FORSCHUNGSDATENZENTRUM
der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung

FDZ-DATENREPORT

Dokumentation zu Arbeitsmarktdaten

11|2026 DE Codebuch und Dokumentation des Websurveys im „Panel Arbeitsmarkt und soziale Sicherung“ (PASS) Datenreport Websurvey 2025

Christian Dickmann, Birgit Jesske, Sabine Schulz (infas Institut für angewandte Sozialwissenschaft GmbH) - Jonas Beste, Nadja Bömmel, Madhu Chauhan, Sandra Dummert, Corinna Frodermann, Sonja Malich, Valentina Prospero, Claudia Wenzig, Mark Trappmann, Stefan Zins, Sebastian Bähr, Matthias Collischon, Stefanie Gundert, Marcel Müller, Jens Stegmaier, Anja Fehn (Institut für Arbeitsmarkt- und Berufsforschung (IAB))

Bundesagentur für Arbeit

Codebuch und Dokumentation des Websurveys im „Panel Arbeitsmarkt und soziale Sicherung“ (PASS)

Datenreport Websurvey 2025

Christian Dickmann, Birgit Jesske, Sabine Schulz (infas Institut für angewandte Sozialwissenschaft GmbH) - Jonas Beste, Nadja Bömmel, Madhu Chauhan, Sandra Dummert, Corinna Frodermann, Sonja Malich, Valentina Prospero, Claudia Wenzig, Mark Trappmann, Stefan Zins, Sebastian Bähr, Matthias Collischon, Stefanie Gundert, Marcel Müller, Jens Stegmaier, Anja Fehn (Institut für Arbeitsmarkt- und Berufsforschung (IAB))

Inhaltsverzeichnis

1 Einleitung 5

1.1 Ziele und Fragestellung des Websurveys 5

1.2 Instrumente und Befragungsprogramm 6

2 Stichprobenumfang 6

3 Datensatzstruktur 8

4 Generierte Variablen 10

4.1 Vercodung offener Restkategorien 10

4.2 Einfache Generierungen 10

5 Datenaufbereitung 12

5.1 Strukturprüfungen und entfernte Interviews 13

5.2 Filterbereinigung und vergebene Missing Codes 15

5.3 Plausibilitätsprüfungen 16

5.4 Anonymisierung 16

6 Gewichtung Websurvey 2025 17

Tabellenverzeichnis

Tabelle 1: Responserate nach Teilstichprobe 7

Tabelle 2: Responserate nach Erhebungssprache der Panelhaupterhebung Welle 19 8

Tabelle 3: Vercodung offener Angaben 10

Tabelle 4: Einfache Generierungen (alphabetische Reihenfolge) 10

Tabelle 5: Übersicht über die Schritte der Datenaufbereitung 12

Tabelle 6: Übersicht über die vergebenen Missing Codes 15

Tabelle 7: Übersicht über die anonymisierten Variablen 16

Tabelle 8: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle 18

Tabelle 9: Logit-Modell zur Teilnahme am Websurvey 20

Tabelle 10: Kenngrößen der Verteilung der vorhergesagten Teilnahmewahrscheinlichkeiten für die Nettostichprobe 22

Tabelle 11: Kenngrößen der Verteilung der Gewichte (Gesamtstichprobe, Personen) 22

Zusammenfassung

Die FDZ-Datenreporte beschreiben die Daten des FDZ im Detail. Diese Reihe hat somit eine doppelte Funktion: Zum einen stellen Nutzerinnen und Nutzer fest, ob die angebotenen Daten für das Forschungsvorhaben geeignet sind, zum anderen dienen sie zur Vorbereitung der Auswertungen. Dieser Datenreport dokumentiert die Aufbereitung des Websurveys des PASS 2025.

Abstract

FDZ-Datenreporte (FDZ data reports) describe FDZ data in detail. As a result, this series of re-ports has a dual function: on the one hand, those using the reports can ascertain whether the data offered is suitable for their research task; on the other, the data can be used to prepare evaluations. This data report documents the data preparation of the PASS websurvey 2025.

Datenverfügbarkeit

Der in diesem Beitrag beschriebene Datensatz ist für die Fachöffentlichkeit zugänglich.

Nähere Informationen dazu auf der Internetseite http://fdz.iab.de/.

1 Einleitung

1.1 Ziele und Fragestellung des Websurveys

Mit dem Panel „Arbeitsmarkt und soziale Sicherung“ (PASS) hat das Institut für Arbeitsmarkt- und Berufsforschung (IAB) 2006 einen Datensatz für die Arbeitsmarkt-, Sozialstaats- und Armutsforschung in Deutschland aufgebaut, der eine neue empirische Grundlage für Wissenschaft und Politikberatung schafft.

Um einen Einstieg in die selbstadministrierte Erhebung im Webmodus zu vollziehen, wurde im Anschluss an Welle 16 im Jahr 2022 erstmalig ein Websurvey im Rahmen des „Panel Arbeitsmarkt und soziale Sicherung“ (PASS) durchgeführt. Das Ziel war dabei insbesondere die Erprobung optimaler Abläufe bei der Durchführung von Webbefragungen in der PASS-Stichprobe. Während für Welle 17 kein Websurvey durchgeführt wurde, erfolgte ein zweiter Websurvey im Anschluss an Welle 18 sowie ein dritter Websurvey im Anschluss an Welle 19. Im Mittelpunkt des Websurveys Welle 19 stand dabei die Erhebung zusätzlicher Fragestellungen, die im Rahmen der Hauptbefragung aus Zeitgründen nicht erhoben werden konnten, wie zum Beispiel ein umfassendes Fragemodul zum Lebensstandard und zur Ressourcenallokation innerhalb von Haushalten. Zudem wurde in dieser Welle erstmalig eine Nachbefragung zur Hauptbefragung der Welle 19 durchgeführt, um die Belastung der Teilnehmenden zu messen.

Als zusätzliche Erhebung im Nachgang zu den jährlichen Panelerhebungswellen im PASS ist der Websurvey grundsätzlich an die dort erfolgreiche Realisierung eines Personeninterviews gekoppelt. Damit wird für jede Welle des Websurveys jeweils eine neue Stichprobe bestimmt. Die panelbereiten Teilnehmerinnen und Teilnehmer der Hauptwelle werden einzeln kontaktiert und die Durchführung der Erhebung erfolgte unabhängig von den anderen Haushaltsmitgliedern. Basis der Stichprobe waren für Welle 19 alle Teilnehmerinnen und Teilnehmer der Hauptstichprobe, für die ein gültig realisiertes Interview im PASS Welle 19 vorlag (11.366 Personen). Eine ausführliche Beschreibung des Designs für den Websurvey und das Vorgehen für die Erhebung findet sich im Methodenbericht PASS Websurvey Haupterhebung 2025 (Jesske et al. 2026). Dieser Datenreport gibt einen Überblick über den Websurvey Welle 19, für den vom 18. Oktober bis zum 20. November 2025 5.264 Personen vollständig befragt wurden.

In Kapitel 1.2 findet sich ein Überblick über das Befragungsprogramm. Der Umfang der Stichprobe des Websurveys wird in Kapitel 2 dargelegt. In den darauffolgenden Kapiteln geht es um die eigentlichen Daten und deren Aufbereitung. In Kapitel 3 wird die Datensatzstruktur und in Kapitel 4 werden die generierten Variablen beschrieben. Weiterhin werden die Schritte der Datenaufbereitung und die dabei getroffenen Entscheidungen in Kapitel 5 erläutert. In Kapitel 6 wird dann das Gewichtungsverfahren vorgestellt.

1.2 Instrumente und Befragungsprogramm

Im PASS Websurvey wurden die Zielpersonen – im Unterschied zur Panelhaupterhebung – ausschließlich auf Ebene der Personen kontaktiert und Interviews realisiert.

Die insgesamt 26 Fragen im Websurvey verteilen sich auf die folgenden Module:

- Soziodemografische Informationen
- Nachbefragung zur Hauptbefragung
- Lebensstandard im Allgemeinen und im Haushalt
- Ressourcenverwaltung und Entscheidungsmacht
- Rolleneinstellungen
- Einstellungen zur Kindergrundsicherung
- Einstellungen Lebensstandard

Die Steuerung einzelner Module erfolgte insgesamt durch drei Preloadvariablen, die zum einen die Teilnahme in den Vorwellen abbildeten (Teilnahme Websurvey Welle 18, Teilnahme Haupterhebung Welle 17) und zum anderen die Erhebungsmethode in Welle 19 widerspiegelten. Die Erhebungsmethode von Welle 19 war für die Steuerung des Moduls *Nachbefragung zur Haupterhebung* maßgeblich. In der Frage selbst ging es um die empfundene Dauer des Interviews und darum, ob diese als „überhaupt nicht belastend“ bzw. „sehr belastend“ eingeschätzt wurde.

Die Fragen in den Blöcken *Ressourcenverwaltung und Entscheidungsmacht*, *Rolleneinstellungen* sowie *Einstellungen Lebensstandard* wurden anhand der Teilnahme bzw. Nichtteilnahme am Websurvey der Welle 18 sowie teilweise an der Nichtteilnahme an der Haupterhebung der Welle 17 gesteuert.

Einige Fragen aus dem Modul *soziodemografische Informationen* (Geschlecht und Geburtsjahr) wurden auch für die Datenprüfungen herangezogen. Abweichungen an dieser Stelle deuteten möglicherweise darauf hin, dass der Online-Fragebogen von einer anderen Person beantwortet wurde. Im Rahmen der Datenprüfung wurden solche Fälle identifiziert und die Interviews als ungültig deklariert (vgl. Kapitel 5).

2 Stichprobenumfang

Ausgangsbasis für den Stichprobenumfang im PASS Websurvey 2025 bilden die befragten Personen der Panelhaupterhebung aus der Welle 19. Dabei muss berücksichtigt werden, dass nur solche Personen für den Websurvey zur Verfügung stehen, für die die Panelbereitschaft[1] vorliegt.

In der Panelhaupterhebung wurden 11.366 Interviews mit panelbereiten Personen realisiert.

[1] Die Panelbereitschaft wird durch den oder die Haushaltsbevollmächtigte erteilt und zählt damit für alle Haushaltsmitglieder. Panelbereite Haushalte haben der Speicherung ihrer Adresse zum Zweck von Wiederholungsbefragungen im Rahmen der Studie zugestimmt.

Diese wurden zwischen Feldende der Panelhaupterhebung (Mitte September 2025) und Feldstart des Websurveys (Mitte Oktober 2025) um kurzfristige Ausfälle (zum Beispiel Verweigerungen, Todesfälle) reduziert. Die bereinigte Personenzahl (n=11.366) bildet den Stichprobenumfang für die Haupterhebung des PASS Websurvey (Bruttoeinsatzstichprobe).

Insgesamt haben am Websurvey Welle 19 5.443 Personen teilgenommen. 5.264 Online-Fragebögen wurden als vollständig auswertbar deklariert.

Tabelle 1 weist die Responseraten differenziert nach der Teilstichprobe sowie die Brutto- und Nettofallzahlen aus. Die Responserate insgesamt liegt bei 0,46 und liegt damit auf dem Niveau der letzten Websurvey Welle. Die Responserate differenziert nach Teilstichprobe variiert von 0,14 (Teilstichprobe 42 Bestandsauffrischung (SGB II[2]) (syrische/irakische HH) Welle 19) bis 0,67 für die Teilstichprobe 21 (Zugangsstichprobe (SGB II) (syrische/irakische HH) Welle 13).

Tabelle 1: Responserate nach Teilstichprobe

Teilstichprobe	Bruttostichprobe	Nettostichprobe[3]	Responserate
	absolut	*absolut*	
Gesamt	**11.366**	**5.264**	**0,46**
1 SGB-II-Stichprobe	657	261	0,40
2 Microm-Stichprobe	1235	730	0,59
3 Zugangsstichprobe (SGB II) Welle 2	124	54	0,44
4 Zugangsstichprobe (SGB II) Welle 3	162	67	0,41
5 Zugangsstichprobe (SGB II) Welle 4	110	44	0,40
6 Bestandsauffrischung (EWO) Welle 5	444	246	0,55
7 Bestandsauffrischung (SGB II) Welle 5	200	89	0,45
8 Zugangsstichprobe (SGB II) Welle 5	103	51	0,50
9 Zugangsstichprobe (SGB II) Welle 6	176	94	0,53
10 Zugangsstichprobe (SGB II) Welle 7	167	80	0,48
11 Zugangsstichprobe (SGB II) Welle 8	189	82	0,43
12 Zugangsstichprobe (SGB II) Welle 9	185	75	0,41
13 Zugangsstichprobe (SGB II) Welle 10	122	73	0,60
14 Zugangsstichprobe (SGB II) (syrische/irakische HH) Welle 10	51	23	0,45
15 Bestandsauffrischung (EWO) Welle 11	492	305	0,62
16 Zugangsstichprobe (SGB II) Welle 11	120	44	0,37
17 Zugangsstichprobe (SGB II) (syrische/irakische HH) Welle 11	50	26	0,52
18 Zugangsstichprobe (SGB II) Welle 12	158	75	0,47
19 Zugangsstichprobe (SGB II) (syrische/irakische HH) Welle 12	63	10	0,16
20 Zugangsstichprobe (SGB II) Welle 13	215	100	0,47
21 Zugangsstichprobe (SGB II) (syrische/irakische HH) Welle 13	12	8	0,67
22 Zugangsstichprobe (SGB II) Welle 14	243	115	0,47
23 Zugangsstichprobe (SGB II) (syrische/irakische HH) Welle 14	12	4	0,33
24 Bestandsauffrischung (SGB II) (syrische/irakische HH) Welle 14	60	25	0,42

[2] In Vorgängerversionen des PASS Datenreports Websurvey war von BA-Stichprobe die Rede. Ab Welle 19 ändert sich diese Bezeichnung zu SGB-II-Stichprobe. Mit dieser sprachlichen Anpassung geht aber keine inhaltliche Änderung einher.

[3] Die Nettostichprobe enthält alle vollständig auswertbaren Online-Fragebögen

Teilstichprobe	Bruttostich-probe	Nettostich-probe[3]	Responserate
	absolut	*absolut*	
25 Zugangsstichprobe (SGB II) Welle 15	486	247	0,51
26 Zugangsstichprobe (SGB II) (syrische/irakische HH) Welle 15	23	10	0,43
27 Bestandsauffrischung (SGB II) (syrische/irakische HH) Welle 15	52	21	0,40
28 Zugangsstichprobe (SGB II) Welle 16	371	183	0,49
29 Zugangsstichprobe (SGB II) (syrische/irakische HH) Welle 16	27	9	0,33
30 Bestandsauffrischung (EWO) Welle 17	1093	660	0,60
31 Zugangsstichprobe (SGB II) Welle 17	218	95	0,44
32 Zugangsstichprobe (SGB II) (syrische/ irakische HH) Welle 17	32	8	0,25
33 Zugangsstichprobe (SGB II) (ukrainische HH) Welle 17	436	173	0,40
34 Zugangsstichprobe (SGB II) Welle 18	247	115	0,47
35 Zugangsstichprobe (SGB II) (syrische/ irakische HH) Welle18	49	10	0,20
36 Zugangsstichprobe (SGB II) (ukrainische HH) Welle 18	358	150	0,42
37 Bestandsauffrischung (EWO) Welle 17/ 2	339	191	0,56
38 Zugangsstichprobe (SGB II) Welle 19	648	218	0,34
39 Zugangsstichprobe (SGB II) (syrische/irakische HH) Welle 19	180	47	0,26
40 Zugangsstichprobe (SGB II) (ukrainische HH) Welle 19	294	87	0,30
41 Bestandsauffrischung (SGB II) Welle 19	879	257	0,29
42 Bestandsauffrischung (SGB II) (syrische/irakische HH) Welle 19	85	12	0,14
43 Bestandsauffrischung (SGB II) (ukrainische HH) Welle 19	199	90	0,45

Die Responserate nach Erhebungssprache zeigt, dass Zielpersonen, die in Welle 19 in einer der angebotenen Fremdsprachen teilgenommen haben, eine niedrigere Responserate aufweisen, wobei sie für Personen, die auf Arabisch teilgenommen haben, am geringsten ausfällt.

Tabelle 2: Responserate nach Erhebungssprache der Panelhaupterhebung Welle 19

Erhebungssprache[4]	Bruttostichprobe	Nettostichprobe	Responserate
	absolut	*absolut*	
Gesamt	11.366	5.264	0,46
Deutsch	9.609	4.631	0,48
Arabisch	566	165	0,29
Russisch	841	332	0,39
Ukrainisch	350	136	0,39

3 Datensatzstruktur

Der Scientific Use File des Websurveys besteht in Welle 19 aus zwei Datensätzen. Sofern nicht in diesem Kapitel anders erläutert, wurden den Variablennamen ein kleines „w“ vorangestellt, um die Variable als aus dem Websurvey stammend zu kennzeichnen. Generierte Variablen werden in Kapitel 4.2 erläutert.

[4] Zugewiesene Sprache aus der Haupterhebung Welle 19.

Im Datensatz *wPENDDAT* liegt für jede durchgeführte Personenbefragung eine Datenzeile vor. Dort enthalten sind die Fragen der in der Fragebogenvorlage dokumentierten Module *Soziodemografische Informationen* (ohne die Angaben zum Alter *wPHH0400* sowie zur verwandtschaftlichen Beziehung zu den anderen Haushaltsmitgliedern *wPHH0500/wPHH0501*), *Nachbefragung zur Hauptbefragung*, *Ressourcenverwaltung und Entscheidungsmacht*, *Rolleneinstellungen*, *Einstellungen zur Kindergrundsicherung* und *Einstellungen Lebensstandard*. Hinzu kommen die Identifikationsvariablen *pnr*, *hnr* und *uhnr*, sowie eine Wellenkennung und eine Stichprobenkennung. Bei den Identifikationsvariablen *pnr*, *hnr* und *uhnr* sowie bei der Wellenkennung *welle* und der Stichprobenkennung *sample* wurden die aus der Panelhaupterhebung bekannten Variablennamen verwendet. Die Variable *NoWelle17* kennzeichnet das Teilnahmeverhalten der Person in der zugehörigen Panelhaupterhebung der Welle 17 und wurde zur Steuerung der Module *Ressourcenverwaltung und Entscheidungsmacht* und *Rolleneinstellungen* genutzt. Die Variable *webW18* dokumentiert das Teilnahmeverhalten der Person am Websurvey der Welle 18 und wurde zur Steuerung der Module *Ressourcenverwaltung und Entscheidungsmacht*, *Rolleneinstellungen* und *Einstellungen Lebensstandard* genutzt. Die Variable *ModeW19* gibt wieder, ob die Person in der zugehörigen Panelhaupterhebung der Welle 19 im CAPI-Feld oder im CATI-Feld befragt wurde und wurde in einem Textfilter einer Frage im Modul *Nachbefragung zur Hauptbefragung* genutzt. Ein Teil der in der Befragung erhobenen Items ist auch Teil des *PENDDAT*-Datensatzes der Panelhaupterhebung. Die Variablen können sich hinsichtlich der Antwortkategorien grundsätzlich allerdings voneinander unterscheiden zwischen Websurvey und Panelhaupterhebung.

Im Datensatz *wRESSOURCEN* sind platzsparend im Long-Format jene Angaben abgelegt, bei denen die befragte Person nicht nur zu sich selbst, sondern auch zu den anderen Haushaltsmitgliedern Auskunft gibt. Der Datensatz enthält für alle durchgeführten Befragungen so viele Zeilen, wie es der abgefragten Haushaltsgröße entspricht. Für jedes Haushaltsmitglied wurde eine Datenzeile abgelegt. Falls es sich entweder um einen Ein-Personen-Haushalt handelt oder die Frage nach der Haushaltsgröße *wHA0100* inhaltlich nicht beantwortet wurde, wurden ausschließlich Angaben zur befragten Person selbst abgelegt. Dort enthalten sind neben den Identifikationsvariablen *pnr* und *hnr*, der Stichprobenkennung *sample* die Fragen des Moduls *Lebensstandard im Allgemeinen und im Haushalt* sowie die Angaben zum Alter *wpalter* und zur verwandtschaftlichen Beziehung zu den anderen Haushaltsmitgliedern *wPHH0500/wPHH0501*. Um die Befragungsdauer in sehr großen Haushalten zu begrenzen, wurden in Haushalten mit mehr als fünf Mitgliedern, zu denen valide Alters- und Verwandtschaftsangaben vorlagen, neben der befragten Person selbst noch vier weitere Personen ausgewählt, für die Fragen des Moduls *Lebensstandard im Allgemeinen und im Haushalt* gestellt wurden. Dadurch liegen für 244 der insgesamt 11.642 Haushaltsmitglieder keine Angaben zu den Deprivationsfragen vor.

Die erhobenen Items wurden anonymisiert (siehe Kapitel 5.4), erhobene offene Textangaben wurden codiert (siehe Kapitel 4.1).

In der Labelung des Websurvey-Scientific Use File werden Umlaute und Sonderzeichen aufgelöst, um eine einheitliche Anzeige der Labeltexte unabhängig vom bei den einzelnen Nutzerinnen und Nutzern eingestellten Zeichensatz zu gewährleisten.

4 Generierte Variablen

4.1 Vercodung offener Restkategorien

Zwei Items der Befragung (*wPHH0500* und *wPEW0100*) wurden als geschlossene Items mit offener Restkategorie erhoben. In beiden Fällen wurden – in Anlehnung an das Vorgehen zur Datenaufbereitung der Haupterhebung – zusätzliche Variablen generiert, die sich von der ursprünglichen Version der Variable lediglich dahingehend unterscheiden, dass die offenen Angaben soweit möglich den entsprechenden Kategorien zugecodet wurden. Es wurden keine neuen Kategorien auf Basis der offenen Angaben gebildet. Die Bezeichnung dieser Variablen unterscheidet sich von der ursprünglichen Variable nur dadurch, dass die „0“ in der letzten Ziffer durch eine „1“ ersetzt wurde. Die nachfolgende Tabelle gibt einen Überblick über die offenen Angaben, die vercodet wurden.

Tabelle 3: Vercodung offener Angaben

Fester Variablenname	zugecodet zu Variable	Bezeichnung (Datensatz)
wPHH0500	*wPHH0501*	Erwerbsstatus (*wPENDDAT*)
wPEW0100	*wPEW0101*	Verwandtschaftliche Beziehung (*wRESSOURCEN*)

4.2 Einfache Generierungen

Die einfachen Generierungen umfassen Variablen, die als Preloadvariablen auf Basis der Panelhaupterhebung oder auf Basis des Websurveys der Vorwelle erzeugt worden sind, die als Systeminformation während des Ausfüllens des Websurveyfragebogens gemessen oder die auf Basis der im Websurvey gemachten Angaben generiert worden sind. Die einfachen generierten Variablen sind in der folgenden Tabelle dargestellt. Die einzelnen Variablen werden dabei jeweils kurz beschrieben. Darüber hinaus werden die für die Generierung der Variable notwendigen Quellvariablen aufgeführt.

Tabelle 4: Einfache Generierungen (alphabetische Reihenfolge)

Variable	Label und Beschreibung (Datensatz)	Quellvariablen für die Generierung
ModeW19	*Modus (CATI/CAPI) Haupterhebung Welle 19 (wPENDDAT)* Gibt an, ob die befragte Person in Welle 19 der Panelhaupterhebung im CAPI-Modus oder im CATI-Modus befragt wurde	*pintmod (PENDDAT)*
NoWelle17	*Person wurde (nicht) in Welle 17 befragt (wPENDDAT)* Gibt an, ob die befragte Person in Welle 17 der Panelhaupterhebung ein Personeninterview gegeben hat	*pnettok17 (p_register)*
walterssplitt	*Gruppierung Kinderalter (wPENDDAT)* Die *PLS*50*-Fragen wurden split-half entweder für ein 5 oder für ein 13 Jahre altes Kind gestellt. Die Variable gibt an, welches Alter im Fragentext eingeblendet wurde	vom Erhebungsinstitut geliefert
webW18	*Teilnahme Websurvey W18 (wPENDDAT)* Gibt an, ob die befragte Person in Welle 18 ein Interview im Rahmen des Websurveys gegeben hat	*pnr (wPENDDAT im Websurvey der Welle 18)*

Variable	Label und Beschreibung (Datensatz)	Quellvariablen für die Generierung
wformauf	*Indikator: Formale Auffälligkeit (wPENDDAT)* Gibt an, ob ein Fall im Zuge der Strukturprüfungen auffällig war, wegen einer plausiblen Erklärung der Abweichung jedoch als gültig realisiert gewertet wurde. Das Vorgehen bei den Prüfungen wird in Kapitel 5.1 detailliert erläutert	vom Erhebungsinstitut geliefert
wgebhalbj	*Geburtshalbjahr, generiert (wPENDDAT)* Indikator, ob Geburtstag im ersten oder zweiten Halbjahr des Geburtsjahres liegt	Informationen zum Geburtsmonat
wgeraet	Eingabegerät (wPENDDAT) Aus dem User Agent String des von der befragten Person verwendeten Browsers wurde die Information generiert, ob für das Ausfüllen des Fragebogens ein Gerät mit Android bzw. iOS als Betriebssystem verwendet wurde oder ein Gerät mit einem anderen Betriebssystem. Die Betriebssysteme Android bzw. iOS wurden der Kategorie „Smartphone / Tablet“ zugeordnet, alle anderen Betriebssysteme der Kategorie „PC / Laptop“. 27 befragte Personen haben im Laufe der Befragung das genutzte Endgerät gewechselt mit der Konsequenz, dass sich auch das genutzte Betriebssystem geändert hat. In diesen Fällen wird das Betriebssystem herangezogen, dass bei der Mehrheit der beantworteten Fragen genutzt wurde	vom Erhebungsinstitut geliefert
wintdatende	*Datum Befragungsende (letzte Eingabe) (wPENDDAT)* Die Variable zeigt im Format YYMMTT an, an welchem Tag die befragte Person die Beantwortung des Fragebogens beendet hat	vom Erhebungsinstitut geliefert
wintdatstart	*Datum Befragungsstart (erste Eingabe) (wPENDDAT)* Die Variable zeigt im Format YYMMTT an, an welchem Tag die befragte Person die Beantwortung des Fragebogens begonnen hat	vom Erhebungsinstitut geliefert
wintdauer	*Gesamtdauer (in Sekunden) (wPENDDAT)* Gibt die Gesamtdauer an, die die befragte Person auf das Ausfüllen des Fragebogens verwendet hat. Gemessen wird ausschließlich das aktive Ausfüllen des Fragebogens. Wenn das Ausfüllen des Fragebogens unterbrochen wurde, ist die Dauer der Unterbrechung nicht Teil der gemessenen Gesamtdauer	vom Erhebungsinstitut geliefert
wpalter	*Alter (wRESSOURCEN)* Für die befragte Person selbst wurden im Interview Geburtsmonat und -jahr in Frage *wPD0100* erhoben. Aus diesen Angaben und dem Befragungsdatum wurde das Alter in Jahren generiert. Für die weiteren Haushaltsmitglieder wurde das Alter in Jahren abgefragt	*wPD0100 (wPENDDAT)*
wpsprache	*Sprachauswahl (wPENDDAT)* Gibt die im Interview genutzte Befragungssprache an. 75 befragte Personen haben im Interviewverlauf die Sprache gewechselt. In diesen Fällen wird die Sprache herangezogen, die bei der Mehrheit der beantworteten Fragen genutzt wurde	vom Erhebungsinstitut geliefert
wzplfd	*Durchlaufzähler Haushaltsmitglied (wRESSOURCEN)* Durchlaufzähler für die Haushaltsmitglieder, wobei die 1 immer an die befragte Person vergeben werden soll und ab Person 6 die Haushaltsmitglieder genannt werden, die nicht ausgewählt wurden, um die Fragen zur Deprivation zu beantworten	vom Erhebungsinstitut geliefert

5 Datenaufbereitung

Die Datenaufbereitung wird von infas durchgeführt. Wichtige Entscheidungen bspw. zur Bereinigung von Strukturproblemen bei den befragten Personen wurden in Abstimmung mit dem IAB getroffen. Darüber hinaus stand das IAB auch für die Beantwortung von Fragen während der Zeit der Datenaufbereitung zur Verfügung.

Die in den Websurveyinterviews erhobenen Informationen liegen bei infas zunächst als ASCII-Daten vor. In einer ersten Stufe werden die Rohdaten als Querschnittsdatensatz in eine Statistiksoftware eingelesen[5]. Dabei werden die unvollständig realisierten Interviews von den vollständigen Interviews abgetrennt. Als vollständig realisiert gelten nur die Fälle, bei denen der Fragebogen vollständig bis zur letzten Frage des Fragebogens durchlaufen wurde. In einer zweiten Stufe werden ausschließlich die Daten der vollständigen Interviews einer ausführlichen formalen und inhaltlichen Prüfung unterzogen und zum Scientific Use File aufbereitet.

Die sich anschließende Datenprüfung bei infas unterteilt sich in zwei Schritte, die in den folgenden Abschnitten genauer dargestellt werden. Zunächst wurden Strukturprüfungen und ggf. Bereinigungen vorgenommen. Bei schweren Strukturproblemen wurden die entsprechenden Interviews entfernt (siehe dazu Kapitel 5.1). Im Anschluss erfolgten eine detaillierte Überprüfung und ggf. Bereinigung der Filterführung. Dabei wurden spezifische Codes für Missing Values gesetzt (siehe dazu Kapitel 5.2).

Die folgende Tabelle gibt eine Übersicht aller im Rahmen der Datenaufbereitung durchgeführten Arbeitsschritte und ihre Abfolgen:

Tabelle 5: Übersicht über die Schritte der Datenaufbereitung

Nr.	Arbeitsschritt
1	Einlesen der erhobenen Rohdaten in Arbeitsdatensätze
2	Neugenerierung von Steuerungsvariablen auf Basis der filtergeprüften Datensätze der Panelhaupterhebung der Welle 19 und des Websurveys der Welle 18
3	Strukturprüfungen (siehe Kapitel 5.1)
4	Entfernen problematischer Interviews (siehe Kapitel 5.1)
5	Korrektur der Struktur (siehe Kapitel 5.1)
6	Filterbereinigung (siehe Kapitel 5.2)
7	Zucodung offener Angaben (siehe Kapitel 4.1)
8	Einfache Generierungen (siehe Kapitel 4.2)
9	Generierung der Datenstruktur für Scientific Use File
10	Anonymisierung (siehe Kapitel 5.4)

[5] Im Rahmen der Datenaufbereitung kamen die Softwarepakete Stata Version 19 sowie SPSS Version 28 zum Einsatz.

5.1 Strukturprüfungen und entfernte Interviews

Der Filterbereinigung vorgelagert fand eine Strukturprüfung statt. Diese bestand in einem Abgleich der im Websurvey gemachten Angaben zu Geschlecht sowie Geburtsmonat und -jahr mit den aktuellen aus der Panelhaupterhebung bekannten Angaben, um Fälle erkennen und ggf. korrigieren zu können, bei denen eine andere als die hierzu eingeladene Person das Websurveyinterview gegeben hat. Bei einer schriftlichen Einladung zur Interviewteilnahme mit Verteilung von individualisierten Zugangscodes zum Befragungsinstrument besteht die Möglichkeit, dass die Zugangscodes innerhalb eines Haushalts bewusst oder versehentlich durch andere Haushaltsmitglieder verwendet werden. Für die wellenübergreifende Betrachtung der erhobenen Daten ist die konsistente Zuordnung und eindeutige Identifizierbarkeit der Personen über die Wellen hinweg zentral. Eine feste Personennummer darf nicht in verschiedenen Wellen unterschiedlichen Personen zugeordnet sein. Wurde auf der Personenebene ein Interview mit der falschen Person geführt, wurde dieses Interview entweder entfernt oder es wurde die feste Personennummer korrigiert. Es wird im Folgenden erläutert, in welchen Konstellationen eine der beiden möglichen Entscheidungen getroffen wurde. Fälle, bei denen eine Korrektur erforderlich war und umgesetzt werden konnte, sind im Datensatz in der Variablen *wformauf* gekennzeichnet.

Beim Abgleich von Geschlecht sowie Geburtsmonat und -jahr wurde ohne pauschalen Ermessensspielraum geprüft. Eine Abweichung wurde also z.B. auch dann im Detail geprüft, wenn lediglich der Geburtsmonat um einen Monat abwich und das Geschlecht sowie das Geburtsjahr identisch waren mit den aus der Panelhaupterhebung bekannten Angaben. Es wurden im Haushaltsnetzwerk *uhnr* der von der Abweichung betroffenen Person die Angaben zum Geburtsdatum bzw. zum Alter in Jahren sowie zum Geschlecht aller Personen betrachtet, die in der Panelhaupterhebung jemals erfasst worden sind. Zudem wurden die aus der Panelhaupterhebung bekannten Vornamen zur Prüfung herangezogen. Alle Interviews im Websurvey, bei denen Abweichungen bestanden und die als gültig gewertet wurden, sind in der Variablen *wformauf* mit Code 1 gekennzeichnet. Korrekturen an den im Websurvey erhobenen Angaben zu Geschlecht sowie Geburtsmonat und -jahr wurden in keinem Fall vorgenommen.

Das weitere Vorgehen hing davon ab, ob die im Websurvey gemachten Angaben zu Geschlecht sowie Geburtsmonat und -jahr einer anderen Person im Haushaltsnetzwerk *uhnr* zugeordnet werden konnten. Wenn dies nicht der Fall war und die Abweichung geringfügig war, wurde das Interview ohne weitere Eingriffe als gültig gewertet. Eine Geringfügigkeit wurde beim Geburtsdatum dann angenommen, wenn nur eines der beiden geprüften Merkmale Monat oder Jahr abwich und die Abweichung beim Geburtsdatum als Tippfehler interpretiert werden konnte oder wenn mit den Angaben im Websurvey eine Angabe aus der Panelhaupterhebung repliziert wurde, die vor der dort aktuellen Angabe in einer früheren Welle gemacht worden war. Beim Geschlecht wurde eine Abweichung als unproblematisch eingestuft, wenn der Vorname für beide Geschlechter verwendet wird oder wenn anhand des Vornamens entschieden werden konnte, dass das in der Panelhaupterhebung erfasste Geschlecht eindeutig falsch ist. Zudem wurde eine Abweichung beim Geschlecht auch dann als unproblematisch eingestuft, wenn die Angaben zum Geburtsdatum nicht abwichen und sich aus der Betrachtung aller Personen im Haushaltsnetzwerk keine Hinweise darauf ergaben, dass das Interview durch eine andere Person gegeben worden sein könnte. Eine weitere unproblematische Abweichung in der Geschlechtsangabe besteht in der Selbsteinordnung als

divers im Websurvey, während in der Panelhaupterhebung das männliche oder das weibliche Geschlecht erfasst ist. Wenn die Abweichung beim Geburtsdatum hingegen erheblich war, weil sie mehrere Altersjahre betrug und sich nicht auf einen Tippfehler zurückführen ließ, wurde das Interview als ungültig realisiert gewertet und entfernt. Gleiches gilt bei einem unklaren Prüfergebnis, wenn z.B. das Geburtsdatum zu Person A und das Geschlecht zu Person B passte.

Wenn die im Websurvey gemachten Angaben zu Geschlecht sowie Geburtsmonat und -jahr einer anderen Person im Haushaltsnetzwerk *uhnr* zugeordnet werden konnten, bestand grundsätzlich die Möglichkeit, die *pnr* dieses Interviews zu korrigieren. Sofern die identifizierte Person in einem anderen Haushalt *hnr* innerhalb des Haushaltsnetzwerks *uhnr* lebte, wurde zudem der Korrekturbedarf an der Haushaltsnummer *hnr* berücksichtigt. Diese Korrektur ist allerdings von weiteren Voraussetzungen abhängig. Zur Teilnahme am Websurvey eingeladen wurden alle Personen, die in der Panelhaupterhebung der Welle 19 ein vollständiges und gültiges Interview auf der Personenebene gegeben haben und die zum Feldbeginn des Websurveys ihre Teilnahmebereitschaft nicht zurückgezogen haben. Wenn die Person, die ausweislich der Angaben zu Geschlecht sowie Geburtsmonat und -jahr tatsächlich das Interview gegeben hat, in Welle 19 kein in den *PENDDAT*-Datensatz aufgenommenes Interview gegeben hat, wurde das Websurveyinterview entfernt. Auch wenn für die Person, die tatsächlich das Interview gegeben hat, ein gültiges Interview aus der Panelhaupterhebung der Welle 19 vorliegt, wurde ihr Websurveyinterview unter bestimmten Umständen entfernt. Vereinzelt hat eine Person sowohl unter Nutzung des eigenen Zugangscodes als auch zusätzlich unter Nutzung des Zugangscodes eines weiteren Haushaltsmitglieds ein Interview gegeben. Dann wurde das Interview, das unter Nutzung des Zugangscodes eines weiteren Haushaltsmitglieds gegeben wurde, entfernt. Das Interview, das unter Nutzung des eigenen Zugangscodes gegeben wurde, wurde in der Variablen *wformauf* mit Code 1 gekennzeichnet. Die Variable *wformauf* kennzeichnet somit nicht ausschließlich Abweichungen bei den im Websurvey gemachten Angaben zu Geschlecht sowie Geburtsmonat und -jahr im Vergleich zur Panelhaupterhebung, sondern auch Auffälligkeiten im Teilnahmeverhalten. Das Interview wurde auch dann entfernt, wenn die auf der Basis der Panelhaupterhebung der Welle 19 und des Websurveys der Welle 18 generierten Steuerungsvariablen sich zwischen den beiden Personen, die an der Vertauschung des Zugangscodes beteiligt waren, unterschieden haben. Denn in diesem Fall hat die Person, die das Interview gegeben hat, teilweise Fragen erhalten, die ihr bei Nutzung ihres eigenen Zugangscodes nicht vorgelegt worden wären.

Zusammenfassend gilt, dass die *pnr* und ggf. auch die *hnr* korrigiert wurden, wenn die Person, die den Zugangscode eines anderen Haushaltsmitglieds genutzt hat, ebenfalls Teil der Einsatzstichprobe des Websurveys war, den Websurveyfragebogen nur ein einziges Mal ausgefüllt hat und ihre Steuerungsvariablen identisch waren im Vergleich mit der anderen an der Vertauschung des Zugangscodes beteiligten Person.

Von den 5.264 realisierten Interviews waren 273 formal auffällig. In 195 Fällen wurde die Auffälligkeit als unproblematisch bewertet. In den restlichen 78 Fällen lag ein ungültig realisierter Fall vor, so dass die Datensätze des Scientific Use Files 5.186 Fälle umfassen.

Es wurde darauf verzichtet, den weiteren Haushaltsmitgliedern, zu denen die befragte Person in den im Datensatz *wRESSOURCEN* abgelegten Fragen Auskünfte gegeben hat, eine *pnr* zuzuordnen. Für eine solche Zuordnung müssten Vorname, Alter und Geschlecht genutzt werden. Die Zuordnung wird häufig nicht ohne weiteres möglich sein. Ein Match, bei dem Namen als Matchschlüssel

verwendet werden, muss mit dem Problem unterschiedlicher Schreibweisen umgehen. Zudem ist die Altersangabe kein fixes Merkmal, sondern kann sich seit der Panelhaupterhebung der Welle 19 verändert haben. Darüber hinaus müsste geprüft werden, ob man den Abgleich auch mit Personen vornimmt, die vor Welle 19 irgendwann einmal Teil des Haushalts waren und ihn seither verlassen haben. Weitere Sonderfälle sind denkbar. So könnten Personen ermittelt werden, die zum Zeitpunkt der Panelhaupterhebung noch nicht Teil des Haushalts waren. Oder es könnten mehrere Personen gleiche Merkmale aufweisen, insbesondere bei Haushalten, in denen Zwillinge leben. Es muss daher angenommen werden, dass es eine relevante Anzahl von Fällen gibt, bei denen keine eindeutige Zuweisung einer *pnr* erfolgen kann.

5.2 Filterbereinigung und vergebene Missing Codes

In der Filterbereinigung wurde die im Instrument vorgesehene Filterführung mithilfe eines Statistikprogramms nachvollzogen. Wurden bestimmte Fragen gestellt, obwohl der Wert der relevanten Filtervariable etwas anderes verlangt hätte (wurden also z.B. detaillierte Angaben zu Entscheidungen in Partnerschaften erhoben, obwohl die/der Befragte angegeben hatte, nicht mit einem Partner bzw. einer Partnerin zusammenzuleben), wurden die Variablen auf den Missing Code „-3“ (trifft nicht zu) gesetzt, den sie auch bei einer korrekten Ausführung der Filterführung erhalten hätten[6]. Des Weiteren wurde geprüft, ob in Einzelfällen Items nicht erhoben wurden, obwohl dies laut der entsprechenden Filtervariable notwendig gewesen wäre. In diesen Fällen wurde der spezifische Missing Code „-4“ (Frage irrtümlich nicht gestellt) vergeben. Die Vergabe des Codes „-4“ war nicht notwendig.

Daneben wurden im Rahmen der Filterprüfung auch die vom Feldinstitut vergebenen Missing Codes sowie die System Missings durch für alle Variablen einheitliche Werte ersetzt. Die vergebenen Werte sind in nachfolgender Tabelle im Überblick dargestellt. „-1“ und „-2“ sind die einheitliche Umkodierung der in der Befragung erhobenen „WN“- und „KA“-Werte. „-3“ ist der generelle „trifft nicht zu“-Code für aufgrund der Filterführung nicht gestellte Fragen. Bei „-5“ und „-6“ handelt es sich um fragespezifische Codes. Hierbei kann es sich sowohl um spezifische Missing Codes als auch um Sonderkategorien für gültige Werte handeln. Diese Codes werden nur im Bedarfsfall vergeben, wobei in dieser Erhebung kein entsprechender Bedarf bestand. Mit dem Code „-11“ wird gekennzeichnet, dass die befragte Person die Frage verlassen hat, ohne eine Antwort anzugeben.

Tabelle 6: Übersicht über die vergebenen Missing Codes

Code	Erläuterung
-1	„weiß nicht“
-2	„keine Angabe“
-3	„trifft nicht zu (Filter)“ (Frage wurde aufgrund der Filterführung nicht gestellt)
-4	„Frage irrtümlich nicht gestellt“ (Frage hätte jedoch gestellt werden sollen)
-5	Fragespezifischer Code Nr. 1, wurde nur bei Bedarf vergeben
-6	Fragespezifischer Code Nr. 2, wurde nur bei Bedarf vergeben
-11	„Weiterklicken ohne Angabe“

[6] Die Bereinigung erfolgte dabei, wie in solchen Fällen üblich, immer von den zuerst erhobenen zu den später erfragten Items.

5.3 Plausibilitätsprüfungen

Anders als in der Panelhaupterhebung wird im Websurvey der Welle 19 auf Plausibilitätsprüfungen verzichtet. Die Liste der in der Panelhaupterhebung angewendeten Prüfungen wurde mit dem Fragebogen des Websurveys abgeglichen mit dem Ergebnis, dass im Websurvey keine prüffähigen Sachverhalte erhoben werden. Eine Prüfung im Längsschnitt, also zwischen den im Websurvey und den in der Panelhaupterhebung gemachten Angaben, wurde nicht durchgeführt.

5.4 Anonymisierung

Vom IAB als besondere Dienststelle der Bundesagentur für Arbeit (BA) erhobene Daten stellen immer Sozialdaten dar, weshalb die Ansprüche an den Datenschutz entsprechend hoch sind. Daher wurde es notwendig, zwei Variablen nur in vergröberter Form in den Scientific Use File aufzunehmen. Die nachfolgende Tabelle gibt einen Überblick über die betroffenen Variablen und die vorgenommenen Anonymisierungsschritte.

Tabelle 7: Übersicht über die anonymisierten Variablen

Varname	Variablenlabel	Vorgehen
wPD0100	Geburtsjahr	Der zusätzlich erhobene Geburtsmonat wurde nicht aufgenommen.
wgebhalbj	Geburtshalbjahr, generiert	Der Monat des Geburtsdatums wurde zu einem Indikator für das 1. bzw. 2. Halbjahr verkürzt.

6 Gewichtung Websurvey 2025

Da nur ein Teil der Personen, die an Welle 19 von PASS teilgenommen haben, auch am Websurvey teilgenommen hat, ist ein zusätzlicher Gewichtungsfaktor (*propweb19*) notwendig, um den Websurvey zu gewichten. Die PASS-Personen-Hochrechnungsfaktoren aus Welle 19 (*wqp, wqmip, wqbap*) können mit diesem zusätzlichen Gewichtungsfaktor multipliziert werden. Daraus resultieren die Hochrechnungsfaktoren für den Websurvey (wqpweb, wqmipweb, wqbapweb). Die Hochrechnungsfaktoren basieren auf den invertierten vorhergesagten Wahrscheinlichkeiten eines Logit-Regressionsmodells zur Schätzung der Teilnahmeneigung am Websurvey.

Unter der Berücksichtigung aller Panelpersonen, die potenziell die Möglichkeit hatten, an der Web-Befragung teilzunehmen (11.366 Fälle), wurde ein Logit-Modell für die Teilnahme geschätzt. Die unabhängige Variable *teilnahme_V1*setzt sich dabei wie folgt zusammen:

- Ausprägung 1 – teilgenommen: Personen, die den Websurvey selbst und vollständig ausgefüllt haben
- Ausprägung 0 – nicht teilgenommen: Personen, die gar nicht teilgenommen haben sowie Personen, die die Befragung abgebrochen haben (*teilnahme*=2) oder bei denen aufgrund von Plausibilitätsprüfungen vermutet wurde, dass eine andere Person den Websurvey ausgefüllt hat (*wraus*=2)

Die Prädiktorvariablen wurden bereits mit dem ersten PASS-Websurvey in Welle 16 ermittelt und für Welle 18 und Welle 19 erneut überprüft. In Welle 16 wurden die Prädiktorvariablen aus einem großen Pool an potenziell relevanten Merkmalen aus dem Personen-, Haushalts- und Methodendatensatz der Welle 16, auf Basis theoretischer Überlegungen sowie vorhandener Literatur ausgewählt.

Dementsprechend wurden die Prädiktorvariablen inklusive der Referenzkategorien bei kategorialen Variablen aus Welle 16 für Welle 18 und Welle 19 übernommen, falls sie vorhanden waren, um zwischen den Wellen ein konsistentes und vergleichbares Vorgehen zu haben.

Wie in Welle 16 wurden einige Variablen, die theoretisch einflussreich erschienen, aufgrund von Bedenken bezüglich Multikollinearität aus dem Modell ausgenommen, wie beispielsweise die Staatsbürgerschaft (*PMI0400*). Zudem wurden in Welle 16 wie auch in Welle 18 wie auch in Welle 19 Ausprägungskategorien von Variablen, wie das Geschlecht (*zpsex*), den Erwerbsstatus (*statakt*), den höchsten Schulabschluss (*schul2*) und die Ortsgrößenklasse (*bik*), aufgrund von Multikollinearität oder aufgrund geringer Fallzahlen, zusammengefasst.

Das Alter der befragten Person (*zpalthh*) wurde wie in Welle 16 in eine kategoriale Variable mit drei Ausprägungen (< 30, 30-74, >74) transformiert.

Berücksichtigt wurden zum einen demographische Variablen, die einen hohen Einfluss auf die Teilnahme vermuten lassen. Hierbei handelt es sich um Altersgruppen (*alterKat*), Geschlecht (*zpsex*), Familienstand (*famstand*), Migrationshintergrund (*migration*), Erwerbstatus (*statakt*), höchster Schulabschluss (*schul2*), Zahl der eigenen Kinder (*kindzges*), ALG-II-Bezug (*alg2abez*), Deprivationsindex (*depindg3*), Ortsgrößenklasse (*bik*), Art des Mietverhältnisses (*HW0300*) sowie Wohnort in alten oder neuen Bundesländern (*blneualt*). Zudem wurden die Anzahl der Kontaktversuche (*ANZKON*) der Welle 19 sowie die Anzahl der bisherigen Wellen mit Teilnahme des Haushalts (*HHTLN*)

und eine Variable, die indiziert, ob eine Person ursprünglich aus einer allgemeinen Bevölkerungsstichprobe stammt (*sampleType*), in das Modell miteinbezogen. Eine weitere Variable (*INR*) gibt die Zahl der im Personeninterview der Welle 19 mit „weiß nicht“ oder „keine Angabe“ beantworteten Fragen wieder. Für die Generierung der Variable *INR* wurden nur Merkmale berücksichtigt, die alle Befragten erhalten haben und die nicht zum Pool der potenziellen Prädiktorvariablen gehörten. Es wurde weiterhin eine Variable aufgenommen, die die Zeit zwischen Durchführung des Personeninterviews und Beginn des Websurveys (18.10.2025) in Tagen misst (*vergangeneZeit*).

Zudem wurden Prädiktorvariablen aufgenommen, die im Zusammenhang mit dem Zugang zum Internet stehen: zunächst, ob die Person Zugang zu einem PC mit Internetanschluss hat (*HLS1100a*) und ob sie grundsätzlich einen Internetzugang hat (*PTS2000*) sowie die tägliche Nutzung von Laptop, PC, Smartphone und/oder Tablet (*taeglNutzGer*).

Mittels Stepwise Selection (Funktion step (direction=“both“) in R) wurde überprüft, welche Variablen aus dem Modell entfernt werden könnten, wenn sie im Rahmen des Auswahlalgorithmus die Güte des Modells nicht verbessern konnten. Im Gegensatz zu den vorherigen Wellen wurden in Welle 19 keine Variablen entfernt.

Fehlende Werte wurden in den finalen Modellen multipel imputiert, um für jede teilnehmende Person eine Teilnahmewahrscheinlichkeit schätzen zu können.

Tabelle 8 enthält eine Übersicht über alle im Modell verwendeten Variablen (Variablenname und Erläuterung), Tabelle 9 enthält die Modellkoeffizienten, Fallzahlen und Anpassungsgüte des verwendeten Modells.

Tabelle 8: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle

Variablenkürzel und Referenzkategorie	Erläuterung
alterKatbis29	Person jünger als 30 Jahre
alterKatab75	Person 75 Jahre und älter
Referenzkategorie	Person 30 – 74 Jahre alt
zpsexW	Person Weiblich
Referenzkategorie	Person Männlich oder Divers
Kindzges	Anzahl eigene Kinder gesamt (inner- und außerhalb HH)
statakt_s_Anderes	Erwerbsstatus der Person: Anderes
statakt_s_Arbeitslos gemeldet	Erwerbsstatus der Person: Arbeitslos gemeldet
statakt_s_Ausbildung/Lehre/Studium	Erwerbsstatus der Person: Ausbildung/Lehre/Studium
statakt_s_Hausfrau/Hausmann	Erwerbsstatus der Person: Hausfrau/Hausmann
statakt_s_Rentner/Pensionaer/Vorruhestand	Erwerbsstatus der Person: Rentner/Pensionär/Vorruhestand
statakt_s_Schueler	Erwerbsstatus der Person: Schüler*in
Referenzkategorie	Erwerbsstatus der Person: Erwerbstätig
schul2_sNoch_Schüler_In	Schulabschluss Person: noch Schüler*In
schul2_sKein_Abschluss	Schulabschluss Person: ohne Abschluss
schul2_sVolks_Hauptschule	Schulabschluss Person: Volks-/Hauptschule
schul2_sMittlere_Reife	Schulabschluss Person: Realschule/mittlerer Abschluss
schul2_sFachhochschule	Schulabschluss Person: Fachhochschulreife
schul2_sSonstiges	Schulabschluss Person: Sonder-/Förderschule/Ausland/Anderes
Referenzkategorie	Schulabschluss Person: Hochschulreife
famstandVerheiratet..z.l.	Familienstand Person: Verheiratet/eing. Lebensp., zus. lebd.

Variablenkürzel und Referenzkategorie	Erläuterung
famstandVerheiratet..g.l.	Familienstand Person: Verheiratet/eing. Lebensp., getr. lebd.
famstandGeschieden	Familienstand Person: Geschieden
famstandVerwitwet	Familienstand Person: Verwitwet
Referenzkategorie	Familienstand Person: Ledig
migrationSelbst	Migrationshintergrund Person: Selbst zugezogen
migrationEltern	Migrationshintergrund Person: Mind. 1 Elternteil zugezogen
migrationGroßeltern	Migrationshintergrund Person: Mind. 1 Großelt. zugez., Elt. in D geb.
Referenzkategorie	Migrationshintergrund Person: Kein Migrationshintergrund
depindg3	Deprivationsindex, gewichtet
HLS1100aKein PC mit Internet	Person hat keinen Zugang zu einem PC mit Internetanschluss
Referenzkategorie	Person hat Zugang zu einem PC mit Internetanschluss
alg2abez_HH bezieht aktuell kein ALG2	Kein Leistungsbezug nach dem SGB II
Referenzkategorie	Leistungsbezug nach dem SGB II
bikCatbis2000	BIK-Größenklasse der Gemeinde: unter 2.000 Einwohner
bikCat2.000_4.999	BIK-Größenklasse der Gemeinde: 2.000 bis u. 5.000 Einwohner
bikCat5.000_19.999	BIK-Größenklasse der Gemeinde: 5.000 bis u. 20.000 Einwohner
bikCat20.000_49.999	BIK-Größenklasse der Gemeinde: 20.000 bis u. 50.000 Einwohner
bikCat50.000_99.999	BIK-Größenklasse der Gemeinde: 50.000 bis u. 100.000 Einwohner
bikCat100.000_499.999	BIK-Größenklasse der Gemeinde: 100.000 bis u. 500.000 Einwohner
Referenzkategorie	BIK-Größenklasse der Gemeinde: 500.000 und mehr Einwohner
blneualtNeu	Neue Bundesländer inkl. Berlin
Referenzkategorie	Alte Bundesländer
HW0300Untermiete	Art des Mietverhältnisses: Untermieter
HW0300Eigentum	Art des Mietverhältnisses: Eigentümer
HW0300Mietfrei	Art des Mietverhältnisses: Wohne mietfrei
HW0300Anderes	Art des Mietverhältnisses: Wohngemeinschaft/Wohnheim
Referenzkategorie	Art des Mietverhältnisses: Hauptmieter
PTS2000 Nein	Zugang zum Internet: Nein
Referenzkategorie	Zugang zum Internet: Ja
ANZKON	Anzahl der Kontaktversuche in Welle 19
pspracheRussisch	Interviewsprache des Personeninterviews: Russisch
pspracheArabisch	Interviewsprache des Personeninterviews: Arabisch
pspracheUkrainisch	Interviewsprache des Personeninterviews: Ukrainisch
Referenzkategorie	Interviewsprache des Personeninterviews: Deutsch
vergangeneZeit	Tage zwischen Datum der Personenbefragung W19 und Start Websurvey (18.10.2025)
sampleTypeBestandsauffrischungen (SGB-II-Stichprobe)	Bestandsauffrischung SGB-II-Stichprobe
sampleTypeBestandsauffrischungen (SGB-II-Stichprobe) (syr./irq. HH)	Bestandsauffrischung SGB-II-Stichprobe (syrisch/irakische Haushalte)
sampleTypeBestandsauffrischungen (SGB-II-Stichprobe) (ukr. HH)	Bestandsauffrischung SGB-II-Stichprobe (ukrainische Haushalte)
sampleTypeZugangsstichproben (SGB-II-Stichprobe)	Zugangsstichproben (SGB-II-Stichprobe)
sampleTypeZugangsstichproben (SGB-II-Stichprobe) (syr./irq. HH)	Zugangsstichproben (SGB-II-Stichprobe) (syrisch/irakische Haushalte)
sampleTypeZugangsstichproben (SGB-II-Stichprobe) (ukr.)	Zugangsstichproben (SGB-II-Stichprobe) (ukrainische Haushalte)
Referenzkategorie	Microm-Stichprobe (inkl. EWO-Auffrischung)
HHTLN	Anzahl der Panelteilnahmen des Haushalts

Variablenkürzel und Referenzkategorie	Erläuterung
INR	Item Nonresponse (Weiß Nicht und K.A.) in Personenbefragung der W19

Tabelle 9: Logit-Modell zur Teilnahme am Websurvey

	Teilnahme	
	Coef.	p
(Intercept)	0,0984341	0,488
alterKatbis29	-0,4079771	0,000
alterKatab75	-0,2780004	0,008
zpsexW	0,1793527	0,000
kindzges	-0,0888090	0,000
statakt_s_Anderes	0,0304507	0,788
statakt_s_Arbeitslos gemeldet	0,0212627	0,774
statakt_s_Ausbildung/Lehre/Studium	0,2237900	0,024
statakt_s_Hausfrau/Hausmann	-0,0885323	0,419
statakt_s_Rentner/Pensionaer/Vorruhestand	0,1053978	0,158
statakt_s_Schueler	-0,0413139	0,853
schul2_sNoch_Schüler_In	-0,0165662	0,948
schul2_sKein_Abschluss	-0,7817597	0,000
schul2_sVolks_Hauptschule	-0,6097436	0,000
schul2_sMittlere_Reife	-0,3038692	0,000
schul2_sFachhochschule	-0,2011967	0,011
schul2_sSonstiges	-0,4556624	0,019
famstandVerheiratet..z.l.	0,3121435	0,000
famstandVerheiratet..g.l.	0,0082702	0,942
famstandGeschieden	0,0644728	0,367
famstandVerwitwet	-0,2011410	0,083
migrationSelbst	-0,6957268	0,000
migrationEltern	-0,0337209	0,667
migrationGroßeltern	0,0482501	0,665
depindg3	-0,0159198	0,579
bikCatbis2000	-0,2845749	0,295
bikCat2.000_4.999	-0,1556031	0,366
bikCat5.000_19.999	-0,1524104	0,050
bikCat20.000_49.999	-0,0707531	0,320
bikCat50.000_99.999	-0,0205866	0,782
bikCat100.000_499.999	-0,1261257	0,011
blneualtNeu	-0,0862397	0,079
HW0300Untermiete	-0,2637309	0,059
HW0300Eigentum	0,0254835	0,656
HW0300Mietfrei	-0,2383536	0,020

	Teilnahme	
	Coef.	p
HW0300Anderes	-0,0409198	0,614
HLS1100aKein PC mit Internet	-0,6984297	0,000
ANZKON	-0,0032232	0,228
pspracheRussisch	0,3602486	0,019
pspracheArabisch	-0,1771019	0,393
pspracheUkrainisch	0,3725310	0,055
vergangeneZeit	0,0031615	0,000
sampleTypeBestandsauffrischungen (SGB II-Stichprobe)	-0,5123261	0,000
sampleTypeBestandsauffr. (SGB II-Stichprobe) (syr./irq. HH)	0,2374951	0,344
sampleTypeBestandsauffr. (SGB II-Stichprobe) (ukr. HH)	0,3827147	0,071
sampleTypeZugangsstichproben (SGB II-Stichprobe)	-0,3348530	0,000
sampleTypeZugangsstichproben (SGB II-Stichprobe) (syr./irq. HH)	0,3571552	0,079
sampleTypeZugangsstichproben (SGB II-Stichprobe) (ukr. HH)	-0,1348255	0,411
taeglNutzGer_neu_impNein	-0,0190083	0,819
alg2abez_HH bezieht aktuell kein ALG2	0,0569146	0,425
PTS2000_Nein	-0,5582754	0,000
HHTLN	0,0064124	0,100
INR	-0,1125602	0,000
N	11.366	
Log likelihood	-7.130,6	
Pseudo R^2	0,0898	

Um den PASS-Websurvey auf die Wohnbevölkerung in Deutschland ab 15 Jahren hochzurechnen, sind zwei weitere Schritte erforderlich:

Das PASS-Personengewicht (*wqp*) der dazugehörigen PASS-Welle (Panel Haupterhebung Welle 19) wird mit der reziproken vorhergesagten Teilnahmewahrscheinlichkeit (*wteilwahr*) am Websurvey multipliziert. Da dieser Schritt allein einerseits nicht gewährleistet, dass die Größe der Zielpopulation (71,062 Mio.) durch die Gewichte replizierbar ist und andererseits starke Ausreißer in der Verteilung der Gewichte entstehen können, wird im Anschluss daran eine Kalibrierung des neuen Gewichts durchgeführt. Neben der Populationsgröße wird dabei lediglich vorgegeben, dass die Gewichte in den Schranken des Minimums und Maximums des Personengewichts von Welle 19 (*wqp*) dividiert durch die Teilnahmequote des Websurveys bleiben. Es wird damit also insbesondere nicht forciert, dass mit der Gewichtungsvariable *wpweight* Verteilungen aus dem Mikrozensus exakt getroffen werden.

Tabelle 10 gibt Aufschluss über Kennwerte der Verteilung der vorhergesagten Teilnahmewahrscheinlichkeiten (Variable *wteilwahr*) und Tabelle 11 über Kennwerte der Verteilung der berechneten Gewichte (Variable *wpweight*).

Tabelle 10: Kenngrößen der Verteilung der vorhergesagten Teilnahmewahrscheinlichkeiten für die Nettostichprobe

1%-Perzentil	0,1343049
5%-Perzentil	0,2392107
10%-Perzentil	0,3052437
25%-Perzentil	0,4267420
50%-Perzentil	0,5389000
75%-Perzentil	0,6307137
90%-Perzentil	0,6962584
95%-Perzentil	0,7248767
99%-Perzentil	0,7665639
Mittelwert	0,51964
Standardabweichung	0,1474466
Minimum	0,03302
Maximum	0,83067
Fallzahl	**5.186**

Tabelle 11: Kenngrößen der Verteilung der Gewichte (Gesamtstichprobe, Personen)

1%-Perzentil	1,326830
5%-Perzentil	1,420133
10%-Perzentil	1,494044
25%-Perzentil	1,695688
50%-Perzentil	2,092210
75%-Perzentil	3,010162
90%-Perzentil	4,804847
95%-Perzentil	6,574506
99%-Perzentil	12,495018
Mittelwert	2,829
Standardabweichung	2,328106
Minimum	1,204
Maximum	44,373
Fallzahl	**11.366**

Literatur

Jesske, Birgit, Schulz, Sabine & Bangel, Philipp (2026): Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung - PASS Websurvey 2025 (Haupterhebung). FDZ-Methodenreport 04/2026, Nürnberg.

Impressum

FDZ-Datenreport 11|2026

Veröffentlichungsdatum
18. November 2026

Herausgeber
Forschungsdatenzentrum
der Bundesagentur für Arbeit
im Institut für Arbeitsmarkt- und Berufsforschung
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/fdz/reporte/2026/DR_11-26.pdf

Dokumentation Version
PASS-SUF0625_DE_v1_dok1, DOI: 10.5164/IAB.FDZD.2611.de.v1

Datensatz Version
PASS-SUF0625 v1, DOI: 10.5164/IAB.PASS-SUF0625.de.en.v1

Bezugsmöglichkeit aller Veröffentlichungen der Reihe „FDZ-Datenreport“
https://fdz.iab.de/forschung/publikationen/fdz-datenreporte/

Website
https://fdz.iab.de/

Rückfragen zum Inhalt
Jonas Beste
Telefon: 0911 179-2279
E-Mail: jonas.beste@iab.de