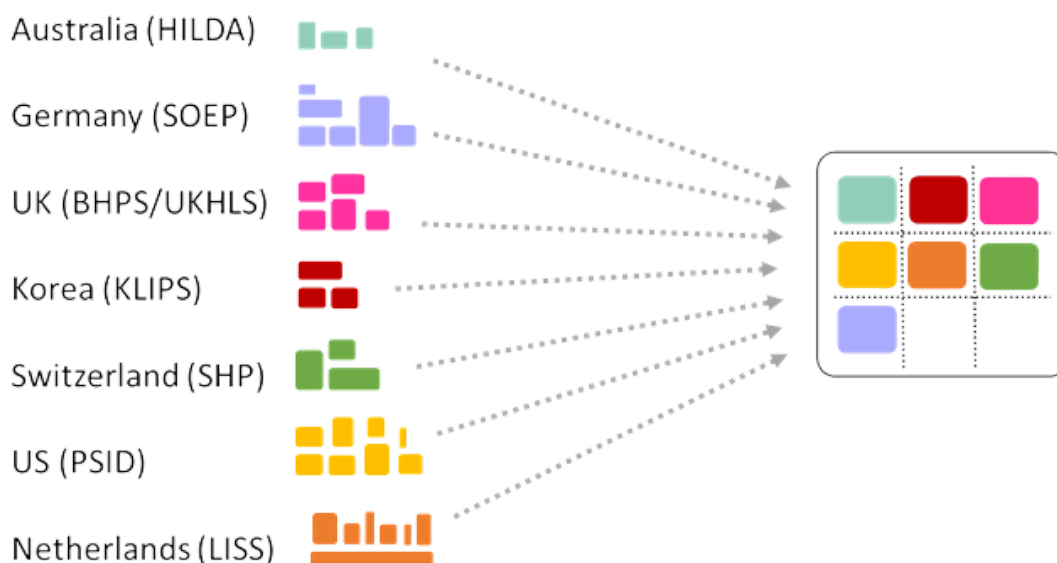


CPF Newsletter

Issue #1
September
2025

- CPF version 2.0 released
- Netherlands (LISS) added
- New waves and extended code
- New website
- Russia (RLMS) excluded



Open Science
Open Source

CPF is an open science project to harmonise the world's major and longest-running household panel surveys from several countries.

CPF Open Harmonization ver. 2.0

We are happy to announce that the CPF Open Harmonization version 2.0 has been released! It is available to **download** (open source code stored at **GitHub**). For detailed information, see the **Manual** and **Codebook**.

This is still a fresh version, so please let us know about any problems or errors: **k.i.turek@tilburguniversity.edu**.

✓ **The CPF version 2.0** encompasses up to 43 waves (between 1968 and 2024), combines data from seven countries, and comprises approximately 3 million observations from more than 400,000 respondents.



✓ The **LISS** panel from the Netherlands was added, offering high-quality annual data across diverse life domains. This was done in collaboration with Priscilla Zhang from Centerdata (the authors of the LISS survey).

✗ The Russian part of the project (**RLMS**) has been excluded from CPF v2.0 in response to Russia's 2022 invasion of Ukraine. See the note **[here](#)**

✓ The code was adjusted to include the latest available **waves** (data collected up to 2023 and 2024)

✓ The **CPF code** was improved with a more transparent syntax structure and some modifications to the harmonization code (based on the team evaluation and users' suggestions).

✓ CPF has a fresh and open-source **[website](#)** beautifully developed by **Xiao Xu**! You can now:

🔔 Subscribe to the newsletter

📄 Submit your **[publications](#)** and projects using CPF



The development of CPF 2.0 was supported by a grant from the NWO (The Dutch Research Council) Open Science Fund (grant ID OSF23.2.017).

About CPF

The Comparative Panel File (CPF) Open Harmonization is an ongoing open-source project that aims to support the community of social researchers. It harmonizes household panel data from seven of the world's most important longitudinal surveys: Australia (HILDA), Germany (SOEP), Great Britain (BHPS/UKHLS), South Korea (KLIPS), Switzerland (SHP), the United States (PSID), and the Netherlands (LISS). These studies offer rich, multi-decade information on individuals and households.

The project aims to support the social science community in analyzing comparative life course data. CPF provides open-source Stata code that links and harmonizes these data into a unified three-level panel structure, enabling comparative life course and social science research. The project is designed for flexibility—users can customize the code to suit their own variables, countries, and samples. It supports both straightforward use and advanced extensions. CPF is not a data product, but a community-driven harmonization tool. Users must download original data from national sources and apply the CPF code to generate the harmonized dataset.

Please note that users must take responsibility for the final harmonization and analytical decisions, while CPF provides flexible tools and a coding framework only. If you find an error, want to suggest an improvement, or propose an extension, please contact us at contact@cpfdata.com / k.l.turek@tilburguniversity.com or suggest the changes on GitHub <https://github.com/cpfdata>.

>> Read more and see the CPF team [**here**](#)

New in CPF v2.0 code

1. The **LISS panel** from the Netherlands was added, offering high-quality annual data across diverse life domains.
2. The **Russian part of the project** (RLMS) has been excluded ([more here](#))
3. The code was adjusted to include the **latest available waves**
4. The **CPF code was improved** with a more transparent syntax structure and some modifications to the harmonization code. The most visible changes include:
 - Updated and improved instructions, comments, and additional information in the syntax files
 - Many minor corrections to the harmonization code (e.g., HH income in UK, kidsn* in KOR and US, relig in US, fedu/medu in GER, parstat6 excluded from US and KOR)
 - More precise definition of yearly income and household income, depending on whether it refers to the previous year (based on respondents' reported value) or the current year (estimated based on monthly income). For example, instead of incjobs_y*, two new variables are created: incjobs_py* (reported annual income for the previous year) and incjobs_cy* (estimated annual income for the current year based on monthly income).
 - Added equivalent household income
 - For Gender (variable female), a category 2 "Other/No answer" was added (available only in some countries)
 - New labels for Self-Rated Health (srh5): (1) Excellent, (2) Very good, (3) Good, (4) Fair, (5) Poor
 - An updated beta version of 'psidtools' has been added to work with the new PSID code. At the moment of preparing the CPF 2.0 (6.2025), Psidtools required a minor update to work with the 2023 wave. The original Psidtools package was compatible only with PSID until wave 2021. The small adjustments made by K. Turek are based on the original code of the Psidtools creator, Prof. Dr. Ulrich Kohler (<https://gitup.uni-potsdam.de/ukohler>), who deserves all the credit for this amazing tool. The original source code is available at: <https://gitup.uni-potsdam.de/ukohler/psidtools/-/blob/main/psid.ado>. The CPF v.2.0 version is included in the CPF 2.0 syntax (in the PSID folder).

Overview of the CPF 2.0 data

The CPF version 2.0 encompasses up to 43 waves (between 1968 and 2024), combines data from seven countries, and comprises approximately 3 million observations from more than 400,000 respondents (Table 1). The oldest survey is the PSID, which began in 1968 and has collected 43 waves. The second oldest is SOEP, which started in 1984 and has collected 40 waves to date. The youngest panel study in CPF is HILDA, with 23 waves since 2001 (however, only 20 waves are included in CPF 2.0, the rest will be added after securing data access), and the newly added LISS with 17 waves.

An average respondent participated in 7.6 waves (between 5.7 in the Netherlands and 11.1 in the US). Out of all 403,751 respondents, 122,610 participated in a minimum of 10 waves and 40,197 in a minimum of 20 waves.

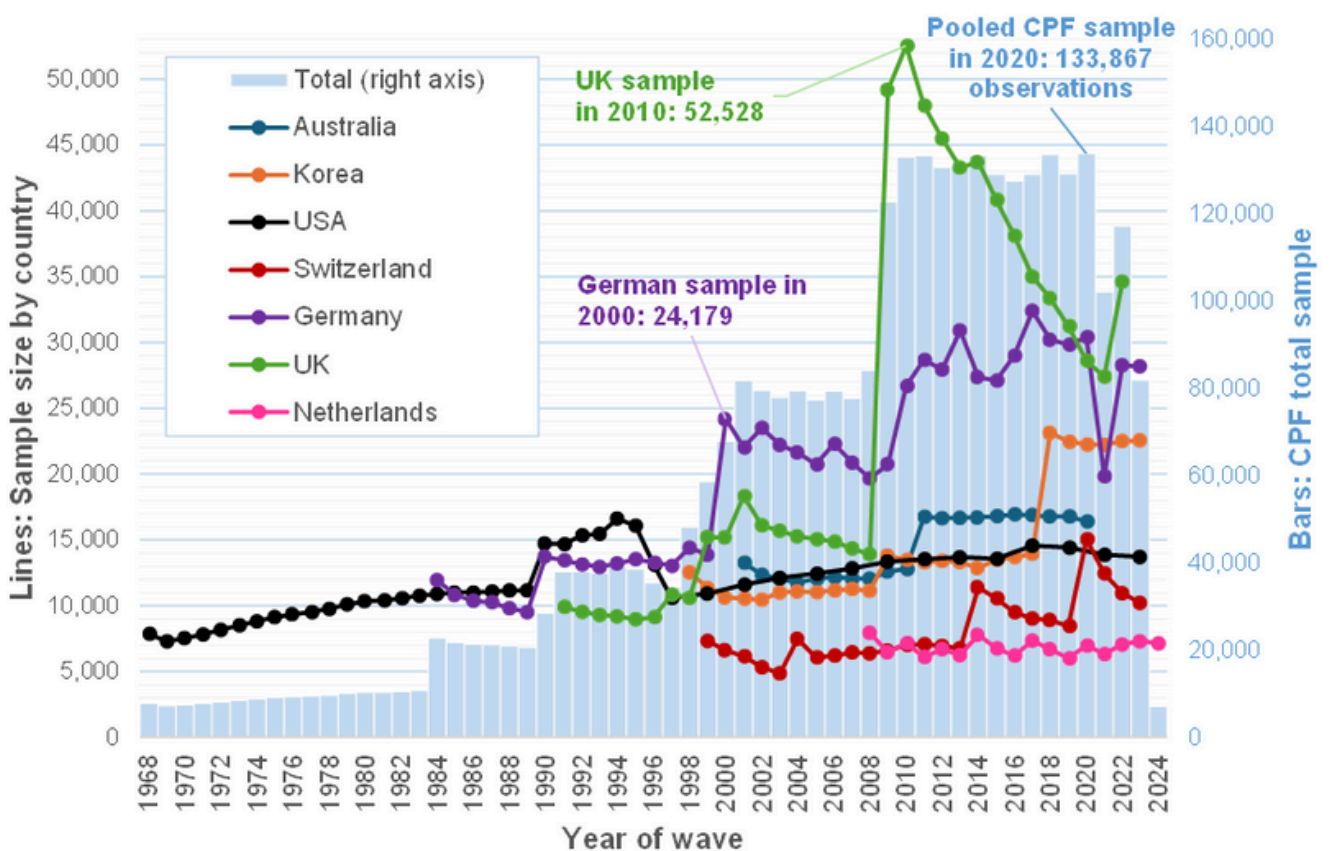
Table 1. *Number of waves, observations, and respondents (CPF v2.0)*

Country	Survey	First wave	No of waves	Observations		Unique respondents	
				n	%	n	%
[1] Australia	HILDA	2001	20	290,616	9.42	31,897	7.9
[2] Korea	KLIPS	1998	26	378,766	12.28	35,744	8.85
[3] USA	PSID	1968	43	499,658	16.2	45,077	11.16
[5] Switzerland	SHP	1999	25	203,914	6.61	30,719	7.61
[6] Germany	SOEP	1984	40	812,135	26.33	125,229	31.02
[7] UK	BHPS/UKHLS	1991	32	782,886	25.38	114,554	28.37
[8] Netherlands	LISS	2008	17	116,355	3.77	20,531	5.09
Total				3,084,330	100	403,751	100

Overview of the CPF 2.0 data (cont.)

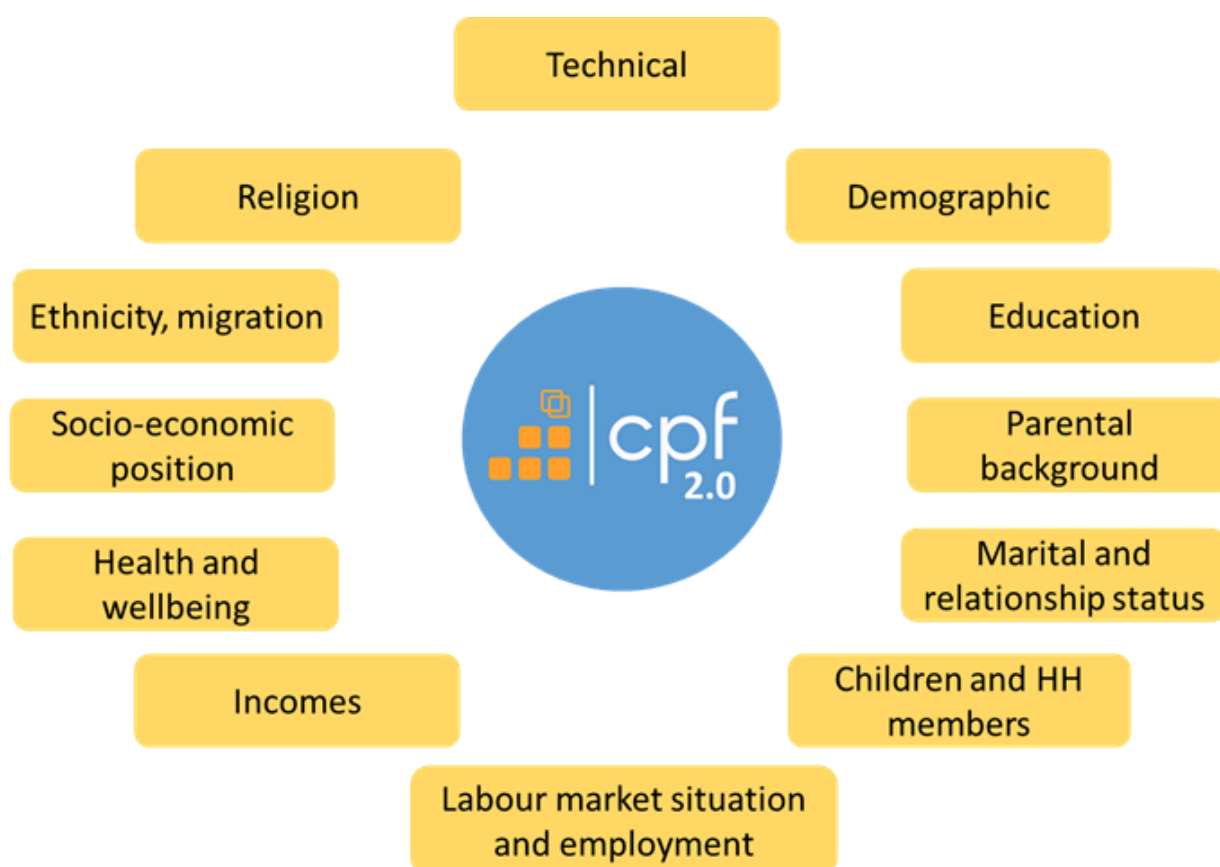
The oldest survey in CPF is PSID, covering the period from 1968 to the present, with 43 waves (Figure 1). The youngest panel study in CPF is the Netherlands, with 17 waves since 2007/2008, and HILDA with 20 waves since 2001. Since the wave of 2000, the number of participants in SOEP has grown significantly. CPF includes three countries since 1991, four countries since 1998, six countries since 2001, and all seven from 2008 (after excluding Russian data, which were available from 1994). A substantial increase observed for the UK sample in 2009 is related to the transition from the BHPS to the UKHLS. For most surveys, data have been collected yearly (after 1997, the PSID switched to 2-year intervals in data collection).

Figure 1. Timeline of the data and sample sizes by wave (CPF v2.0)



Overview of the CPF 2.0 data (cont.)

Figure 2. *Main groups of variable harmonized in CPF 2.0*



CPF v.2.0 team

CPF 2.0 was designed and developed by:

Konrad Turek, Tilburg University & Netherlands Interdisciplinary Demographic Institute, www.konraturek.com

Support for the Dutch LISS data:

Priscilla Zhang, Centerdata

Consultations & general support:

Matthijs Kalmijn, Netherlands Interdisciplinary Demographic Institute

Xiao Xu, Netherlands Interdisciplinary Demographic Institute

Contact: k.i.turek@tilburguniversity.edu

www.cpfdata.com

Project supported by:

