PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

MASTER THESIS N° 2021 – 03

The subjective well-being impact of water insecurity: a quasi-experimental evidence from peri-urban Kinshasa, Democratic Republic of the Congo

Ny Sata Andrianirina

JEL Codes: C83, C93, H41, I10, I31, J01, O12, O15, O22, Q25, R23
Keywords: Water access. Subjective well-being. Policy evaluation. Water policy

The subjective well-being impact of water insecurity: a quasi-experimental evidence from peri-urban Kinshasa, Democratic Republic of the Congo

JEL codes: C83, C93, H41, I10, I31, J01, O12, O15, O22, Q25, R23

Ny Sata Andrianirina
Supervisor: **Flore Gubert**
Referee: **Camille Saint-Macary**

Master thesis in
Public Policies and Development



Master 2, Public Policies and Development 2020-2021
Paris School of Economics
September, 2021

Abstract

Recent researches suggest that insecurity of access to key resources (eg: food, water, electricity) is associated with negative mental health outcomes and signs of anxiety and depression especially in developing countries. In this Master's thesis, we investigate the influence of water insecurity on subjective well-being. The data come from surveys conducted in a municipality in the peri-urban area of Kinshasa called N'Sele, as part of the impact evaluation of the PILAEP 2 project, which aims to build self-sufficient drinking water solutions in the periphery of Kinshasa. The data include 2,345 individuals who self-declared themselves knowledgeable about water activities from different households in 15 neighbourhoods of N'Sele and they are interviewed in two waves (a baseline in 2018 and a follow-up in 2019). In this thesis, we relate three measures of subjective well-being: emotional distress (represented by Cohen's Perceived Stress Scale 'PSS-score'), happiness measure and life satisfaction index, while water insecurity is approximated by the amount of water collected by the household in the week prior to the survey. In order to avoid the endogeneity of the amount of water collected induced by the self-selection of households in the area where they settle, our identification strategy is based on 2SLS estimates taking a second-order polynomial form of the rainfall shock as instrument. We also include socio-economic variables such as employment status, economics of the household, asset characteristics, and health status to control for confounding factors. The first stage indicates a non-zero impact of our instrument on the explanatory variable and our results for the second stage identify a negative and significant relationship between the amount of water collected and the level of stress: namely, a one-liter increase in the amount of water collected during the week preceding the survey decreased on average the PSS score by 0.004, ceteris paribus. For the alternative measures of well-being, our results for the ordered probit regression model indicate a positive and significant relationship between happiness or life satisfaction and the amount of water collected. These results are encouraging for the formulation of policy measures regarding access to water and maximising happiness, but above all a policy agenda to improve water supply infrastructure is vital to avoid water-related diseases and deaths that subsequently influence the depression levels, the happiness and life satisfaction of sufferers and their relatives, knowing that, according to WHO and UNICEF (2017), unsafe water and inadequate sanitation and hygiene are responsible for 88% of the thousand million cases of diarrhoea that result in 1.6 million deaths per year.

Keywords Water access. Subjective well-being. Policy evaluation. Water policy

Remerciements

Je tiens à remercier en premier lieu Flore Gubert qui a été la directrice de ce mémoire. Sa très grande disponibilité, sa compréhension et sa rigueur m'ont permis de travailler dans la sérénité sur cette thématique très intéressante qu'elle m'a proposée.

Je remercie également Camille Saint-Macary pour avoir accepté d'encadrer mon mémoire, pour sa disponibilité et pour m'avoir éclairé tout au long de mes travaux de recherches. Merci également à Virginie Comblon pour le travail qu'elle a initié sur mon dossier et sa disponibilité lorsque j'avais des questions.

Je tiens aussi à exprimer mes plus vifs remerciements à l'équipe du DIAL et à tous les directeurs de recherche et les enseignants-chercheurs qui y sont associés et qui ont initié la recherche académique sur le sujet de mon mémoire. Au sein des locaux du DIAL, je remercie spécialement Anne Legendre et Loic Le Pezennec pour leur accueil les fois où j'y ai travaillé et pour la gestion administrative de mes documents.

Merci aussi à l'ensemble de l'équipe pédagogique de l'Ecole d'Economie de Paris et plus particulièrement du master Politiques Publiques et Développement. Je tiens à porter mon entière reconnaissance aux responsables de la formation, en l'occurence Akiko Suwa-Eisenmann, Marc Gurgand, Luc Behaghel ainsi que tout le corps professoral de la formation de qualité qui m'a permis d'appréhender avec virtuosité tous les aspects de mes travaux de recherches durant la préparation de ce mémoire. Merci également aux responsables pédagogiques, en particulier Lucia Roxana-Ban, Megan Devayeres et Pauline Marmin d'avoir été disponible tout au long de mon cursus académique et d'avoir été là si j'avais des questions par rapport à la scolarité.

Merci à l'ensemble de la promotion du master PPD et à tous les étudiants de la PSE grâce à qui ces trois années ont été pour moi très riches. Mention spéciale à Rose Deperrois et Chiara Loriaux qui ont travaillé sur le même dossier que moi et m'ont aidé à dresser les bases de mon travail.

Enfin, je souhaite remercier ma famille et l'ensemble des personnes de mon entourage qui m'ont soutenu, m'ont encouragé et qui ont contribué de près ou de loin à l'accomplissement de ce mémoire.

Contents

1 Introduction **6**

2 Related Literature **8**

2.1 The three dimensions of water insecurity 8

2.2 Subjective well-being notion: emotional distress and the use of the PSS-10 as psychometric measures 9

2.3 Conceptual framework of the mechanism behind the relationship of water insecurity and subjective well-being 10

2.4 Water-related stress and its impact on the perpetuation of poverty 12

3 Experimental Design and Data **13**

3.1 The PILAEP Project 13

3.1.1 Region of study 13

3.1.2 Randomization process 15

3.2 Data, variables selection and descriptive statistics 16

3.2.1 The EACV1 survey 16

3.2.2 Rainfall data 16

3.2.3 Missing values treatment 17

3.2.4 Individual id problem 17

3.2.5 Final sample and descriptive statistics 18

4 Empirical Strategy **24**

5 Results **26**

5.1 Pooled OLS estimates and Fixed-effects regression 26

5.2 First-Stage and Reduced-Form estimates 27

5.3 2SLS Estimates 29

6 Robustness checks and sensitivity analysis **33**

7 Concluding remarks: analysis of PILAEP 2 potential benefits and policy recommendations, limitations. **36**

7.1 Evaluation of the economic benefits in terms of subjective well-being for future PILAEP II beneficiaries 36

7.2 Discussions and limitations . . . 37
7.2.1 Potential omitted variable bias . . . 37
7.2.2 Discussion about the exclusion restriction of our instrument . . . 38
7.2.3 Water insecurity: a multidimensional notion . . . 39
7.2.4 Sampling design and a potential loss of statistical power . . . 40
7.2.5 Survey design . . . 40
7.2.6 Unaddressed issue: the quality of water collected . . . 40

8 Appendix **42**

9 Bibliography **58**

List of Figures

1 Hierarchy of water requirements after Maslow's hierarchy of needs (Source: WHO (2013)) . . . 8
2 Perceived stress scale by Sheldon Cohen (Source: Cohen and Williamson (1988)) . . . 10
3 The relationship between life satisfaction and poverty (Source: Haushofer and Fehr (2014)) . . . 13
4 Location of PILAEP 1 and 2 Beneficiaries (Source: Boaz et al. (2020)) . . . 15
5 Balanced panel "PANEL I" (id consistent) (Source: Author's data handling) . . . 18
6 Balanced panel "PANEL II" (id & rainfall data consistent) . . . 19
7 Treatment of the changes in names, sex or ages for individuals with same individual id (Source: Author computation) . . . 45
8 Economic benefits arising from water and sanitation improvements (Source: Hutton et al. (2004)) . . . 56
9 Calculation methodology, data sources and values for economic benefits (Source: Hutton et al. (2004)) 57

List of Tables

1 Treated and control areas . . . 14
2 Summary statistics of "PANEL I" . . . 21
3 The effect of quantity of water collected in liters on the PSS-Score . . . 31
4 The effect of quantity of water collected in liters on the PSS-Score taking into account the household size . . . 32

5 Sensitivity analysis considering other measure of subjective well-being (BP3: Taking everything into consideration, how happy you are in your life?) . . . 34

6 Sensitivity analysis considering other measure of subjective well-being (BP4: Taking everything into consideration, how satisfied are you with your life?) . . . 35

7 Theoretical and effective samples and sampling rates . . . 42

8 Summary statistics of "PANEL II" . . . 42

9 Pooled OLS regression for the relationship between water collected and the PSS-score . . . 46

10 Fixed Effects regression for the relationship between water collected and the PSS-score . . . 47

11 2SLS estimates for the relationship between water collected and the PSS-score, considering linear form of rainfall shock as instrument . . . 48

12 2SLS estimates for the relationship between water collected and the PSS-score, considering 2nd-order polynomial form of rainfall shock as instrument . . . 49

13 First-stage regression of the relationship between water collected and the PSS-score, considering linear form of rainfall shock as instrument . . . 50

14 First-stage regression for the relationship between water collected and the PSS-score, considering 2nd-order polynomial form of rainfall shock as instrument . . . 51

15 Reduced-form estimates, considering linear form of rainfall shock as instrument . . . 52

16 Reduced-form estimates, considering 2nd-order polynomial form of rainfall shock as instrument . . . 53

17 2SLS estimates using water collected per household member as independent variables . . . 54

18 2SLS estimates using water collected × household member as independent variables . . . 55

1 Introduction

On 28th July 2010, The General Assembly of the United Nations said that "the right to safe and clean drinking water and sanitation is part of human right that is essential for the full enjoyment of life and all human rights". As of today, water insecurity is still a growing problem with major international health implications especially in developing countries. First, low quality water and lack of clean water should cause illnesses such as diarrhea or cholera that would have subsequent influence on happiness and satisfaction with life of sick people and their relatives. According to UNDP (2006), 5% of total deaths in the world, 80% of ilnesses and 30% of deaths in developing countries are water-related (see Elhance (1999)). Second, the lack of access to safe water and adequate sanitations has also implications for psychosocial and psychological well-being of individuals and households (Bisung and Elliott (2017)). With regard to water, psychosocial and psychological health includes stressful experiences related to water and sanitation. It includes burden of collection, negotiating access, the opportunity cost of buying from informal sources and inequalities in the ownership and distribution of water (Ennis-McMillan (2001); Wutich (2009); Sultana (2011); Bisung and Elliott (2017)). Therefore, studying the well-being, happiness and psychosocial and psychological impacts of the lack of access to water on mental health is important as they may have long-term implications on mental and physical health issues.

Consequently in this master thesis, we are going to investigate the relationship between water and subjective well-being (SWB) especially in developing countries where the conditions of access to water are not good. For now, the literature has not generally paid that much attention on the relationship between water and subjective well-being, except certain exceptions (Bookwalter and Dalenberg (2004); Guardiola et al. (2013); Bisung and Elliott (2017); Wutich and Ragsdale (2008); Sultana (2011); Ennis-McMillan (2001)) which makes our research question innovative and interesting. In the same line of research that analyses the factors that influence well-being, this master thesis estimates the effect of the quantity of water collected on the subjective well-being of inhabitants in the neighborhood districts of Kinshasa in DRC. Those districts are N'Sele, Kimbanseke and Mont Ngafula but the survey was only carried out in N'Sele. The choice of those districts are motivated by the fact that despite high water resources in the DRC, only 42% of the population has access to clean water (according to the WHO and UNICEF (2017)) and this situation is worse in the region of the capital Kinshasa due to higher demographic growth. In this study, as we want to include all aspects of subjective well-being that covers both satisfaction with life and the emotional distress dimension, we use the synthetic index of Perceived Stress Scale (PSS) proposed by Cohen and Williamson (1988).

The data come from a panel of 2,345 individuals each from different household and answers from questions related to water access, water consumption, water storage, sanitation and subjective well-being and other socioeconomic variables. Those surveys were ran into two waves as part of the impact evaluation of PILAEP Project 2 (2017-2020) which aimed to build sustainable water distribution system in the peri-urban neighborhood of Kinshasa (in the district of N'Sele, Kimbanseke and Mont Ngafula). The PILAEP Project has been funded by the Agence Française de Développement sinec 2008 and the socioeconomic impact evaluation of the project is being conducted by the economic research of DIAL and the National Institute of Statistics, and the ministry of Finance in the DRC.

The first wave of the survey was between April and May 2018 (referred to as the Baseline period), and the second wave was in the same period but in 2019 (referred to as the Follow-up period).

The aim of this study is to empirically test the influence of the water access on subjective well-being. In details, we assess the relationship between the quantity of water collected in the past week (in liters) and the PSS. In order to do so, we use a fixed-effect regression in which the dependent variable is the PSS. To avoid the endogeneity issues of the quantity of water collected that is induced by self-selection of individuals into areas where they want to live, we implement an Instrumental Variable (IV) strategy and consider the positive rainfall shock as a relevant instrument. Indeed, in the poor areas inside the peri-urban Kinshasa, people are used to harvest rainwater directly to fulfill their sanitation and consumption needs but also some water sources are available only during rainy period according to the first report written by the DIAL researchers, the INS and AFD in September 2020 (Boaz et al. (2020)). To avoid spurious conclusion on the relationship of water access on subjective well-being, we also include socioeconomic variables such as employment status, assets characteristics, economic status such as access to electricity, health status etc in our model. Indeed, water access could be capturing other shortages in other basic needs, such as lack of income. The results obtained serve to make policy recommendations in order to improve subjective well-being through water policies in peri-urban Kinshasa.

The rest of this thesis is organized as follows. The Section 2 provides a literature review of every aspects of water access and well-being and their relationship using the existing literature as basis. Section 3 presents the region of study, the data sources, the variables used and some descriptive statistics. In Section 4, we present the empirical strategy and research hypothesis. The results will be presented in the Section 5 while some validity checks will be performed in Section 6. In Section 7, we present our conclusions and try to give some monetary equivalence of our results, their policy implications and to discuss some limitations of our study.

2 Related Literature

2.1 The three dimensions of water insecurity

Water insecurity is becoming a growing global concern. Climate change coupled with population growth, economic changes, bad infrastructure and poor governance contribute to the deterioration of water security in terms of quantity and quality (Howard et al. (2016)). According to Wutich and Ragsdale (2008), there are 3 dimensions of water insecurity: (1) inadequate water supply, (2) unequal access to water distribution and (3) seasonal variability of water sources (such as rivers and rainwater).

As per the international health standards definition, "adequate water supply" refers to the fact that an adult needs approximately 50L of water per day to meet minimum health and hygiene requirements (Gleick (1996), WHO (2013)). Consuming less than 50L would then put anyone into the situation of water unsecured. According to the WHO (2013) report, 50L of water per day would insure basic needs such as drinking, cooking, personal washing, washing clothes and cleaning home (see figure 1). Many reasons may induce the lack of water supply: they can be high population growth rates, increasing urbanization, lack of investment, corruption and poor governance (Biswas (2007)).

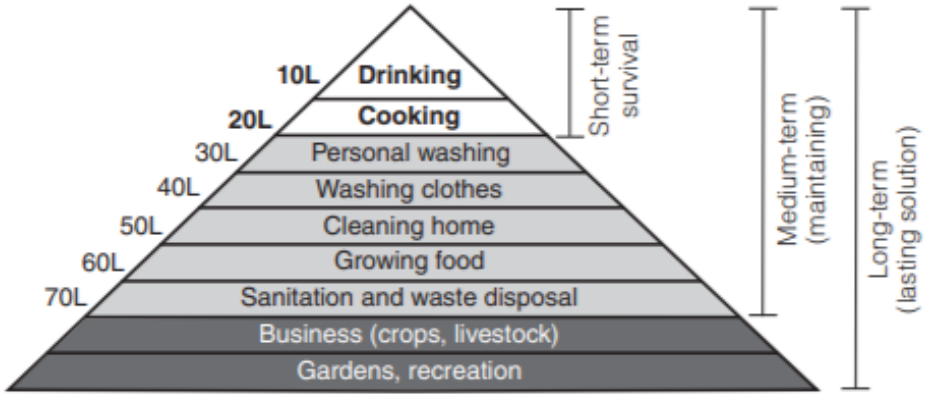


Figure 1: Hierarchy of water requirements after Maslow's hierarchy of needs (Source: WHO (2013))

The second dimension of water insecurity is the "unequal access to water distribution systems". According to Sen's capability theory and Moser's asset vulnerability theory, people do not have the same exposure to food (and water in this context) insecurity based on their initial endowment of assets. Indeed, having more ownership right and greater initial endowment ensure people to acquire sufficient access to resources compared to people with fewer assets (Sen (1982), Moser (1998)). People may all be entitled to same primary goods such as "water" but they do not have same purpose means that include rights, liberties and opportunities, income and wealth, and the social bases of self-respect to help them to promote their ends. In other terms, poor households do not have the same opportunity (or capabilities) to access water distribution system than richer households. Indeed, previous research in Villa Israel by Wutich (2007) has shown that households with greater economic assets and more access

to water vendors were significantly less vulnerable to water insecurity than households with fewer assets and less access. That also supports the idea that water quantity collected by household is endogenous with respect to the initial physical and social assets of the household.

Finally, the third dimension of water insecurity is "seasonal variability in water availability". Given the importance of climate change (and drought) in developing countries and more generally at the international level, seasonal variability of water sources (such as rivers and rainwater) need to be considered as a source of water insecurity. Indeed, availability of water plays vital economic role for agricultural and foraging communities but also for the basic needs of children in the household to make them less vulnerable to water-related diseases.

2.2 Subjective well-being notion: emotional distress and the use of the PSS-10 as psychometric measures

Much of earlier literature have relied on stated preferences, mainly, people's willingness-to-pay to improvements in water supply or to cope with unreliable water supply (Casey et al. (2006), Pattanayak et al. (2005)). They also used revealed preferences such as hedonic pricing (North and Griffin (1993), Des Rosiers et al. (1999)). One limitation of those methodologies is that they require individuals to be able to predict their future utility. However, preferences are not stable and tend to be influenced by past and day-to-day experiences so that it is better to base economic appraisal on "experienced utility" rather than "stated or revealed preference". As a consequence, we decided to use in this thesis the broad measure of subjective well-being as the welfare criterion, as it includes experienced utility and it is changing according to past and actual experiences.

In the literature, there are 3 aspects of subjective well-being: hedonic, evaluative, and eudemonic (Deci and Ryan (2008)). Hedonic measures of well-being generally focus on positive and negative emotions such as happiness, smiles, depression, anxiety, pain. They also refers to shorter periods emotions such as feelings in the past months, weeks, days, or current feelings (Kahneman and Krueger (2006), Tov and Au (2013)). Moreover, evaluative well-being is referring more to global or overall evaluation of one's life. The most common measure is the ordinal life satisfaction. (Tov and Au (2013)). Lucas and Diener (2008) found that the main difference between hedonic measures and evaluatives measures is that the former are more likely to get influenced by immediate life events while the latter is a combination of all experiences and long-term well-being. The last measure of well-being is the Eudemonic measure which goes beyond emotions and satisfaction with life and puts more concern on self-realisation and meaning of life (Deci and Ryan (2008)). Therefore, they include questions such as one best life possible, sense of self-worth, social integration, autonomy etc (Cooke et al. (2016)).

Our focus in this master thesis is mainly on the hedonic and evaluative subjective well-being driven by the data we have. On the hedonic part, we especially focus our attention on the emotional distress aspect and to that purpose we will use the Perceived Stress Scale (PSS) index created by Cohen and Williamson (1988). The PSS is the most widely used psychological instrument for measuring one's perception of stress. Items are designed to tap how unpredictable, uncontrollable, and overloaded respondents find their lives. The PSS-10 are organized around 10 questions about situations in one's life that are appraised as stressful (see figure 2). Of note, question 7 and

PERCEIVED STRESS SCALE

The questions in this scale ask you about your feelings and thoughts during the last month. **In each case, you will be asked to indicate by circling *how often* you felt or thought a certain way.**

Name _____ Date _____

Age _____ Gender (*Circle*): **M F** Other _____

1 = Never 2 = Almost Never 3 = Sometimes 4 = Fairly often 5 = Very Often

1. In the last month, how often have you been upset because of something that happened unexpectedly?	1	2	3	4	5
2. In the last month, how often have you felt that you were unable to control the important things in your life?	1	2	3	4	5
3. In the last month, how often have you felt nervous and "stressed"?	1	2	3	4	5
4. In the last month, how often have you felt confident about your ability to handle your personal problems?	1	2	3	4	5
5. In the last month, how often have you felt that things were going your way?	1	2	3	4	5
6. In the last month, how often have you found that you could not cope with all the things that you had to do?	1	2	3	4	5
7. In the last month, how often have you been able to control irritations in your life?	5	4	3	2	1
8. In the last month, how often have you felt that you were on top of things?	5	4	3	2	1
9. In the last month, how often have you been angered because of things that were outside of your control?	1	2	3	4	5
10. In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?	1	2	3	4	5

Figure 2: Perceived stress scale by Sheldon Cohen (Source: Cohen and Williamson (1988))

8 of the PSS-10 have their item reversed to compute the PSS-10 score due to the formulation of these questions which goes on the opposite direction of the others. Finally, we also use evaluative measures of unhappiness by using unhappiness scale ($unhappiness - 4$) proposed by Lyubomirsky and Lepper (1999): "all things considered, how did you consider yourself: (1) Very happy (2) Somewhat happy (3) Somewhat unhappy (4) Unhappy" and Life Satisfaction Index (LSI): "(BP4) All things considered, how satisfied are you with your life: (1) very unsatisfied ... (10) very satisfied" (Lucas and Diener (2008)). The use of evaluative measures only serves to confirm the results of hedonic measures in this master thesis.

2.3 Conceptual framework of the mechanism behind the relationship of water insecurity and subjective well-being

After defining the notion of water insecurity and subjective well-being, it is now important to establish the relationship between the two concepts. Earlier literature has identified various channels through which water insecurity

may impact subjective well-being.

The first channel is related to health, access to unsafe or inadequate water could cause diseases such as cholera, typhus, hepatitis, yellow fever or malaria (Guardiola et al. (2013)). According to WHO and UNICEF (2017), unsafe water and inadequate sanitation and hygiene causes 88% of the thousand million cases of diarrhoea that results in 1.6 million deaths a year. That would have a direct impact on subjective well-being of the sick children and their relatives. Therefore, improving access to water and water quality could help to avoid unnecessary death and suffering.

The second channel is related to economics: time spent gathering water could be devoted to activities that generate income (James et al. (2002), Larson et al. (2006)) and better access should reduce the workload of women and girls giving them more opportunities to earn more money for the household (Crow et al. (2012)). Also, access to piped water could free up more time to education for children and avoid them being sick. Indeed, if supply and sanitation services are of low quality, children fall ill and cannot go to school. According to UNDP (2006), 433 million school days are lost every year due to water and sanitation-related diseases. For adults, being sick also increases work absenteeism and the amount of money devoted to recovering from illnesses or preventing them could influence the household economy negatively (Prüss et al. (2002), Hutton et al. (2004)). In this vein, we should highlight the relationship between household income and subjective well-being but many papers have already focused on that question and found positive relationship (Boyce et al. (2010); Easterlin et al. (2010); Easterlin (2013)). However, according to Chindarkar et al. (2019), individuals' subjective well-being is not uniformly impacted by the coping costs of inadequate water access. The exact direction between access to water and subjective well-being is likely to be conditioned by the coping behaviour (some may want to spend more time collecting water than spend on water storage for instance), household budget constraint, and other household characteristics.

Finally, Bisung and Elliott (2017) defined 4 water-related categories of stressors: (1) financial stressors, (2) stressors related to physical (lack of) access, (3) social stressors, (4) stressors associated with (perceived) inequities. First, financial stressors reflect pressures associated with access to water. In the absence of reliable sources, households tend to buy water from vendors who charge higher amounts therefore this increases household budget constraint for other necessities (Ennis-McMillan (2001), Wutich and Ragsdale (2008)). Second, with regard to sanitation, insufficient water supply for hygiene resulted in psychosocial outcomes: discomfort at open defecation sites (risk of attack, risk of injury, fear of using unclean latrines etc.), frustration from long waiting times or queues in the water sources and public toilets. Third, the social stressors are problems that affect social interactions at the household and community levels: fear of rape and sexual assault when using open defecation, shame over being unable to provide visitors with toilet facilities or water, domestic abuse over water, fighting each other in water queues etc (Ennis-McMillan (2001), Wutich and Ragsdale (2008), Subbaraman et al. (2015), Bisung and Elliott (2016)). These largely affect psychological well-being and the effect is even more pronounced for women. This finding is also confirmed by the influential RCT experiment ran by Devoto et al. (2012) in Urban Morocco. Being connected to the city's water grid generates important time gains for leisure and social activities. As water is often source of tension between households, connections should improve social integration and reduce conflict. Therefore, subjective well-being improved substantially among households in the treatment group of the RCT within 6 months.

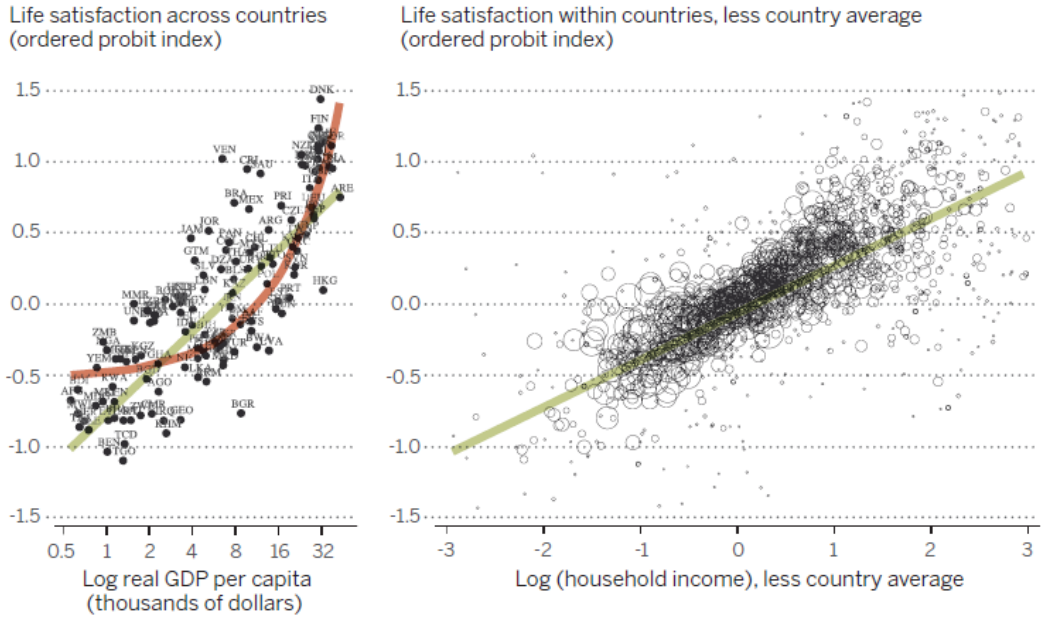
Fourth, people may also be annoyed and angry with structural inequities and geographical patterns of access (see Bolivia example: Wutich (2009)).

2.4 Water-related stress and its impact on the perpetuation of poverty

Affordability of water is essential to ensure water security for households worldwide. Target 6.1 of the Sustainable Development Goals (SDGs) aims to "ensure universal and equitable access to affordable drinking water for all", while target 6.2 focuses on access to sanitation. As discussed in the previous section, previous literature recognises that water insecurity contributes substantially to high reports of perceived stress (e.g. self-reported distress, perceived stress, anxiety/depression symptoms) and deterioration in subjective well-being: limited access to water (Brewis et al. (2019)), experiences of water insecurity (Stevenson et al. (2016)), shameful or conflictual water collection dynamics (Sultana (2011)), unpredictable and unjust water systems (Wutich and Ragsdale (2008)), and social inequality in water systems (Ennis-McMillan (2001)).

Similarly, researchers in the psychology of poverty have recognised that stress is a major factor contributing to the perpetuation of poverty. Haushofer and Fehr (2014) indicates that stress and negative affective states can lead to short-sighted and risk-averse decisions, limiting attention and favouring habitual behaviours over goal-directed ones. Indeed, stress can reduce the willingness to take risks and forego current income in favour of higher future income. In an article by Schmitz and Grillon (2012) and another by Guiso et al. (2013), it is shown that people in high threat conditions showed a significantly higher risk aversion than those in low threat conditions. In addition, stress also manifests itself in an unfavourable attitude towards the adoption of new technologies and long-term investments such as education and health. Therefore, stress (and water-related stress in our context) can foster behaviours that make it difficult to escape poverty. In the paper of Haushofer and Fehr (2014), using data from the Gallup World Poll, they show that life satisfaction (and thus positive affect) is positively associated with the logarithm of real GDP per capita (see figure 3) and the relationship is significant. Thus, measures of perceived stress can provide a valuable overall summary assessment of an individual's socio-economic, cultural and mental health conditions.

Figure 3: The relationship between life satisfaction and poverty (Source: Haushofer and Fehr (2014))



3 Experimental Design and Data

3.1 The PILAEP Project

3.1.1 Region of study

Since 2008, the Agence Française de Développement has created and funded the PILAEP Project ("Promotion de modalités Innovantes pour l'Accès à l'Eau Potable" translated into *Promotion of Innovating Solutions for Access to Drinkable Water*) after the observation that only 42% of the population in the DRC has access to drinking water and this situation is even worse in urban areas, in the capital Kinshasa and in the periphery of Kinshasa. Consequently, PILAEP focuses its intervention in the city's expanding neighbourhoods, which are experiencing strong demographic growth. This is reflected in two phenomena: the inflation of asset prices, which is pushing part of the population of Kinshasa out of the city, and the dynamics of urbanisation, which is bringing households from other rural provinces to the city of Kinshasa. In addition, conflicts in the country also push people to settle in the capital. All this leads to water insecurity for part of the population of Kinshasa, especially those in remote and poor areas.

The first part of the project, PILAEP I (2008-2014), was launched in 2008 with co-financing from AFD and the European Union and cooperation between the Belgian Development Agency (BTC) and several Congolese NGOs. This first part of the project made it possible to build autonomous drinking water solutions in 11 peri-urban neighbourhoods in Kinshasa and 4 neighbourhoods in Bas-Congo. The PILAEP 2 project (2017-2020), a continuation of PILAEP 1, is the one we will study in this thesis. This new component also aims to improve sanitation conditions and access to water for 400,000 inhabitants in 26 other districts of Kinshasa: 15 in the commune of N'Sele, 5 in Kimbanseke, 6 in Kinsenso; and 8 districts in Mont-Ngafula. The aim is to provide autonomous

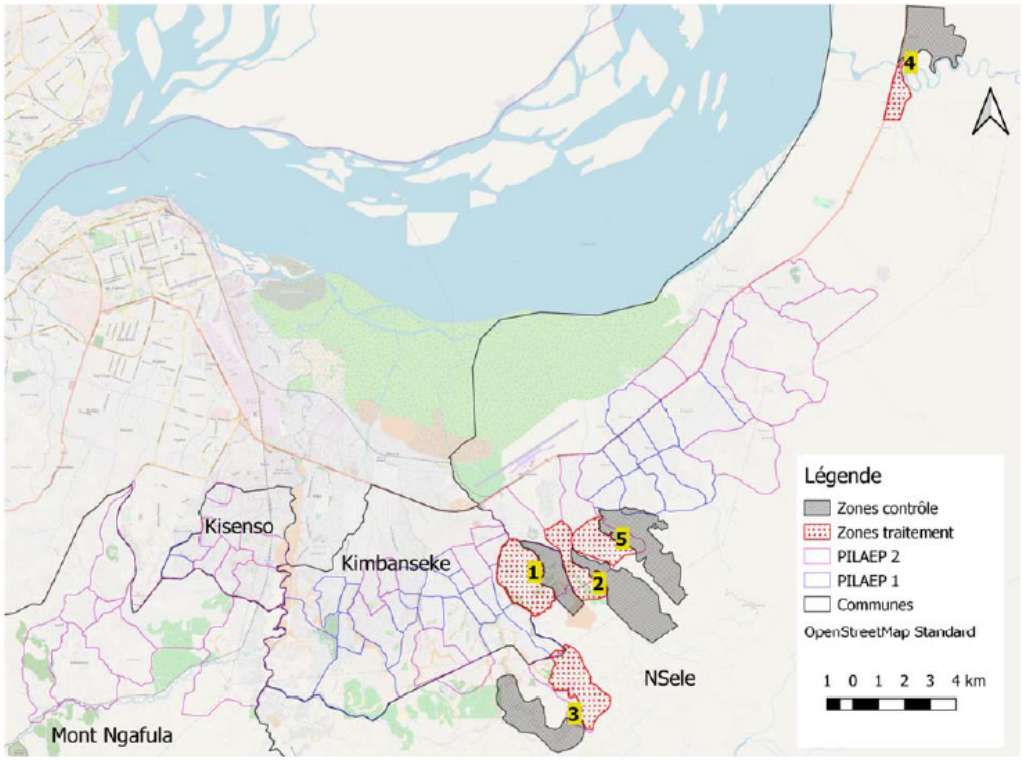
water supply networks in areas that are still not served by the national water companies (REGIDESO). The future beneficiaries of the district would benefit from a borehole, a reservoir, a distribution network, and standpipes that should theoretically supply drinking water to 1,000 inhabitants within a radius of 250 metres. In 2020, the construction of water networks has already begun in the commune of N'Sele despite the COVID-19 epidemic.

In this Master's thesis, we focus only on the 15 districts of N'Sele because these are the areas chosen by the National Statistics Agency of the DRC (INS) and the NGO "ADIR" (Action pour le Développement des Infrastructures en milieu rural) to conduct the survey. Our panel database contains the results of two waves of surveys: the baseline survey (conducted between April and May 2018) and the follow-up survey (conducted between April and May 2019). For future investigations and post-evaluation, it was important for DIAL and ADIR researchers to divide these 15 districts of N'Sele into 5 pairs of districts: 5 of them are the future beneficiaries of the new water supply networks, and 5 of them are the control areas that will be used as the comparison group needed to identify the ex-post treatment impact of the project. These 5 pairs of districts were chosen under the assumption of parallel trends, meaning that the beneficiary and control households move on parallel trends and are equivalent in the absence of the project. Table 1 shows how the areas were divided into 10 districts, and Figure 2 shows their location on the map. According to this figure, the zones extend to the east of Kinshasa, away from the centre. Apart from Kindobo and Mbenzale, which are located along a national road. And apart from Kindobo and Mbenzale, which are along a national road, the other areas are several kilometres from a tarred road. Market gardening remains an important source of income in some of these sparsely populated areas, and the vast majority of these areas are not connected to electricity.

Table 1: Treated and control areas

Pairs	Treated areas (Future Beneficiaries of the new water supply network)	Control areas (Comparison groups)
1	Badara 2	Badara 3
2	Belair (outside Palu and Baluba localities)	Belair (Palu and Baluba)
3	Buma (outside Manvululu localities)	Buma (Manvululu locality)
4	Kindobo	Mbenzale
5	Maba 2	Munke 2

Figure 4: Location of PILAEP 1 and 2 Beneficiaries (Source: Boaz et al. (2020))



It is important to note that in the context of this Master's thesis, the distinction between future beneficiaries and control areas is not mandatory since the objective of this thesis is to provide an ex-ante (as opposed to ex-post) assessment of the impact that access to water would have on subjective well-being regardless of the future treatment status of the district. However, this distinction of districts is important for the use of district fixed effects in the specification which we will explain in section 4.

3.1.2 Randomization process

In order to be representative of the Commune of N'Sele and more generally of the population of the periphery of Kinshasa, the DIAL researchers and the INS carried out a simple random sampling of all households living in the study areas. Since the sample target was 3000 households, they first wanted to draw 300 households in each of the 10 districts. As some areas are over-represented in the sample compared to others due to their size, it was not possible to draw 300 households in each district, so they adjusted the sampling rate for each district to be perfectly balanced between areas (see table 7 in the Appendix). As a result, in the baseline survey they were able to collect data from 2,998 households, representing 14,130 individuals. And during the follow-up survey, the field investigators were only able to follow up on 83.16% of the initial sample, which represents 2,493 households and 13,282 individuals. Thus, the attrition rate is estimated to be about 16.84%, but for simplicity, we assume that the sample of missing households in the follow-up survey is also drawn randomly from the existing set of households and that the remaining households are not systematically different from the missing households. However, to correct for

attrition bias, many methodological approaches can be proposed: inverse probability weighting IPW (Gerber and Green (2012)), non-parametric bounds (Angrist et al. (2002)), or the Heckman correction selection model (Thomas et al. (2012); Maluccio (2004)).

3.2 Data, variables selection and descriptive statistics

3.2.1 The EACV1 survey

To assess the sanitation and water access conditions of the population of N'Sele, the investigators conducted a survey entitled "EACV1". The questionnaire mainly uses the *quantitative* methodology and encompasses four main modules: (1) Household, (2) Health and education of children, (3) Water, (4) Individual. First, the household module includes questions on household living conditions, assets, socio-demographic characteristics, sanitation. Second, the child health and education module includes questions on children's school activities and health and hygiene status. Thirdly, the water module is one of the most important parts of the survey as it collects all the information on water consumption and collection habits, water treatment, storage and finally water consumption within the household. Finally, the individual module gives us the individual characteristics of the respondents such as their personal feelings indicated by their stated subjective well-being but also information on their education, professional and secondary activities, their time use during the week, their health status, and finally some questions on social capital and their perception of governance. However, unlike the baseline survey, the follow-up survey does not cover the questions on social capital, perception of governance and some other questions included in the other modules of the baseline survey. Therefore, in order to frame our panel data set with this pattern in mind, we decided to use the follow-up survey as a base and to extract only those common questions that are present in both the baseline and follow-up surveys. Thus, our final dataset contains the answers to about 200 questions if we do not count individual questions on name, age, gender and home address.

3.2.2 Rainfall data

To link the rainfall data to the corresponding household, the surveyors used the CHIRPS data (rainfall estimates) of the area where the household lives after asking for its geographical coordinates. The reference date for the rainfall estimates was the exact day when the water part of the module was asked (which is called module *ETE* in the survey) and they also extract the rainfall 15 days before this reference day. In this master thesis, we decide to use the value of the positive shock of precipitation during the week (7-days) prior the survey as our main instrument. This variable is truncated as it is only positive if the household's district is exposed to a positive rainfall shock during the week preceding the survey compared to the average rainfall of all districts during the whole survey period and takes the value zero otherwise:

$$Z_{it} = \begin{cases} 0 \text{ when } M_{idt} < \bar{M}_d \\ M_{idt} - \bar{M}_d \text{ when } M_{idt} \geq \bar{M}_d \end{cases}$$

M_{idt} is the average amount of rainfall in the household's i district d during the 7 days prior the survey t. $\bar{M}_d$ is the average quantity of rainfall in all districts during the whole survey period.

3.2.3 Missing values treatment

Missing data is a significant problem found in field research and data analysis. Missing values may lead to lack of information and inconsistent knowledge discovery. On the dataset we have, missing values represent 189,501 cells which represent 18.61% of the total dataset. On the missing data literature, there are many ways to deal with unobserved values but first we need to identify the mechanism that causes the absence (Rubin (1976), Rubin (1996)). Rubin identified 3 mechanisms that generate missing data: **MCAR** (missing completely at random), **MAR** (missing at random), and **NMAR** (not missing at random). The third mechanism exhibits that the probability of the absence for one variable is related to the value of other observed values or unobserved values. Several tests have been proposed to test MCAR versus MAR but these tests are not widely used due to their unclear practical results (see Enders (2010) for an evaluation with two stages). According to Van Buuren (2018), it is not possible to test MAR versus NMAR since the information that is needed for such test is missing. Indeed, we don't observe/measure the other missing values that drives the initial absence. Consequently, one needs to assume that unobserved values are at least MAR so that multiple imputation can be performed.

In this master thesis, after assuming that our unobserved values are MAR, we use the predictive mean approach that is embedded inside the R-package Hmisc: *transcan()* and *aregimpute()* to impute missing data. [1] Nevertheless, we did not apply this multiple imputation methods on our future dependent variables $PSS - score$ and *unhappiness*-4 to avoid the potential misinterpretation of the results although missing values just represent less than 3% for $PSS - score$, for *unhappiness*-4 and for LSI (see the last column of table 2).

3.2.4 Individual id problem

Continuing with the problems encountered in data management, we also note that several individual identifiers change between the baseline and the follow-up. Of the initial sample of 14,130 individuals who are interviewed in the baseline, only 12,129 individual identifiers appear in the follow-up survey because the interviewers were not able to track them all. However, we had a problem with these 12,129 individual identifiers: we found that 11,773 individuals had exactly the same name, identifier, gender and age in both waves of the survey, while 288 of them had names that had changed between the two waves and **68** of them had the same name and identifier but their gender and/or age had changed. Going further, I tried to understand whether the 288 individuals who had different names in the baseline and follow-up surveys are ultimately the same, but whether they were simply subject to a misspelling of their name. In doing so, I found from the subsample of 288 individuals that: (1) 175 names are "quasi-equivalent" (but **10** names of them have their age inconsistent and **6** names have their gender changed), (2) 39 names are "somewhat equivalent" (but we are still not completely convinced that **22** of them are the same individuals), (3) **74** names are "radically different" making it difficult to state unambiguously that they are the same person in the baseline and in the follow-up. In summary, we have 180 "doubtful" individuals (i.e. individuals for whom it is not safe to say that they are the same during the baseline and the survey although they have the

[1] These function use additive regression (to predict all of the original missing and non-missing values), bootstrapping (to approximate the process of drawing predicted values from a full Bayesian predictive distribution) and predictive matching (to impute missing values back into the original data).

same individual identifier):

- 68 have the same name but either their gender or/and age changed.
- 10 individuals have an inconsistent age although their identity and gender have not changed and their names are almost identical over the two waves of the survey.
- 6 people with almost the same name have their gender changed.
- 22 individuals are not sure if they are the same in both waves of the survey, although their identity, age and gender are the same and their names are somewhat equivalent.
- 74 individuals experience a radical change in their name during the two waves.

Therefore, in order to avoid misinterpretation of the results and arbitrary changes of name or sex, we decided to eliminate these 180 doubtful individuals from the sample of 14,130 baseline and 12,129 follow-up individuals, as they represent less than 2% of the total sample. Full details can be found in Figure 7 in the Appendix.

3.2.5 Final sample and descriptive statistics

To create the final panel sample, recall from Section 3.1.2 that 2,493 households are interviewed in the survey follow-up compared to 2,998 in the baseline, representing an attrition rate of 16.84%. In addition, unlike the baseline survey, the follow-up survey asks questions on SWB only of the member of the family who answers the water module questions, most often the head of the household or his/her spouse, and not of all members. Therefore, we only have data on SWB for the 2,493 individuals in each household who are interviewed for the water module in the follow-up survey. In order to obtain balanced panel data, we decide to extract the same 2,493 individuals from the baseline, but we need to eliminate what we called "doubtful individuals" in the previous section with respect to their name, individual identifier, or gender and age. Thus, after removing the 148 individuals considered doubtful from the dataset, we are left with 2,345 individuals in the baseline survey and 2,345 individuals in the follow-up survey, each from different households. We called the corresponding panel "PANEL I" (see figure 5).

Figure 5: Balanced panel "PANEL I" (id consistent) (Source: Author's data handling)



In addition, there is also a major concern regarding rainfall data. The reference date for the rainfall estimates was the exact day on which the water part of the module was asked (which is called *ETE* module in the survey) and they also extract rainfall 15 days before this reference day. However, it seems that in some households they used the interview date of the other modules as the reference date rather than the survey date of the ETE module (due to hardware errors of the tablet used in the field). Obviously, this leads to the association of an inconsistent rain date with certain households, whereas we are only interested in knowing the collection behaviour during rainy periods. This date mismatch concerns 14% of the 2,998 households in the baseline and 54% of the 2,493 households in the follow-up, which corresponds to 2,539 usable consistent observations in the baseline and 1,098 in the follow-up. For simplicity, we assume that the inconsistency of rainfall names and reference date follows a random distribution in order to avoid bias in the interpretation of the results, although a detailed analysis of rainfall data and reference date inconsistency is provided in other support document. As our main objective remains to have balanced panel data, we finally retrieve the 898 households that meet the following two criteria: (1) a consistent rainfall reference date, (2) a consistent name/sex/age or individual identifier. Thus, when we do analyses involving rainfall estimates in the next section of this master thesis, we will use this restricted panel which we have called "PANEL II". (see figure 6).

Figure 6: Balanced panel "PANEL II" (id & rainfall data consistent)



Having explained all the aspects and challenges of data processing, we can finally display the summary statistics table (see table 2). This table has two sub-sections: Wave 1 displays the summary statistics for the baseline survey and Wave 2 for the follow-up survey. In columns 2 to 4, there is a comparison between the means of the male and female samples and the t-test for comparison of means is described in column 5. In the last column, we show the proportion (in %) of missing observations before any imputation of NA is made.

Before commenting, we should remind once again that our panel called "PANEL I" is composed exclusively of the household member who is primarily responsible for water-related activities in each household, since he or she is the one who answers the questions on the water module and the subjective well-being module in the follow-up survey. Thus, women represent 64% of them, that is more than the share of women in the parent dataset[2] which is 51%. The large number of women (1503) compared to men (842) is logical since in households, women are the

[2]The parent dataset is the raw dataset that has not received any processing and includes 14,130 individuals in the baseline survey and 13,282 individuals in the follow-up survey.

ones who either carry out water collection directly or are responsible for and knowledgeable about water-related activities: 73% of the women surveyed were involved in water collection in the home in Wave 1, compared to only 33% of the men. For Wave 2, the share of women in charge of water collection is 84% while men represent only 47%. In terms of age, the women respondents (37 years) are significantly younger than the men (42 years), which also suggests that women are more involved in water-related activities despite being younger than their male counterparts. It is interesting to note that our panel respondents are older (on average 38 years old) compared to the parent database (21.4 years) (see Boaz et al. (2020))). This is due to the fact that our panel is mainly composed of 90% by the head of the household or spouse, as they are the main ones responsible for water-related activities in the household and therefore the ones who answer the water module in the survey.

As far as economic status is concerned, men are clearly more active in the labour market than women: in Wave 1, 44% of women declare that they were not in employment in the week preceding the survey, whereas this proportion is only 19% for men. In Wave 2, the situation for men worsens, with 50% of them declaring that they did not have a job in the week preceding the survey, but the proportion remains higher for women (62%). Similarly, in Wave 1 (Wave 2), 49% (53%) of women also report not having a source of pay or salary in the previous week, compared to only 29% (46%) of men. Also, men are significantly more independent of family support for their livelihoods: in Wave 1, 76% of men report being financially independent of their family (compared to 57% for women); in Wave 2, 42% for men and 28% for women.

Concerning the assets and dwelling characteristics, half of respondents are living in precarious dwelling (housing built from recycled materials: sheet metal, boards, etc.). About their financial conditions, 87% of respondents self-declare themselves not being able to save. Also, only 12% of households are connected to the grid for electricity.

Now, if we look at the PSS score, we have the same score of about 27 for both men and women and the two means are not statistically different from each other. This score has a high magnitude and indicates that men and women are less able to exercise control over their life circumstances. This is also reflected in their Unhappiness score. While the average unhappiness-4 score is 2, indicating that they are rather unhappy, both men and women surveyed have a score around 2.5 in our panel which means that they are unhappier than the average and women are significantly more unhappy than men, especially in the follow-up survey. Similarly, the Life Satisfaction Index (LSI) also indicates that men are significantly more satisfied with their life than women. Nonetheless both LSI are averaging 5 which translates that they are both quite unsatisfied with their life situation.

In terms of the amount of water collected in the week prior to the survey, they were able to collect an average of 579L of water during the baseline period and 729L of water during the follow-up. From our sample calculations, we know that families have an average of 4-5 members, and if we assume that there is no inequality in the water collected, this means that each member has access to an average of 132L per week in the baseline period and 160L per week in the follow-up survey, which is less than 50L per day per individual. It is clear that these quantities don't only allow each individual in the family to meet basic needs. (see figure 1).

Finally, for the rainfall measures, we decided to use the proportion of rainfall that is above the average

rainfall of all districts during the whole survey period (in millimetres), which we can also call positive rainfall shock (in millimetres). This variable is truncated because it is only positive if the household's district is exposed to a positive rainfall shock during the week preceding the survey in relation to all districts rainfall average during the whole survey period. Therefore, there is no point in making any interpretation of the average positive rainfall shock that appears in the summary statistics table because the value of zero pulls the average down.

Note that we comment on the descriptive statistics of the 'Panel I' sample but it is important to note that there is no significant difference between the means of 'Panel I' and 'Panel II'. We present these in the appendix 8. This confirms once again that the inconsistency of the precipitation measurements which leads us to use Panel II is random.

Table 2: Summary statistics of "PANEL I"

Wave 1: Baseline survey summary statistics

Variables	Male	Female	Full Sample	Two.sample.t.test.p.value	Prop_of_NAs...
Dependent Variables					
PSS-score	26.886 (4.997)	27.165 (5.04)	27.065 (5)	0.204	2.77
Unhappiness-4	2.53 (0.752)	2.556 (0.792)	2.547 (1)	0.428	2.77
LSI	5.314 (2.07)	5.166 (2.044)	5.219 (2)	0.095	2.77
Independent Variables					
Quantity of Water Collected (liters)	502.69 (365.06)	622.02 (395.39)	579.17 (389)	0	0
Mean household size	- (-)	- (-)	4.96 (3)	-	0
Mean quantity of water collected per household member (liters)	- (-)	- (-)	132.1 (93.32)	-	0
Socioeconomic characteristics					
Proportion of respondents who directly carry water collection	0.355 (0.479)	0.737 (0.441)	0.6 (0)	0	0
Proportion of respondents responsible for water activities	0.337 (0.473)	0.806 (0.395)	0.638 (0)	0	0
Age	42.063 (14.439)	37.224 (12.987)	38.962 (14)	0	0
Proportion of respondents without any professional activities during the past week	0.195 (0.396)	0.446 (0.497)	0.356 (0)	0	1.49
Proportion of respondents without any of wages or salary during the past week	0.303 (0.46)	0.487 (0.5)	0.421 (0)	0	56.38
Proportion of respondents independant from their families for susbsistence	0.732 (0.443)	0.5 (0.5)	0.583 (0)	0	68.4
Proportion of respondents with precarious housing in the baseline survey	0.538 (0.499)	0.486 (0.5)	0.505 (1)	0.016	0
Monthly rent during the baseline survey	16419 (30758)	18540 (33196)	17778 (32351)	0.128	69.55
Proportion of respondents that are unable to save	0.868 (0.339)	0.87 (0.336)	0.87 (0)	0.886	0
Proportion of respondents with access to electricity	0.095 (0.293)	0.134 (0.34)	0.12 (0)	0.006	0
Proportion of respondents who are subject to chronic disease	0.064 (0.245)	0.073 (0.261)	0.07 (0)	0.41	1.49
Rainfall measures					
Rainfall proportion beyond the district mean (mm in the last 7 days)	- (-)	- (-)	2.839 (2)	-	0
Respondents status in the family					
Proportion of household head (%)	-	-	49.42	-	-
Proportion of the spouse of the head (%)	-	-	43.88	-	-
Proportion of the children of the head or of the spouse(%)	-	-	5.12	-	-
Head or spouse's grandchild/ parent(s)/non-related member (%)	-	-	1.58	-	-
N (obs)	842	1503	2345	-	-

Wave 2: Follow-up survey summary statistics

Variables	Male	Female	Full Sample	Two.sample.t.test.p.value	Prop_of_NAs...
Dependent Variables					
PSS-score	27.53 (5.172)	27.025 (5.144)	27.206 (5)	0.023	0.0426
Unhappiness-4	2.428 (0.733)	2.536 (0.767)	2.497 (1)	0.001	0.0426
LSI	5.423 (1.841)	5.202 (2.08)	5.281 (2)	0.01	0.0426
Independent Variables					
Quantity of Water Collected (liters)	711.47 (1247.09)	740.05 (1124.52)	729.79 (1170)	0.571	0
Mean household size	- (-)	- (-)	5.37 (3)	-	0
Mean quantity of water collected per household member (liters)	- (-)	- (-)	160.65 (367.94)	-	0
Socioeconomic characteristics					
Proportion of respondents who directly carry water collection	0.474 (0.5)	0.846 (0.361)	0.713 (0)	0	0
Proportion of respondents responsible for water activities	0.404 (0.491)	0.916 (0.278)	0.732 (0)	0	0
Age	42.077 (14.433)	37.23 (12.986)	38.971 (14)	0	0
Proportion of respondents without any professional activities during the past week	0.507 (0.5)	0.623 (0.485)	0.581 (0)	0	0
Proportion of respondents without any of wages or salary during the past week	0.432 (0.496)	0.52 (0.5)	0.488 (0)	0	41.83
Proportion of respondents independant from their families for susbsistence	0.359 (0.48)	0.22 (0.414)	0.27 (0)	0	54.16
Proportion of respondents with precarious housing baseline survey	0.538 (0.499)	0.486 (0.5)	0.505 (1)	0.016	0
Monthly rent during the baseline survey	16251 (30679)	17600 (30606)	17116 (30633)	0.306	69.55
Proportion of respondents that are unable to save	0.868 (0.339)	0.87 (0.336)	0.87 (0)	0.886	0
Proportion of respondents with access to electricity	0.095 (0.293)	0.134 (0.34)	0.12 (0)	0.006	0
Proportion of respondents who are subject to chronic disease	0.063 (0.243)	0.073 (0.259)	0.069 (0)	0.381	1.49
Rainfall measures					
Rainfall proportion beyond the district mean (mm in the last 7 days)	- (-)	- (-)	0.652 (1)	-	0
Respondents status in the family					
Proportion of household head (%)	-	-	49.42	-	-
Proportion of the spouse of the head (%)	-	-	43.88	-	-
Proportion of the children of the head or of the spouse(%)	-	-	5.12	-	-
Head or spouse's grandchild/parent(s)/non-related member (%)	-	-	1.58	-	-
N (obs)	842	1503	2345	-	-

4 Empirical Strategy

First, to test the relationship between water insecurity and subjective well-being, we preliminary use (naive) Pooled OLS estimates represented in the following equation (0):

$$(0)\ Y_{ijt} = \beta_0 + \beta_1 D_{ijt} + X_{it}^{'} + u_{ijt}$$
$$u_{ijt} = c_i + d_j + \epsilon_{i,j,t}$$

The index i denotes the respondents id, j denotes the district, and t denotes whether we are in baseline or in the follow-up survey. The sample for equation (0) is the panel we called "PANEL I" of 2,345 households in which we have one respondent per household in the two waves survey. The dependent variable, Y_{ijt}, is the PSS-score of the individual i that belongs to district j in the survey wave t. D_{ijt} is the independent variable of interest, the quantity of water in litres collected by respondent i who belongs to district j during survey wave t.

$X_{it}^{'}$ is a set of covariates that we motivated above and which we will discuss again when we present the results. We include socio-economic variables, as the literature shows that they are spuriously correlated with subjective well-being and also with the amount of water collected, which may bias our fixed effect and 2SLS estimates. (Russell et al. (2020), Karabchuk and Soboleva (2020), Chindarkar et al. (2019), Diener and Tay (2015), Steptoe et al. (2015)). These control variables are economic status in the labor market (working or not in the past week), whether they receive any other wages from other non-professional activities, financial dependency status from families, whether the respondents are the one who directly collect the water or not, whether the respondents are the one primarily responsible of water activities in the household or not, the housing characteristics (availaible only for the baseline survey), whether the household has access to electricity or not (available only for the baseline survey), the monthly rents (available only for the baseline survey), and health status (whether respondents are subject to chronic disease or not). We also introduce interacted terms between the dummy variables *unemployed* and *no_wages*, same as with *unemployed* and *financially_independent* because being unemployed is even more psychologically harmful when one has no other source of wages or cannot depend on families' private support.

As we cannot guarantee that the random effect assumption is not violated, i.e $cov(D_{it}, c_i)$, $cov(D_{it}, \epsilon_{ijt})$, $cov(X'_{it}, c_i)$, $cov(X'_{it}, \epsilon_{ijt})$ are unlikely to be equal to 0, the pooled OLS estimators in equation (0) are likely to be biased. In other words, observable components are likely to be correlated with time-invariant (c_i) and time-varying (ϵ_{ijt}) unobserved components in such a way as to violate the conditional zero mean assumption. To see this, we can for example take the example of an unobserved time-invariant variable c_i which is the education of the head or the personal character. This variable is correlated with the amount of water collected and affects subjective well-being but is unknown. Consequently, we prefer to use fixed effects estimates instead to measure only changes within households over time, including dummy variables to control for any missing or unknown time-invariant characteristics. Equation (1) below is the fixed-effects estimates of the relationship between water insecurity and subjective well-being:

$$Y_{ijt} = \beta_0 + \beta_1 D_{ijt} + X_{it}^{'} + \delta_i + \Gamma_t + \Psi_j + u_{ijt} \quad (1)$$

δ_i, Γ_t, Ψ_j are respectively individual fixed effects, time fixed effects and district fixed effects that allow us to capture any time invariant, time specific and district specific components in our fixed effect regression.

One challenge for estimating the causal effect of water insecurity (represented by water collected in the last week in our context) on SWB is the endogeneity of the location where the household decides to settle in the study area. Indeed, the choice of location may be influenced by other socio-economic factors which, in turn, may have an effect on SWB.

As we know that people choose for themselves the areas where they want to live, households may decide to settle near water sources. Self-location behaviours are likely to be determined by inherent factors that may not appear in the collected data but that explain part of the variation in subjective well-being. To this end, our identification strategy is to use *Two stage least squares (2SLS)* estimator by instrumenting the amount of water collected in litres by a positive rainfall shock. It has been shown by several papers and notably the report of DIAL researchers in 2020 ((Boaz et al. (2020))) that inhabitants in peri-urban Kinshasa use rainwater to supplement the supply of drinking water, as is done in several regions of Sub-Saharian Africa and in Southern Asia.
Therefore, we consider a simple case where we use rainfall positive shock (linear) as an instrument for water collected:

$$Y_{ijt} = \beta_0 + \beta_1 D_{ijt} + X_{it}^{'} + \delta_i + \Gamma_t + \Psi_j + v_{ijt} \tag{2}$$

$$D_{ijt} = \alpha_0 + \alpha_1 Z_{it} + X_{it}^{'} + \delta_i + \Gamma_t + \Psi_j + \epsilon_{ijt} \tag{3}$$

Equation (2) is the second stage of our 2SLS system and equation (3) is the first stage. The sample we analyse in these regressions is the so-called "PANEL II" panel of 898 respondents observed in the two waves of the survey and each of them comes from a different household.
The coefficient of interest, β_1 is the estimated effect of an additional liter of water collected on the PSS score. A negative coefficient, $\beta_1 < 0$ indicates that, on average, an increase in the provision of water decrease the respondent's level of stress.

To flexibly control for non-linear effects of rainfall shock, we also introduce second-order polynomial for positive rainfall shock to see whether the impacts of rainfall on water harvested is concave or convex rather than fully linear:

$$Y_{ijt} = \beta_0 + \beta_1 D_{ijt} + X_{it}^{'} + \delta_i + \Gamma_t + \Psi_j + v_{ijt} \tag{4}$$

$$D_{ijt} = \alpha_0 + \alpha_1 Z_{it} + \alpha_2 Z_{it}^2 + X_{it}^{'} + \delta_i + \Gamma_t + \Psi_j + \epsilon_{ijt} \tag{5}$$

Equation (4) is the second stage of our 2SLS system with polynomial specification and equation (5) is the first stage. Conceptually, causal inference requires the assumption of the exclusion restriction of the instrument, i.e. that the positive precipitation shock influences subjective well-being only through the water collected during rainy periods (conditional on the covariates we choose). A natural concern is that there may be other channels through which the rainfall shock can impact on subjective well-being. On the one hand, for example, people may feel overwhelmed when there is a rainstorm in their neighbourhood: they cannot go to work, they get cold, etc. This has a negative impact on subjective well-being. And on the other hand, it is also possible that people simply like the rain without any specific reason and that influences their affective states when they answer the survey. This

represents a limitation of our study, but we can assume that introducing an individual fixed effect that captures all time-invariant components of individuals (such as personal feelings, activity domain, etc.) would help us to partially get rid of these confounders in the 2SLS estimates. In addition, we can also capture the presence of this impact by considering reduced-form estimates, which compare the difference in subjective well-being after a rainfall event. The reduced-form estimates take the instrument directly as an independent variable:

$$Y_{ijt} = \eta_0 + \eta Z_{it} + X_{it}^{'} + \delta_i + \Gamma_t + \Psi_j + v_{ijt} \tag{6}$$

$$Y_{ijt} = \eta_0 + \eta_1 Z_{it} + \eta_2 Z_{it}^2 + X_{it}^{'} + \delta_i + \Gamma_t + \Psi_j + v_{ijt} \tag{7}$$

Equation (6) is the reduced form estimates with linear specification of instrument which is now assumed to be the independent variable and equation (7) is for polynomial specification of the instrument.

Finally, we will also examine the effect found with the 2SLS estimates on some restricted samples, namely the sample of the head of household or spouse only, the sample of individuals who are neither head of household nor spouse, the male sample, the female sample, which will be referred to as equations (8), (9), (10), (11) in the next section. Similarly, we also need to include household size in the analysis because the water collected per household member depends on the size of the family. To do this, we introduce equation (12) and equation (13) which use respectively water collected per household member and water collected interacting with family size as independent variables:

$$(12)\ Y_{ijt} = \beta_0 + \beta_1 \frac{\hat{D_{ijt}}}{hhsize_{it}} + X_{it}^{'} + \delta_i + \Gamma_t + \Psi_j + v_{ijt}$$

$$(13)\ Y_{ijt} = \beta_0 + \beta_1 D_{ijt} \hat{\times} hhsize_{it} + X_{it}^{'} + \delta_i + \Gamma_t + \Psi_j + v_{ijt}$$

5 Results

Before presenting the results, it should be noted that the full regression for each specification is also displayed in the appendix. They are respectively represented by the table (9), (10), (11), (12) (14), (15), (16) for equation (0), (1), (2), (3), (4), (5), (6) , (7).

5.1 Pooled OLS estimates and Fixed-effects regression

We begin the analysis by presenting first the pooled OLS estimates of equation (0) and then the fixed effects estimates of equation (1), which are presented in the first and second rows of the table 3. Note that EC6b represents the amount of water collected in litres in the week prior to the survey.

Column (1) reports the correlation estimates between the amount of water collected in litres and the PSS score for a specification without covariate for equation (0) and includes only the individual FE, the district FE and the survey year FE for equation (1). This is because time-invariant individual characteristics, such as personal feelings and personality, are expected to influence emotional distress. This is also the case for time-invariant characteristics of districts such as economic activities, age category of inhabitants, demographic patterns, etc., which may also affect stress levels. And finally, there are systematic differences between the two observed time

units of the survey that can also have an impact on well-being: one of them can be, for example, the presidential election at the end of 2018 in DRC. Presidential elections often cause general enthusiasm throughout the country and have an impact on the reported level of stress.

Column (2) adds age covariates, a dummy variable "being responsible for water harvesting" and "being responsible for water-related activities in the household". It is generally accepted that age is an important determinant of subjective well-being. In the literature, age is often associated with a deterioration in health leading to a decline in mood and quality of life. (Tomini et al. (2016), Marengoni et al. (2011)). It has also been shown that being primarily in charge of or responsible for fetching water is a factor that can explain who is most vulnerable to the emotional and psychological stress of water insecurity. (Adams et al. (2020), Geere et al. (2018), Mushavi et al. (2020)). For women, the distress and depression related to water insecurity is explained partly by the burden of carrying water and partly by their fear of not being able to fulfil the domestic roles (e.g. washing, cleaning and cooking) and agricultural and other professional activities of men.

Finally, column (3) introduces the socio-economic control variables which include labour market status, health status, asset type and financial situation, as it has also been shown in the literature that socio-economic factors, somatic health and the quality of social relationships and social integration are important determinants of subjective well-being. (George (2010)). Unfortunately, it is difficult to capture the value of the individual's social relationships from the data available to us because questions on social capital and community life are not asked in the follow-up survey.

The estimate is close to zero and insignificant for all three columns, indicating that we potentially have no relationship between collected water and emotional distress when we examine only the pooled OLS and fixed effects estimates.

5.2 First-Stage and Reduced-Form estimates

By observing the sign of the pooled OLS and the fixed effects regression point estimates above which is negative (which is intended), we suspect that the reason for the non-significance is that the amount of water collected by the household is endogenous to several unobserved socio-economic characteristics. To deal with this problem, we use two-stage regressions taking the positive rainfall shock as an instrument: the positive rainfall shock is truncated as it takes the value of the difference between the average amount of rainfall in the neighborhood of the respondent (in millimetres) during the week before the survey and the average rainfall of all districts during the whole survey period if the difference is positive, otherwise the positive rainfall shock takes the value zero if the difference is negative. The first-stage and reduced-form estimates are presented in equation (3) and equation (6) when we introduce a linear form of the rainfall shock as an instrument or in equation (5) and equation (7) when we take a second-order polynomial form of the rainfall shock instead.

The first concern is that the weather conditions that affect farmers in the study area may be correlated with the source of variation in their income, which in turn may affect the amount of water collected. Therefore, in the first step, we should have added control variables on the sector of activity in which the respondent works

and identify those that are directly affected by weather conditions such as agriculture, livestock, forestry activities, etc. For example, a drought event results in the inability to meet crop water demand, which affects the household's agricultural income and therefore the ability to collect or purchase water in the future. From the data in hand, it was impossible to identify precisely the nature of principal and secondary occupations and whether it is impacted by rainfall or not. To avoid such ambiguity, we prefer to add control variables on labour market status by asking the respondent whether he or she has done any paid work or activity in the past week and whether one of these activities is agriculture. Furthermore, as the follow-up survey is only conducted one year after the baseline survey, we assume that the respondent does not change the sector of activity drastically in one year. Therefore, the introduction of individual fixed effects should control for the type of activity of the respondent, as it is invariant over time.

A second concern is that weather conditions may also be correlated with health status and chronic diseases, which in itself will affect the ability to transport water from water sources to the home. Therefore, we include the health status of individuals in the first stage regression and the reduced form regression.

Finally, to control for other potential confounders that are correlated with the rainfall shock and meanwhile influence the amount of water collected, we include district age, fixed effects, year fixed effects, socio-economic characteristics, asset type. Finally, we end up using all covariates in the fixed effect regression in this first stage and the reduced form estimates.

In doing so, our results for equation (3) indicate that introducing a positive precipitation shock in a linear form has no impact on the amount of water collected: the point estimates are positive (which is intended) but are not significant from column (1) to column (3) of the specification. However, when we use a second-order polynomial form of the positive precipitation shock, we obtain a positive and significant coefficient associated with the positive precipitation shock variable and a negative and significant coefficient associated with the positive precipitation shock squared value. Also, the F-statistics for the excluded instrument is between 3.1 and 3.3. Thus, it is very unlikely that our estimates are biased by weak instruments. This result indicates that the impact of a positive precipitation shock event has an inverted U-shaped effect on the amount of water collected, which is consistent. Rainfall shock first increases amount of water collected as some sources becomes available (rainwater, river). However, there is an inflection point in the amount of rainfall beyond which it becomes difficult to collect more water. In particular, an increase of one millimetre in the difference between the average rainfall in the neighbourhood for the past week and the average rainfall in all districts during the whole survey period leads to an increase in collected water of about 102 to 106 litres. However, once the positive difference in rainfall reaches about 3 millimetres and holding all other covariates constant, the amount of water collected starts to decrease. All these impacts are significant at the 10% level.

With regards to reduced-form results, our results in equation (6) and equation (7) show no significant coefficients for the three specifications from column (1) to column (3). This suggests that there is no externality in the impact of rain on the SWB apart from its impact (if any) on the amount of water collected. This also suggests that the exclusion restriction assumption of our instrument is not violated. Note that a coefficient associated with rainfall shock is negative and significant for equation 7 in column (1), however when we include the other covariates in column (2) and column (3), the significance level disappears.

5.3 2SLS Estimates

The lines associated with equation (4) and equation (6) report the 2SLS estimates for the linear form of the instrument and the second-order polynomial form respectively. The results indicate a significant negative relationship between the amount of water collected and the stress level for the second-order polynomial form of the instrument and no significant relationship for the linear form. In detail, we have that an increase of 1L in the amount of water collected during the week preceding the survey would lead to a decrease in the PSS-score of 0.004 points, the other variables remaining constant. In other words, being able to fill five 220 litre barrels in one week would decrease the PSS-score by 4.4 points. Our results are closely related to a mixed-effects tobit regression conducted by Stoler et al. (2020) in a panel of 20 low- and middle-income countries in 2017 and 2018. In the three specifications they used, they found that higher water insecurity scores[3] were significantly associated with higher PSS-scores (their point estimates for the three mixed-effects tobit regression models were 0.06 and are significantly different from 0 at 1% level).

After that, we want to understand in which sub-sample the impact of the collected water is most pronounced. Therefore, we look at the impact on the subsample of the head or spouse who responds to the survey and compare it to the subsample of any other member who is not the head or spouse but who appears to respond to the survey. We do this in equation (8) and equation (9) and our results suggest that the head or the head's spouse is the one who feels most upset/angry about water problems in the family. The impact of one litre of water collected significantly decreases the PSS score by -0.04 points, as in the full sample, ceteris paribus, while the impact is not significant in the sub-sample of non-heads/spouses of heads.

We perform the same analysis for gender in equation (10) and equation (11) to see if our results confirm previous findings that women are more likely to be most affected by water-related emotional distress. The water literature shows that due to their disproportionate responsibilities, women are more likely than men to incur costs related to water acquisition, including lost time, lost income and health problems. (Ray (2007), Cooper-Vince et al. (2018), Tsai et al. (2016), Graham et al. (2016)). It is interesting to note that, in our sample, the change in the perceived stress score is mainly due to the male sub-sample rather than the female sub-sample. One explanation for this counter-intuitive result is the composition of our sample. Indeed, we have to keep in mind that our sample is mainly composed of the head of the household or the spouse at 90% because they are responsible for water-related activities in the household and, of course, they may also have responsibilities in other roles within the household. This is confirmed by our baseline statistics which indicate that men are statistically more employed than women in the sample and that they are mainly responsible for covering the needs of their families (half of women depends on family's help for their subsistence and only 30% of men). This means that our results may reflect other shortages than lack of access to water for which men are more affected or aware than women (e.g. food insecurity issues,

[3]The water insecurity score used by the authors here is the number of times in the previous four weeks that the household experienced problems *water availability* (supply interrupted, no water availability at all), *quantity* (not enough to wash clothes, having to change foods eaten, not having as much to drink as liked, going to sleep thirsty), *hygiene* (inadequate water for bathing, inadequate water for handwashing) and *psychosocial dimensions.* (worrying about having enough water, having one's day interrupted because of water problems, feeling upset/angry about the water situation) to finally obtain a score ranging between 0-33 where higher scores indicate greater water insecurity

household economy, health, etc.). Unfortunately, we were not able to disentangle all these domains of life concern from the data at hand, but we hope that the control variables used should partially capture these confounding effects on stress levels. Also, one paper published by Social and Science Medecine and written by Mushavi et al. (2020) supports this result and indicate that in Uganda, water insecurity was associated with depression symptom severity and that the estimated association was larger among men than among women. According to them, one potential explanation of this discrepant finding is that women were better able to draw on social and emotional support, thereby more effectively buffering themselves against the deleterious effects of water insecurity. Indeed, sociological literature shows that women are more able to mobilize social support to cope with psychosocial stressors (Barnett et al. (1987), Kawachi and Berkman (2001)).

Table 3: The effect of quantity of water collected in liters on the PSS-Score

			Dependent variables			
			PSS-Score	PSS-score	PSS-score	
Sample	Baseline specification	Independent variables	(1)	(2)	(3)	Observations
Panel I	Equation (0). Pooled OLS	EC6b	-0.0001	-0.0001	-0.0001	4,624
			(0.0002)	(0.0002)	(0.0002)	
Panel I	*Equation (1). Fixed-effects regression*	EC6b	-0.0001	-0.0001	-0.0002	4,624
			(0.0002)	(0.0002)	(0.0002)	
Panel II	*Equation (2). 2SLS estimates*	EC6b	-0.005	-0.005	-0.002	1,776
	IV: Linear rainfall shock		(0.036)	(0.034)	(0.031)	
Panel II	*Equation (4). 2SLS estimates*	EC6b	−0.004*	−0.004*	−0.004*	1,776
	IV: 2nd-order polynomial rainfall shock		(0.002)	(0.002)	(0.002)	
Panel II	*Equation (6). Reduced-form estimates*	rainfall_shock	-0.022	-0.021	-0.010	1,776
	IV: Linear rainfall shock		(0.139)	(0.141)	(0.142)	
Panel II	*Equation (7). Reduced-form estimates*	rainfall_shock	-0.474*	-0.468*	-0.446	1,776
			(0.278)	(0.281)	(0.284)	
	IV: 2nd-order polynomial rainfall shock	$(\text{rainfall_shock})^2$	0.080	0.074	0.074	
			(0.049)	(0.051)	(0.051)	
Panel II	*Equation (8). 2SLS estimates*	EC6b	-0.004*	-0.004*	-0.003	1,678
"Head or spouse only"	*IV: 2nd-order polynomial rainfall shock*		(0.002)	(0.002)	(0.002)	
Panel II	*Equation (9). 2SLS estimates*	EC6b	-0.010	-0.009	-0.013	98
"Neither head nor spouse"	*IV: 2nd-order polynomial rainfall shock*		(0.010)	(0.009)	(0.010)	
Panel II	*Equation (10). 2SLS estimates*	EC6b	−0.003*	−0.003**	−0.002*	668
"Male sample"	*IV: 2nd-order polynomial rainfall shock*		(0.001)	(0.001)	(0.001)	
Panel II	*Equation (11). 2SLS estimates*	EC6b	-0.001	-0.001	-0.0004	1,108
"Female sample"	*IV: 2nd-order polynomial rainfall shock*		(0.003)	(0.003)	(0.003)	

Sample:			Dependent variable: EC6b			
Panel II	*Equation(3). First-stage estimates*	rainfall_shock	1.948	1.831	1.985	1,796
	IV: Linear rainfall shock		(28.712)	(28.807)	(29.237)	
		F-statistic (excl instr.)	0.01863	0.04346	0.0368	
Panel II	*Equation(5) First-stage estimates*	rainfall_shock	102.781**	106.232**	103.754**	1,796
			(45.719)	(46.169)	(45.876)	
	IV: 2nd-order polynomial rainfall shock	$(\text{rainfall_shock})^2$	−17.375*	−17.986*	−17.350*	
			(9.806)	(9.872)	(9.685)	
		F-statistic (excl instr.)	3.127	3.248	3.307	
	Controls (for all panels):					
	Individual fixed-effects		Yes	Yes	Yes	
	District fixed-effects		Yes	Yes	Yes	
	Year fixed-effects		Yes	Yes	Yes	
	Dummy ïn charge of collecting waters"		No	Yes	Yes	
	Dummy "responsible for water activities"		No	Yes	Yes	
	Age		No	Yes	Yes	
	Labor market status		No	No	Yes	
	Health status		No	No	Yes	
	Type of assets and financial situation		No	No	Yes	
	Notes:			*p0.1;	**p0.05;	***p0.01

Household size

By considering household size in our analysis, we can see that the point estimates are negative as expected. This means that the amount of water collected per household member is negatively correlated with the PSS score. However, when we look at the table 4, introducing household size into the denominator of the amount of water collected, the significance level of the coefficient is only between 10 and 15 percent (equation 12) while the interaction variable between the amount of water collected and household size produces a statistically significant coefficient at the 10 percent level (equation 13). This result is stable over the three specifications in columns (1), (2) and (3). Therefore, our results are robust to household size. Note that the full regressions are also displayed in appendix: table (17) for equation (12) and table (18) for equation (13).

Table 4: The effect of quantity of water collected in liters on the PSS-Score taking into account the household size

			Dependent variables			
			PSS-Score	PSS-score	PSS-score	
Sample	Baseline specification	Independent variables	(1)	(2)	(3)	Observations
Panel II	*Equation (12). 2SLS estimates*	EC6b / Household size	-0.020	-0.019	-0.018	1,776
	IV: 2nd-order polynomial rainfall shock		(0.013)	(0.013)	(0.013)	
Panel II	*Equation (13). 2SLS estimates*	EC6b * Household size	-0.001*	-0.001*	-0.001*	1,776
	IV: 2nd-order polynomial rainfall shock		(0.0004)	(0.0004)	(0.0004)	
	Controls (for all panels):					
	Individual fixed-effects		Yes	Yes	Yes	
	District fixed-effects		Yes	Yes	Yes	
	Year fixed-effects		Yes	Yes	Yes	
	Dummy ïn charge of collecting waters"		No	Yes	Yes	
	Dummy "responsible for water activities"		No	Yes	Yes	
	Age		No	Yes	Yes	
	Labor market status		No	No	Yes	
	Health status		No	No	Yes	
	Type of assets and financial situation		No	No	Yes	
	Notes:˜			*p0.1;	**p0.05;	***p0.01

6 Robustness checks and sensitivity analysis

In this section we check the robustness of our 2SLS estimates. For this purpose, we use an alternative measure of SWB such as the happiness measure or the life satisfaction scale. To this end, we first use evaluative measures of unhappiness using the unhappiness scale proposed by Lyubomirsky and Lepper (1999) and which we have named "*unhappiness-4*": "all things considered, how did you rate yourself: (1) Very happy (2) Somewhat happy (3) Somewhat unhappy (4) Unhappy". The higher the "*unhappiness-4*", the lower the level of happiness. As the answers are discrete and have an ordered level, we use a generalized linear model. Digging deeper, we use an ordered probit model that follows a latent model:

$$(14)\begin{cases} y^* = x'\beta + \epsilon \\ y = j \in \{1,2,3,4\} \text{ iff } \alpha_{j-1} < y^* < \alpha_j \\ \epsilon | x \sim N(0,1) \end{cases}$$

y^* is the level of *happiness-4*. β is the coefficient of interest that indicates the relationship between the level of happiness and the amount of water collected. Therefore, after assuming normality between the variables and the error terms, we obtain that the conditional probability for alternative j is :

$$Pr(y = j|x) = Pr(\alpha_{j-1} < y^* < \alpha_j|x) = \Phi(\alpha_j - x'\beta) - \Phi(\alpha_{j-1} - x'\beta)$$

The conditional likelihood:

$$L(\beta, \alpha|y; x) = \prod_{i=1}^{n}\prod_{j=1}^{4}[\Phi(\alpha_j - x_i'\beta) - \Phi(\alpha_{j-1} - x_i'\beta)]^{1[y_i=j]}$$

The regression results for equation (14) are displayed in columns (1) and (2) of the table(5) for the ordered probit model and the 2SLS estimates are displayed in columns (3) and (4). The point estimates associated with "ec6b" indicate a negative and statistically significant relationship at the 1% level between the level of happiness and the amount of water collected in the past week. In other words, an increase in the amount of water collected in the week prior to the survey leads to a decrease in unhappiness, ceteris paribus, and this result is stable for all specifications.

We also decide to reproduce the same methodology of equation (14) by taking for y^* another measure of subjective well-being represented by the life satisfaction scale which we will call *LSI*: "(BP4) All things considered, how satisfied are you with your life: (1) very unsatisfied ... (10) very satisfied". Results are displayed in table 6. The positive and highly significant regression coefficients associated with ec6b indicate that increasing the amount of water collected is likely to increase overall life satisfaction. This result is stable across various specifications, but more importantly, it confirms the initial hypothesis tested in this Master thesis, namely "water insecurity (proxied here by the amount of water collected) negatively affects the subjective well-being of individuals".

Table 5: Sensitivity analysis considering other measure of subjective well-being (BP3: Taking everything into consideration, how happy you are in your life?)

	Dependent variable: BP3			
	Ordered probit model		*2SLS estimates*	
			(IV: 2nd-order polynomial rainfall shock)	
	(1)	(2)	(3)	(4)
ec6b	-0.0002^{***}	-0.0002^{***}	-0.0002^{***}	-0.0002^{***}
	(0.00003)	(0.00003)	(0.00004)	(0.00004)
collect		0.109^{**}		0.229^{***}
		(0.048)		(0.079)
resp		0.012		-0.044
		(0.050)		(0.080)
unempl		-0.278^{***}		-0.102
		(0.085)		(0.145)
no_wage		0.099		0.231^{*}
		(0.074)		(0.121)
independent		0.085		0.182^{**}
		(0.056)		(0.091)
unempl x independent		0.245^{***}		0.145
		(0.088)		(0.140)
unempl x no_wage		0.269^{**}		-0.067
		(0.105)		(0.175)
Constant	1.902^{***}	1.457^{***}	1.809^{***}	1.366^{***}
	(0.046)	(0.086)	(0.069)	(0.125)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 6: Sensitivity analysis considering other measure of subjective well-being (BP4: Taking everything into consideration, how satisfied are you with your life?)

	Dependent variable: BP4			
	Ordered probit model		*2SLS estimates*	
			(IV: 2nd-order polynomial rainfall shock)	
	(1)	(2)	(3)	(4)
ec6b	0.0001***	0.0001***	0.0001***	0.0001***
	(0.00002)	(0.00002)	(0.00003)	(0.00003)
collect		−0.099**		−0.187**
		(0.046)		(0.075)
resp		−0.024		−0.004
		(0.047)		(0.077)
unempl		−0.040		−0.203
		(0.080)		(0.137)
no_wage		0.180**		−0.196*
		(0.070)		(0.115)
independent		−0.044		−0.250***
		(0.053)		(0.086)
unempl x independent		−0.183**		0.170
		(0.084)		(0.133)
unempl x no_wage		−0.188*		0.218
		(0.099)		(0.165)
Constant	1.703***	1.859***	1.719***	1.991***
	(0.041)	(0.085)	(0.067)	(0.126)

Note: *p<0.1; **p<0.05; ***p<0.01

7 Concluding remarks: analysis of PILAEP 2 potential benefits and policy recommendations, limitations.

7.1 Evaluation of the economic benefits in terms of subjective well-being for future PILAEP II beneficiaries

The negative impacts of water insecurity include an increase in mental illness due to psychological stress and an increase in the burden of water-related diseases (such as diarrhoea, cholera, typhus, hepatitis, malaria, etc.) that would worsen the life satisfaction of the sick and their families. Additionally, water insecurity may also create conflicts between water users and food insecurity. Therefore, managing and coping with water insecurity is having great economic, social, and policy significance as it concerns health and survival of people.

This section aims to estimate the monetary value of water supply infrastructure improvements that correspond to our results on the SWB. According to Hutton et al. (2004), the calculation of the total societal economic benefit of interventions to improve water and sanitation services is the sum of: (1) Health sector benefit due to avoided illness, (2) Patient expenses avoided due to avoided illness, (3) Value of deaths avoided, (4) Value of time savings due to access to water and sanitation, (5) Value of productive days gained of those with avoided illness, (6) Value of days of school attendance gained of those with avoided illness, (7) Value of child days gained of those with avoided illness. Therefore, the costs saved may accrue to the health service, the patient, the employer of the patient, consumers, children and agricultural and industrial sectors (see details in figure 8 in appendix). Furthermore, the same paper also provides a set of reference values for each unit of benefit (see details in figure 9 in appendix).

This methodology, despite being exhaustive to evaluate the socioeconomic benefits of PILAEP 2 project for future beneficiaries, is not adapted to our context. Indeed, we only need to capture the economic benefits of subjective well-being rather than the overall benefits of the improved future water infrastructure. To do this, we will use a non-market valuation technique proposed by del Saz-Salazar et al. (2020) which itself uses the Life Satisfaction Approach (LSA) to value environmental attributes. Although we primarily use the PSS index rather than the life satisfaction index in our first empirical strategy, the methodology remains the same. As we know, responses to the emotional distress questions for the creation of the PSS-10 score can be interpreted as a retrospective general judgement influenced in part by the immediate context and the respondent's current mood and memory (Kahneman and Krueger (2006)). Therefore, the measure of reported emotional distress is then used as an indicator of individual utility according to the following additive function:

$$SWB_{i,t} = \alpha + \beta Y_i, t + \sum_{K=1}^{n} \gamma_k X_{k,i,t} + \sum_{K=1}^{n} \delta_K Q_{k,i,t} + \epsilon_{i,t}$$

which is equivalent to the identification equation (1) where $SWB_{i,t}$ is the PSS-score of respondent i at date t, Y is the respondent's income, X_k is a set of explanatory socioeconomic variables other than income, Q_k is a set of variables related to the environment (water insecurity is among environment-related variables), and ϵ represents the idiosyncratic errors. Then the principle is simple: suppose that in this previous equation β represents the marginal utility of income while δ_1 is the marginal utility of the quantity of water, assuming that among the

different environmental variables considered, Q_1 is the new quantity of environmental water to which the household has access with the policy. We found from our results that δ_1 is equal to the ATT = -0.004 $\times$ the quantity of quality water that the new autonomous water systems planned by the PILAEP2 can provide to the household (L). After that we should calculate the utility-constant marginal rate of substitution (MRS) between both variables which represents the economic benefit:

$$MRS = \frac{\frac{\partial SWB}{\partial Q_1}}{\frac{\partial SWB}{\partial Y}}$$

$$MRS = \frac{\delta_1}{\beta}$$

This income-equivalent value (MRS)is the amount of income needed to keep the individual's well-being constant when a change in environmental conditions occurs. The MRS can be positive or negative depending on whether environmental conditions improve or deteriorate. In the first case, it would be the marginal willingness to pay (WTP) for improvements and in the second case, the marginal willingness to accept compensation (WTA) for allowing environmental degradation. Therefore, in order to estimate the social benefits derived from a policy aimed at improving water supply infrastructure (such as PILAEP 2), we should compute the MRS and the implicit WTP obtained should be aggregated over the population of all municipalities who are the main beneficiaries of this policy.

7.2 Discussions and limitations

7.2.1 Potential omitted variable bias

We know that including covariates allows the researcher to answer questions about whether the effect of one independent variable depends on the level of another. In the literature, water scholars recognized that socioeconomic factors, somatic health, the quality of social relationships and social integration are important determinants of SWB level(George (2010), Lukaschek et al. (2017)). Therefore, in this analysis we should have performed a stepwise regression analysis by including an increasing number of factors. Unfortunately, with the survey data, we may be missing crucial information without which the results may be biased when covariates vary over time (which cannot be eliminated using the individual fixed effect). This information includes the following:(1) water storage available before the week of water collection called for in the survey, (2) income, (3) share of water expenditure in total income, (4) somatic health and health status of the family, (5) food insecurity, (6) social capital. Note that these omitted variables could enter the regression by interacting with the independent variable or linearly as additional covariates.

First, to fully identify the impact of the water collected 1 week before the survey on SWB measures, it is essential to know how much water was already stored in the household exactly the day before the reference week in which we ask them how much water they collect. Indeed, not being able to collect less water does not necessarily lead to an increase in frustration or stress levels when the amount of water stored was already high during the week before the harvest. Therefore, the interaction of this information on stored water with EC6b (the amount of water

collected) should have provided us with a more accurate estimate.

Second, the main variable missing from the survey is the respondent's earnings. As already explained, the literature indicates that income is closely associated with measures of subjective well-being (Frank (2012); Veenhoven (1991); Tay et al. (2018)). If we disregard the debate on relative versus absolute income and on self-reported versus administrative income, income still plays an important role in determining subjective well-being in many aspects: income is a resource that serves as a buffer against major life events (e.g. illness, unemployment), daily life hassles and inconveniences. It is also a means to obtain goods and services. Therefore, income can mediate or moderate the association between water insecurity and subjective well-being. Therefore, total earnings (including agricultural income, remittances, wages and other income) should have been introduced as an additional control variable in a specification.

Third, for the same reason as income, the share of water expenditure in the household budget is also a source of frustration and should be introduced into the specification as an additional covariate.

Fourth, it is also important to collect information on somatic health: the health status of the respondent has a considerable impact on his/her subjective well-being. We have therefore included somatic multimorbidity variables (which describes "the coexistence of two or more chronic diseases") as covariates in our specification, but they are only available for the baseline survey and are not sufficient. Previous literature has also recognised that the health of close relatives (children, spouse, other family member) has a subsequent influence on the respondent's level of anxiety/depression, happiness and life satisfaction, especially when illnesses (and in particular water-related illnesses) result in death (Guardiola et al. (2013)).

Fifth, we should also include the food insecurity status of the household as it affects levels of depression/anxiety symptoms. In theory, it is very likely that water insecurity worsens food insecurity, but not necessarily the other way around (Brewis et al. (2020)), so the omission of food insecurity status could bias our coefficient of interest.

Finally, social integration and the quality of social relationships are also major determinants of subjective well-being. For example, living alone, loneliness and social disintegration, not being part of community life (especially in poor countries) have been shown to significantly affect the level of stress and depression (Pinquart and Sörensen (2000)). It should be noted that the introduction of additional control variables is not intended to fully estimate any variation in the SWB score and thus increase the R^2, but rather to obtain a consistent specification and unbiased estimates, even if the degrees of freedom of our regression would be significantly affected.

7.2.2 Discussion about the exclusion restriction of our instrument

Since we take rainfall as an instrument, we suspect that the effect of rainfall on subjective well-being is not only through its effect on collected water. Indeed, precipitation can have an impact on other economic activities (e.g. agricultural production) and thus on welfare. Since we are not able to control for the type of occupation of the respondent, it is likely that the exclusion restriction is violated. In this study, we assume that the type of occupation does not vary between the two waves of the survey, which are only one year apart; therefore, the individual fixed

effect dummy variable should control for this time invariant variable. The insignificant results for reduced-form estimates (equation (6, 7) also suggest that there is no externality in the impact of rainfall on SWB. In other terms, the only impact of rainfall on SWB go through its impact on amount of water collected.

7.2.3 Water insecurity: a multidimensional notion

In this master thesis, we have only addressed one aspect of water insecurity, namely access to water (represented by water collected in a week). However, according to Wutich and Ragsdale (2008), there are 3 dimensions of water insecurity: inadequate water supply, unequal access to water distribution and seasonal variability of water sources. In practice Young et al. (19a) and also Young et al. (19b) recently constructed much exhaustive index for water insecurity experiences which they called HWISE Scale (Households Water Insecurity Experiences). In total, 12 items are collected to construct water insecurity scores. The 11 items relate to the number of times in the previous four weeks that the household has experienced a problem related *water availability* (interrupted supply, no water availability at all), *quantity* (not enough to wash clothes, having to change foods eaten, not having as much to drink as liked, going to sleep thirsty), *hygiene* (inadequate water for bathing, inadequate water for handwashing) and *psychosocial dimensions.* (worrying about having enough water, having one's day interrupted because of water problems, feeling upset/angry about the water situation). And the twelfth item on 'feelings of shame about the water situation' is only collected in the second wave of sampling in the original study. Responses were individually scored from 0 to 3 as follows: 0 = never, 1 = rarely (1-2 times in the previous four weeks), 2 = sometimes (3-10 times), 3 = often (11-20 times) or always (> 20 times). They finally obtain a score ranging from 0 to 36, where higher scores indicate greater water insecurity. From the PILAEP database, it was difficult to construct such an index because PILAEP only collects numerical variables (number of litres, number of days, total expenditure on water-related activities) or indicator variables (binary "yes" or "no" questions) instead of the number of times respondents experienced water insecurity. But also, PILAEP collects the necessary information within a maximum of one week before the survey rather than 4 weeks.

Therefore, this raises the question of whether each dimension of water insecurity simultaneously co-determines subjective well-being or whether it should instead enter the regression model singularly. In other words, it is interesting to know whether we should introduce all water-related variables (water availability, number of water cuts, water expenditure, water for sanitation, etc.) at once since they are correlated with each other and affect SWB, or whether each of them should be devoted to a single model. In an article by Guardiola et al. (2013), introducing several water-related variables into the models allows the effect of each of them to be disentangled. In this specific paper, they found that perceived water quality has a positive effect and that spending on bottled water/total income has a negative effect on subjective well-being, while the perceived number of cuts, having a tap inside the houses or access to a well have no influence. The main limitation of this methodology is that due to the multiplicity of independent variables, they were not able to use the 2SLS identification strategy to address the endogeneity of each of them.

7.2.4 Sampling design and a potential loss of statistical power

From 14,139 individuals in the baseline survey and 13,282 individuals in the follow-up survey, we end up with only one eligible adult per household who identified themselves as responsible for or knowledgeable about the water situation in the household, which corresponds to 2,345 individuals consisting mainly of the head of household or his/her spouse. Furthermore, when we have to use consistent rainfall estimates for the IV identification strategy, our sample is reduced to only 898 individuals. This leads to a potential loss of statistical power and even to attrition bias if we assume that the inconsistency of the rainfall estimates is not completely random. In our master thesis, we assume that it is random.

7.2.5 Survey design

When we look at the interviewer's manual, we realise that the questions on SWB come after the water modules in the questionnaire. This raises concerns about the order of the questions, as there could be a risk of inconsistent experimental knowledge discovery. Many previous studies on subjective well-being have pointed out that questions on emotional distress, life satisfaction or happiness are sensitive to wording, response scale and question order. The framing effect (the way you frame the question, the questions that precede it, etc.) can change what people focus on and thus lead to different responses (Yap et al. (2017), Lucas et al. (2016)).

7.2.6 Unaddressed issue: the quality of water collected

One of the main omitted variables that is correlated with collected water is the quality of collected water which affects subjective well-being through its impact on water-related diseases. But if we want to include both the quantity and the quality aspect in our regression, we will not be able to use the IV identification strategy. There are two ways to assess water quality: the first is to use perceived water quality by asking the respondent how he/she perceives the quality of water in a specific water source[4] and the second is to use objective data that analyses the site manager and the chemical composition of the water. To properly assess the water quality of each source in PILAEP 2, the project team made a census of all the water sources in the study area as indicated by the respondents in which they assigned the quality index afterwards. However, the process of knowing unambiguously whether or not the household is consuming poor quality water is complex: even if the respondent states that the household collects water from a source that is not safe and secure in terms of quality and potability, they can still treat the water afterwards, making it good quality (or at least average quality) water. Therefore, one way to capture the quality of water used by the household from a source is to interact an indicator variable equal to 1 if the water source is of poor quality and 0 if the water source is of good quality with an indicator variable equal to 1 if the household does not treat the water from the source and equal to 0 if the household does. Therefore, the interaction variables will take the value 1 for poor quality water and 0 for good (or average) quality water. Because of the the wide range of water sources used by each household and the different uses made by the household for each water harvested, it is difficult to have a clear picture of whether poor quality water is used for drinking, sanitation, household chores

[4]The perceived quality question is asked in (ESA15): "According to you, does this water source provide drinking water or not?"

that would be detrimental to health rather than for crops, livestock or other economic activities that would not. As we are not able to use this quality index, we assume that the quantity of water collected captures both the effect of quality and quantity on subjective well-being, as they are correlated with each other.

8 Appendix

Table 7: Theoretical and effective samples and sampling rates

		Theoretical Sample		Sampling method with replacement		Final Sampling	
	Household numbers in the district	N	Sampling rate (%)	N	Sampling rate (%)	N	Sampling rate (%)
Badara 2 (T)	3652	300	8.2	350	9.6	300	8.2
Badara 3 (C)	1189	300	25.2	528	44.4	478	40.2
Belair (outside Palu and Baluba localities - T)	1870	300	16.0	350	18.7	300	16.0
Belair (Palu and Baluba - C)	188	300	159.6	188	100.0	169	89.9
Buma (outside Manvululu locality - T)	1603	300	18.7	350	21.8	298	18.6
Buma (Manvululu locality - C)	176	300	170.5	176	100.0	158	89.9
Maba 2 (T)	1551	300	19.3	350	22.6	301	19.4
Munke 2 (C)	526	300	57.0	397	75.5	346	65.8
Kindobo (T)	1027	300	29.2	350	34.1	300	29.2
Mbenzale (C)	592	300	50.7	398	67.2	348	58.8
Total Treated (T)	9703	1500	15.5	1750	18.0	1499	15.4
Total Control (C)	2671	1500	56.2	1687	63.2	1499	56.1
Total (T + C)	12374	3000	24.2	3437	27.8	2998	24.2

Table 8: Summary statistics of "PANEL II"

Wave 1: Baseline survey summary statistics 'Panel II'

Variables	Male	Female	Full Sample	Two.sample.t.test.p.value	Prop_of_NAs...
Dependent Variables					
PSS-score	27.167 (5.025)	27.456 (5.13)	27.347 (5)	0.415	2.77
Unhappiness-4	2.553 (0.762)	2.586 (0.787)	2.573 (1)	0.545	2.77
LSI	5.405 (2.019)	5.15 (1.997)	5.246 (2)	0.065	2.77
Independent Variables					
Quantity of Water Collected (liters)	519.18 (366.8)	631.42 (375.01)	589.17 (376)	0	0
Mean household size	- (-)	- (-)	4.98 (3)	-	0
Mean quantity of water collected per household member (liters)	- (-)	- (-)	135.44 (96.41)	-	0
Socioeconomic characteristics					
Proportion of respondents who directly carry water collection	0.331 (0.471)	0.716 (0.451)	0.571 (0)	0	0
Proportion of respondents responsible for water activities	0.328 (0.47)	0.816 (0.388)	0.633 (0)	0	0
Age	42.917 (14.224)	37.039 (12.66)	39.252 (14)	0	0
Proportion of respondents without any professional activities during the past week	0.189 (0.392)	0.454 (0.498)	0.354 (0)	0	1.49
Proportion of respondents without any of wages or salary during the past week	0.293 (0.456)	0.5 (0.5)	0.422 (0)	0	56.38
Proportion of respondents independant from their families for susbsistence	0.769 (0.422)	0.516 (0.5)	0.611 (0)	0	68.4
Proportion of respondents with precarious housing in the baseline survey	0.55 (0.498)	0.5 (0.5)	0.519 (0)	0.144	0
Monthly rent during the baseline survey	17476 (33362)	20234 (36614)	19196 (35432)	0.259	69.55
Proportion of respondents that are unable to save	0.861 (0.347)	0.889 (0.314)	0.879 (0)	0.208	0
Proportion of respondents with access to electricity	0.062 (0.242)	0.114 (0.318)	0.095 (0)	0.01	0
Proportion of respondents who are subject to chronic disease	0.05 (0.219)	0.061 (0.239)	0.057 (0)	0.514	1.49
Rainfall measures					
Rainfall proportion beyond the district mean (mm in the last 7 days)	- (-)	- (-)	2.839 (2)	-	0
Respondents status in the family					
Proportion of household head (%)	-	-	51.11	-	-
Proportion of the spouse of the head (%)	-	-	42.76	-	-
Proportion of the children of the head or of the spouse(%)	-	-	4.9	-	-
Head or spouse's grandchild/ parent(s)/non-related member (%)	-	-	1.22	-	-
N (obs)	338	560	898	-	-

Wave 2: Follow-up survey summary statistics 'Panel II'

Variables	Male	Female	Full Sample	Two.sample.t.test.p.value	Prop_of_NAs...
Dependent Variables					
PSS-score	27.47 (5.548)	27.389 (5.375)	27.42 (5)	0.829	0.0426
Unhappiness-4	2.358 (0.738)	2.495 (0.775)	2.443 (1)	0.009	0.0426
LSI	5.595 (1.829)	5.4 (2.018)	5.473 (2)	0.147	0.0426
Independent Variables					
Quantity of Water Collected (liters)	813.08 (1507.81)	790.18 (1274.48)	798.8 (1366)	0.808	0
Mean household size	- (-)	- (-)	5.39 (3)	-	0
Mean quantity of water collected per household member (liters)	- (-)	- (-)	178.53 (434.58)	-	0
Socioeconomic characteristics					
Proportion of respondents who directly carry water collection	0.464 (0.499)	0.852 (0.356)	0.706 (0)	0	0
Proportion of respondents responsible for water activities	0.388 (0.488)	0.9 (0.3)	0.707 (0)	0	0
Age	42.935 (14.217)	37.048 (12.664)	39.264 (14)	0	0
Proportion of respondents without any professional activities during the past week	0.556 (0.498)	0.639 (0.481)	0.608 (0)	0.013	0
Proportion of respondents without any of wages or salary during the past week	0.482 (0.5)	0.564 (0.496)	0.533 (0)	0.017	41.83
Proportion of respondents independant from their families for susbsistence	0.308 (0.462)	0.207 (0.406)	0.245 (0)	0.001	54.16
Proportion of respondents with precarious housing baseline survey	0.55 (0.498)	0.5 (0.5)	0.519 (0)	0.144	0
Monthly rent during the baseline survey	15728 (28071)	17921 (30696)	17095 (29739)	0.285	69.55
Proportion of respondents that are unable to save	0.861 (0.347)	0.889 (0.314)	0.879 (0)	0.208	0
Proportion of respondents with access to electricity	0.062 (0.242)	0.114 (0.318)	0.095 (0)	0.01	0
Proportion of respondents who are subject to chronic disease	0.05 (0.219)	0.059 (0.236)	0.056 (0)	0.585	1.49
Rainfall measures					
Rainfall proportion beyond the district mean (mm in the last 7 days)	- (-)	- (-)	0.652 (1)	-	0
Respondents status in the family					
Proportion of household head (%)	-	-	51.11	-	-
Proportion of the spouse of the head (%)	-	-	42.76	-	-
Proportion of the children of the head or of the spouse(%)	-	-	4.9	-	-
Head or spouse's grandchild/parent(s)/non-related member (%)	-	-	1.22	-	-
N (obs)	338	560	898	-	-

Figure 7: Treatment of the changes in names, sex or ages for individuals with same individual id (Source: Author computation)

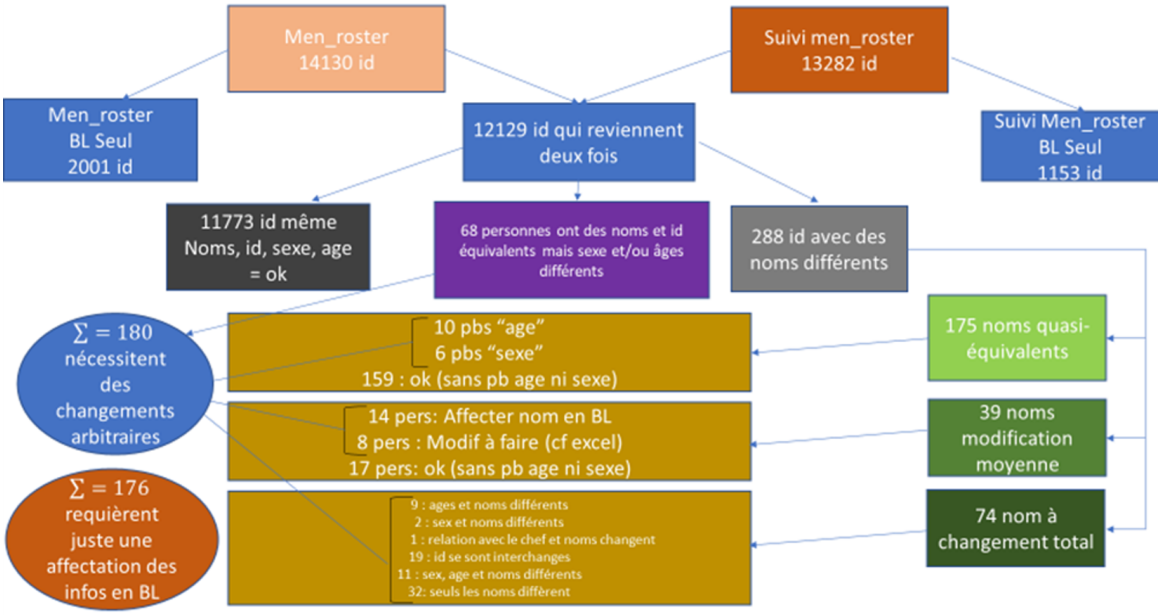


Table 9: Pooled OLS regression for the relationship between water collected and the PSS-score

	Dependent variable:		
	well_being		
	(1)	(2)	(3)
ec6b	−0.0001	−0.0001	−0.0001
	(0.0002)	(0.0002)	(0.0002)
collecteur		−0.138	−0.193
		(0.236)	(0.237)
resp		−0.599**	−0.499**
		(0.236)	(0.239)
age		0.009	0.010
		(0.007)	(0.007)
loyer			0.00000
			(0.00000)
hprecaire			0.024
			(0.178)
acces_elec			0.316
			(0.246)
s02			0.211
			(0.359)
zero_epargne			0.707***
			(0.243)
chomage			−1.231***
			(0.395)
zero_salaire			0.515
			(0.374)
autonome			0.191
			(0.247)
I(chomage ∗autonome)			0.622
			(0.420)
I(chomage ∗zero_salaire)			1.438***
			(0.522)
Constant	28.223***	28.393***	27.273***
	(0.127)	(0.357)	(0.468)
Observations	4,624	4,624	4,624
R^2	0.0001	0.004	0.015
Adjusted R^2	−0.0001	0.003	0.012
Residual Std. Error	5.978 (df = 4622)	5.969 (df = 4619)	5.942 (df = 4609)
F Statistic	0.585 (df = 1; 4622)	4.567*** (df = 4; 4619)	4.994*** (df = 14; 4609)

Table 10: Fixed Effects regression for the relationship between water collected and the PSS-score

	Dependent variable:		
	well_being		
	(1)	(2)	(3)
ec6b	−0.0001	−0.0001	−0.0002
	(0.0002)	(0.0002)	(0.0002)
collecteur		−0.138	−0.081
		(0.236)	(0.337)
resp		−0.599**	0.066
		(0.236)	(0.401)
age		0.009	−0.056
		(0.007)	(0.530)
loyer			−0.00000
			(0.00001)
chomage			−0.536
			(0.481)
zero_salaire			0.660
			(0.542)
autonome			0.331
			(0.329)
I(chomage ∗autonome)			−0.640
			(0.575)
I(chomage ∗zero_salaire)			0.563
			(0.703)
Observations	4,624	4,624	4,624
R^2	0.609	0.004	0.610
Adjusted R^2	0.202	0.003	0.202
Residual Std. Error	5.340 (df = 2268)	5.969 (df = 4619)	5.340 (df = 2259)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 11: 2SLS estimates for the relationship between water collected and the PSS-score, considering linear form of rainfall shock as instrument

	Dependent variable:		
	well_being		
	(1)	(2)	(3)
ec6b(fit)	−0.005	−0.005	−0.002
	(0.036)	(0.034)	(0.031)
collecteur		0.025	0.070
		(1.689)	(1.600)
resp		−1.121	−0.593
		(5.419)	(4.472)
age		−0.198	−0.302
		(2.151)	(4.696)
chomage			−2.951
			(2.064)
zero_salaire			1.840
			(4.954)
autonome			0.011
			(7.055)
chomage:autonome			1.693
			(7.399)
chomage:zero_salaire			1.751
			(10.192)
Observations	1,776	1,776	1,776
R^2	0.273	0.313	0.544
Adjusted R^2	−0.474	−0.396	0.057
Residual Std. Error	7.313 (df = 876)	7.118 (df = 873)	5.849 (df = 859)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 12: 2SLS estimates for the relationship between water collected and the PSS-score, considering 2nd-order polynomial form of rainfall shock as instrument

	Dependent variable:		
	well_being		
	(1)	(2)	(3)
‘ec6b(fit)‘	−0.004*	−0.004*	−0.004*
	(0.002)	(0.002)	(0.002)
collecteur		0.042	0.009
		(0.714)	(0.715)
resp		−1.059	−0.957
		(0.911)	(0.894)
age		−0.176	−0.522
		(0.966)	(1.007)
chomage			−2.799**
			(1.116)
zero_salaire			0.507
			(1.133)
autonome			1.181*
			(0.655)
I(chomage ∗autonome)			−0.001
			(0.914)
I(chomage ∗zero_salaire)			4.131**
			(1.761)
Observations	1,776	1,776	1,776
R^2	0.337	0.357	0.387
Adjusted R^2	−0.342	−0.320	−0.266
Residual Std. Error	6.979 (df = 876)	6.922 (df = 864)	6.777 (df = 859)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 13: First-stage regression of the relationship between water collected and the PSS-score, considering linear form of rainfall shock as instrument

	Dependent variable:		
	ec6b		
	panel linear	*felm*	
	(1)	(2)	(3)
choc_positif	1.948	1.831	1.985
	(28.712)	(28.807)	(29.237)
collecteur		−43.812	−40.567
		(66.537)	(66.613)
resp		−167.564	−150.869
		(119.218)	(120.601)
age		−55.540	−96.434
		(78.076)	(97.365)
s02			−563.996**
			(254.671)
chomage			−200.595
			(131.832)
zero_salaire			52.358
			(132.407)
autonome			51.991
			(87.554)
loyer			−0.001
			(0.001)
chomage:autonome			−48.315
			(77.913)
chomage:zero_salaire			426.409**
			(193.741)
Observations	1,776	1,776	1,776
R^2	0.521	0.525	0.534
Adjusted R^2	0.030	0.035	0.037
Residual Std. Error	997.069 (df = 876)	994.116 (df = 875)	993.260 (df = 859)
F-statistic (excl instr.)	0.01863	0.04346	0.0368
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 14: First-stage regression for the relationship between water collected and the PSS-score, considering 2nd-order polynomial form of rainfall shock as instrument

	Dependent variable:		
	ec6b		
	panel *linear*	*felm*	
	(1)	(2)	(3)
choc_positif	102.781**	106.232**	103.574**
	(45.719)	(46.169)	(45.876)
I(choc_positif^2)	−17.375*	−17.986*	−17.350*
	(9.806)	(9.872)	(9.685)
collecteur		−42.670	−42.878
		(66.255)	(66.575)
resp		−179.087	−168.216
		(119.121)	(118.734)
age		−64.937	−155.701*
		(67.656)	(84.587)
s02			−287.095***
			(82.583)
chomage			−89.352
			(151.564)
zero_salaire			160.235
			(125.675)
autonome			213.044*
			(121.763)
I(chomage ∗autonome)			−212.365*
			(122.657)
I(chomage ∗zero_salaire)			338.727*
			(183.241)
Observations	1,776	1,776	1,776
R^2	0.525	0.526	0.535
Adjusted R^2	0.035	0.026	0.038
Residual Std. Error	994.116 (df = 875)	998.937 (df = 863)	992.995 (df = 858)
F-statistic (excl instr.)	3.127	3.248	3.307
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 15: Reduced-form estimates, considering linear form of rainfall shock as instrument

	Dependent variable:		
	well_being		
	panel	*felm*	
	linear		
	(1)	(2)	(3)
choc_positif	−0.022	−0.021	−0.010
	(0.139)	(0.141)	(0.142)
collecteur		0.222	0.161
		(0.624)	(0.629)
resp		−0.389	−0.309
		(0.645)	(0.648)
age		0.064	0.003
		(0.693)	(0.607)
chomage			−2.829***
			(0.884)
zero_salaire			1.510*
			(0.882)
autonome			−0.439
			(0.599)
loyer			−0.00000
			(0.00001)
chomage:autonome			2.164**
			(0.895)
chomage:zero_salaire			1.109
			(1.165)
Observations	1,776	1,776	1,776
R^2	0.00004	0.575	0.583
Adjusted R^2	−1.026	0.137	0.147
Residual Std. Error		5.597 (df = 873)	5.563 (df = 867)
F Statistic	0.032 (df = 1; 876)		
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 16: Reduced-form estimates, considering 2nd-order polynomial form of rainfall shock as instrument

	Dependent variable:		
	well_being		
	panel linear	*felm*	
	(1)	(2)	(3)
choc_positif	−0.474*	−0.468*	−0.446
	(0.278)	(0.281)	(0.284)
I(choc_positif^2)	0.080	0.079	0.077
	(0.049)	(0.050)	(0.051)
collecteur		0.182	0.151
		(0.624)	(0.624)
resp		−0.328	−0.331
		(0.643)	(0.647)
age		0.104	−0.089
		(0.750)	(0.727)
chomage			−1.908**
			(0.847)
zero_salaire			0.271
			(0.936)
autonome			0.996*
			(0.565)
I(chomage ∗autonome)			0.166
			(0.861)
I(chomage ∗zero_salaire)			2.348*
			(1.236)
Observations	1,776	1,776	1,776
R^2	0.004	0.577	0.583
Adjusted R^2	−1.020	0.139	0.146
Residual Std. Error		5.589 (df = 872)	5.566 (df = 867)
F Statistic	1.895 (df = 2; 875)		
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 17: 2SLS estimates using water collected per household member as independent variables

	Dependent variable:		
	well_being		
	(1)	(2)	(3)
collecteur		0.146	0.119
		(0.751)	(0.735)
resp		−0.980	−0.935
		(0.986)	(0.960)
age		−0.391	−0.762
		(0.850)	(0.951)
chomage			−2.861**
			(1.246)
zero_salaire			0.365
			(1.125)
autonome			2.213*
			(1.271)
I(chomage ∗autonome)			−0.524
			(1.161)
I(chomage ∗zero_salaire)			4.373**
			(1.991)
‘eau_par_membre(fit)‘	−0.020	−0.019	−0.018
	(0.013)	(0.013)	(0.013)
Observations	1,776	1,776	1,776
R^2	0.145	0.178	0.249
Adjusted R^2	−0.751	−0.688	−0.553
Residual Std. Error	7.972 (df = 867)	7.826 (df = 864)	7.506 (df = 859)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 18: 2SLS estimates using water collected × household member as independent variables

	Dependent variable:		
	well_being		
	(1)	(2)	(3)
familysize		0.558	
		(0.459)	
collecteur		0.059	0.036
		(0.728)	(0.727)
resp		−0.874	−0.951
		(0.967)	(0.938)
age		0.060	−0.471
		(1.027)	(1.110)
loyer			−0.00001
			(0.00001)
chomage			−2.605**
			(1.016)
zero_salaire			0.564
			(1.091)
autonome			0.842
			(0.642)
I(chomage ∗autonome)			0.092
			(0.904)
I(chomage ∗zero_salaire)			3.845**
			(1.588)
‘I(ec6b ∗familysize)(fit)‘	−0.001*	−0.001*	−0.001*
	(0.0004)	(0.0004)	(0.0004)
Observations	1,776	1,776	1,776
R^2	0.185	0.237	0.283
Adjusted R^2	−0.669	−0.569	−0.484
Residual Std. Error	7.783 (df = 867)	7.544 (df = 863)	7.339 (df = 858)

Note: *p<0.1; **p<0.05; ***p<0.01

Figure 8: Economic benefits arising from water and sanitation improvements (Source: Hutton et al. (2004))

BENEFICIARY	Direct economic benefits of avoiding diarrhoeal disease	Indirect economic benefits related to health improvement	Non-health benefits related to water and sanitation improvement
Health sector	▪ Less expenditure on treatment of diarrhoeal disease	▪ Value of less health workers falling sick with diarrhoea	▪ More efficiently managed water resources and effects on vector bionomics
Patients	▪ Less expenditure on treatment of diarrhoeal disease and less related costs ▪ Less expenditure on transport in seeking treatment ▪ Less time lost due to treatment seeking	▪ Value of avoided days lost at work or at school ▪ Value of avoided time lost of parent/ caretaker of sick children ▪ Value of loss of death avoided	▪ More efficiently managed water resources and effects on vector bionomics
Consumers			▪ Time savings related to water collection or accessing sanitary facilities ▪ Labour-saving devices in household ▪ Switch away from more expensive water sources ▪ Property value rise ▪ Leisure activities and non-use value
Agricultural and industrial sectors	▪ Less expenditure on treatment of employees with diarrhoeal disease	▪ Less impact on productivity of ill-health of workers	▪ Benefits to agriculture and industry of improved water supply, more efficient management of water resources – time-saving or income-generating technologies and land use changes

Figure 9: Calculation methodology, data sources and values for economic benefits (Source: Hutton et al. (2004))

Benefit by sector	Variable	Data source	Data values (+ range)
1. Health sector			
Direct expenditures avoided, due to less illness from diarrhoeal disease	Unit cost per treatment	WHO regional unit cost data	US$4.3-US$9.7 (cost per visit) US$16.1-US$39.7 (cost per day) *Varying by WHO region*
	Number of cases	WHO BoD data	Variable by region
	Visits or days per case	Expert opinion	1 outpatient visit per case (0.5-1.5) 5 days for hospitalised cases (3-7)
	Hospitalisation rate	WHO data	91.8% of cases ambulatory 8.2% of cases hospitalised
2. Patients			
Direct expenditures avoided, due to less illness from diarrhoeal disease	Transport cost per visit	Assumptions	US$0.50 per visit
	% patients use transport	Assumptions	50% of patients use transport (0-100%)
	Non-health care patient costs	Assumptions	US$0.50 ambulatory (US$0.25-1.00) US$2.00 hospitalisation (US$1.0-3.0)
	Number of cases	WHO BoD data	Variable by region
	Visits or days per case	Expert opinion	1 outpatient visit per case (0.5-1.5) 5 days for hospitalised cases (3-7)
	Hospitalisation rate	WHO data	91.8% of cases ambulatory 8.2% of cases hospitalised
Income gained, due to days lost from work avoided	Days off work/ episode	Expert opinion	2 days (1-4)
	Number of people of working age	WHO population data 2002	Variable by region
	Opportunity cost of time	World Bank data	Minimum wage rate (GNP per capita – value added in manufacturing)
Days of school absenteeism avoided	Absent days / episode	Expert opinion	3 (1-5)
	Number of school age children (5-14)	WHO population data 2002	Variable by region
	Opportunity cost of time	World Bank data	Minimum wage rate (GNP per capita – value added in manufacturing)
Productive parent days lost avoided, due to less child illness	Days sick	Expert opinion	5 (3-7)
	Number of babies (0-4)	WHO population data 2002	Variable by region
	Opportunity cost of time	World Bank data	50% minimum wage rate (50% GNP per capita – 50% value added in manufacturing)
Value of loss-of-life avoided (life expectancy, discounting future years at 3%)	Discounted productive years lost (0 – 4 years)	WASH study [16]	16.2 years (9.5 – 29.1)
	Discounted productive years lost (5 – 14 years)	WASH study [16]	21.9 years (15.2 – 33.8)
	Discounted productive years lost (15+ years)	WASH study [16]	19.0 years (16.3 – 22.7)
	Opportunity cost per year of life lost	World Bank data	Minimum wage rate
3. Consumers			
'Convenience' – time savings	Water collection time saved per household per day for better external access	Expert opinion	0.5 hours (0.25-1.0)
	Water collection time saved per household per day for piped water	Expert opinion	1.5 hours (1.0-2.0)
	Sanitation access time saved per person	Expert opinion	0.5 hours (0.25-0.75)
	Average household size	WHO population data 2002	6 people (4-8)
	Opportunity cost of time	World Bank data	Minimum wage rate (GNP per capita – value added in manufacturing)

9 Bibliography

References

Adams, E. A., Stoler, J., and Adams, Y. (2020). Water insecurity and urban poverty in the global south: Implications for health and human biology. *American Journal of Human Biology*, 32(1):e23368.

Angrist, J., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). Vouchers for private schooling in colombia: Evidence from a randomized natural experiment. *American economic review*, 92(5):1535–1558.

Barnett, R. C., Biener, L. E., and Baruch, G. K. (1987). *Gender and stress.* Free Press.

Bisung, E. and Elliott, S. J. (2016). 'everyone is exhausted and frustrated': exploring psychosocial impacts of the lack of access to safe water and adequate sanitation in usoma, kenya. *Journal of Water, Sanitation and Hygiene for Development*, 6(2):205–214.

Bisung, E. and Elliott, S. J. (2017). Psychosocial impacts of the lack of access to water and sanitation in low-and middle-income countries: a scoping review. *Journal of Water and Health*, 15(1):17–30.

Biswas, A. K. (2007). Water as a human right in the mena region: challenges and opportunities. *International Journal of Water Resources Development*, 23(2):209–225.

Boaz, M., Gubert, F., Makabu, T., Munkunda, J. M., Roubaud, F., Saint-Macary, C., Zanuso, C., and Macias, L. (2020). Eau, assainissement, conditions de vie au congo. In *Eau, assainissement, conditions de vie au Congo*, pages 1–80. Agence française de développement.

Bookwalter, J. T. and Dalenberg, D. (2004). Subjective well-being and household factors in south africa. *Social indicators research*, 65(3):333–353.

Boyce, C. J., Brown, G. D., and Moore, S. C. (2010). Money and happiness: Rank of income, not income, affects life satisfaction. *Psychological science*, 21(4):471–475.

Brewis, A., Choudhary, N., and Wutich, A. (2019). Low water access as a gendered physiological stressor: Blood pressure evidence from nepal. *American Journal of Human Biology*, 31(3):e23234.

Brewis, A., Workman, C., Wutich, A., Jepson, W., Young, S., (HWISE-RCN), H. W. I. E.-R. C. N., Adams, E., Ahmed, J. F., Alexander, M., Balogun, M., et al. (2020). Household water insecurity is strongly associated with food insecurity: Evidence from 27 sites in low-and middle-income countries. *American Journal of Human Biology*, 32(1):e23309.

Casey, J. F., Kahn, J. R., and Rivas, A. (2006). Willingness to pay for improved water service in manaus, amazonas, brazil. *Ecological Economics*, 58(2):365–372.

Chindarkar, N., Chen, Y. J., and Gurung, Y. (2019). Subjective well-being effects of coping cost: Evidence from household water supply in kathmandu valley, nepal. *Journal of Happiness Studies*, 20(8):2581–2608.

Cohen, S. and Williamson, G. (1988). Perceived stress in a probability sample of the united states. *Social Psychology of Health*.

Cooke, P. J., Melchert, T. P., and Connor, K. (2016). Measuring well-being: A review of instruments. *The Counseling Psychologist*, 44(5):730–757.

Cooper-Vince, C. E., Arachy, H., Kakuhikire, B., Vořechovská, D., Mushavi, R. C., Baguma, C., McDonough, A. Q., Bangsberg, D. R., and Tsai, A. C. (2018). Water insecurity and gendered risk for depression in rural uganda: A hotspot analysis. *BMC Public Health*, 18(1):1–9.

Crow, B., Swallow, B., and Asamba, I. (2012). Community organized household water increases not only rural incomes, but also men's work. *World Development*, 40(3):528–541.

Deci, E. L. and Ryan, R. M. (2008). Hedonia, eudaimonia, and well-being: An introduction. *Journal of happiness studies*, 9(1):1–11.

del Saz-Salazar, S., González-Gómez, F., and Guardiola, J. (2020). Valuing water supply infrastructure improvements using life satisfaction data as a complement to contingent valuation. *Water and Environment Journal*, 34:401–413.

Des Rosiers, F., Bolduc, A., and Thériault, M. (1999). Environment and value does drinking water quality affect house prices? *Journal of Property Investment & Finance*.

Devoto, F., Duflo, E., Dupas, P., Parienté, W., and Pons, V. (2012). Happiness on tap: Piped water adoption in urban morocco. *American Economic Journal: Economic Policy*, 4(4):68–99.

Diener, E. and Tay, L. (2015). Subjective well-being and human welfare around the world as reflected in the gallup world poll. *International Journal of Psychology*, 50(2):135–149.

Easterlin, R. A. (2013). Happiness, growth, and public policy. *Economic Inquiry*, 51(1):1–15.

Easterlin, R. A., McVey, L. A., Switek, M., Sawangfa, O., and Zweig, J. S. (2010). The happiness–income paradox revisited. *Proceedings of the National Academy of Sciences*, 107(52):22463–22468.

Elhance, A. P. (1999). *Hydropolitics in the Third World: Conflict and cooperation in international river basins*. US Institute of Peace Press.

Enders, C. K. (2010). *Applied missing data analysis*. Guilford press.

Ennis-McMillan, M. C. (2001). Suffering from water: social origins of bodily distress in a mexican community. *Medical Anthropology Quarterly*, 15(3):368–390.

Frank, R. H. (2012). The easterlin paradox revisited. *Emotion*, 12(6):1188.

Geere, J.-A. L., Cortobius, M., Geere, J. H., Hammer, C. C., and Hunter, P. R. (2018). Is water carriage associated with the water carrier's health? a systematic review of quantitative and qualitative evidence. *BMJ Global Health*, 3(3):e000764.

George, L. K. (2010). Still happy after all these years: Research frontiers on subjective well-being in later life. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 65(3):331–339.

Gerber, A. S. and Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.

Gleick, P. H. (1996). Basic water requirements for human activities: meeting basic needs. *Water international*, 21(2):83–92.

Graham, J. P., Hirai, M., and Kim, S.-S. (2016). An analysis of water collection labor among women and children in 24 sub-saharan african countries. *PloS one*, 11(6):e0155981.

Guardiola, J., González-Gómez, F., and Grajales, Á. L. (2013). The influence of water access in subjective well-being: Some evidence in yucatan, mexico. *Social Indicators Research*, 110(1):207–218.

Guiso, L., Sapienza, P., and Zingales, L. (2013). Time varying risk aversion (no. w19284).

Haushofer, J. and Fehr, E. (2014). On the psychology of poverty. *science*, 344(6186):862–867.

Howard, G., Calow, R., Macdonald, A., and Bartram, J. (2016). Climate change and water and sanitation: likely impacts and emerging trends for action. *Annual review of environment and resources*, 41:253–276.

Hutton, G., Haller, L., Water, S., Organization, W. H., et al. (2004). Evaluation of the costs and benefits of water and sanitation improvements at the global level. Technical report, World Health Organization.

James, A., Verhagen, J., Van Wijk, C., Nanavaty, R., Parikh, M., and Bhatt, M. (2002). Transforming time into money using water: a participatory study of economics and gender in rural india. 26(3):205–217.

Kahneman, D. and Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *Journal of Economic perspectives*, 20(1):3–24.

Karabchuk, T. and Soboleva, N. (2020). Temporary employment, informal work and subjective well-being across europe: Does labor legislation matter? *Journal of Happiness Studies*, 21(5):1879–1901.

Kawachi, I. and Berkman, L. F. (2001). Social ties and mental health. *Journal of Urban health*, 78(3):458–467.

Larson, B., Minten, B., and Razafindralambo, R. (2006). Unravelling the linkages between the millennium development goals for poverty, education, access to water and household water use in developing countries: evidence from madagascar. *The Journal of Development Studies*, 42(1):22–40.

Lucas, R. E. and Diener, E. (2008). Can we learn about national differences in happiness from individual responses? a multilevel approach. *Multilevel analysis of individuals and cultures*, pages 223–248.

Lucas, R. E., Oishi, S., and Diener, E. (2016). What we know about context effects in self-report surveys of well-being: comment on deaton and stone. *Oxford Economic Papers*, 68(4):871–876.

Lukaschek, K., Vanajan, A., Johar, H., Weiland, N., and Ladwig, K.-H. (2017). "in the mood for ageing": determinants of subjective well-being in older men and women of the population-based kora-age study. *BMC geriatrics*, 17(1):1–9.

Lyubomirsky, S. and Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social indicators research*, 46(2):137–155.

Maluccio, J. A. (2004). Using quality of interview information to assess nonrandom attrition bias in developing-country panel data. *Review of Development Economics*, 8(1):91–109.

Marengoni, A., Angleman, S., Melis, R., Mangialasche, F., Karp, A., Garmen, A., Meinow, B., and Fratiglioni, L. (2011). Aging with multimorbidity: a systematic review of the literature. *Ageing research reviews*, 10(4):430–439.

Moser, C. O. (1998). The asset vulnerability framework: reassessing urban poverty reduction strategies. *World development*, 26(1):1–19.

Mushavi, R. C., Burns, B. F., Kakuhikire, B., Owembabazi, M., Vořechovská, D., McDonough, A. Q., Cooper-Vince, C. E., Baguma, C., Rasmussen, J. D., Bangsberg, D. R., et al. (2020). "when you have no water, it means you have no peace": a mixed-methods, whole-population study of water insecurity and depression in rural uganda. *Social Science & Medicine*, 245:112561.

North, J. H. and Griffin, C. C. (1993). Water source as a housing characteristic: hedonic property valuation and willingness to pay for water. *Water Resources Research*, 29(7):1923–1929.

Pattanayak, S. K., Yang, J.-C., Whittington, D., and Bal Kumar, K. (2005). Coping with unreliable public water supplies: averting expenditures by households in kathmandu, nepal. *Water Resources Research*, 41(2).

Pinquart, M. and Sörensen, S. (2000). Influences of socioeconomic status, social network, and competence on subjective well-being in later life: a meta-analysis. *Psychology and aging*, 15(2):187.

Prüss, A., Kay, D., Fewtrell, L., and Bartram, J. (2002). Estimating the burden of disease from water, sanitation, and hygiene at a global level. *Environmental health perspectives*, 110(5):537–542.

Ray, I. (2007). Women, water, and development. *Annu. Rev. Environ. Resour.*, 32:421–449.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American statistical Association*, 91(434):473–489.

Russell, H., Leschke, J., and Smith, M. (2020). Balancing flexibility and security in europe? the impact of unemployment on young peoples' subjective well-being. *European Journal of Industrial Relations*, 26(3):243–261.

Schmitz, A. and Grillon, C. (2012). Assessing fear and anxiety in humans using the threat of predictable and unpredictable aversive events (the npu-threat test). *Nature protocols*, 7(3):527–532.

Sen, A. (1982). *Poverty and famines: an essay on entitlement and deprivation*. Oxford university press.

Steptoe, A., Deaton, A., and Stone, A. A. (2015). Subjective wellbeing, health, and ageing. *The Lancet*, 385(9968):640–648.

Stevenson, E., Ambelu, A., Caruso, B., Tesfaye, Y., and Freeman, M. (2016). Community water improvement, household water insecurity, and women's psychological distress: an intervention and control study in ethiopia. *PloS one*, 11(4):e0153432.

Stoler, J., Pearson, A. L., Staddon, C., Wutich, A., Mack, E., Brewis, A., Rosinger, A. Y., Adams, E., Ahmed, J. F., Alexander, M., et al. (2020). Cash water expenditures are associated with household water insecurity, food insecurity, and perceived stress in study sites across 20 low-and middle-income countries. *Science of the Total Environment*, 716:135881.

Subbaraman, R., Nolan, L., Sawant, K., Shitole, S., Shitole, T., Nanarkar, M., Patil-Deshmukh, A., and Bloom, D. E. (2015). Multidimensional measurement of household water poverty in a mumbai slum: Looking beyond water quality. *PloS one*, 10(7):e0133241.

Sultana, F. (2011). Suffering for water, suffering from water: Emotional geographies of resource access, control and conflict. *Geoforum*, 42(2):163–172.

Tay, L., Zyphur, M., and Batz, C. (2018). Income and subjective well-being: Review, synthesis, and future research. *Handbook of well-being. Salt Lake City, UT: DEF Publishers. DOI: nobascholar. com.*

Thomas, D., Witoelar, F., Frankenberg, E., Sikoki, B., Strauss, J., Sumantri, C., and Suriastini, W. (2012). Cutting the costs of attrition: Results from the indonesia family life survey. *Journal of Development Economics*, 98(1):108–123.

Tomini, F., Tomini, S. M., and Groot, W. (2016). Understanding the value of social networks in life satisfaction of elderly people: a comparative study of 16 european countries using share data. *BMC geriatrics*, 16(1):1–12.

Tov, W. and Au, E. (2013). Comparing well-being across nations: Conceptual and empirical issues.

Tsai, A. C., Kakuhikire, B., Mushavi, R., Vořechovská, D., Perkins, J. M., McDonough, A. Q., and Bangsberg, D. R. (2016). Population-based study of intra-household gender differences in water insecurity: reliability and validity of a survey instrument for use in rural uganda. *Journal of water and health*, 14(2):280–292.

UNDP (2006). Human development report 2006-beyond scarcity: Power, poverty and the global water crisis. *UNDP Human Development Reports (2006)*.

Van Buuren, S. (2018). *Flexible imputation of missing data*. CRC press.

Veenhoven, R. (1991). Is happiness relative? *Social indicators research*, 24(1):1–34.

WHO (2013). How much water is needed in emergencies. *Technical notes on drinking-water, sanitation and hygiene in emergencies*, 9:1–4.

WHO and UNICEF (2017). Progress on drinking water, sanitation and hygiene: 2017 update and sdg baselines.

Wutich, A. (2007). 2.1 vulnerability, resilience, and robustness to urban water scarcity: A case from cochabamba, bolivia. *Perspectives on Social Vulnerability*, 6:62–71.

Wutich, A. (2009). Intrahousehold disparities in women and men's experiences of water insecurity and emotional distress in urban bolivia. *Medical anthropology quarterly*, 23(4):436–454.

Wutich, A. and Ragsdale, K. (2008). Water insecurity and emotional distress: coping with supply, access, and seasonal variability of water in a bolivian squatter settlement. *Social science & medicine*, 67(12):2116–2125.

Yap, S. C., Wortman, J., Anusic, I., Baker, S. G., Scherer, L. D., Donnellan, M. B., and Lucas, R. E. (2017). The effect of mood on judgments of subjective well-being: Nine tests of the judgment model. *Journal of Personality and Social Psychology*, 113(6):939.

Young, S. L., Boateng, G. O., Jamaluddine, Z., Miller, J. D., Frongillo, E. A., Neilands, T. B., Collins, S. M., Wutich, A., Jepson, W. E., and Stoler, J. (/19a). The household water insecurity experiences (hwise) scale: Development and validation of a household water insecurity measure for low-income and middle-income countries. *BMJ Global Health*, 4(5):e001750.

Young, S. L., Collins, S. M., Boateng, G. O., Neilands, T. B., Jamaluddine, Z., Miller, J. D., Brewis, A. A., Frongillo, E. A., Jepson, W. E., Melgar-Quiñonez, H., et al. (/19b). Development and validation protocol for an instrument to measure household water insecurity across cultures and ecologies: The household water insecurity experiences (hwise) scale. *BMJ open*, 9(1):e023558.