähigkeit an das IAB-Betriebspanel erscheint möglich, jedoch nicht ohne Bruch in den Zeitreihen.

Die amtliche Lieferverpflichtung an Eurostat kann über ausreichend hohe Fallzahlen erfüllt werden, da es sich beim rollierenden Panel um eine quartalsweise Befragung handelt. Es ist dafür jedoch erforderlich, ein auf das Erhebungsdesign angepasstes Hochrechnungsverfahren entwickeln zu lassen.

Die Befragung soll weiterhin als Komplementärstatistik zu den amtlichen Daten der gemeldeten offenen Stellen der BA Informationen zu Einstellungsprozessen erheben. Grundsätzlich kann ein rollierendes Panel solche Fragen abbilden. Die Fragen zu den Einstellungsprozessen sind jedoch sehr umfangreich, bei gleicher Länge des Fragebogens müssten also entweder diese oder andere Fragen gekürzt werden.

Beratungs- und forschungsrelevante Informationen sollen in höherer Frequenz als jährlich zur Verfügung stehen. Diese Anforderung erfüllt ein rollierendes Panel.

Anknüpfung an und Auswirkungen auf bestehende Datenprodukte

Als weiteres Projektziel ist zu diskutieren, welche Auswirkungen ein rollierendes Panel auf die Datensätze hätte, die eng mit dem IAB-Betriebspanel verbunden sind. Zu nennen sind hier die Anschlussfähigkeit an das IAB-Betriebspanel, die Anschlussfähigkeit an die IAB-Stellenerhebung, das Linked-Personal-Panel (LPP) sowie die Linked Employer-Employee (LIAB)-Daten.

In Bezug auf Anknüpfbarkeit an das IAB-Betriebspanel bleiben für einen Teil der Betriebe Fragen und Befragungszeitpunkte ähnlich. Je nachdem, in welchem Quartal ein Betrieb befragt wird, sind jedoch insbesondere zeitpunktbezogene Fragen, nicht ohne Bruch fortführbar (vgl. Kapitel 2.2.1). Es entsteht ein Bruch in den Zeitreihen, der sich bei einer geeigneten Gestaltung der Fragen jedoch nachvollziehen und quantifizieren lassen sollte. Bei den Panel-Fragen besteht die Möglichkeit einer Verknüpfung mit dem Gesamtdatensatz. Die Quartalsfragen sind gesondert zu betrachten und vermutlich nicht an den jetzigen Betriebspanel-Datensatz anschlussfähig.

Bezüglich der IAB-Stellenerhebung müsste im Rahmen einer möglichen Fortführung der Quartalsfragen zu Arbeitszeitkonten (hausinterne Zulieferung von 5 Fragen zur IAB-Arbeitszeitrechnung) ein denkbarer Zeitreihenbruch in allen erhobenen Dimensionen möglichst verhindert oder zumindest minimiert werden.

Im LPP werden alle zwei Jahre mittelgroße und große Betriebe aus der Privatwirtschaft im Anschluss an die Erhebung zum IAB-Betriebspanel befragt. Im rollierenden Panel könnte man auch für das LPP vier Mal im Jahr befragen und die Daten anschließend zusammenfassen. Das bedeutet aber, dass man vier Befragungszeitpunkte hat, die nur eingeschränkt vergleichbar sind, was als Nachteil zu interpretieren ist. Allerdings sind bei den LPP-Fragen kaum große Änderungen innerhalb eines Jahres zu erwarten.

Beim LIAB-Querschnittsmodell (LIAB-QM) werden die Erwerbsbiographien jeweils zum Stichtag 30.06. abgebildet und mit dem IAB-Betriebspanel verknüpft. Das LIAB-QM umfasst Beobachtungen aller Personen, die im Jahr eines Betriebspanel-Interviews in einem Betriebspanelbetrieb tätig waren. Das LIAB-Längsschnittmodell umfasst die gesamten zeitraumbezogenen Erwerbsbiographien von Personen einer Unterstichprobe, der im IAB-Betriebspanel wiederholt befragten Betriebe (sog. Panel-Betriebe). Bei einer Umstellung des Betriebspanels auf ein rollierendes Panel mit vier verschiedenen Erhebungszeiträume wäre also für beide LIAB-Modelle eine Verknüpfung möglich. Die Voraussetzung ist aber, dass es einen konsolidierten Gesamtdatensatz pro Jahr gibt, in dem der Kern der Panelfragen (des jetzigen Betriebspanels) erhalten bleibt. Zu bedenken ist dabei, dass bisher die Verknüpfung für den 30.06. erfolgt und zu dem Zeitpunkt passt, auf den sich einige Fragen im IAB-Betriebspanel beziehen. Sollten im rollierenden Panel verschiedene Stichtage umgesetzt werden, z.B. 30.06. und 31.12., wäre es naheliegend, auch zu diesen Stichtagen die Beschäftigten zu ziehen oder das Stichtagsprinzip aufzugeben und alle Beschäftigungsspeils in den Betrieben im Kalenderjahr zu ziehen.

4.1.5 Fazit

Mit einem rollierenden Panel können Inhalte häufiger als einmal jährlich abgefragt und ein für die Forschung attraktiver Paneldatensatz bereitgestellt werden, ohne die gleichen Betriebe mehrmals in einem Jahr zu befragen.

Eine Integration der Fragebögen der beiden Erhebungen erscheint grundsätzlich möglich. Allerdings sollte aufgrund der sinkenden Teilnahmebereitschaft seitens der Betriebe der Fragebogen die bisherige Länge nicht überschreiten oder experimentell geprüft werden, wie sich ein längerer Fragebogen auf die Teilnahmebereitschaft auswirkt. Eine Integration der bisherigen Fragebögen erfordert es daher, Inhalte, die bisher erhoben werden, ganz wegzulassen oder seltener zu erheben.

Mit einem rollierenden Panel wäre es grundsätzlich möglich, die Lieferverpflichtung an Eurostat zu erfüllen. Hierfür wäre es allerdings notwendig, ein speziell darauf angepasstes Hochrechnungsverfahren entwickeln zu lassen. Die Lieferverpflichtung an Eurostat bringt aber aufgrund der hohen Fallzahlen, die nur durch CAPI Interviews zu realisieren sind, verschiedene Schwierigkeiten mit sich. Vor dem Hintergrund der sinkenden Teilnahmebereitschaft der Betriebe einerseits und der sinkenden Anzahl an Interviewenden seitens des Erhebungsinstitutes andererseits erscheint dies herausfordernd. Zudem besteht aufgrund der kurzen Feldphasen und der gleichzeitig hohen Fallzahlen eine gewisse Anfälligkeit für Erhebungsprobleme. So gibt es, sollten die Rücklaufquoten nicht erreicht werden, wenig Spielraum für Anpassungsmaßnahmen wie einer Verlängerung der Feldzeit.

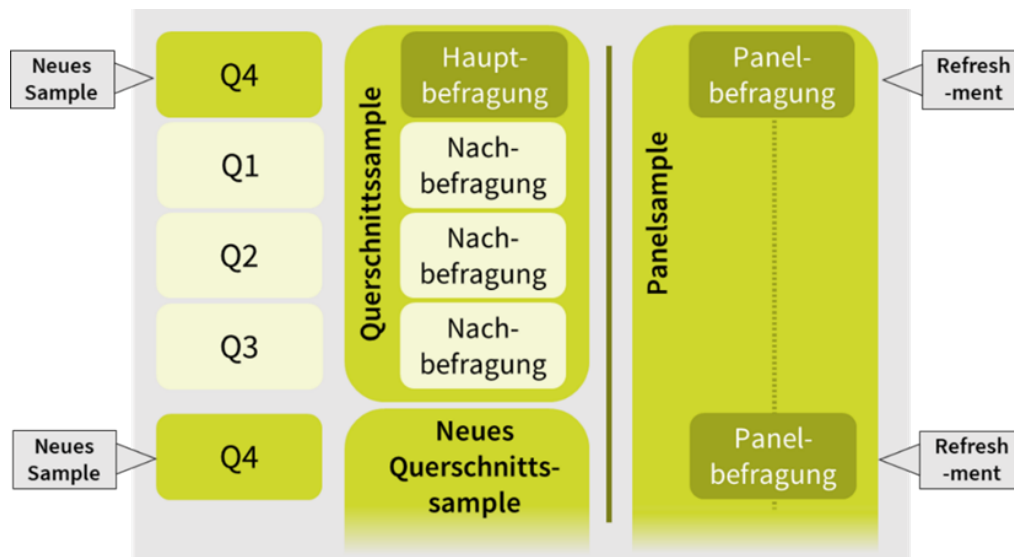
Im Vergleich zum bisherigen IAB-Betriebspanel würde ein rollierendes Panel zudem substantiell zusätzliche Sach- und Personalmittel erfordern. Die zusätzlichen Sachmittel wären nötig, weil deutlich mehr Betriebe befragt werden und eingekaufte Tätigkeiten bis zu fünf Mal statt einmal pro Jahr anfallen. Die höhere Frequenz erfordert auch zusätzliche Personalmittel, weil dadurch zahlreiche interne Aufgaben ebenfalls mehrmals pro Jahr anfallen.

Ein rollierendes Panel könnte grundsätzlich an die bestehenden Datenprodukte Betriebspanel, Stellenerhebung und an den LIAB sowie das LPP anknüpfen, allerdings wohl nicht ohne Brüche in bestehenden Zeitreihen.

4.2 Integration ausgehend von IAB-Stellenerhebung

Für ein weiteres Szenario ist vorgesehen, eine Variante der Integration zu diskutieren, welche die IAB-Stellenerhebung und das IAB-Betriebspanel unter dem Dach eines Erhebungsinstituts in einer kombinierten Querschnitts- und Längsschnittbefragung in eine Gesamterhebung integriert. Abbildung 21 illustriert hierzu das vorgesehene Erhebungsdesign, welches auf einer in eine Querschnitts- und Längsschnittstichprobe gesplitteten Gesamtstichprobe basiert. Wie in der Abbildung 21 dargestellt, wird jedes Jahr zu einem festen Zeitpunkt eine neue, repräsentative Querschnittsstichprobe gezogen, die den Start des Erhebungszyklus in viertes Quartal markiert. Die resultierende Nettostichprobe dient als Grundlage, um Betriebe einzuladen, in den drei nachfolgenden Quartalen an kleineren Folgerhebungen teilzunehmen. Diese Folgerhebungen ermöglichen die Sammlung von zeitnahen Verlaufsdaten im Rahmen eines unterjährigen Panels, die Veränderungen während dieser Quartale unter den Teilnehmer aufzeigen können.

Abbildung 21: Integration des Betriebspanels ausgehend von der IAB-Stellenerhebung



Quelle: Eigene Darstellung. © IAB

Auf Basis der Panelstichprobe soll - wie derzeit im IAB-Betriebspanel - eine jährliche wiederholte Befragung von Betrieben durchgeführt werden. Die Grundlage für diese Stichprobe sollen die Betriebe bilden, die aktuell am IAB-Betriebspanel teilnehmenden, um sicherzustellen, dass ihre langfristigen Daten weiter genutzt werden. Diese Vorgehensweise erlaubt es, die bestehenden langfristigen Beziehungen der Betriebe zum Betriebspanel auszunutzen und deren Längsschnittdaten weiterhin nutzen zu können (IAB-Betriebspanel-Wiederholerstichprobe – siehe Abschnitt 2.1.1). Dieser Teil der Erhebung liefert damit Längsschnittdaten, die dieselben Betriebe über einen längeren Zeitraum hinweg verfolgen. Die Panelstichprobe muss jedes Jahr analog zu der IAB-Betriebspanel Ergänzungsstichprobe um eine Auffrischungstichprobe (Refreshment) ergänzt werden, um sicherzustellen, dass die Daten repräsentativ bleiben und Ausfälle durch die Befragung neuer teilnehmender Betriebe nach Bedarf ersetzt werden können (siehe Abschnitt 2.1.1). Wie in Abbildung 21 dargestellt, soll der Befragungszyklus nach der letzten Nachbefragung im Querschnittsteil der Befragung enden. Im nächsten Quartal ist vorgesehen, wieder eine neue Querschnittsprobe zu ziehen.

Das oben beschriebene Integrationsszenario weist folgende zentrale Eigenschaften auf:

- **Kombination eines unterjährigen mit einem jährlichen Panels:** Die Kombination eines unterjährigen Panels mit einem jährlichen Panel erlaubt einerseits die Analyse kurzfristiger Verhaltensänderungen von Betrieben in Reaktion z.B. auf exogene Schocks wie etwa die Covid-19 Pandemie sowie den russischen Angriffskrieg gegen die Ukraine sowie andererseits die Analyse langfristiger Verhaltensänderungen in Reaktion auf säkularere Entwicklungen, wie etwa technologischen und demographischen Wandel.
- **Hochrechnung:** Die vorgeschlagene Integration ermöglicht die weitere Erhebung der offenen Stellen auf Basis einer Querschnittsstichprobe, die mit Quartalsnachbefragungen kombiniert wird. Die Hochrechnung auf Basis einer Querschnittsstichprobe hat den Vorteil, dass sie anders als eine Hochrechnung der offenen Stellen auf Basis von Betrieben, die bereits langjährig an einem Panel teilnehmen, weniger stark durch die Selektivität der Teilnahme

bzw. umgekehrt die Panelmortalität betroffen ist. Mit der Beibehaltung der Hochrechnung auf Basis einer Querschnittsstichprobe kann das Szenario somit auf die Erfahrungen bei der Hochrechnung der offenen Stellen im Rahmen der IAB-Stellenerhebung zurückgreifen. So wurde auf Basis dieser Hochrechnung, wie in Abschnitt 2.2.5 dargestellt, in den vergangenen Jahren eine zufriedenstellende Kreuzvalidierung der Hochrechnungsergebnisse der gemeldeten offenen Stellen, die sich aus den Befragungsinformationen ergeben, mit den amtlichen Daten der BA Statistik der gemeldeten offenen Stellen erreicht.

- **Inhaltliche Flexibilität durch Hochfrequenz:** Die hochfrequenten Nachbefragungen im unterjährigen Panel ermöglichen eine inhaltliche Flexibilität hinsichtlich möglicher Beratungs- und Informationsbedarfe, indem auch unterjährig Fragen in die Befragung eingespeist werden können. Um abschätzen zu können, inwiefern sich durch eine Erweiterung des Fragenkatalogs Rückwirkungen auf die Teilnahmebereitschaft der Betriebe und die für die amtliche Datenlieferung an Eurostat zur Verfügung stehende Fallzahl ergeben, können konkret die Befunde aus dem Längenexperiment mit der IAB-Stellenerhebung genutzt werden, welche im ersten Teil dieses Projektberichts dokumentiert wurden.
- **Fragebogenumfang:** Aus der Restriktion einer Maximalbelastung des einzelnen Betriebs pro Fragebogen ergeben sich durch dieses Integrationskonzept keine zwingenden Kürzungen im bestehenden Fragebogenumfang für die IAB-Stellenerhebung und das IAB-Betriebspanel, unbeschadet einer möglichen Konsolidierung der Befragungsinstrumente.
- **Voraussetzungen für die Kostenermittlung:** Auf Basis der derzeit bestehenden extramuralen Vergaben können – unter bestimmten Annahmen und Voraussetzungen – die Kosten für das vorgeschlagene Design ermittelt und mit den Kosten des Status-Quos verglichen werden. Eine zentrale Voraussetzung hierfür, deren Folgen im weiteren Projektverlauf noch diskutiert werden müssen (siehe Abschnitt 3.2.1), ist jedoch die Vereinheitlichung des Erhebungsmodus auf den derzeit in der IAB-Stellenerhebung verwendeten *Push-to-Web* Modus (vgl. Abschnitt 2.2.2).

4.2.1 Offene noch zu diskutierende Punkte

Auch wenn das oben beschriebene Integrationsszenario die „minimalinvasivste“ Variante der Integration der derzeit bestehenden Betriebsbefragungen darstellt, erfordert und ermöglicht das von der IAB-Stellenerhebung ausgehende Szenario zahlreiche Anpassungen, die Folgen für die Teilnahmebereitschaft und Repräsentativität der Gesamterhebung haben können. Die Diskussion möglicher Anpassungen sowie deren Folgen für den so genannten *Total Survey Error* stellt den zentralen Bestandteil der Untersuchungen im Rahmen des noch ausstehenden Projektverlaufs bis Ende 2025 dar - nicht zuletzt auch deswegen, da z.B. ohne eine Untersuchung der Folgen für die Teilnahmebereitschaft der Betriebe über den gesamten Erhebungszyklus, keine zuverlässige Kostenabschätzung möglich sein wird. Der vorliegende Abschnitt gibt einen Überblick über die offenen, im weiteren Projektverlauf noch zu diskutierenden Anpassungen des hier dargestellten Integrationsszenarios:

- **Implikationen eines einheitlichen Erhebungsmodus:** Derzeit setzen das IAB-Betriebspanel und die IAB-Stellenerhebung unterschiedliche Erhebungsmodi ein. Das Betriebspanel führt überwiegend persönliche (*Computed-Assisted Personal Interviewing*, CAPI) Interviews durch und hat diese durch die Nutzung von selbstadministrierten Modi erweitert, während die IAB-

Stellenerhebung aktuell ausschließlich einen selbstadministrierten *Push-to-Web* Modus einsetzt (CAWI kombiniert mit PAPI). Um eine realistische Kostenabschätzung des hier vorgeschlagenen Szenarios abgeben zu können, wird - wie oben dargestellt - für die weitere Diskussion eine Vereinheitlichung des Erhebungsmodus auf den derzeit in der IAB-Stellenerhebung verwendeten *Push-to-Web* Modus zugrunde gelegt. Diese Vereinheitlichung stützt sich einerseits auf den derzeit in der Befragungslandschaft zu beobachtenden Trend hin zu Online-Befragungen (vor allem auch durch den Einbruch von CAPI Interviews in Folge der Covid-19 Pandemie) (Maslovskaya et al., 2022), ist andererseits auch der aktuell bestehenden extramuralen Vergabe geschuldet, welche eine eigene Durchführung – und somit Kostenkalkulation - von CAPI Interviews nicht erlauben würde. Als Alternative zu einer Umstellung auf einen vollständig einheitlichen Erhebungsmodus ist vorgesehen, zusätzlich die Kombination einer selbstadministrierten Befragung (CAWI und PAPI) mit einer CATI-Erhebung, die z.B. im Rahmen von Auffrischungsbefragungen genutzt werden kann, zu diskutieren.

Der Übergang zu einem weitestgehend einheitlichen Modus stellt für die Integration beider Befragungen eine erhebliche Herausforderung dar, da diese Umstellung weitreichende Implikationen für die Teilnahmebereitschaft (Rücklaufquote, Panel Attrition), Repräsentativität und Messfehler haben kann. So können etwa modus-spezifische Messdifferenzen zu Variationen im Messfehler führen, die die Implementierungskomplexität erhöhen, die Kosten steigern oder zu einer Stichprobenverzerrung führen (für eine vollständige Beschreibung siehe z. B. (Schouten et al., 2022). Da die Folgen dieser Umstellung im weiteren Projektverlauf nicht experimentell untersucht werden können, soll die Abschätzung der Implikationen der Modus-Vereinheitlichung bzw. – umstellung auf Basis einer umfassenden Analyse der bisher in der survey-methodologischen publizierten Ergebnisse von Modusexperimenten erfolgen.

Die empirische Forschung zu Erhebungsmethoden in Betriebsbefragungen zeigt gemischte Ergebnisse hinsichtlich der Rücklaufquoten und der Datenqualität. Papierbefragungen führen im Allgemeinen zu höheren Rücklaufquoten im Vergleich zu webbasierten Methoden (Fraze et al., 2003; Sakshaug et al., 2019). Webbefragungen können jedoch die Teilnahmebelastung verringern, ohne die wahrgenommene Belastung oder die Antwortzeit erheblich zu beeinflussen (Haas et al., 2021). Eine Kombination aus Papier- und E-Mail-Kontakten kann die Rücklaufquote optimieren und den *Nonresponse*-Bias reduzieren. Dabei erzielen Papiereinladungen, gefolgt von Papiererinnerungen, die höchste Rücklaufquote (Sakshaug et al., 2019).

Studien, die Daten aus den Betriebsbefragungen des IAB – dem Betriebspanel und der Stellenerhebung – verwenden, zeigen jedoch andere Ergebnisse. Weder die Rücklaufquoten noch der *Nonresponse*-Bias werden durch die Einführung von webbasierten Erhebungsmethoden signifikant beeinflusst. Beide Studien (Gleiser et al., 2022; Küfner et al., 2024) zeigen jedoch, dass größere Betriebe eher auf webbasierte Befragungen antworten. Ein weiterer wichtiger Punkt, den beide Studien belegen, ist, dass durch die Einführung von Webbefragungen erhebliche Kosteneinsparungen erzielt werden können.

- **Stichprobendesign:** Die für das Integrationsszenario vorgesehene Kombination von Querschnitt- und Panelstichprobe eröffnet grundsätzlich Möglichkeiten, die jährliche Ziehungen der Querschnittsstichprobe sowie der Auffrischungsstichprobe (Refreshment) für

das jährliche Panel zu konsolidieren und miteinander zu kombinieren. So ist etwa denkbar, unter den teilnehmenden Betrieben der quartalsweisen Nachbefragungen die Teilnahmebereitschaft an der jährlichen Panelbefragung in der darauffolgenden Welle zu erheben und diejenigen Betriebe, die ihre Teilnahmebereitschaft signalisiert haben, gezielt für die Auffrischungsstichprobe zu befragen. Dies hat den Vorteil, dass die für die Auffrischungsstichprobe des jährlichen Panels erforderliche Bruttostichprobe reduziert werden kann. Eine entsprechend geringere Bruttostichprobe kann zu einer Reduzierung der Befragungslast der Grundgesamtheit der Betriebe beitragen. Hier wäre jedoch zu untersuchen, inwiefern dies die Möglichkeit einschränkt, die Panelstichprobe über die Aufstockungsstichproben repräsentativ zu halten.

- **Konsolidierung der Befragungsinstrumente (Fragebögen):** Das oben dargestellte Integrationsszenario liefert weiterhin Potenziale für die Konsolidierung der Befragungsinstrumente insbesondere in den Hauptbefragungen. Derzeit wird in der IAB-Stellenerhebung in der Hauptbefragung im vierten Quartal ein 8-seitiger Fragebogen mit rund 77 Fragen verwendet, während der Fragebogen des IAB-Betriebspanels 89 Fragen aufweist. Als mögliche Szenarien für die Konsolidierung der Hauptfragebögen kommen vor allem die mögliche Vereinheitlichung von Kernfragen zu den Charakteristika der Betriebe sowie mögliche Kürzungspotenziale bei potenziellen Überlappungen in Frage. Eine weitere Konsolidierungsmöglichkeit umfasst die Modularisierung der Fragebögen (wie sie derzeit bereits teilweise praktiziert wird), jedoch mit der möglichen Modifikation, dass bestimmte Module nicht mehr in jedem Jahr abgefragt werden. Da die Folgen einer potenziellen Konsolidierung der Fragebögen im weiteren Projektverlauf nicht experimentell untersucht werden können, muss die Abschätzung der Implikationen einer solchen Konsolidierung (ähnlich wie bei der Modus-Vereinheitlichung) auf Basis einer umfassenden Analyse der bisher in der survey-methodologischen publizierten Ergebnisse zu Fragebogenlängen-Experimenten in Betriebsbefragungen erfolgen. In Bevölkerungsumfragen zeigt die Forschung durchgehend, dass kürzere Fragebögen tendenziell zu höheren Rücklaufquoten führen, insbesondere bei traditionellen PAPI-Formaten (Guo et al., 2016; Iglesias und Torgerson, 2000). Der Effekt der Länge des Fragebogens ist jedoch in webbasierten Umfragen weniger ausgeprägt, bei denen die Rücklaufquoten nicht so stark in Abhängigkeit von der Fragebogenlänge variieren (Sandelin, 2022).
Trotz der leicht niedrigeren Rücklaufquoten bei längeren Umfragen scheint die Datenqualität nicht zu leiden. Mehrere Studien haben herausgefunden, dass längere Fragebögen die Validität oder Konsistenz der Antworten nicht beeinträchtigen. So stellten Burchell and Marsh (1992) fest, dass die Teilnehmer selbst bei längeren Umfragen detaillierte und konsistente Antworten auf offene Fragen gaben und die prädiktive Validität über den gesamten Fragebogen hinweg erhalten blieb. In ähnlicher Weise fand Sandelin (2022) keine signifikante Verschlechterung der Datenqualitätskennzahlen wie *Item-Nonresponse* oder *Breakoff*, selbst wenn den Befragten längere Fragebögen vorgelegt wurden. Daher können kürzere Umfragen zwar die Teilnahmebereitschaft erhöhen, doch längere Fragebögen – insbesondere in Online-Formaten – beeinträchtigen die Datenqualität nicht.
Im Gegensatz dazu reagieren Betriebsumfragen empfindlicher auf die Länge des Fragebogens. Studien zeigen, dass längere Fragebögen, insbesondere solche, die 30 Minuten oder länger dauern, zu einem deutlichen Rückgang der Teilnahme führen. Die Forschung von

Seshadri and Broekemier (2022) bestätigte, dass die Länge des Fragebogens ein entscheidender Faktor für Führungskräfte ist und die Rücklaufquoten sinken, je länger der Fragebogen wird.

Darüber hinaus hat die Komplexität der Umfrage einen erheblichen Einfluss auf die Bearbeitungszeit und die Teilnahme. Beispielsweise stellte Žmuk (2017) fest, dass in Betriebsumfragen die Anzahl der Aufgaben und der erforderliche kognitive Aufwand, insbesondere bei Fragebögen mit mehr offenen oder vielfältigen Fragen, die Zeit, die für die Beantwortung benötigt wird, erheblich verlängerten. Für die Abschätzung der Folgen für den Rücklauf und die entsprechenden Kosten ist vorgesehen, mit verschiedenen Rücklaufszenarien zu kalkulieren.

Eine wichtige inhaltliche Frage, die im Zusammenhang mit der Konsolidierung der Befragungsinstrumente diskutiert werden sollte, betrifft die Abfrage der offenen Stellen in beiden Befragungsteilen. Derzeit werden die offenen Stellen zwar im IAB-Betriebspanel erhoben, jedoch unterscheiden sich das Erhebungskonzept und hier insbesondere der Erfassungszeitraum grundlegend von der entsprechenden Fragestellung in der IAB-Stellenerhebung.²⁰ Eine vereinheitlichte Abfrage in beiden Befragungsteilen (gepaart mit der Vereinheitlichung der Befragungszeitpunkte, siehe nächster Abschnitt) böte prinzipiell das Potenzial, die Informationen aus der Querschnitts- sowie der Längsschnittbefragung in die Datenlieferung für das Quartal der Hauptbefragung an Eurostat einfließen zu lassen und – abhängig vom erwarteten Rücklauf in den Nachbefragungen – möglicherweise Einsparungen bei der Querschnitts-Bruttostichprobe zu realisieren. Dies würde jedoch Anpassungen bei der Hochrechnung erfordern, da für einen Teil der Befragung Querschnittsgewichte und für den anderen Teil Längsschnittgewichte zum Einsatz kämen.

- **Vereinheitlichung der Feldphasen:** Die Hauptbefragung der IAB-Stellenerhebung findet derzeit im vierten Quartal eines jeden Jahres statt, während die Feldphase des IAB-Betriebspanels derzeit vom dritten bis ins vierte Quartal eines jeden Jahres andauert (siehe Kapitel 3.1.1). Die oben dargestellte Konsolidierung der Befragungsinstrumente wirft die Frage nach der Vereinheitlichung der Feldphasen und damit einhergehenden Befragungszeitpunkte sowie die damit verbundenen Implikationen für den Total Survey Error auf. Da die Folgen einer potenziellen Vereinheitlichung der Feldphasen im weiteren Projektverlauf nicht experimentell untersucht werden können, muss die Abschätzung der Implikationen einer solchen Vereinheitlichung (analog zu der Modus-Vereinheitlichung sowie Fragebogenkonsolidierung) auf Basis einer umfassenden Analyse der bisher in der survey-methodologischen publizierten Ergebnisse zur Variation von Befragungszeitpunkten in Betriebsbefragungen erfolgen. Für die Abschätzung der Folgen für den Rücklauf und die entsprechenden Kosten ist vorgesehen, mit verschiedenen Rücklaufszenarien zu kalkulieren. Ein weiterer zentraler Punkt, der in diesem Zusammenhang diskutiert werden muss, betrifft

²⁰ Im Rahmen des IAB-Betriebspanels lauten die Fragen: Suchen Sie derzeit Arbeitskräfte (ohne Auszubildende), die sofort – also zum nächstmöglichen Einstellungstermin – eingestellt werden sollen? [Wenn ja:] Wie viele Arbeitskräfte suchen Sie für sofort? [Anzahl insgesamt sowie differenziert nach drei Beschäftigtengruppen]

Im Rahmen der IAB-Stellenerhebung lauten die Fragen: Suchen Sie derzeit aktiv neue Arbeitskräfte? Bitte berücksichtigen Sie bei der Angabe zu den offenen Stellen KEINE Ausbildungsverhältnisse, KEINE Entfristungen oder Vertragsverlängerungen, KEINE Arbeitskräfte, die Sie von Zeit-/Leiharbeitsfirmen entleihen wollen, KEINE öffentlich geförderte Beschäftigung (z.B. Ein-Euro-Jobs). Wenn ja, wie viele? [Anzahl offene Stellen + davon der Arbeitsagentur gemeldet, jeweils insgesamt sowie differenziert nach fünf Beschäftigtengruppen (-> Berufliche Anforderungsniveaus gem. KldB.2010)]. Wie viele der gesamten offenen Stellen sind sofort zu besetzen (d.h. derzeit unbesetzt) bzw. später zu besetzen? [Anzahl insgesamt]

die Anschlussfähigkeit der Panelbefragung an die derzeit bestehenden Erweiterungen des IAB-Betriebspanels (LIAB, LPP).

5 Fazit auf Basis der bisherigen Ergebnisse und Ausblick

Als ein Teilprojekt des vom BMAS finanzierten Projekts „Förderung innovativer Ansätze zur Stärkung von Dateninfrastruktur und Methoden“ führt das IAB die „Machbarkeitsstudie zur Weiterentwicklung der Betriebsbefragungen am IAB“ durch. In dem vorliegenden Projektbericht werden erste Ergebnisse – jedoch noch keine abschließenden Befunde – zu den folgenden Kernfragestellungen präsentiert: Ist es möglich, die beiden bestehenden Betriebsbefragungen des IAB (IAB-Stellenerhebung und IAB-Betriebspanel) um hochfrequente Elemente zu erweitern? Ist es möglich, die beiden Befragungen in eine Befragung zu integrieren? Zur zweiten Fragestellung wurden und werden im weiteren Projektverlauf Machbarkeitsstudien für unterschiedliche Integrationsvarianten durchgeführt. Als Nebenbedingungen sind dabei zu beachten, dass weiterhin eine Panelbefragung von Betrieben erfolgen soll, dass die Lieferverpflichtungen an Eurostat erfüllt werden, dass weiterhin Informationen zu den Einstellungsprozessen erhoben werden und dass beratungs- und forschungsrelevante Informationen in höherer Frequenz als einmal jährlich zur Verfügung stehen.

Zur ersten Fragestellung, der Integration hochfrequenter Elemente in die bestehenden Befragungen, wurden in beiden Befragungen Experimente durchgeführt, um die Effekte hochfrequenter Elemente vor allem auf den erwarteten Rücklauf zu untersuchen. Im Rahmen des IAB-Betriebspanels wurde im Jahr 2024 eine unterjährige, telefonische Befragung zwischen den Haupterhebungen der Jahre 2023 und 2024 durchgeführt. Auf dieser Basis lassen sich die Teilnahmebereitschaft von Betrieben an einer zusätzlichen Erhebung sowie die Umsetzbarkeit beurteilen. Um auch den Effekt der unterjährigen Zusatzerhebung auf die Teilnahmebereitschaft an zukünftigen Erhebungen untersuchen zu können, wurden nur zufällig ausgewählte erst- und wiederholtbefragte Betriebe zur unterjährigen Befragung eingeladen. Die bisherigen Analysen zeigen, dass eine unterjährige Befragung in diesem Umfang prinzipiell umsetzbar ist. Allerdings willigt nur ein relativ geringer Teil der befragten Betriebe ein, für eine zusätzliche unterjährige Befragungen kontaktiert zu werden. Bei den kontaktierten Betrieben bewegt sich die tatsächliche Teilnahmebereitschaft jedoch im erwarteten Rahmen. Eine Aussage über den Einfluss der unterjährigen Befragung auf die Teilnahmebereitschaft an zukünftigen Erhebungen kann zum jetzigen Zeitpunkt noch nicht getroffen werden, da die Daten der Haupterhebung 2024 noch ausstehen. Dieser Aspekt wird daher im weiteren Projektverlauf untersucht.

In der IAB-Stellenerhebung werden im Status-Quo bereits quartalsweise Nachbefragungen durchgeführt. Im Rahmen dieser Nachbefragungen wurde ein Survey-Experiment implementiert, um die Einführung weiterer hochfrequenter Fragen und die Auswirkungen verlängerter Fragebögen auf die Rücklaufquote und weitere Ergebnisvariablen zu untersuchen. Nach Einteilung einer Experimentalstichprobe in zwei Gruppen, von der jeweils eine in den Quartals-Nachbefragungen einen 2-seitigen Fragebogen und die andere einen 4-seitigen Fragebogen

erhielt, zeigen die Ergebnisse, dass hinsichtlich der Rücklaufquote keine signifikanten Unterschiede zwischen den Gruppen nachgewiesen werden konnten. Auch im Vergleich zum Status-Quo, in dem die Betriebe in der Nachbefragung einen 1-seitigen Fragebogen erhalten, fallen die Unterschiede gering aus. Bezüglich der *Response Burden* deuten die Ergebnisse jedoch auf leichte Vorteile des 2-seitigen im Vergleich zu einem 4-seitigen Fragebogen hin. Dies deutet darauf hin, dass die kürzere Variante etwas besser geeignet sein könnte, die Teilnahmebereitschaft langfristig aufrechtzuerhalten, ohne die Belastung der Befragten zu erhöhen. Die Experimental-Ergebnisse aller Nachbefragungen konnten in diesem Bericht noch nicht berücksichtigt werden, da die dazu erforderlichen Daten erst nach dem Fälligkeitsdatum dieses Berichts verfügbar sein werden. Die Analyse dieser Daten, die für den weiteren Projektverlauf vorgesehen ist, wird Aufschluss darüber liefern, ob die oben dargestellten Ergebnisse weiterhin Gültigkeit behalten. Insbesondere ermöglicht die Verfügbarkeit der Daten über alle Quartale, die langfristigen Auswirkungen unterschiedlicher Fragebogenlängen auf die Datenqualität abzuschätzen und hiermit eine valide Schätzung der Kosten für die Implementierung hochfrequenter Fragen vorzunehmen.

Mit Blick auf die Integration der beiden Befragungen wurde zunächst die Machbarkeit eines rollierenden Panels untersucht. Hier wird eine jährliche Panelstichprobe in vier Substichproben aufgeteilt, welche zu vier unterschiedlichen Zeitpunkten im Jahr (quartalsweise) als Panelerhebung befragt werden. Aufgrund der Anforderungen für die Meldung der offenen Stellen wird dabei von quartalsweisen Stichproben von 7.000 Betrieben ausgegangen. Durch diese Gestaltung können Inhalte häufiger als einmal jährlich abgefragt und ein für die Forschung attraktiver Paneldatensatz bereitgestellt werden. Die Lieferverpflichtung an Eurostat kann mit einem rollierenden Panel grundsätzlich erfüllt werden, allerdings wäre ein speziell darauf angepasstes Hochrechnungsverfahren nötig. Eine Integration der Fragebögen der beiden Erhebungen erscheint grundsätzlich möglich, erfordert aber, Inhalte, die bisher erhoben werden, ganz wegzulassen oder nicht in jeder Befragung zu erheben. Im Vergleich zum bisherigen IAB-Betriebspanel würde ein rollierendes Panel zudem substantiell zusätzliche Sach- und Personalmittel erfordern. Die zusätzlichen Sachmittel wären nötig, weil deutlich mehr Betriebe befragt werden und eingekaufte Tätigkeiten bis zu fünf Mal statt einmal pro Jahr anfallen. Die höhere Frequenz erfordert auch zusätzliche Personalmittel, und die kurzen Feldphasen mit gleichzeitig hohen Fallzahlen bringen eine gewisse Anfälligkeit für Erhebungsprobleme mit sich. Schließlich erscheint eine Anknüpfung an die bestehenden Befragungen nicht ohne Brüche in Zeitreihen möglich.

Zur Frage, ob die beiden bestehenden Betriebsbefragungen des IAB um hochfrequente Elemente erweitert werden können, zeigen die bisherigen Ergebnisse, dass eine solche Erweiterung grundsätzlich durchführbar wäre. Allerdings stehen sowohl beim IAB-Betriebspanel als auch bei der IAB-Stellenerhebung noch Analysen zu den Effekten auf die Teilnahmebereitschaft aus. Diese sind jedoch für eine abschließende Beantwortung der Frage nötig und werden im weiteren Projektverlauf durchgeführt.

Hinsichtlich des Projektziels, die Machbarkeit einer Integration der beiden Befragungen in eine Befragung zu untersuchen, liegt mit diesem Bericht eine Analyse für einen ersten Ansatz zur Integration vor. Hier sollen im weiteren Projektverlauf noch mindestens die in Abschnitt 4.2 skizzierte Variante ausgehend von der IAB-Stellenerhebung untersucht werden. Abhängig von

den Ergebnissen zu den Effekten unterjähriger zusätzlicher Erhebungen auf die Teilnahmebereitschaft am IAB-Betriebspanel soll auch eine Integration über Zusatzerhebungen zum IAB-Betriebspanel evaluiert werden. Schließlich soll im weiteren Projektverlauf untersucht werden, ob sich die Hochrechnung der Zahl der offenen Stellen optimieren lässt, um mit einer geringeren Fallzahl die Meldeverpflichtungen gegenüber Eurostat zu erfüllen. Gelingt dies, kann die geringere benötigte Fallzahl die Kosten bei einer Integration der beiden Befragungen reduzieren.

Literaturverzeichnis

American Association for Public Opinion Research. (2023). 2023 Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.

Bächmann, Ann-Christin; Bellmann, Lisa; Gensicke, Miriam; Kohaut, Susanne; Möller, Iris; Schwengler, Barbara; Tschersich, Nikolai; Umkehrer, Matthias (2023): [IAB-Betriebspanel \(IAB-BP\) 1993-2022 \(16/2023\)](#), FDZ-Datenreport. Abruf am 04.12.2024.

Backhaus, Nils; Bellmann, Lutz; Gleiser, Patrick; Hensgen, Sophie; Kagerl, Christian; Koch, Theresa; König, Corinna; Kleifgen, Eva; Kuhn, Moritz; Leber, Ute; Moritz, Michael; Pohlan, Laura; Robelski, Swantje; Roth, Duncan; Schierholz, Malte; Sommer, Sabine; Stegmaier, Jens; Tisch, Anita; Umkehrer, Matthias; Aminian, Armin (2022): Panel [“Establishments in the Covid-19 Crisis” – 20/21/22. A longitudinal study in German establishments – waves 1 - 24](#). FDZ-Datenreport 09/2022, Nürnberg.

Bavdaž, Mojca (2010): [Sources of Measurement Errors in Business Surveys](#). In: Journal of Official Statistics, Jg. 26, H. 1, S. 25–42. Abruf am 04.12.2024.

Bellmann, Lisa; Möller, Iris; Voll, Maximilian (2024): [Nutzungshäufigkeit der Fragen des IAB-Betriebspanels. \(23\)](#). IAB-Forschungsbericht. Abruf am 04.12.2024.

Blumenberg, Cauane, Menezes, Ana M. B.; Gonçalves, Helen; Assunção, Maria Cecília Formoso; Wehrmeister, Fernando C.; Barros, Fernando C.; Barros, Aluisio J. D. (2019): [The role of questionnaire length and reminders frequency on response rates to a web-based epidemiologic study: a randomised trial](#). In: International Journal of Social Research Methodology, Jg. 22, H. 6, S. 625–635.

Börschlein, Erik-Benjamin; Diegmann, André (2024, 21. Oktober 2024): [14 Euro Mindestlohn: Rund in Fünftel der Betriebe erwartet einen Beschäftigungsrückgang](#)., Abruf am 04.12.2024.

Bosler, Mario; Gürtzgen, Nicole; Kubis, Alexander; Küfner, Benjamin; Lochner, Benjamin (2020): [The IAB Job Vacancy Survey: design and research potential](#). Journal for Labour Market Research, Jg. 54, H. 1, S. 13.

Brenzel, Hanna; Czepek, Judith; Kiesl, Hans; Kriechel, Ben; Kubis, Alexander; Moczall, Andreas; Rebien, Martina; Röttger, Christof; Szameitat, Jörg; Warning, Anja; Weber, Enzo (2016): [Revision der IAB-Stellenerhebung: Hintergründe, Methode und Ergebnisse. IAB-Forschungsbericht 201604 \(2016\). Revision der IAB-Stellenerhebung: Hintergründe, Methode und Ergebnisse \(04/2016\)](#). IAB-Forschungsbericht, Abruf am 04.12.2024.

- Burchell, Brendan; Marsh, Catherine (1992): [The effect of questionnaire length on survey response](#). In: Quality and Quantity, Jg. 26, H. 3, S. 233–244.
- Callegaro, Mario; Yang, Yongwei; Bhola, Dennison S.; Dillman, Don A.; Chin, Tzu-Yun (2009): [Response latency as an indicator of optimizing in online questionnaires](#). In: Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, Vol. 103, S. 5–25.
- Cernat, Alexandru; Sakshaug, Joseph; Christmann, Pablo; Gummer, Tobias (2022): [The Impact of Survey Mode Design and Questionnaire Length on Measurement Quality](#). In: Sociological Methods und Research, Jg. 53, H. 4, S. 1873–1904.
- Dillman, Don A. (2017): [The promise and challenge of pushing respondents to the Web in mixed-mode surveys](#). In: Survey Methodology Jg. 43, H. 1, S. 3–30. Abruf am 04.12.2024.
- Dillman, Don A.; Sinclair, Michael D.; Clark, Jon R. (1993): [Effects of Questionnaire Length, Respondent-Friendly Design, and a Difficult Question on Response Rates For Occupant-Addressed Census Mail Surveys](#). In: Public Opinion Quarterly, Jg. 57, H. 3, S. 289–304.
- Dillman, Don A.; Smyth, Jolene D.; Christian, Leah M. (2014): Internet, phone, mail, and mixed mode surveys: The tailored design method (4th ed. ed.). John Wiley und Sons Inc.
- Erikson, Johan; Giesen, Deirdre; Houben, Leanne; Saraiva, Paulo (2023): [Managing Response Burden for Official Statistics Business Surveys – Experiences and Recent Developments at Statistics Netherlands, Statistics Portugal, and Statistics Sweden](#). In: Ger Snijkers, Mojca Bavdaž, Stefan Bender, Jacqui Jones, Steve MacFeely, Joseph W. Sakshaug, Katherine J. Thompson, Arnout van Delden (Eds.), Advances in Business Statistics, Methods and Data Collection, S. 193–223.
- Eurostat. (2024a). [ESS Guidelines on Seasonal Adjustment](#).
- Eurostat. (2024b). [Offene Stellen](#). Abruf am 04.12.2024.
- Eurostat. (2024c). [Statistical requirements compendium](#).
- Faul, Franz; Erdfelder, Edgar; Lang, Albert-Georg; Buchner, Axel (2007): G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. In: Behavior Research Methods Jg. 39, S. 175–191.
- Fraze, Steve D.; Hardin, Kelly K.; Brashears, M. Todd; Haygood, Jacqui L.; Smith, James H. (2003): [The Effects Of Delivery Mode Upon Survey Response Rate And Perceived Attitudes Of Texas Agri-Science Teachers](#). In: Journal of Agricultural Education, JG. 44, S. 27–37.
- Galesic, Mirta; Bosnjak, Michael (2009): [Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey](#). In: Public Opinion Quarterly, Jg. 73, H. 2, S. 349–360.
- Giesen, Deirdre (2012, 11-14 June 2012): Exploring Causes and Effects of Perceived Response Burden. Fourth International Conference on Establishment Surveys (ICES IV), Montreal, Canada.
- Gleiser, Patrick; Sakshaug, Joseph W.; Volkert, Marieke; Ellguth, Peter; Kohaut, Susanne; Möller, Iris (2022): [Introducing Web in a Mixed-Mode Establishment Survey: Effects on Nonresponse](#). In: Journal of the Royal Statistical Society Series A: Statistics in Society, Jg. 185, H. 3, S. 891–915.
- Groves, Robert M.; Fowler, Floyd J.; Couper, Mick P.; Lepkowski, James M.; Singer, Eleanor; Tourangeau, Roger (2004): Survey Methodology. John Wiley und Sons, Inc.

- Groves, Robert M.; Peytcheva, Emilia (2008): [The impact of nonresponse rates on nonresponse bias: A meta-analysis](#). In: Public Opinion Quarterly, Jg. 72 H. 2; S. 167–189.
- Guo, Yimeng; Kopec, Jacek A.; Cibere, Jolanda; Li, Linda C.; Goldsmith, Charles H. (2016): [Population Survey Features and Response Rates: A Randomized Experiment](#). In: Am J Public Health, Jg. 106, H. 8, S. 1422–1426.
- Haas, Georg-Christoph; Eckman, Stephanie; Bach, Ruben (2021): [Comparing the Response Burden between Paper and Web Modes in Establishment Surveys](#). In: Journal of Official Statistics, Jg. 37, H. 4, S. 907–930.
- Haraldsen, Gustav; Jones, Jacqui; Giesen, Deirdre; Zhang, Li-Chun (2013): [Understanding and Coping with Response Burden](#). In: In Designing and Conducting Business Surveys, S. 219–252.
- Hartmann, Michael; Reimer, Kim (2010): [Umstellung der Statistik der gemeldeten Arbeitsstellen](#). Methodenbericht der Statistik der BA. Abruf am 04.12.2024.
- Höhne, Jan K.; Schlosser, Stephan (2018): [Investigating the adequacy of response time outlier definitions in computer-based web surveys using paradata](#) Survey Focus. In: Social Science Computer Review, Jg. 36, S. 369–378.
- Hutter, Christian (2020): [A new indicator for nowcasting employment subject to social security contributions in Germany](#). In: Journal for Labour Market Research, Jg. 54, H. 1, S. 1–10.
- Iglesias, Cynthia; Torgerson, David (2000): [Does length of questionnaire matter? A randomised trial of response rates to a mailed questionnaire](#). In: Journal of Health Services Research und Policy, Jg. 5, H. 4, S. 219–221.
- Koitsalu, Marie; Eklund, Martin; Adolfsson, Jan; Grönberg, Henrik; Brandberg, Yvonne (2018): [Effects of pre-notification, invitation length, questionnaire length and reminder on participation rate: a quasi-randomised controlled trial](#). In: BMC Medical Research Methodology, Jg. 18, H. 1, S. 3.
- Kreuter, Frauke (2013): Improving surveys with paradata: introduction. In: Frauke Kreuter (Ed.), Improving Surveys with Paradata: Analytic Uses of Process Information, S. 1-9. Wiley.
- Krosnick, Jon A.; Holbrook, Allyson L.; Berent, Matthew K.; Carson, Richard T.; Hanemann, Michael W.; Kopp, Raymond J.; Cameron Mitchell, Robert; Presser, Stanley; Ruud, Paul A.; Smith, Kerry V.; Moody, Wendy R.; Green, Melanie C.; Conaway, Michael (2002): [The Impact of "No Opinion" Response Options on Data Quality: Non-Attitude Reduction or an Invitation to Satisfice?](#) In: Public Opinion Quarterly, Jg. 66, H. 3, S. 371–403.
- Küfner, Benjamin; Sakshaug, Joseph W.; Zins, Stefan (2022): [Analysing Establishment Survey Non-Response Using Administrative Data and Machine Learning](#). In: Journal of the Royal Statistical Society Series A: Statistics in Society, Vol. 185, Supplement 2, S. S310–S342.
- Küfner, Benjamin; Sakshaug, Joseph W.; Zins, Stefan; Globisch, Claudia (2024): [The Impact of Mail, Web, and Mixed-Mode Data Collection on Participation in Establishment Surveys](#). In: Journal of Survey Statistics and Methodology.
- Lenth, Russell V. (2024): [emmeans: Estimated Marginal Means, aka Least-Squares Means](#). In (Version R package version 1.10.1) Abruf am 04.12.2024.

- Maslovskaya, Olga; Struminskaya, Bella; Durrant, Gabriele (2022): [The Future of Online Data Collection in Social Surveys: Challenges, Developments and Applications](#). In: Journal of the Royal Statistical Society Series A: Statistics in Society, Vol. 185, H. 3, S. 768–772.
- McClain, Colleen A.; Couper, Mick P.; Hupp, Andrew L.; Keusch, Florian; Peterson, Gregg; Piskowski, Andrew D.; West, Brady (2019): [A Typology of Web Survey Paradata for Assessing Total Survey Error](#). In: Social Science Computer Review, Jg. 37, H. 2., S. 196–213.
- McPhee, Cameron; Barlas, Frances; Brigham, Nancy; Darling, Jill; Dutwin, David; Jackson, Chris; Jackson, Mickey; Kirzinger, Ashley; Little, Roderick; Lorenz, Emily; Marlar, Jenny; Mercer, Andrew; Scanlon, Paul J.; Weiss, Steffen; Wronski, Laura (2022): [Data Quality Metrics for Online Samples: Considerations for Study Design and Analysis](#). Abruf am 04.12.2024.
- Mittereder, Felicitas; West, Brady T. (2021): [A dynamic survival modeling approach to the prediction of web survey breakoff](#). In: Journal of Survey Statistics and Methodology, Jg. 10, H. 4, S. 945–978.
- Olson, Kristen; Peytchev, Andy (2007): [Effect of Interviewer Experience on Interview Pace and Interviewer Attitudes](#). In: Public Opinion Quarterly, Jg. 71, H. 2, S. 273–286.
- Presser, Stanley; Couper, Mick P.; Lessler, Judith T.; Martin, Elizabeth; Martin, Jean; Rothgeb, Jennifer M.; Singer, Eleanor (2004): [Methods for Testing and Evaluating Survey Questions](#). In: Public Opinion Quarterly, Jg. 68, H. 1, S. 109–130.
- Revilla, Melanie; Ochoa, Carlos (2017): [Ideal and Maximum Length for a Web Survey](#). In: International Journal of Market Research, Jg. 59, H. 5, S. 557–565.
- Sakshaug, Joseph W.; Vicari, Basha; Couper, Mick P. (2019): [Paper, E-mail, or Both? Effects of Contact Mode on Participation in a Web Survey of Establishments](#). In: Social Science Computer Review, Jg. 37, H. 6, S. 750–765.
- Sandelin, Frida (2022): [The effects of questionnaire length on response rate, non-response bias, and data quality The SOM Institute's Notes on Survey Methodology](#). Abruf am 04.12.2024.
- Schouten, Barry; van den Brakel, Jan; Buelens, Bart; Giesen, Deirdre; Luiten, Annemieke; Meertens, Vivian (2022): [Mixed-Mode Official Surveys: Design and Analysis \(1st ed.\)](#). Chapman and Hall/CRC.
- Seshadri, Srivatsa; Broekemier, Greg M. (2022): [Small Business Executives' Online Survey Response Intentions: The Effects of Incentives and Survey Length](#). In: Small Business Institute Journal, Jg. 18, H. 2, S. 1–9.
- Snijkers, Ger; Demedash, Susan; Andrews, Jessica (2023): [Using Paradata in Electronic Business Survey Questionnaires](#). In: Ger Snijkers; Mojca Bavdaž; Stefan Bender; Jacqui Jones; Steve MacFeely; Joseph W. Sakshaug; Katherine J. Thompson; Arnout v. Delden (Eds.), Advances in Business Statistics, Methods and Data Collection, S. 413–435. John Wiley and Sons, Inc.
- Wagner, James (2012): [A Comparison of Alternative Indicators for the Risk of Nonresponse Bias](#). In: Public Opinion Quarterly, Jg. 76, H. 3, S. 555–575.
- Willimack, Diane K.; Snijkers, Ger (2013): [The business context and its implications for the survey response process](#). In: Ger Snijkers; Gustav Haraldsen; Jacqui Jones; Diane K. Willimack (Eds.), Designing and Conducting Business Surveys, S. 39–82. Wiley.

- Winters, Jutta; Kargus, Andrea (2012): [Die ProIAB: Stets zur Stelle im Dienste der Wissenschaft. Ein Gespräch mit Christine Hense und Christian Sprenger](#). In: IAB-Forum, H. 2, S. 34–39.
- Yan, Ting; Tourangeau, Roger (2008): [Fast times and easy questions: The effects of age, experience and question complexity on web survey response times](#). In: Applied Cognitive Psychology, Jg. 22, S. 51–68.
- Yan, Ting; Williams, Douglas (2022): [Response Burden – Review and Conceptual Framework](#). In: Journal of Official Statistics, Jg. 38, H. 4, S. 939–961.
- Žmuk, Berislav (2017): [Impact of questionnaire length and complexity on survey time: Comparison of two business web questionnaire versions](#). In: Advances in Methodology and Statistics, 14.

Anhang 1: Pretest Ergebnisse

Abbildung 22: ProlAB Pretest Interviewleitfaden

Frage 1 und 2

Sonderfragen Energie	
1. Hat ihr Betrieb in den letzten drei Monaten Energie für den eigenen Bedarf erzeugt?	
Ja ☐	Nein ☐
2. Hat ihr Betrieb in den letzten drei Monaten Energie auf dem Energiemarkt vertrieben oder verkauft?	
Ja ☐	Nein ☐

Quelle: : IAB-Stellenerhebung 4. Quartal 2023 Fragebogen. © IAB

Rückfragen (Reihenfolge beachten):

- Welche Schwierigkeiten haben Sie bei der Beantwortung der Fragen?
- Wissen Sie, wo Sie die Informationen zur Beantwortung dieser Frage finden können?
- Würden Sie im Bedarfsfall die Informationen von anderen Stellen in ihrem Betrieb einholen?
- Ist die Bedeutung der Begriffe „vertrieben“ und „verkauft“ klar?
- Sind Ihnen besondere Umstände bekannt, die die Beantwortung dieser Frage erschweren?
- Erschwert Ihnen die Eingrenzung auf die letzten drei Monate die Beantwortung der Frage?

Rückmeldung der Betriebe:

Frage 1: Die Frage ist für IP klar verständlich. Zur Beantwortung der Frage müsste IP aber eine andere Abteilung befragen.

Frage 2: Die Frage ist für IP klar verständlich. Zur Beantwortung der Frage müsste IP aber eine andere Abteilung befragen. IP vermutet, dass es sich bei dieser Frage nur um das Vorhandensein von einer Photovoltaikanlage handeln könne.

Anhang 2: Teilnahmewahrscheinlichkeit multivariaten Modell

Tabelle 25: Ergebnisse der logistischen Regression zur Teilnahme an der Befragung im vierten Quartal 2023 nach Gruppe, Wirtschaftszweig und Betriebsgröße

	Modell 0			Modell 1			Modell 2			Modell 3		
	Odds ratio	KI	p	Odds ratio	KI	p	Odds ratio	KI	p	Odds ratio	KI	p
(Intercept)	0,17	0,17–0,17	<0,001	0,18	0,17–0,18	<0,001	0,24	0,22–0,26	<0,001	0,24	0,22–0,27	<0,001
Gruppe [Experiment]	0,99	0,95–1,02	0,468	0,98	0,95–1,02	0,411	0,98	0,95–1,02	0,415	0,92	0,75–1,14	0,448
Referenz Betriebsgröße (1-9)												
Betriebsgröße (10-19)				0,92	0,88–0,96	<0,001	0,93	0,89–0,97	0,001	0,91	0,86–0,96	<0,001
Betriebsgröße (20-49)				1,18	1,13–1,23	<0,001	1,20	1,15–1,26	<0,001	1,18	1,13–1,24	<0,001
Betriebsgröße (50-249)				0,99	0,94–1,03	0,525	1,01	0,97–1,06	0,660	1,01	0,96–1,06	0,716
Betriebsgröße (>250)				0,51	0,48–0,55	<0,001	0,51	0,47–0,54	<0,001	0,50	0,46–0,53	<0,001
Referenz WZ Land- u. Forstwirtschaft												
WZ Bergbau/Energie/Wasser/Abfall							0,72	0,65–0,80	<0,001	0,72	0,64–0,80	<0,001
WZ Verarbeitendes Gewerbe							0,89	0,81–0,97	0,008	0,88	0,80–0,98	0,015
WZ Baugewerbe							0,99	0,89–1,10	0,852	0,98	0,87–1,11	0,809
WZ Handel und KFZ-Reparatur							0,61	0,54–0,67	<0,001	0,59	0,52–0,67	<0,001
WZ Verkehr und Lagerei							0,71	0,63–0,79	<0,001	0,70	0,62–0,80	<0,001
WZ Sonstige Dienstleistungen							0,73	0,67–0,80	<0,001	0,73	0,66–0,81	<0,001
WZ Information und kommunikation							0,58	0,52–0,64	<0,001	0,56	0,50–0,63	<0,001
WZ Finanz- und Versicherungsdienstleistungen							0,42	0,38–0,47	<0,001	0,42	0,37–0,47	<0,001
WZ Unternehmensnahe Dienstleistungen							0,63	0,57–0,69	<0,001	0,63	0,57–0,69	<0,001

	Modell 0			Modell 1			Modell 2			Modell 3		
	Odds ratio	KI	p	Odds ratio	KI	p	Odds ratio	KI	p	Odds ratio	KI	p
WZ Öffentliche Verwaltung/Sozialversicherung							0,93	0,83–1,03	0,174	0,95	0,84–1,07	0,385
Referenz Experimentalgruppe Betriebsgröße (1-9)												
Experimentalgruppe * Betriebsgröße (10-19)										1,10	0,98–1,22	0,092
Experimentalgruppe *: Betriebsgröße (20-49)										1,10	0,99–1,21	0,082
Experimentalgruppe * Betriebsgröße (50-249)										1,00	0,90–1,12	0,946
Experimentalgruppe * Betriebsgröße (>250)										1,09	0,93–1,29	0,290
Referenz Experimentalgruppe WZ Land u. Forstwirtschaft												
Experimentalgruppe * WZ Bergbau/Energie/Wasser/Abfall										1,05	0,82–1,34	0,688
Experimentalgruppe * WZ Verarbeitendes Gewerbe										1,01	0,82–1,25	0,917
Experimentalgruppe * WZ Baugewerbe										1,02	0,78–1,34	0,869
Experimentalgruppe * WZ Handel und KFZ-Reparatur										1,11	0,85–1,44	0,443
Experimentalgruppe * WZ Verkehr und Lagerei										1,01	0,78–1,32	0,932
Experimentalgruppe * WZ Sonstige Dienstleistungen										1,00	0,81–1,24	0,997
Experimentalgruppe * WZ Information u. Kommunikation										1,13	0,87–1,46	0,356
Experimentalgruppe * WZ Finanz- und Versicherungs-DL										1,07	0,82–1,41	0,611
Experimentalgruppe * WZ Unternehmensnahe DL										1,00	0,80–1,24	0,991
Experimentalgruppe * WZ Öffentl. Verwaltung/Sozialvers.										0,90	0,69–1,18	0,444
Beobachtungen	139.987			139.987			139.987			139.987		
R ² / R ² adjusted	0,000 / -0,000			0,006 / 0,006			0,012 / 0,012			0,012 / 0,012		

Anmerkung: Die Ergebnisse wurden mit Inklusionswahrscheinlichkeitsgewichte gewichtet. Konfidenzintervalle wurden auf dem 95%-Niveau berechnet.

Quelle: : IAB-Stellenerhebung; eigene Berechnungen. © IAB

Tabelle 26: Paarweise Kontrastergebnisse der emmeans-Interaktionen nach Gruppe und Wirtschaftszweig im vierten Quartal 2023

Kontrast	Wirtschaftszweig	Estimate	Standard Error	95% Konfidenzintervall		Z-Ratio	p-Wert
				Untere Grenze	Obere Grenze		
Standard - Experiment	Land- und Forstwirtschaft	1,027	0,106	0,838	1,258	0,257	0,796
Standard - Experiment	Bergbau/Energie/Wasser/Abfall	0,976	0,070	0,847	1,125	-0,322	0,746
Standard - Experiment	Verarbeitendes Gewerbe	1,015	0,040	0,939	1,097	0,388	0,697
Standard - Experiment	Baugewerbe	1,004	0,092	0,838	1,202	0,043	0,964
Standard - Experiment	Handel und KFZ-Reparatur	0,927	0,078	0,784	1,095	-0,889	0,373
Standard - Experiment	Verkehr und Lagerei	1,015	0,090	0,853	1,208	0,170	0,864
Standard - Experiment	Sonstige Dienstleistungen	1,026	0,041	0,948	1,111	0,652	0,514
Standard - Experiment	Information und kommunikation	0,909	0,075	0,772	1,071	-1,136	0,255
Standard - Experiment	Finanz- und Versicherungsdienstleistungen	0,956	0,091	0,792	1,154	-0,463	0,643
Standard - Experiment	Unternehmensnahe Dienstleistungen	1,028	0,047	0,938	1,126	0,602	0,546
Standard - Experiment	Öffentliche Verwaltung/Sozialversicherung	1,140	0,103	0,954	1,363	1,444	0,148

Quelle: : IAB-Stellenerhebung; eigene Berechnungen. © IAB

Tabelle 27: Paarweise Kontrastergebnisse der emmeans-Interaktionen nach Gruppe und Betriebsgröße im vierten Quartal 2023

Kontrast	Betriebsgröße	Estimate	Standard Error	95% Konfidenzintervall		Z-Ratio	p-Wert
				Untere Grenze	Obere Grenze		
Standard - Experiment	0-9	1,057	0,043	0,976	1,146	1,370	0,170
Standard - Experiment	10-19	0,963	0,040	0,886	1,046	-0,881	0,377
Standard - Experiment	20-49	0,965	0,037	0,895	1,041	-0,904	0,365
Standard - Experiment	50-249	1,053	0,046	0,967	1,148	1,196	0,231
Standard - Experiment	> 250	0,968	0,074	0,833	1,125	-0,421	0,673

Quelle: : IAB-Stellenerhebung; eigene Berechnungen. © IAB

Tabelle 28: Ergebnisse der logistischen Regression zur Teilnahme an der Befragung im ersten Quartal 2024 nach Gruppe, Wirtschaftszweig und Betriebsgröße

	Modell 0			Modell 1			Modell 2			Modell 3		
	Odds ratio	KI	p	Odds ratio	KI	p	Odds ratio	KI	p	Odds ratio	KI	p
(Intercept)	2,34	2,26–2,43	<0,001	2,00	1,87–2,13	<0,001	1,86	1,57–2,21	<0,001	1,76	1,45–2,14	<0,001
Gruppe [2-Seiter]	0,90	0,82–1,00	0,050	0,90	0,81–1,00	0,045	0,90	0,82–1,00	0,052	0,93	0,52–1,63	0,791
Gruppe [4-Seiter]	0,91	0,82–1,01	0,072	0,91	0,82–1,00	0,057	0,91	0,82–1,00	0,062	1,50	0,83–2,70	0,176
Referenz Betriebsgröße (1-9)												
Betriebsgröße (10-19)				1,20	1,09–1,31	<0,001	1,20	1,10–1,32	<0,001	1,18	1,06–1,31	0,002
Betriebsgröße (20-49)				1,33	1,22–1,45	<0,001	1,34	1,22–1,46	<0,001	1,29	1,17–1,42	<0,001
Betriebsgröße (50-249)				1,30	1,19–1,43	<0,001	1,30	1,18–1,43	<0,001	1,28	1,15–1,42	<0,001
Betriebsgröße (>250)				0,88	0,76–1,01	0,067	0,86	0,74–0,99	0,036	0,91	0,77–1,07	0,249
Referenz WZ Land- u. Forstwirtschaft												
WZ Bergbau/Energie/Wasser/Abfall							1,35	1,10–1,67	0,005	1,45	1,15–1,84	0,002
WZ Verarbeitendes Gewerbe							1,12	0,94–1,34	0,220	1,18	0,96–1,44	0,115
WZ Baugewerbe							1,24	0,98–1,55	0,069	1,45	1,12–1,88	0,005
WZ Handel und KFZ-Reparatur							0,87	0,70–1,09	0,227	0,86	0,67–1,10	0,228
WZ Verkehr und Lagerei							1,07	0,86–1,34	0,545	1,23	0,96–1,59	0,106
WZ Sonstige Dienstleistungen							0,99	0,83–1,19	0,953	1,09	0,89–1,33	0,422
WZ Information und Kommunikation							0,92	0,74–1,14	0,443	0,99	0,78–1,27	0,966
WZ Finanz- und Versicherungsdienstleistungen							1,02	0,81–1,29	0,851	1,08	0,83–1,41	0,561
WZ Unternehmensnahe Dienstleistungen							1,04	0,86–1,24	0,704	1,08	0,88–1,32	0,486
WZ Öffentliche Verwaltung/Sozialversicherung							1,38	1,09–1,74	0,007	1,64	1,26–2,13	<0,001
Referenz Gruppe 2/4-Seiter: Betriebsgröße (1-9)												
Gruppe 2-Seiter * Betriebsgröße (10-19)										1,17	0,86–1,58	0,310
Gruppe 4-Seiter * Betriebsgröße (10-19)										1,04	0,77–1,41	0,800

	Modell 0			Modell 1			Modell 2			Modell 3		
	Odds ratio	KI	p	Odds ratio	KI	p	Odds ratio	KI	p	Odds ratio	KI	p
Gruppe 2-Seiter * Betriebsgröße (20-49)										1,22	0,92–1,61	0,172
Gruppe 4-Seiter * Betriebsgröße (20-49)										1,13	0,85–1,51	0,389
Gruppe 2-Seiter * Betriebsgröße (50-249)										1,17	0,86–1,59	0,316
Gruppe 4-Seiter * Betriebsgröße (50-249)										1,02	0,75–1,39	0,905
Gruppe 2-Seiter * Betriebsgröße (>250)										1,08	0,68–1,72	0,739
Gruppe 4-Seiter * Betriebsgröße (>250)										0,56	0,35–0,89	0,015
Referenz Gruppe 2/4-Seiter WZ Land u. Forstwirtschaft												
Gruppe 2-Seiter * WZ Bergbau/Energie/Wasser/Abfall										0,75	0,38–1,49	0,415
Gruppe 4-Seiter * WZ Bergbau/Energie/Wasser/Abfall										0,67	0,33–1,35	0,260
Gruppe 2-Seiter * WZ Verarbeitendes Gewerbe										1,02	0,56–1,84	0,960
Gruppe 4-Seiter * WZ Verarbeitendes Gewerbe										0,60	0,33–1,10	0,098
Gruppe 2-Seiter * WZ Baugewerbe										0,47	0,23–0,97	0,042
Gruppe 4-Seiter * WZ Baugewerbe										0,50	0,23–1,07	0,074
Gruppe 2-Seiter * WZ Handel und KFZ-Reparatur										1,45	0,69–3,05	0,322
Gruppe 4-Seiter * WZ Handel und KFZ-Reparatur										0,76	0,36–1,62	0,482
Gruppe 2-Seiter * WZ Verkehr und Lagerei										0,60	0,29–1,25	0,171
Gruppe 4-Seiter * WZ Verkehr und Lagerei										0,45	0,21–0,94	0,034
Gruppe 2-Seiter * WZ Sonstige Dienstleistungen										0,83	0,46–1,51	0,550
Gruppe 4-Seiter * WZ Sonstige Dienstleistungen										0,52	0,28–0,95	0,032
Gruppe 2-Seiter * WZ Information u. Kommunikation										0,65	0,32–1,31	0,225
Gruppe 4-Seiter * WZ Information u. Kommunikation										0,72	0,35–1,49	0,378
Gruppe 2-Seiter * WZ Finanz- u. Versicherungs-DL										0,81	0,38–1,73	0,588
Gruppe 4-Seiter * WZ Finanz- u. Versicherungs-DL										0,71	0,33–1,55	0,391
Gruppe 2-Seiter * WZ Unternehmensnahe Dienstl.										1,00	0,55–1,84	0,991

	Modell 0			Modell 1			Modell 2			Modell 3		
	Odds ratio	KI	p	Odds ratio	KI	p	Odds ratio	KI	p	Odds ratio	KI	p
Gruppe 4-Seiter * WZ Unternehmensnahe Dienstl.										0,69	0,37–1,29	0,246
Gruppe 2-Seiter * WZ Öffentl. Verwaltung/Sozialvers.										0,59	0,28–1,25	0,166
Gruppe 4-Seiter * WZ Öffentl. Verwaltung/Sozialvers.										0,35	0,16–0,74	0,007
Beobachtungen	18.268			18.268			18.268			18.268		
R ² / R ² adjusted	0,000 / 0,000			0,003 / 0,003			0,006 / 0,005			0,008 / 0,005		

Anmerkung: Die Ergebnisse wurden mit Inklusionswahrscheinlichkeitsgewichte gewichtet. Konfidenzintervalle wurden auf dem 95 Prozent-Niveau berechnet.

Quelle: : IAB-Stellenerhebung; eigene Berechnungen. © IAB

Tabelle 29: Paarweise Kontrastergebnisse der emmeans-Interaktionen nach Gruppe und Wirtschaftszweig im ersten Quartal 2024

Kontrast	Wirtschaftszweig	Estimate	Standard Error	95% Konfidenzintervall		Z-Ratio	p-Wert
				Untere Grenze	Obere Grenze		
Standard – Seiter-2	Land- und Forstwirtschaft	0,959	0,274	0,483	1,903	-0,143	0,918
Standard – Seiter-4	Land- und Forstwirtschaft	0,721	0,210	0,358	1,452	-1,117	0,791
Seiter-2 – Seiter-4	Land- und Forstwirtschaft	0,751	0,289	0,298	1,892	-0,740	0,918
Standard – Seiter-2	Bergbau/Energie/Wasser/Abfall	1,276	0,270	0,768	2,121	1,153	0,746
Standard – Seiter-4	Bergbau/Energie/Wasser/Abfall	1,080	0,234	0,642	1,818	0,358	1,000
Seiter-2 – Seiter-4	Bergbau/Energie/Wasser/Abfall	0,846	0,241	0,428	1,673	-0,585	1,000
Standard – Seiter-2	Verarbeitendes Gewerbe	0,945	0,106	0,721	1,237	-0,500	0,617
Standard – Seiter-4	Verarbeitendes Gewerbe	1,201	0,132	0,921	1,564	1,657	0,292
Seiter-2 – Seiter-4	Verarbeitendes Gewerbe	1,270	0,188	0,891	1,811	1,616	0,292
Standard – Seiter-2	Baugewerbe	2,048	0,507	1,132	3,705	2,896	0,011
Standard – Seiter-4	Baugewerbe	1,440	0,375	0,772	2,687	1,401	0,322
Seiter-2 – Seiter-4	Baugewerbe	0,703	0,235	0,315	1,565	-1,053	0,322

Standard - Seiter_2	Handel und KFZ-Reparatur	0,659	0,165	0,361	1,204	-1,653	0,294
Standard – Seiter-4	Handel und KFZ-Reparatur	0,944	0,239	0,515	1,735	-0,225	0,821
Seiter-2 – Seiter-4	Handel und KFZ-Reparatur	1,431	0,481	0,639	3,200	1,066	0,572
Standard – Seiter-2	Verkehr und Lagerei	1,596	0,387	0,892	2,853	1,926	0,161
Standard – Seiter-4	Verkehr und Lagerei	1,602	0,391	0,892	2,877	1,928	0,161
Seiter-2 – Seiter-4	Verkehr und Lagerei	1,003	0,321	0,465	2,163	0,012	0,990
Standard – Seiter-2	Sonstige Dienstleistungen	1,150	0,129	0,878	1,505	1,242	0,388
Standard – Seiter-4	Sonstige Dienstleistungen	1,395	0,157	1,064	1,830	2,946	0,009
Seiter-2 – Seiter-4	Sonstige Dienstleistungen	1,213	0,181	0,849	1,734	1,297	0,388
Standard – Seiter-2	Information und kommunikation	1,484	0,334	0,865	2,547	1,752	0,238
Standard – Seiter-4	Information und kommunikation	1,002	0,239	0,566	1,774	0,009	0,992
Seiter-2 – Seiter-4	Information und kommunikation	0,674	0,207	0,323	1,406	-1,281	0,400
Standard – Seiter-2	Finanz- und Versicherungsdienstleistungen	1,183	0,317	0,622	2,250	0,629	1,000
Standard – Seiter-4	Finanz- und Versicherungsdienstleistungen	1,012	0,275	0,527	1,944	0,046	1,000
Seiter-2 – Seiter-4	Finanz- und Versicherungsdienstleistungen	0,855	0,305	0,363	2,011	-0,437	1,000
Standard – Seiter-2	Unternehmensnahe Dienstleistungen	0,956	0,128	0,693	1,318	-0,332	1,000
Standard – Seiter-4	Unternehmensnahe Dienstleistungen	1,041	0,139	0,756	1,433	0,302	1,000
Seiter-2 -Seiter-4	Unternehmensnahe Dienstleistungen	1,088	0,193	0,711	1,667	0,477	1,000
Standard – Seiter-2	Öffentliche Verwaltung/Sozialversicherung Deutschland	1,633	0,420	0,883	3,023	1,910	0,112
Standard – Seiter-4	Öffentliche Verwaltung/Sozialversicherung Deutschland	2,079	0,531	1,128	3,832	2,866	0,012
Seiter-2 -Seiter-4	Öffentliche Verwaltung/Sozialversicherung Deutschland	1,272	0,430	0,566	2,858	0,713	0,475

Quelle: : IAB-Stellenerhebung; eigene Berechnungen. © IAB

Tabelle 30: Paarweise Kontrastergebnisse der emmeans-Interaktionen nach Gruppe und Betriebsgröße im ersten Quartal 2024

Kontrast	Betriebsgröße	Estimate	Standard Error	95% Konfidenzintervall		Z-Ratio	p-Wert
				Untere Grenze	Obere Grenze		
Standard-Seiter-2	0-9	1,357	0,146	1,048	1,757	2,833	0,013
Standard-Seiter-4	0-9	1,091	0,121	0,835	1,426	0,783	0,433
Seiter-2-Seiter-4	0-9	0,803	0,117	0,567	1,139	-1,496	0,268
Standard-Seiter-2	10-19	1,161	0,142	0,865	1,558	1,215	0,672
Standard – Seiter-4	10-19	1,049	0,126	0,786	1,401	0,401	1,000
Seiter-2-Seiter-4	10-19	0,904	0,145	0,614	1,329	-0,625	1,000
Standard-Seiter-2	20-49	1,115	0,121	0,859	1,448	1,003	0,929
Standard-Seiter-4	20-49	0,962	0,106	0,739	1,254	-0,342	0,929
Seiter-2-Seiter-4	20-49	0,863	0,125	0,609	1,221	-1,015	0,929
Standard-Seiter-2	50-249	1,159	0,147	0,854	1,572	1,161	0,736
Standard-Seiter-4	50-249	1,071	0,135	0,791	1,450	0,542	1,000
Seiter-2-Seiter-4	50-249	0,923	0,155	0,617	1,382	-0,469	1,000
Standard-Seiter-2	>250	1,254	0,271	0,747	2,107	1,047	0,294
Standard-Seiter-4	>250	1,948	0,424	1,157	3,280	3,064	0,006
Seiter-2-Seiter-4	>250	1,552	0,446	0,780	3,089	1,531	0,251

Quelle: : IAB-Stellenerhebung; eigene Berechnungen. © IAB

Tabelle 31: Ergebnisse der logistischen Regression zur Teilnahme an der Befragung im zweiten Quartal 2024 nach Gruppe, Wirtschaftszweig und Betriebsgröße

	Modell 0			Modell 1			Modell 2			Modell 3		
	Odds ratio	KI	p	Odds ratio	KI	p	Odds ratio	KI	p	Odds ratio	KI	p
(Intercept)	2,04	1,97–2,11	<0,001	1,74	1,64–1,85	<0,001	1,75	1,47–2,07	<0,001	1,82	1,50–2,21	<0,001
Gruppe [2-Seiter]	0,96	0,87–1,06	0,432	0,96	0,87–1,06	0,408	0,96	0,87–1,06	0,429	0,71	0,41–1,24	0,227
Gruppe [4-Seiter]	0,88	0,80–0,98	0,015	0,88	0,80–0,97	0,011	0,88	0,80–0,97	0,012	0,84	0,49–1,45	0,529
Referenz Betriebsgröße (1-9)												
Betriebsgröße (10-19)				1,22	1,11–1,33	<0,001	1,23	1,12–1,35	<0,001	1,17	1,06–1,30	0,002
Betriebsgröße (20-49)				1,31	1,20–1,42	<0,001	1,32	1,21–1,43	<0,001	1,31	1,19–1,44	<0,001
Betriebsgröße (50-249)				1,29	1,18–1,42	<0,001	1,30	1,19–1,43	<0,001	1,30	1,17–1,44	<0,001
Betriebsgröße (>250)				0,89	0,78–1,03	0,109	0,87	0,76–1,01	0,063	0,93	0,79–1,09	0,382
Referenz WZ Land- u. Forstwirtschaft												
WZ Bergbau/Energie/Wasser/Abfall							1,22	0,99–1,50	0,058	1,22	0,96–1,53	0,100
WZ Verarbeitendes Gewerbe							1,09	0,91–1,30	0,351	1,03	0,85–1,26	0,747
WZ Baugewerbe							1,01	0,81–1,26	0,951	0,98	0,77–1,26	0,903
WZ Handel und KFZ-Reparatur							0,83	0,67–1,03	0,091	0,79	0,62–1,01	0,063
WZ Verkehr und Lagerei							1,00	0,80–1,24	0,976	0,98	0,77–1,27	0,902
WZ Sonstige Dienstleistungen							0,90	0,76–1,08	0,262	0,89	0,73–1,09	0,261
WZ Information und kommunikation							0,87	0,70–1,08	0,201	0,82	0,65–1,05	0,120
WZ Finanz- und Versicherungsdienstleistungen							1,08	0,86–1,36	0,522	0,99	0,76–1,29	0,938
WZ Unternehmensnahe Dienstleistungen							0,92	0,77–1,10	0,379	0,88	0,72–1,08	0,223
WZ Öffentliche Verwaltung/Sozialversicherung							1,19	0,95–1,49	0,132	1,21	0,94–1,56	0,137
Referenz Gruppe 2/4-Seiter: Betriebsgröße (1-9)												
Gruppe 2-Seiter * Betriebsgröße (10-19)										1,44	1,06–1,95	0,018
Gruppe 4-Seiter * Betriebsgröße (10-19)										1,05	0,78–1,41	0,758

Gruppe 2-Seiter * Betriebsgröße (20-49)										1,14	0,87–1,51	0,340
Gruppe 4-Seiter * Betriebsgröße (20-49)										0,90	0,68–1,19	0,452
Gruppe 2-Seiter * Betriebsgröße (50-249)										1,15	0,85–1,56	0,366
Gruppe 4-Seiter * Betriebsgröße (50-249)										0,89	0,66–1,21	0,466
Gruppe 2-Seiter * Betriebsgröße (>250)										0,81	0,51–1,27	0,350
Gruppe 4-Seiter * Betriebsgröße (>250)										0,70	0,44–1,11	0,132
Referenz Gruppe 2/4-Seiter WZ Land u. Forstwirtschaft												
Gruppe 2-Seiter * WZ Bergbau/Energie/Wasser/Abfall										0,93	0,48–1,80	0,829
Gruppe 4-Seiter * WZ Bergbau/Energie/Wasser/Abfall										1,08	0,56–2,07	0,820
Gruppe 2-Seiter * WZ Verarbeitendes Gewerbe										1,24	0,70–2,21	0,455
Gruppe 4-Seiter * WZ Verarbeitendes Gewerbe										1,27	0,72–2,22	0,409
Gruppe 2-Seiter * WZ Baugewerbe										0,99	0,48–2,03	0,976
Gruppe 4-Seiter * WZ Baugewerbe										1,24	0,61–2,52	0,554
Gruppe 2-Seiter * WZ Handel und KFZ-Reparatur										1,61	0,79–3,30	0,192
Gruppe 4-Seiter * WZ Handel und KFZ-Reparatur										0,96	0,48–1,92	0,907
Gruppe 2-Seiter * WZ Verkehr und Lagerei										1,04	0,51–2,14	0,908
Gruppe 4-Seiter * WZ Verkehr und Lagerei										1,06	0,53–2,13	0,873
Gruppe 2-Seiter * WZ Sonstige Dienstleistungen										1,17	0,66–2,07	0,599
Gruppe 4-Seiter * WZ Sonstige Dienstleistungen										0,98	0,56–1,71	0,944
Gruppe 2-Seiter * WZ Information u. Kommunikation										1,25	0,63–2,49	0,525
Gruppe 4-Seiter * WZ Information u. Kommunikation										1,27	0,64–2,52	0,490
Gruppe 2-Seiter * WZ Finanz- u. Versicherungs-DL										1,86	0,86–4,01	0,116
Gruppe 4-Seiter * WZ Finanz- u. Versicherungs-DL										1,18	0,57–2,46	0,657
Gruppe 2-Seiter * WZ Unternehmensnahe Dienstl.										1,17	0,65–2,11	0,592
Gruppe 4-Seiter * WZ Unternehmensnahe Dienstl.										1,29	0,73–2,30	0,379
Gruppe 2-Seiter * WZ Öffentl. Verwaltung/Sozialvers.										1,18	0,56–2,47	0,668

Gruppe 4-Seiter * WZ Öffentl. Verwaltung/Sozialvers.										0,71	0,35–1,45	0,349
Beobachtungen	18.268			18.268			18.268			18.268		
R ² / R ² adjusted	0,000 / 0,000			0,003 / 0,003			0,005 / 0,004			0,006 / 0,004		

Anmerkung: Die Ergebnisse wurden mit Inklusionswahrscheinlichkeitsgewichte gewichtet. Konfidenzintervalle wurden auf dem 95 Prozent-Niveau berechnet.

Quelle: : IAB-Stellenerhebung; eigene Berechnungen. © IAB

Tabelle 32: Paarweise Kontrastergebnisse der emmeans-Interaktionen nach Gruppe und Wirtschaftszweig im zweiten Quartal 2024

Kontrast	Wirtschaftszweig	Estimate	Standard Error	95% Konfidenzintervall		Z-Ratio	p-Wert
				Untere Grenze	Obere Grenze		
Standard – Seiter-2	Land-und Forstwirtschaft	1,289	0,355	0,666	2,494	0,923	0,894
Standard – Seiter-4	Land- und Forstwirtschaft	1,322	0,354	0,695	2,514	1,040	0,894
Seiter-2 – Seiter-4	Land- und Forstwirtschaft	1,025	0,368	0,433	2,426	0,069	0,944
Standard – Seiter-2	Bergbau/Energie/Wasser/Abfall	1,387	0,285	0,847	2,270	1,590	0,335
Standard – Seiter-4	Bergbau/Energie/Wasser/Abfall	1,225	0,251	0,750	2,002	0,992	0,641
Seiter-2 – Seiter-4	Bergbau/Energie/Wasser/Abfall	0,883	0,240	0,460	1,693	-0,455	0,648
Standard – Seiter-2	Verarbeitendes Gewerbe	1,036	0,114	0,795	1,349	0,323	1,000
Standard – Seiter-4	Verarbeitendes Gewerbe	1,044	0,115	0,801	1,361	0,392	1,000
Seiter-2 – Seiter-4	Verarbeitendes Gewerbe	1,007	0,148	0,709	1,432	0,052	1,000
Standard – Seiter-2	Baugewerbe	1,304	0,325	0,717	2,369	1,065	0,860
Standard – Seiter-4	Baugewerbe	1,066	0,267	0,585	1,943	0,257	1,000
Seiter-2 – Seiter-4	Baugewerbe	0,817	0,271	0,369	1,810	-0,606	1,000
Standard – Seiter-2	Handel und KFZ-Reparatur	0,800	0,195	0,445	1,436	-0,911	0,362
Standard – Seiter-4	Handel und KFZ-Reparatur	1,377	0,325	0,782	2,425	1,357	0,349
Seiter-2 – Seiter-4	Handel und KFZ-Reparatur	1,722	0,548	0,802	3,693	1,705	0,264
Standard – Seiter-2	Verkehr und Lagerei	1,235	0,305	0,683	2,235	0,855	1,000

Kontrast	Wirtschaftszweig	Estimate	Standard Error	95% Konfidenzintervall		Z-Ratio	p-Wert
				Untere Grenze	Obere Grenze		
Standard – Seiter-4	Verkehr und Lagerei	1,248	0,301	0,700	2,225	0,919	1,000
Seiter-2 – Seiter-4	Verkehr und Lagerei	1,010	0,327	0,465	2,193	0,031	1,000
Standard – Seiter-2	Sonstige Dienstleistungen	1,105	0,123	0,846	1,444	0,898	0,369
Standard – Seiter-4	Sonstige Dienstleistungen	1,349	0,148	1,036	1,755	2,722	0,019
Seiter-2 – Seiter-4	Sonstige Dienstleistungen	1,220	0,178	0,859	1,733	1,358	0,348
Standard – Seiter-2	Information und kommunikation	1,030	0,232	0,601	1,767	0,134	1,000
Standard – Seiter-4	Information und kommunikation	1,039	0,239	0,598	1,803	0,166	1,000
Seiter-2 – Seiter-4	Information und kommunikation	1,008	0,303	0,490	2,072	0,026	1,000
Standard – Seiter-2	Finanz- und Versicherungsdienstleistungen	0,694	0,198	0,350	1,379	-1,270	0,582
Standard – Seiter-4	Finanz- und Versicherungsdienstleistungen	1,119	0,298	0,591	2,120	0,423	0,671
Seiter-2 – Seiter-4	Finanz- und Versicherungsdienstleistungen	1,611	0,591	0,668	3,882	1,298	0,582
Standard – Seiter-2	Unternehmensnahe Dienstleistungen	1,098	0,141	0,807	1,494	0,731	1,000
Standard – Seiter-4	Unternehmensnahe Dienstleistungen	1,021	0,132	0,749	1,393	0,167	1,000
Seiter-2 -Seiter-4	Unternehmensnahe Dienstleistungen	0,930	0,159	0,617	1,401	-0,422	1,000
Standard – Seiter-2	Öffentliche Verwaltung/Sozialversicherung	1,096	0,284	0,589	2,039	0,356	0,721
Standard – Seiter-4	Öffentliche Verwaltung/Sozialversicherung	1,855	0,447	1,041	3,305	2,564	0,031
Seiter-2 – Seiter-4	Öffentliche Verwaltung/Sozialversicherung	1,692	0,561	0,764	3,746	1,584	0,226

Quelle: : IAB-Stellenerhebung; eigene Berechnungen. © IAB

Tabelle 33: Paarweise Kontrastergebnisse der emmeans-Interaktionen nach Gruppe und Betriebsgröße im zweiten Quartal 2024

Kontrast	Betriebsgröße	Estimate	Standard Error	95% Konfidenzintervall		Z-Ratio	p-Wert
				Untere Grenze	Obere Grenze		
Standard - Seiter-2	0-9	1,173	0,126	0,905	1,519	1,440	0,420
Standard - Seiter-4	0-9	1,101	0,120	0,847	1,433	0,883	0,754
Seiter-2 - Seiter-4	0-9	0,939	0,136	0,663	1,329	-0,431	0,754
Standard - Seiter-2	10-19	0,814	0,101	0,604	1,096	-1,653	0,294
Standard - Seiter-4	10-19	1,051	0,123	0,794	1,392	0,429	0,667
Seiter-2-Seiter-4	10-19	1,291	0,207	0,879	1,896	1,596	0,294
Standard - Seiter-2	20-49	1,025	0,109	0,794	1,323	0,239	0,810
Standard - Seiter-4	20-49	1,225	0,128	0,953	1,574	1,939	0,157
Seiter-2 - Seiter-4	20-49	1,194	0,166	0,855	1,667	1,273	0,405
Standard - Seiter-2	50-249	1,020	0,127	0,756	1,375	0,160	0,872
Standard - Seiter-4	50-249	1,233	0,150	0,920	1,651	1,716	0,258
Seiter-2 - Seiter-4	50-249	1,208	0,197	0,816	1,788	1,156	0,495
Standard - Seiter-2	>250	1,455	0,306	0,879	2,411	1,782	0,149
Standard - Seiter-4	>250	1,570	0,335	0,941	2,619	2,109	0,104
Seiter-2 - Seiter-4	>250	1,078	0,302	0,550	2,112	0,269	0,787

Quelle : IAB-Stellenerhebung; eigene Berechnungen. © IAB

Abbildungsverzeichnis

Abbildung 1:	Anzahl gemeldete offene Stellen ohne Arbeitnehmerüberlassung* im Vergleich zur BA-Registerstatistik	23
Abbildung 2:	Aufbau des Experiments.....	27
Abbildung 3:	Detaillierter Ablauf des Experiments	28
Abbildung 4:	Rücklauf vollständiger Interviews und Ausfälle während der Feldzeit.....	33
Abbildung 5:	Experimentelles Design der Hochfrequenz im Rahmen der IAB-Stellenerhebung	40
Abbildung 6:	Rücklaufquoten nach Gruppe und Erhebungsmodus in der Hauptbefragung	47
Abbildung 7:	Rücklaufquoten nach Gruppe und Erhebungsmodus in der ersten Nachbefragung..	49
Abbildung 8:	Rücklaufquoten nach Gruppe und Erhebungsmodus in der zweiten Nachbefragung	51
Abbildung 9:	Teilnahmewahrscheinlichkeit nach Wirtschaftszweigen (11 Kategorien) im vierten Quartal 2023	59
Abbildung 10:	Teilnahmewahrscheinlichkeit nach 5 Betriebsgrößenklassen im vierten Quartal 2023.....	60
Abbildung 11:	Teilnahmewahrscheinlichkeit nach Wirtschaftszweigen (11 Kategorien) im ersten Quartal 2024	61
Abbildung 12:	Teilnahmewahrscheinlichkeit nach 5 Betriebsgrößenklassen im ersten Quartal 2024.....	62
Abbildung 13:	Teilnahmewahrscheinlichkeit nach Wirtschaftszweigen (11 Kategorien) im zweiten Quartal 2024.....	63
Abbildung 14:	Teilnahmewahrscheinlichkeit nach 5 Betriebsgrößenklassen im zweiten Quartal 2024.....	64
Abbildung 15:	Response Burden im vierten Quartal 2023	68
Abbildung 16:	Response Burden im ersten Quartal 2024	70
Abbildung 17:	Response Burden im zweiten Quartal 2024.....	72
Abbildung 18:	Gruppenspezifischer Anteil Teilnehmer mit fehlenden Angaben im ersten Quartal 2024.....	76
Abbildung 19:	Gruppenspezifischer Anteil Teilnehmer mit fehlenden Angaben im zweiten Quartal 2024	78
Abbildung 20:	Rollierendes Panel – Überblick	82
Abbildung 21:	Integration des Betriebspanels ausgehend von der IAB-Stellenerhebung	95
Abbildung 22:	ProIAB Pretest Interviewleitfaden	107

Tabellenverzeichnis

Tabelle 1:	Auswertbare Fälle des IAB-Betriebspanels	13
Tabelle 2:	Geplante EU-Präzisionsziele am Beispiel der Erhebungswelle 2023/2024.....	21
Tabelle 3:	Aufschlüsselung der Fallzahlen für die Experimentalgruppe	32
Tabelle 4:	Ausfälle in der unterjährigen Befragung.....	34
Tabelle 5:	Average Absolute Relative Nonresponse Bias	35
Tabelle 6:	Abweichungen bei der Anzahl Beschäftigten für Betriebe mit <= 20 Beschäftigten in 2023 (N=1.260). Absolute Werte	36
Tabelle 7:	Abweichungen bei den Beschäftigten für Betriebe mit >20 Beschäftigten in 2023 (N=1.040). Werte in Prozent.....	37
Tabelle 8:	Anzahl der treatmentspezifischen Fragen in den drei Nachbefragungen.....	41
Tabelle 9:	Anzahl der Neueinreichungen aus dem IAB aufgrund des Call for Questions und ihre umgesetzten Themen	42
Tabelle 10:	Vergleich der Rücklaufquoten zwischen Standardgruppe und Experimentalgruppe in der Hauptbefragung	48
Tabelle 11:	Vergleich der Rücklaufquoten zwischen Standardgruppe und Experimentalgruppen in der ersten Nachbefragung	50
Tabelle 12:	Vergleich der Rücklaufquoten zwischen Standardgruppe und Experimentalgruppe in der zweiten Nachbefragung	53
Tabelle 13:	Kumulative Rücklaufquoten (KRR)	54
Tabelle 14:	Vergleich ausgewählter Rücklaufquoten nach 24 Wirtschaftszweigen sowie Cohen's h	55
Tabelle 15:	Vergleich ausgewählter Rücklaufquoten nach 7 Betriebsgrößenklassen sowie Cohen's h	57
Tabelle 16:	Vergleich ausgewählter Rücklaufquoten in Ost- und Westdeutschland sowie Cohen's h	57
Tabelle 17:	Average Absolute Nonresponse Bias im ersten und zweiten Quartal 2024.....	66
Tabelle 18:	Ergebnisse von Standardgruppe und Experimentalgruppe zu den Perceived Response Burden Items im vierten Quartal 2023, differenziert nach dem Erhebungsmodus	69
Tabelle 19:	Ergebnisse der Treatmentgruppen zu den Perceived Response Burden Items im ersten Quartal 2024, differenziert nach dem Erhebungsmodus.....	71
Tabelle 20:	Ergebnisse der Treatmentgruppen zu den Perceived Response Burden Items im zweiten Quartal 2024, differenziert nach dem Erhebungsmodus	73
Tabelle 21:	Item-Nonresponse-Rate im ersten Quartal 2024 nach Gruppe und Erhebungsmodus, differenziert nach dem Erhebungsmodus	74
Tabelle 22:	Item-Nonresponse-Rate im zweiten Quartal 2024 nach Gruppe und Erhebungsmodus, differenziert nach dem Erhebungsmodus	77
Tabelle 23:	Breakoff Rates und Breakoff - Durchschnittliche Dauer nach Gruppe und Quartal....	80
Tabelle 24:	Skizze eines Erhebungszyklus des rollierenden Panels	90
Tabelle 25:	Ergebnisse der logistischen Regression zur Teilnahme an der Befragung im vierten Quartal 2023 nach Gruppe, Wirtschaftszweig und Betriebsgröße	108
Tabelle 26:	Paarweise Kontrastergebnisse der emmeans-Interaktionen nach Gruppe und Wirtschaftszweig im vierten Quartal 2023	110

Tabelle 27:	Paarweise Kontrastergebnisse der emmeans-Interaktionen nach Gruppe und Betriebsgröße im vierten Quartal 2023	110
Tabelle 28:	Ergebnisse der logistischen Regression zur Teilnahme an der Befragung im ersten Quartal 2024 nach Gruppe, Wirtschaftszweig und Betriebsgröße.....	111
Tabelle 29:	Paarweise Kontrastergebnisse der emmeans-Interaktionen nach Gruppe und Wirtschaftszweig im ersten Quartal 2024	113
Tabelle 30:	Paarweise Kontrastergebnisse der emmeans-Interaktionen nach Gruppe und Betriebsgröße im ersten Quartal 2024.....	115
Tabelle 31:	Ergebnisse der logistischen Regression zur Teilnahme an der Befragung im zweiten Quartal 2024 nach Gruppe, Wirtschaftszweig und Betriebsgröße.....	116
Tabelle 32:	Paarweise Kontrastergebnisse der emmeans-Interaktionen nach Gruppe und Wirtschaftszweig im zweiten Quartal 2024.....	118
Tabelle 33:	Paarweise Kontrastergebnisse der emmeans-Interaktionen nach Gruppe und Betriebsgröße im zweiten Quartal 2024	120

Impressum

IAB-Forschungsbericht 2|2025

Veröffentlichungsdatum

13. Februar 2025

Herausgeber

Institut für Arbeitsmarkt- und Berufsforschung
der Bundesagentur für Arbeit
Regensburger Straße 104
90478 Nürnberg

Nutzungsrechte

Diese Publikation ist unter folgender Creative-Commons-Lizenz veröffentlicht:
Namensnennung – Weitergabe unter gleichen Bedingungen 4.0 International (CC BY-SA 4.0)
<https://creativecommons.org/licenses/by-sa/4.0/deed.de>

Bezugsmöglichkeit dieses Dokuments

<https://doku.iab.de/forschungsbericht/2025/fb0225.pdf>

Bezugsmöglichkeit aller Veröffentlichungen der Reihe „IAB-Forschungsbericht“

<https://iab.de/publikationen/iab-publikationsreihen/iab-forschungsbericht/>

Website

<https://iab.de>

ISSN

2195-2655

DOI

[10.48720/IAB.FB.2502](https://doi.org/10.48720/IAB.FB.2502)

Rückfragen zum Inhalt

Nicole Gürtzgen

Telefon: 0911 179-6736

E-Mail: Nicole.Guertzgen@iab.de

Michael Oberfichtner

Telefon: 0911 179-4473

E-Mail: Michael.Oberfichtner@iab.de