Research and Innovation Action

Social Sciences & Humanities **Open Cloud**

Project Number: 823782   Start Date of Project: 01/01/2019   Duration: 40 months

Deliverable 4.15 Report on integrating API into GGP

Assessment of data quality and scalability of integrating the social policy API in a social survey

Dissemination Level	PU
Due Date of Deliverable	31/12/2021 (M36)
Actual Submission Date	30/11/2021
Work Package	WP 4 - Innovations in Data Production
Task	Task 4.5 Social policy APIs for social surveys
Type	Report
Approval Status	Waiting EC approval
Version	V1.0
Number of Pages	p.1 – p.17

Abstract:

This deliverable assesses the data quality and scalability of integrating a social policy API in an online survey. The assessment is based on data collected via the WageIndicator Survey in the Netherlands.

History

Version	Date	Reason	Revised by
V0.1	15/11/2021	Submission to WP Lead	Judith C. Koops
V1.0	29/11/2021	Final version for submission	Judith C. Koops

Author List

Organisation	Name	Contact Information
KNAW	Judith C. Koops	Koops@nidi.nl
UVA-WWI	Kea G. Tijdens	K.G.Tijdens@uva.nl

Executive Summary

Task 4.5 Social policy APIs for social surveys in the Social Sciences and Humanities Open Cloud project aims to demonstrate the application of social policy APIs in a social sciences survey infrastructure. A proof of concept was prepared by the Generations and Gender Programme and the WageIndicator Survey in the form of an experiment in which a Social Policy Module was integrated in the Dutch WageIndicator Survey. The experiment can be used as a template for future applications which aim to link social policy information with information of individuals or households.

The Social Policy Module was collected from June until September 2021 among Dutch men and women. The response rate of the Social Policy Module was about 30%. In total 86 respondents judged the estimated family benefits, and 50 respondents judged the estimated housing benefits.

Integrating social policy information via an API into a survey can facilitate the integration of complex social policy information captured in complex algorithms. The API approach is scalable, it can be used with a wide range of social policy information, and it can serve different purposes such as validating an existing database, improving a survey instrument and collecting substantive information about the effect of social policies on people's perceptions. Implementing a social policy API in an existing social sciences infrastructure can substantially reduce costs for social policy data collection, without substantially increasing interview time or respondent burden. Researchers preparing a social policy API are advised to follow the 'FAIR Guiding Principles for scientific data management and stewardship' in order to make the product findable, accessible, interoperable and reusable for others.

Abbreviations and Acronyms

API	Application Programming Interface
ESS	European Social Survey
GGP	Generations & Gender Programme
GGS	Generations and Gender Survey
IDSC	International Data Service Center
IZA	Institute of Labour Economics
KNAW	Royal Netherlands Academy of Arts and Sciences
NIDI	Netherlands Interdisciplinary Demographic Institute
OECD	Organisation for Economic Co-operation and Development
SHARE	Survey of Health Ageing & Retirement in Europe
SSHOC	Social Science & Humanities Open Cloud
UVA	University of Amsterdam

Table of Contents

1.Introduction

Understanding individual's perception of family support can generate significant insight of opinions about social policy and how social policies affect people's behaviour. Traditionally, social policy information are made available to the public via use files. However, this format may make it difficult to link policy information directly to individuals, because of the lack of detail as well as the size of these files.

Task 4.5 Social policy APIs for social surveys in the Social Sciences and Humanities Open Cloud (SSHOC) project aims to demonstrate the usefulness of capturing social policy information in an Application Programming Interface (API) by integrating the estimates based on the Organisation for Economic Co-operation and Development (OECD) Family Database in an existing online social sciences survey. The experiment was organized by KNAW-NIDI and UVA-WWI. A technical description of the Social Policy API tool based on the OECD Family Database was provided in the first deliverable associated with *SSHOC Deliverable 4.14 Policy API Tool[1]*.

The current deliverable is the second associated with Task 4.5. It provides information about the implementation of the Social Policy Module in the Dutch Wage Indicator survey (Chapter 2), the archiving strategy of the generated social policy information (Chapter 3), information about the response rate, drop-off rate, and quality of the estimated family and housing benefits (Chapter 4), and a reflection on scalability of the project (Chapter 5). Conclusions and recommendations for future policy research are provided in Chapter 6.

2. Changes to Data Collection Approach

Deliverable 4.14 provided a description of the Social Policy API tool based on the OECD Family Policy Indicator. The intention was to demonstrate the potential of the API by integrating a module in the Generations and Gender Survey (GGS) in the first half of 2021. However, due to the COVID-19 pandemic, data collection of GGS was postponed, which made it unsuitable to collect data for this task. Instead, the Social Policy module was collected via the WageIndicator Survey. Because the WageIndicator Survey is an online survey, data collection was not interrupted by the COVID-19 pandemic.

The aim of WageIndicator[2] is to provide "More labour market transparency for the benefit of all employers, employees and workers worldwide by sharing and comparing information on wages, Labour Law and career". It accomplishes this by maintaining a website with general information about work and labour law for 196 countries. The WageIndicator Survey is a non-probability survey, which targets

---

[1] Emery, Tom. (2020). SSHOC D4.14 Policy API Tool (v1.0). Zenodo. https://doi.org/10.5281/zenodo.3725926

[2] WageIndicator: https://wageindicator.org/Wageindicatorfoundation/researchlab/wageindicator-survey-and-data; accessed on 9/11/2021

employees, the self-employed, informal workers, job seekers and the unemployed. Respondents of the web-survey are volunteers recruited through the national WageIndicator websites. The WageIndicator web-survey is implemented in the website of more than 80 countries. There is a prize incentive for survey completion. The topics of the web-survey relate to wages and working conditions. In addition, the survey regularly includes project-specific survey questions for a limited number of countries and a limited period of time. The partners made use of this option to field a Social Policy Module.

2.1   Implementation of Social Policy Module

The Social Policy Module was collected in the Dutch WageIndicator survey, respondents were recruited through Loonwijzer.nl - the Dutch WageIndicator survey[3], which was deemed a good test case for the following reasons: The Netherlands is an OECD country with a rather complicated social policy benefit scheme. In these circumstances, the use of an API has an added benefit in particular. In addition, the Dutch survey has a high data output in general, which allows for a relatively short period of data collection. The use of the Dutch language facilitated the preparation of the module and data analyses for the Dutch collaborators of Task 4.5. To keep the module short, it was decided to focus on two types of benefits only: family benefits and housing benefits.

In order to calculate family and housing benefits with the use of the OECD Family Database, the following information of the individual had to be captured:

- Number of children (0, 1, 2, 3, 4)
- Income of respondent as a % of median income (50, 100, 150)
- Income of partner as a % of median income (50, 100, 150, NA)
- Age of youngest child in months
- Couple status (single or not)

Part of these questions were already captured with the WageIndicator survey, the remaining questions were added to the Social Policy Module. During the online interview all relevant information needed to estimate family and housing benefits were captured. This information was then linked to the information of the OECD Family Database. The corresponding family and housing benefits, unique for that combination of information that the respondent provided, was presented to the respondent. The family and housing benefits estimates were presented per month and per year. Respondents were then asked if they thought these estimates were correct. If this was not the case, respondents were invited to provide a correction of the benefits they receive. Respondents were asked to provide the corrected benefits received per month, because this estimate was deemed more accurate than benefits received per year.

---

[3] www.loonwijzer.nl (accessed Nov 2021)

To capture family benefits, two questions were asked, one about received child benefit "*Kinderbijslag*" and another about received child budget "*Kindgebonden budget*". Child benefit and child budget are two different schemes through which caregivers of children can receive family benefits in the Netherlands. Child benefit is dependent on the age of the child, while child budget depends on the number of children and the income of care givers.

More concretely, the questions asked (translated from Dutch) were the following:

Based on a general calculation, we estimate that you are eligible for the following allowances. Could you indicate to what extent these general estimates are correct?

*Housing benefit: … per month or **…** per year.* {Estimate based on information of OECD Family Database}

This is:
- Certainly not correct
- Probably not correct
- Probably correct
- Certainly correct

{If Certainly not correct or Probably not correct}
How much housing benefit do you receive?
… per month {Information provided by respondent}

Child benefit and child budget: … per month or … per year. {Estimate based on information of OECD Family Database}

This is:
- Certainly not correct
- Probably not correct
- Probably correct
- Certainly correct

{If Certainly not correct or Probably not correct}
How much child benefit do you receive?
… per month {Information provided by respondent}

{If Certainly not correct or Probably not correct}
How much child budget do you receive?
… per month {Information provided by respondent}

---

3. Data Storage

The collected survey information will be archived at the data repository of the Institute of Labour Economics (IZA) called the International Data Service Center (IDSC)[4] and be made available to researchers for scientific purposes. WageIndicator data will be reposited in March or April 2022. In addition, a dataset with the data collected in the Social Policy Module will be made available in Zenodo before March 2022.

4. Data Quality

4.1   Response rate

Questions about received housing and family benefits were asked to men and women with the following characteristics:

- Lived in the Netherlands
- Had a co-resident partner
- Had at least one child who lived in their household
- Their youngest (or only) child was born after 2002 (younger than 18 years old)
- They had a paid job, or their partner had a paid job at the time of the interview
- Rent their home (only conditional for questions about housing benefit)

The Social Policy Module was fielded from Sunday 20 June until the end of September 2021. During this period, in total 1147 respondents participated in the Dutch WageIndicator Survey. Of these respondents, 333 were eligible to take part in the Social Policy Module. Figure 1 provides details on the response rate to the different questions within the Social Policy Module. The daily response rates are shown for each day in the period from Sunday 20 June 2021 (day 171) until Thursday 30 September (day 273). On average 3 respondents a day were eligible to answer questions about Family benefits and on average 2 respondents a day were eligible to answer questions about Housing benefits.

---

[4] International Data Service Center: https://legacy.iza.org/en/webcontent/research/izadsc.html; accessed on 10/11/2021
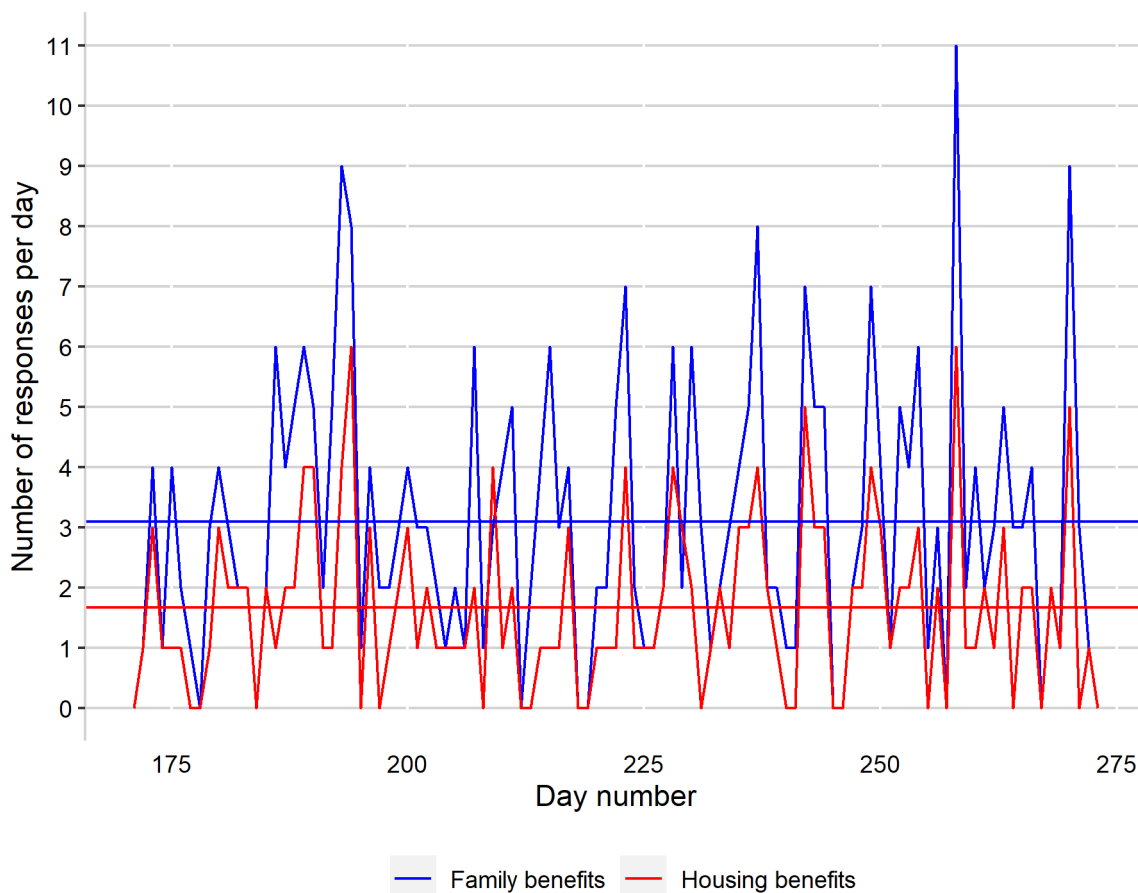
Figure 1. Daily number of responses to Social Policy Module over the period from Sunday 20 June 2021 (day 171) until Thursday 30 September (day 273).

Response rates fluctuate from day to day. Figure 2 shows the response rates per weekday. The pattern for response to the Social Policy Module is similar to the overall participation rate in the survey. Participation is highest at the start of the week and lowest on Fridays and during weekends. Participation in the Social Policy Module shows a small peak on Wednesdays, which is not found in the overall participation to the WageIndicator survey. It is possible that by selecting on respondents with a partner and children, more part-time workers are captured. In the Netherlands it is common that part-timers take Wednesdays off, because young children who go to primary school are often free from school on Wednesday afternoons.
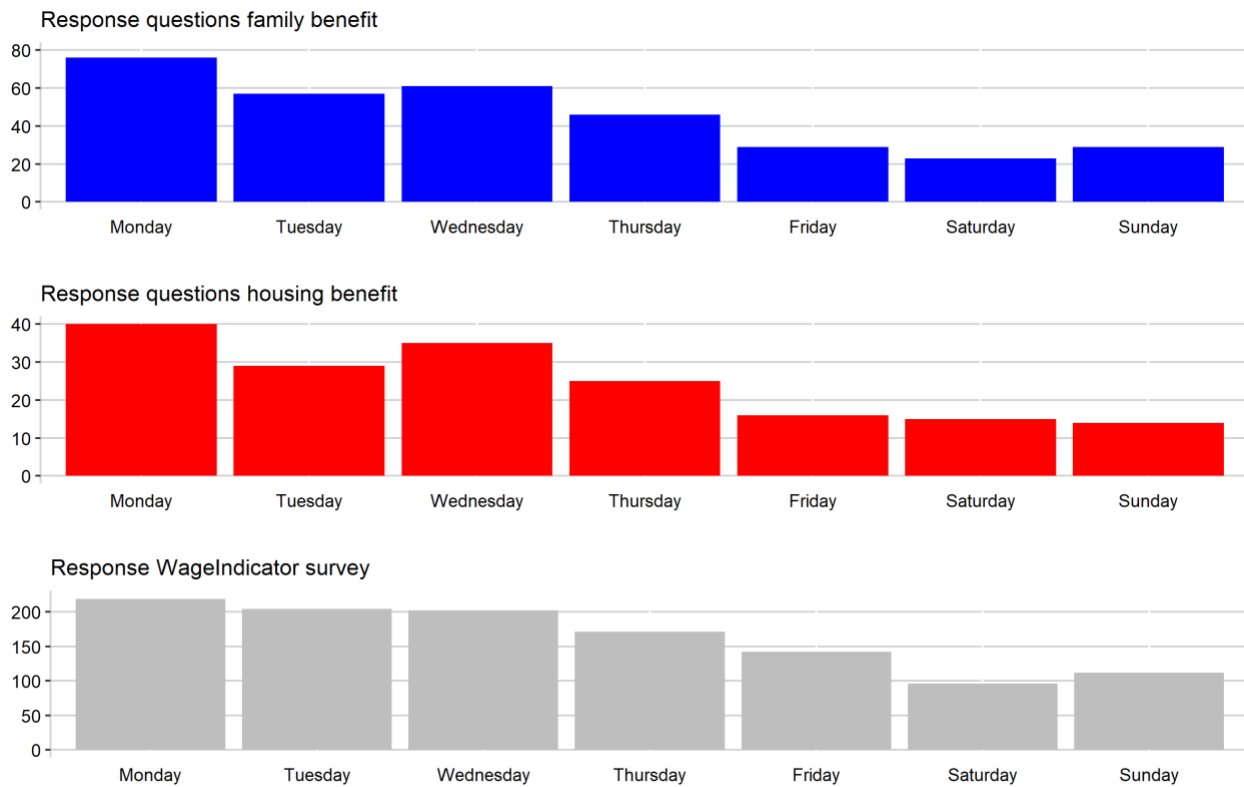
Figure 2. Response rate per week day.

4.2   Drop-off rate

In the period from June till September 2021, a total of 321 respondents were eligible to participate in the Social Policy Module and were presented with an estimate of their family benefits (see Figure 3). Of these respondents, 86 judged the family benefits, which equals a response rate of 27%. Housing benefits were estimated for 174 respondents who rented their home. 50 of them judged the estimates, which equals a response rate of 29%.
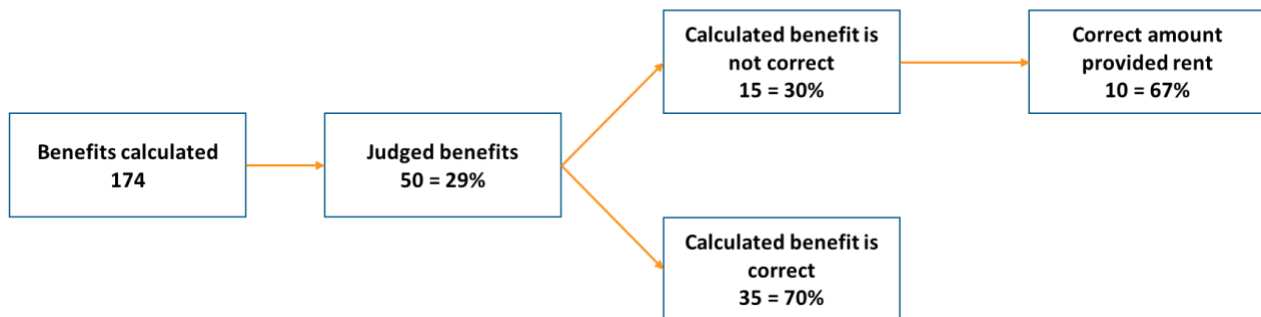
Family benefits
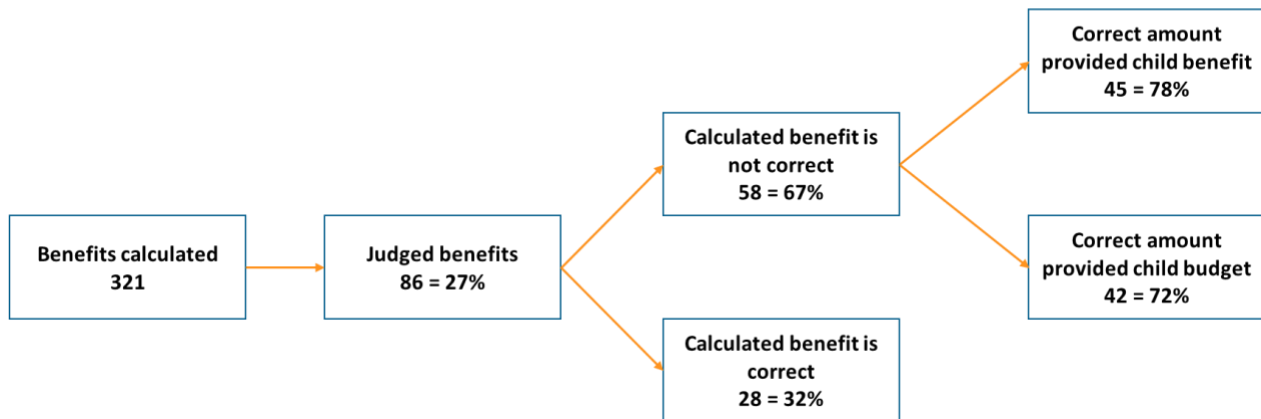
Housing benefits

Figure 3. Response rate to questions related with family benefits and housing benefits.

4.1   Quality of the estimates

32% of the respondents indicated that the estimated family benefit was correct (N=19) or probably correct (N=9). The majority indicated that the estimate was incorrect (N=46) or probably incorrect (N=12). Housing benefits were more often calculated correctly. 70% believed their housing benefit was correct (N=30) or probably correct (N=5), while 30% indicated the estimate was incorrect (N=11) or probably incorrect (N=4).

Respondents who thought their estimated family or housing benefit was incorrect or probably incorrect, were asked to provide the correct amount themselves. 78% of respondents who indicated that the benefit estimates was incorrect, provided information on actual received child benefit, 72% provide information on received child budget, and 71% provided information on received rent benefit.

The distribution of the estimated yearly family and housing benefit (in grey) as well as the corrected family benefit (in blue) and housing benefit (in red) reported by the respondent is shown in Figure 4. The graph only captures the cases for which respondents indicated that the estimated yearly benefit was incorrect or probably incorrect and the respondent provided information about their actual received benefits. The median estimate of family benefits based on the OECD family calculator of €1987 per year / €166 per month was lower than the benefits respondents said they actually received (€2820 per year / €235 per month). For housing benefits, the opposite was the true. Here the estimated housing benefit based on the OECD was higher than the corrected benefits provided by the respondent.
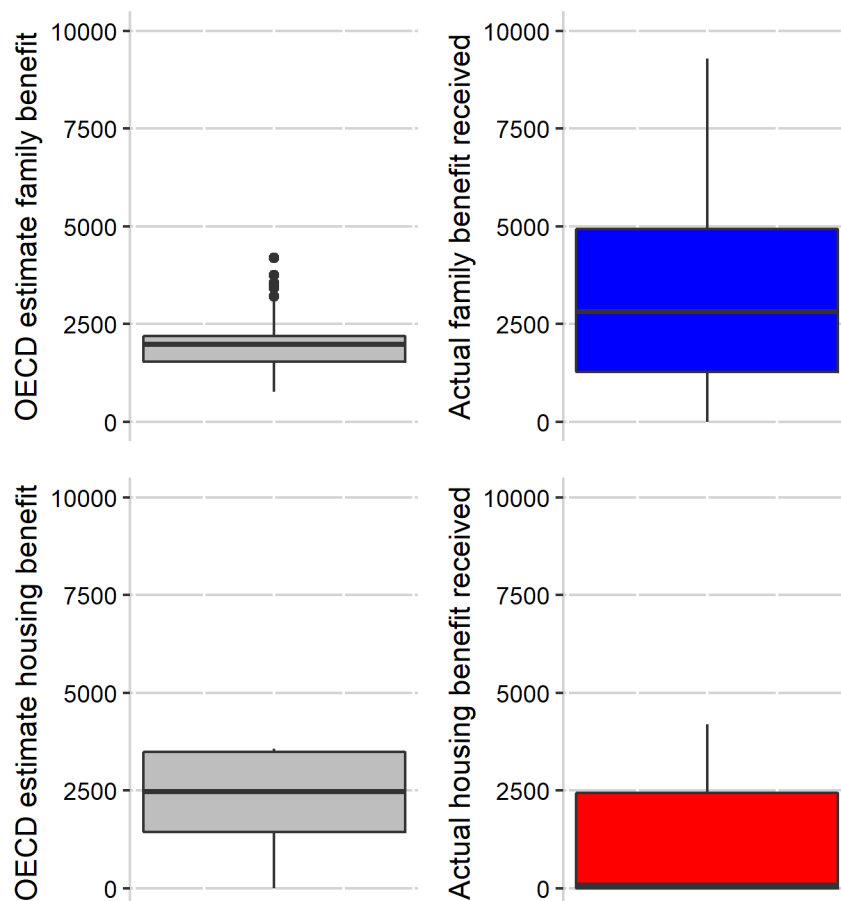

Figure 4. Distribution estimated yearly family and housing benefit based on the OECD calculator and provided by the respondent.

Because fewer respondents were eligible to participate in the housing benefit section and the estimate based on the OECD calculator was more accurate, only a few respondents (N = 10) provided information on the corrected amount of housing benefits received. This makes it difficult to compare the two boxplots of Figure 4. An easier comparison is provided in Figure 5 where the difference between the estimated yearly family and housing benefits based on the OECD calculator and the actual yearly benefits received by the respondent is portrayed.

The difference between estimated and actual housing benefits is positive. This means that when respondents indicated that the estimated housing benefit was incorrect the estimate was generally higher than what respondents reported to receive. In fact, of the 10 respondents who provided information on the housing benefits they received, 7 respondents reported to receive benefits which were lower than those estimated with the OECD Family Database, while 3 respondents reported to receive benefits which were higher than those estimated with the OECD Family Database. The median difference was €1466 per year, or €122 a month.

The difference between estimated and actual family benefits is negative, which means that when the estimated family benefit was incorrect, the estimate was generally lower than what respondents reported to have received. Of the 46 respondents who provided information on the family benefits they received, 10 reported to receive family benefits that were lower than those based on the OECD Family Database, while 36 reported to receive family benefits that were higher than those based on the OECD Family Database. The median difference is €-839 per year, or -€70 a month.
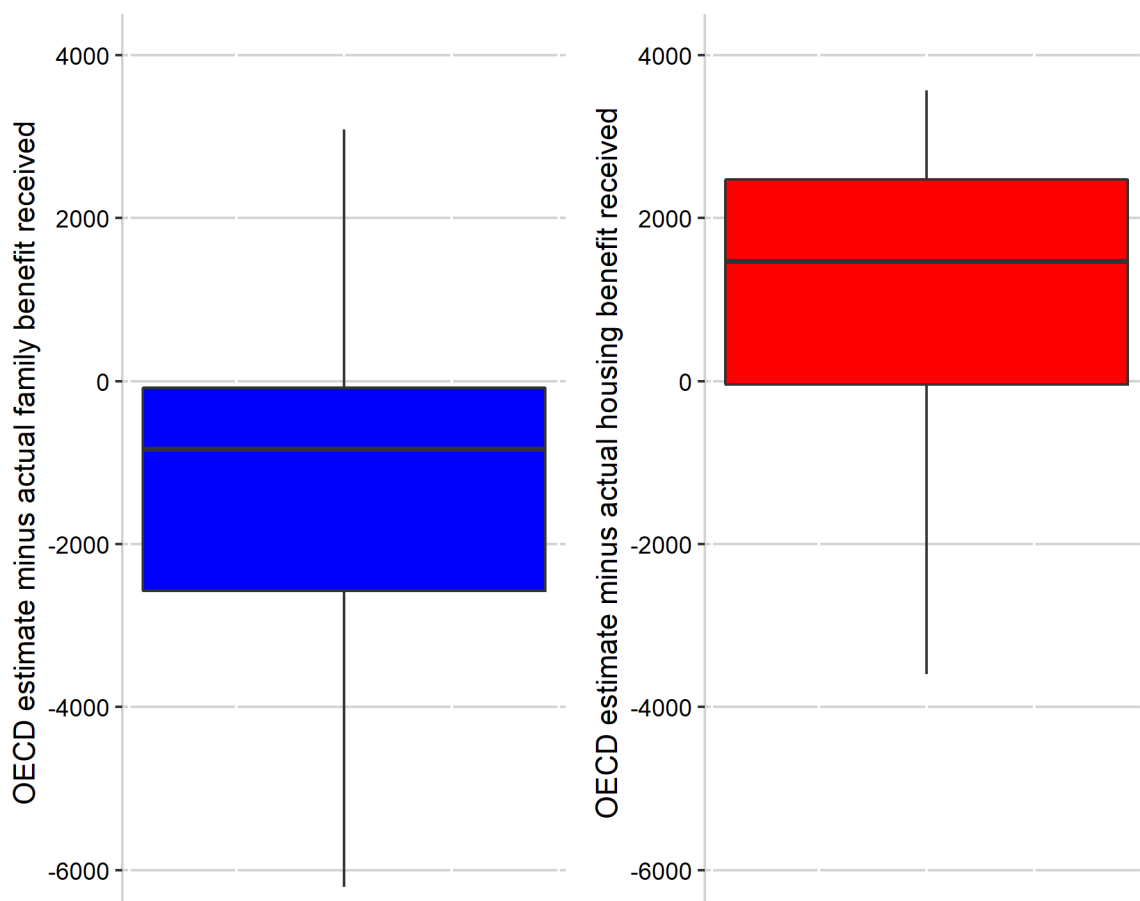

Figure 5. Difference between estimated yearly family and housing benefit based on the OECD calculator and the yearly benefit provided by the respondent.

5. Scalability

There are different ways in which the current project is scalable. Firstly, the implementation of social policy information in a survey can serve different purposes. In the current experiment, one of the purposes of implementing information about personalized housing and family benefits was to validate the estimates of the OECD algorithm. However, once it is ascertained that an algorithm can provide reasonable estimates, implementing a social policy API could serve other goals. Respondents often struggle to provide financial information (such as received benefits), which can lead to high non-response on these types of questions. Although it is not certain and future research should confirm if this is true, it is possible that respondents are more engaged when shown an estimate and asked to correct this than if they are just asked to provide a number themselves. If this is true, implementing social policy information in the survey via an API could be used to improve the survey instrument. Social policy APIs could also be used to understand individual's perception of support for social policies. Using an API for the creation of different scenarios at random (but within a certain range of values), could provide more detailed information than just presenting a few scenarios to respondents. A second way in which the current project is scalable is by extending the types of policy information captured in an API. In the current experiment, information of an algorithm providing estimates of housing and family benefit were integrated in a survey. However, other information - for example related with poverty reduction, health care and pension schemes - could be implemented in a similar way.

Using a social policy API to include social policy information in a survey is most effective when the policy information is part of an algorithm or is based on a combination of information. When information can be captured accurately in a simple table, other ways, such as directly coding the information in the survey software, is probably less time-consuming. In SSHOC task 4.5, the API was published through Plumber (an R package). However, most programming languages can be used to create APIs. The use of an API to import social policy information, is therefore very flexible and can be accommodated to the survey software used and the expertise in the team.

It is difficult, if not impossible, to predict the amount of effort and financial resources needed to implement a social policy API in a survey, because this depends for a large part on the questionnaire software used, the IT capabilities in the team, and prior experience with the implementation of an API. However, overall costs for data collection can be kept to a minimum by implementing a Social Policy Module in an existing survey infrastructure. Sharing costs for the payment of interviewers, incentives for respondents, sample preparation, data preparation etc. will make the data collection more cost-efficient. This would permit more time and resources in the preparation and implementation of the API. Moreover, because many European social surveys, such as WageIndicator survey, GGS, ESS and SHARE, already capture socio-economic and demographic information needed to estimate benefits, implementing a social policy module would not increase interview time or respondent burden substantially. In the current experiment, the API was hosted on a server and accompanied with documentation. Such an approach allows other researchers to query and use the API and implement it into their own research lifecycle. This

approach aligns well with the 'FAIR Guiding Principles for scientific data management and stewardship'[5] and will reduce the efforts and costs for other researchers interested in using the same policy information.

6. Conclusion and Recommendations

Social policies are generally algorithmic in nature and are based on a set of household or personal indicators. Social policy information, such as eligibility to family, housing, or pension benefits, is mostly made available to the public via use files. Disadvantages of use files are that they are not able to provide the same detail as an algorithm and they are often too large to efficiently distribute and load. In Task 4.5 *Social policy APIs for social surveys* of the SSHOC project, the algorithm of the OECD Family Database was published as an API. This approach allowed to capture the social policies rules in an algorithm without having to load and surge a whole database when linking the policy information with information of individuals.

In the current experiment, the functionality of the API was demonstrated by using the database to estimate family and housing benefits of respondents participating in a social survey. The Social Policy Module was implemented in the Dutch WageIndicator Survey, an existing social sciences infrastructure which collects information about the labour market and salaries via online web interviewing. Over the period from June to September 2021, information about family benefits was collected among 86 respondents and information of housing benefits among 50 respondents. Most questions needed to estimate family and housing benefits, such as family situation and employment status, were already part of the Dutch WageIndicator Survey. For this reason, implementation of the Social Policy Module did not increase respondent burden nor interview length by much. As such, the implementation of the Social Policy Module in the WageIndicator Survey was deemed successful and could be used as a template for further development of linking social policy information with social sciences surveys.

The OECD Family Database was created in 2014 and might therefore be outdated. In order to judge the accuracy, the family and housing estimates were presented during the online interview to respondents for verification. The verification showed that the housing benefits based on the OECD Family Database algorithm were rather accurate. However, the OECD Family Database algorithm was less successful in predicting family benefits. Only a third of respondents judged the estimated family benefit as accurate or probably accurate. In instances where the estimate was not accurate, family benefits were mostly underestimated.

Based on the experiment of Task 4.5 of the SSHOC project, the partners conclude that APIs are a way facilitate the implementation of complex social policy information in a survey context. Social policy APIs

---

[5] FAIR Guiding Principles: https://www.go-fair.org/fair-principles/; accessed on 29/11/2021

can be used to validate a database, to improve a survey instrument and to collect substantive information such as perceptions of social policies. The effort and finances needed to implement a social policy API in a survey depends on the questionnaire software used, the IT capabilities in the team, and prior experience with the implementation of an API. However, overall fieldwork costs can be reduced by implementing the social policy API in an existing survey, such as WageIndicator survey, GGS, ESS and SHARE. Following the 'FAIR Guiding Principles for scientific data management and stewardship' when preparing a social policy API, can reduce the effort and costs for future research projects interested in using the same policy information.

List of Figures