The economic value of improved air quality in urban Africa: a contingent valuation survey in Douala, Cameroon

HERMANN PYTHAGORE PIERRE DONFOUET

CREM, UMR 6211, University of Rennes I, 7 Place Hoche, 35 065 Rennes Cedex, France. Email: donfouetz@yahoo.fr

JOSEPH COOK

Evans School of Public Affairs, University of Washington, USA. Email: jhcook@u.washington.edu

P. WILNER JEANTY

Kinder Institute for Urban Research & Hobby Center for the Study of Texas, Rice University, USA. Email: pwjeanty@rice.edu

ABSTRACT. This study investigates the effects of 'time to think' and 'ballot box' on willingness-to-pay, while providing the first empirical evidence on assessing the benefits of an air quality improvement program in urban Africa. Our hypothetical referendum scenario proposes to reduce the air pollution related morbidity rate in Douala, Cameroon by 25 per cent in exchange for a one-time surcharge on the electricity bill of each respondent. We find that on average WTP decreases by nearly one-fourth when allocating respondents time to think but markedly increases when we use a 'ballot box' approach allowing respondents to state their willingness privately. Our results suggest that on average households are willing to pay US$0.42 per month (0.2 per cent of household annual income). We conclude that total city-wide benefits are unlikely to exceed the costs of implementing such a program at this point, although this situation may change quickly with increasing economic growth.

1. Introduction

Urban air pollution remains an important health concern in many cities in the developing world. In sub-Saharan Africa, air quality in many cities has been deteriorating with increasing traffic volume and the use of firewood in densely populated urban neighborhoods, increasing levels of premature mortality and respiratory illness (Robinson and Hammitt, 2009). Although air quality monitoring data is scarce in these cities, it is clear that pollution levels often far exceed those in urban areas of industrialized countries. The World Health Organisation's (WHO) guideline value for mean annual PM10 concentration (suspended particulate matter less than 10 microns in diameter) is $20 \mu g/m^3$. As a point of comparison, the US and France had annual averages of 18 and $27 \mu g/m^3$ in 2008. While we do not have monitoring data for Douala, our study site, or even the country of Cameroon, two other countries in West Africa (Senegal and Ghana) had average levels five to eight times higher, at 145 and $98 \mu g/m^3$ (WHO, 2012).

In Douala, the commercial capital of and the largest city in Cameroon, vehicular traffic is the major source of air pollution (World Bank, 2004). Pollutants such as carbon monoxide, nitrogen oxide and sulphur dioxide are believed to be far from the accepted levels. Although the World Bank Clean Air Initiative program which was launched in 1998 in sub-Saharan Africa spurred some interest among policy makers for mitigation policies, particularly around vehicular traffic, there is no existing information on household preferences or willingness-to-pay (WTP) for air quality improvements. Although this information would be very useful in the context of a cost-benefit analysis of potential mitigation strategies, we know of no such study in any urban area in Africa.[1] We used the dichotomous-choice contingent valuation method (CVM) and asked 496 randomly selected respondents in the city of Douala, Cameroon whether they would be willing to pay a one-time fee on their electricity bill to reduce air pollution related morbidity by 25 per cent. In addition to providing the first empirical evidence on benefits provided by an air pollution reduction program in urban Africa, our paper contributes to the literature in two important ways.

First, we explore a long-standing concern that stated preference (SP) approaches overestimate true WTP and suffer from 'hypothetical bias' (Champ *et al.*, 1997; List and Gallet, 2001; Harrison and Rutström, 2002; Blumenschein *et al.*, 2008). SP researchers have put forward a number of techniques to obtain conservative welfare estimates, including the use of well-crafted, neutral scenario scripts, well-trained enumerators, plausible payment vehicles and 'cheap talk' scripts (Cummings and Taylor, 1999). Welfare estimates can also be adjusted *ex post* based on the certainty respondents felt in their responses (Blumenschein *et al.*, 2008). Another approach

---

[1] Because reducing mortality risk is often the most economically important component of air quality improvements, a related estimate is how households trade off income with mortality risk. Robinson and Hammitt (2009) and Yaduma *et al.* (2013) also report that little is known about this tradeoff in sub-Saharan African countries, requiring analysts to transfer benefits estimates from elsewhere.

to obtaining conservative WTP estimates has been to allow respondents 'time to think' before they state their WTP (Whittington *et al.*, 1992). These studies have generally shown that it reduces welfare estimates. For instance, in a recent cross-country study of private demand for vaccines, Cook *et al.* (2012) found that welfare estimates fell by 30–40 per cent. Another approach designed to minimize pro-social bias during in-person interviews is the 'ballot box' method, where the respondent's vote is kept private from the interviewer. We test these latter two approaches against the standard elicitation procedure. Second, we explore the stability and reliability of welfare estimates. According to Carson *et al.* (2001), reliability is an index of the reproducibility and stability of a measure. For policy purposes, it seems therefore relevant to investigate the stability of welfare estimates over time. We compare the effects of giving time to think versus these 'test-retest' approaches by giving some respondents an opportunity to revise their answers overnight. The difference between giving time to think and test-retest is clear. In the latter, enumerators are asking respondents to make a choice and then return to see if they 'changed their mind', whereas in the former they allow them time to reflect on their decision before making their choice. The two will differ more when people 'anchor' more on their first decision and do not wish to reconsider.

The next section of the paper provides a brief review of studies on air quality improvements in low-income countries. Section 3 describes the sampling procedure, experimental design and empirical model. Section 4 reports the empirical results of the study. Section 5 discusses the findings and concludes.

2. Literature review

2.1. Willingness-to-pay for air quality improvements in low-income countries

We begin with a review of the few studies measuring the value of air quality improvements to households in low-income countries. Although most of these studies involved split-sample experimental treatments, we focus here on the payment vehicles, management scenarios and best estimates of household WTP. In an air quality improvement valuation study in Sofia, Bulgaria, Wang and Whittington (1999) asked respondents whether they would vote to increase utility bills to 'implement an environmental program'. Based on the stochastic payment card approach, households in Sofia were willing to pay up to 0.35 per cent of their monthly income for a program to improve air quality; the income elasticity of WTP for air quality improvement was about 27 per cent. Afroz *et al.* (2005) estimated WTP for air quality improvements in Klang Valley, Malaysia using three elicitation approaches including open-ended question, dichotomous choice and a payment card. The air quality management program[2] funded by an

---

[2] The program includes the following measures: (i) strict enforcement of traffic regulation; (ii) creation of non-traffic areas in some parts of Klang Valley; (iii) reducing traffic congestion sectors; (iv) decreasing the use of gasoline and diesel; (v) increasing the use of public transportation in urban areas; (vi) installation of

increase in fuel prices proposed a 20 per cent reduction in the concentration of particulate matter to make the air quality consistent with the Malaysian air quality guidelines. Mean WTP value for air quality improvement was a US$0.03 per liter increase in the fuel price. Wang *et al.* (2006) analyzed WTP for a 25 per cent reduction in 'harmful substances' in Beijing, China using a referendum format elicitation approach. Average household WTP was estimated at US$22.94 per household per year, or 0.7 per cent of household annual income. In Jinan, China, Wang *et al.* (2007) used a hypothetical scenario that aimed to change 'class 3' air quality standards in Jinan to 'class 2' national standards of air quality. Using a sample of 1,500 residents and the open-ended elicitation format, they estimated average WTP at US$16.05 per person per year, or 0.74 per cent of household annual income.

We know of no peer-reviewed study of household WTP for air quality in Africa. The only available report is Gbinlo (2006) in Cotonou, Benin. One hundred and twenty respondents were purposively chosen along heavily congested roads, compromising the study's ability to extrapolate to all of Cotonou. Furthermore, the study used an open-ended valuation question that was framed as a voluntary contribution, an elicitation approach with incentive compatibility problems. In addition, the report provided no detail on the actual air quality scenario used.

2.2. *'Time to think' and interviewer effects*

Several studies have examined the effect of time to think on WTP by using a split-sample survey. In a study on WTP for public and private tap connections in Anambra State (southeastern Nigeria), Whittington *et al.* (1992) found that giving respondents time to think decreased WTP by approximately 37 per cent for public taps and 32 per cent for private tap connections. In contrast, Whittington *et al.* (1993) did not find that giving respondents time to think reduced demand for sanitation services in Ghana. On the other hand, Svedsater (2007) found that giving respondents time to think about hypothetical donations to an environmental program in London reduced the respondents' uncertainty and WTP. In a multi-country study of household demand for cholera and typhoid vaccines, Cook *et al.* (2012) found that giving respondents time to think reduced the probability that a respondent said he or she would buy the hypothetical vaccines. As a result, average WTP fell by approximately 40 per cent. Respondents who were given time to think were also more certain of their answers. No studies thus far have examined the potential effect of time to think on the likelihood of respondents rejecting the scenario or giving 'protest' answers. The occurrence of protest responses may fail to determine the correct economic value of the good or policy being valued (Meyerhoff and Liebe, 2008) and may bias welfare estimates if included in or excluded from the analysis.

A number of test-retest studies have measured whether responses are stable when the elicitation approach is repeated at a later time. Kealy *et al.* (1990), Loomis (1989) and McConnell *et al.* (1998) found that respondents' preferences do not change over time. However, Brouwer and Bateman

catalytic converters on all cars; and (v) increasing the use of natural gas in the transport.

(2005) in a test-retest study about WTP for flood control and wetland conservation found that WTP estimates significantly changed over time. Recently, Bedate *et al.* (2010) found a mixed result in their test-retest study of values for cultural goods (a new museum of contemporary art) in Spain: preferences were stable for visitors to the museum but not for residents of the city.

Carson *et al.* (1994) examined interviewer effects in a US damage assessment study and found no statistical difference in responses between in-person surveys administered with and without a ballot box. Leggett *et al.* (2003) had one group of respondents fill out the CV questionnaire alone and another group complete the CV questionnaire with the help of an interviewer but place their stated WTP in a ballot box. They found that WTP (for a visit to a national monument in the US) was approximately 23 per cent higher when surveys were conducted through in-person interviews with a ballot box rather than being self-administered. Subade (2005) tested a variation of the self-administered survey approach in the Philippines termed the 'drop-off' protocol. Interviewers in this approach spend time with respondents explaining the purpose of the survey and the basics of the valuation scenario before leaving the survey instrument with the respondents to complete. Interviewers return later that day or within a few days, pick up the questionnaire, and answer any questions that the respondents may have had while filling in their answers. Subade (2005) found that the WTP estimates of the respondents who received the drop-off protocol were approximately 1.5–57 per cent lower than estimates from respondents who completed conventional in-person, single-session surveys.

3. Sampling procedure, experimental design and empirical model
3.1. Sampling
Data for the study come from a face-to-face survey conducted in April 2011.[3] The survey took on average three weeks to complete and due to budget constraints we did not compensate households in any way or give them any gifts for participating in the survey. We used a three-stage cluster sampling procedure to select respondents. At the first stage, we chose three subdivisions (Douala III, Douala IV and Douala V) by weighting the probability of choosing a subdivision by its population. At the second stage, we randomly chose 18 blocks,[4] proportionate to the total number of blocks per subdivision. In the absence of a household-level census or voter registration roll, we used a systematic sampling approach to select households at the third stage: enumerators were told to walk in a given direction from a central point in the block and attempt to interview the head of household in every eighth household they encountered. If the household head was not

---

[3] We do not have monitoring data over the period of survey work. Thus, we cannot correlate survey responses to air pollution levels on the day of the survey, although that may be a very interesting point.

[4] A block is a limited portion of the territory by visible details containing in principle 700–1,100 inhabitants on average.

at home, the enumerators revisited at a different time on a different day. In total, 496 heads of household completed the survey with a refusal rate of 2 per cent.

3.2. Scenario

We split the sample of respondents into three experimental treatments, discussed in more detail below. The management scenario in all the three experimental designs was the same. The payment vehicle used was a one-time surcharge on a household's electricity bill. During the focus group discussions and pre-test, we devoted a considerable amount of time to the choice of an appropriate payment vehicle. Since vehicular traffic is believed to be a major contributor to air pollution in Douala, we investigated the possibility of using an increase in fares for taxis and minibuses (the most common form of transportation in Douala) as a payment vehicle. We found that an increase in the monthly electricity bill was the most credible vehicle for participants, and almost all households in Douala have electricity connections. It also avoided the problem of asking private car owners about fare increases for services they do not use and the problem of potential quantity changes in minibus trips that would complicate our welfare calculations. Based on results from the focus group discussions and the pre-test, we chose the following four bid levels: 200, 350, 500 and 1,000 CFA francs.[5] These bids represent an increase of 4, 6, 9 and 18 per cent of the population's average monthly electricity bill (5,658 CFA francs). Each bid was randomly assigned to the respondent.

The management scenario involved a number of activities to improve air quality in the city, including tree planting, buying old polluting vehicles at market value to remove them from the vehicle fleet, providing subsidies for the purchase of new fuel-efficient, low-emission minibuses, and reducing traffic congestion (the full text is provided in the online appendix, available at http://journals.cambridge.org/EDE). The program also included funding for new air quality monitoring stations. A new committee elected by residents would administer the money collected from the one-time surcharge on electricity bills. If implemented, respondents were told that with the program 'the number of people who get sick because of breathing problems will be decreased by 25 per cent'. We chose this metric because we know of no current data on ambient air quality in the city (the most recent report was conducted by the World Bank, 2004) and because focus group participants had an easier time understanding and believing a morbidity-based metric than a pollutant-concentration metric. The questionnaire also asked about household demographics, health status and attitudes towards air pollution. Guidelines provided by Arrow *et al.* (1993) and Whittington (1998, 2002) were followed, and a 'consequentialism' script was integrated in the CV scenario (Bulte *et al.*, 2005).[6] The full bilingual

---

[5] US$1 = 534 CFA francs.

[6] This script is: 'Lastly, before you tell us how you would vote on a possible referendum, you should consider that the results of the study will be made available to policymakers, and could serve as a guide for future decisions. In other words,

text of the survey instrument is available at the second author's faculty website.

3.3. Experimental treatments

The first subsample of respondents, whom we refer to as the 'control', answered a single-bounded dichotomous choice (SBDC) question about the air quality management program (figure 1). This was followed by a follow-up certainty question with two categories: 'probably sure' and 'definitely sure'. They described the main reason for their votes, which we use to identify 'protest' votes. They also completed a payment card exercise that used the colors of traffic lights to communicate uncertainty (Whittington *et al.*, 2008; Cook *et al.*, 2012; also see the online appendix to this paper for more details on stoplight), although we will not discuss the results from this exercise in the paper.[7]

A second subsample of respondents received the same management program but were given overnight to think about their votes. These 'time to think' (TTT) respondents were also given a specific referendum price and encouraged to discuss it with their spouse, friends and neighbors (the full text of the 'time to think' text is given in the online appendix). Enumerators recorded households' decisions during the second interview the next day as well as the main reason behind their votes. Because of an oversight in the survey implementation, TTT respondents were not asked how certain they were of their responses. We describe the effects of this decision on our analysis below.

Respondents in the 'ballot box' subsample were given the same scenario but asked to mark their responses to the dichotomous choice question on a card that was put in a sealed envelope to avoid pro-social interviewer bias. The enumerators informed subjects that their responses would be kept private. Respondents in this subsample also marked their responses to the certainty question. Because enumerators could not observe the votes and not all respondents could read, these respondents were not asked for the main reason why they voted the way they did and we therefore cannot identify 'protest' votes for this subsample. This 'ballot box' subsample also did the stoplight exercise and provided their responses in a sealed envelope. More specifically, the enumerators explained to respondents how the air quality in Douala could be improved, the ballot box exercise which contains the two formats (SBDC and stoplight) and they gave them two printed cards[8] with the prices: one printed card for the SBDC format and the other

---

the results from the study will have an actual effect, and you should consider this when answering the questions below.'

[7] Although all respondents should have been asked to complete the stoplight experiment, in two of three experimental subsamples only respondents who said 'yes' to the SBDC question were asked to do so. These results are thus less useful for policy, and comparisons across subsamples are more complicated. We report these results in the online appendix for interested readers.

[8] For the respondents who could not read, the two printed cards contain colors and enumerators had to explain the meaning of these colors for these respondents. They also provided felt-tip pens to them and asked them to put their response in
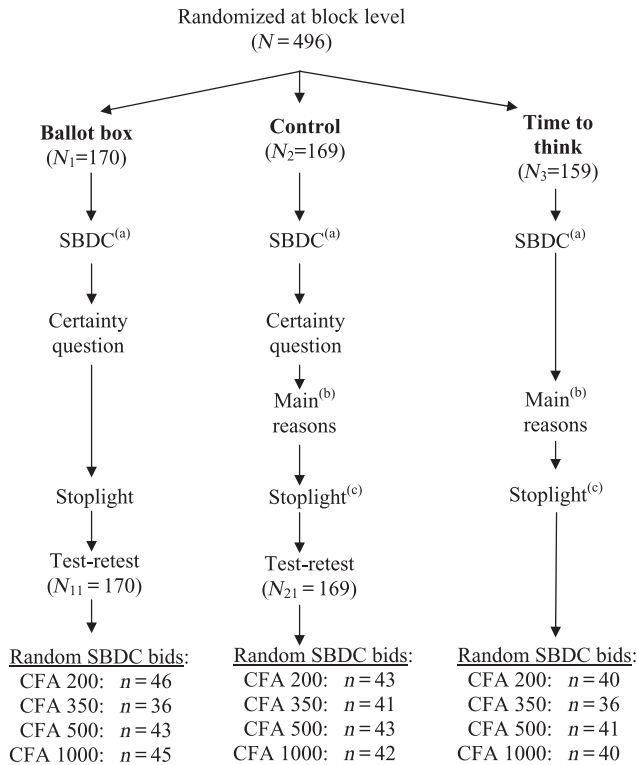
Randomized at block level
($N = 496$)

Ballot box	Control	Time to think
(N_1=170)	(N_2=169)	(N_3=159)
SBDC[a]	SBDC[a]	SBDC[a]
Certainty question	Certainty question	
	Main[b] reasons	Main[b] reasons
Stoplight	Stoplight[c]	Stoplight[c]
Test-retest (N_{11} = 170)	Test-retest (N_{21} = 169)	

Random SBDC bids:	Random SBDC bids:	Random SBDC bids:
CFA 200: $n = 46$	CFA 200: $n = 43$	CFA 200: $n = 40$
CFA 350: $n = 36$	CFA 350: $n = 41$	CFA 350: $n = 36$
CFA 500: $n = 43$	CFA 500: $n = 43$	CFA 500: $n = 41$
CFA 1000: $n = 45$	CFA 1000: $n = 42$	CFA 1000: $n = 40$

Figure 1. *Research design*
Notes: (*a*) and (*b*) refer to the single bounded-dichotomous choice and the main reasons to pay and not to pay, respectively. (*c*) Only respondents who answered 'yes' to the SBDC question were asked to complete the stoplight exercise for the control and TTT subgroups (unintentionally).

printed card for the stoplight. They were then asked to mark it themselves and put their answers in the sealed envelope.

To explore the stability of WTP over time, enumerators returned to interview the 'control' and 'ballot box' respondents the next day. They were asked the same dichotomous choice question with the same referendum price. In total, 169, 157 and 170 heads of household received the control, TTT and ballot box treatments. These three experimental designs were randomized at the block level rather than the household level for administrative simplicity; all households within a given block completed the same experimental treatment. Because each subsample was split by four referendum surcharges, the sample size in each cell varied from 36 to 46 (see figure 1).

a sealed envelope once they had finished. We included the two printed cards in the full bilingual text of the survey.

3.4. Empirical model

We analyze responses to the dichotomous choice question following Cameron and James (1987) and Haab and McConnell (2002). We elect to use a linear rather than a lognormal specification to allow for zero or negative WTP arising from protest respondents. We also address the issue of negative WTP by restricting mean WTP to lie between zero and the highest bid presented to respondents. Thus, we report the truncated mean WTP in the text (table 4) and the untruncated mean WTP in the online appendix. The untruncated mean is generally less than or equal to the truncated mean (Haab and McConnell, 1997; Johansson *et al.*, 1989). We also recode 'don't know' responses as a 'no' (Carson *et al.*, 1998; Groothuis and Whitehead, 2002). The standard errors of mean WTP were calculated using the delta method.

Drawing upon Cameron (1988) and Cameron and Huppert (1989), we estimate mean WTP from the stoplight/payment card exercise using a parametric interval regression model. We also estimate lower bound and midpoint non-parametric WTP measures following Turnbull (1976) and Kristrom (1990), and calculate standard errors of non-parametric WTP using the method of Vaughan and Rodriguez (2001).

The presence of protest responses in the data warrants special attention. As mentioned above, inclusion of protest responses in the analysis may bias average welfare estimates downwards if respondents are answering 'no' because of problems with the scenario rather than because they are not willing to pay for the specified change. If the protest responses are non-random, their exclusion may also yield biased estimates. Exclusion of protest responses is only appropriate when the protest respondents are not significantly different from the rest of the sample (Strazzera *et al.*, 2003). We carry out a sensitivity analysis to explore the change in welfare estimates when protesters are included in or dropped from the analysis. We report the results from estimation with the protest respondents in the text, but relegate results from the estimation without the protest respondents to the online appendix (see tables A8, A9, A11, A12, A14 and A15).

4. Results
4.1. Raw data

Table 1 provides the background characteristics of the respondents in the three experimental groups. A representative respondent in our sample is a 45-year-old male in a household with five members with a mean household income of approximately 140,000 CFA francs (US$262) per month. He is 'very concerned' about air pollution in the city but is unlikely to have a household member with respiratory problems. By random chance we have some statistically significant differences between subsamples in age, education, household size and the presence of respiratory problems. We expect these characteristics to be significant predictors of WTP for air quality improvements. Although we control for these differences in the multivariate regression analysis, they do complicate comparisons of responses and non-parametric WTP estimates.

Table 1. *Background characteristics*

Variables	Description	Control (n = 169)	Time to think (n = 159)	Ballot box (n = 170)	p-value of control vs. TTT	p-value of control vs. ballot
Male	Percent male	0.72	0.68	0.74	0.50	0.60
Age	Age[a]	46 (11.79)	45 (12.11)	43 (10.43)	0.49	0.00***
Income	Reported monthly income, CFA francs[a]	132,367 (61.16)	141,927 (59.64)	137,618 (58.55)	0.15	0.42
HHnumber	Number of people in household	5.36 (2.51)	4.02 (2.65)	4.53 (2.93)	0.00***	0.00***
Education low	= 1 if the respondent has completed 0–7 years of school; otherwise 0	0.18	0.09	0.08	0.00***	0.01**
Education mid	= 1 if the respondent has completed 8–14 years of school; otherwise 0	0.56	0.48	0.53	0.11	0.55
Education high	= 1 if the respondent has been to university; otherwise 0	0.25	0.43	0.39	0.00***	0.00***
Respiratory	Respiratory diseases reported in household (1 if found; otherwise 0)	0.20	0.10	0.18	0.01**	0.66
ConcernAirpoll	Concern of the respondent about air pollution (1 if the respondent is very concern; otherwise 0)	0.85	0.88	0.82	0.38	0.48

Notes: Standard deviations are in parentheses. *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.
[a] At the pre-test stage, many respondents were hostile in declaring their age and income, so we used ranges. We report the midpoint here. The total sample size is $N = 496$.

Table 2. *Percentage of 'yes' responses, by bid and experimental treatment*

	Percentage of 'yes' responses			p-value	
Bid (FCA)	Control	Time to think	Ballot box	Control versus TTT	Control versus ballot box
200	51	35	54	0.07*	0.6
350	51	33	66	0.06*	0.09*
500	30	27	51	0.4	0.02**
1000	7	8	29	0.5	0.00***

Notes: *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$. Respondents answering 'don't know' recoded as 'no'. Responses are *not* adjusted for certainty (e.g. probably sure, definitely sure, not sure at all).

We begin by comparing the percentage of respondents who voted 'yes' to the SBDC question across the three subgroups. We start without adjusting for respondent uncertainty (the TTT group was not asked the uncertainty questions) or protest responses (the ballot-box group could not complete the question on the main reason for the vote). Table 2 shows that the percentage of respondents voting yes declines monotonically with price for both the control and TTT groups, and is generally higher among those in the 'control' group than those given time to think. The differences are not, however, statistically significant at each price. Among ballot box respondents, the percent yes declines with higher bids except for the second bid level (350 CFA francs). Contrary to our prior expectations, it is higher than the control group for three of the four bid levels, and substantially different at the highest bid (29 vs. 7 per cent yes in the control).

We refine these comparisons by adjusting for protest responses in the control and TTT subsamples. (Because illiterate ballot box respondents were unable to report the 'main reason' for their vote we cannot identify protest responses in that subsample.) We consider protest responses as those where the respondent answered 'no' to the offered bid and gave a reason that we believe indicated scenario rejection. The pre-coded options for respondents answering 'no' were: '(1) I don't trust the people that will manage the fund; (2) The environment is clean enough; (3) I am afraid of the repercussions of the program (it will increase my expenditures); (4) I don't want such policy; (5) The government must search for another policy; (6) I would vote if the fixed surcharge is lower ; (7) I really want to vote but I don't think that my vote will count; (8) I don't know; and (9) Other reasons (please specify).' We consider responses (1), (4), (5) and (7) to be potential protest responses, although no one answered (7) in the field. Assuming responses (1) and (4) are protests, 24 per cent of responses in the control treatment were protests compared to 11 per cent in the TTT subsample, a statistically significant difference.[9] After dropping those protest responses, the percent 'yes' remains monotonically declining with increasing bids for

---

[9] To further explore whether TTT may be associated with the probability of a respondent protesting the scenario, we pooled the two data sets (control and TTT)

Table 3. *Percentage of 'yes' responses in the control and ballot box treatments after recoding 'probably sure' yes votes to 'no' votes*

Bid	Control	Ballot box	Control versus ballot box (*p-value*)
200	51	54	0.81
350	51	44	0.51
500	30	40	0.36
1000	7	4	0.6

the TTT sample and lower at every price than the control group (details are provided in online appendix table A1). The difference in the percent 'yes' between the control and TTT groups is statistically significant at the two middle bid levels. If we expand the definition of protest to include 'the government must search for another policy', the percentage of protest votes rises to 33 and 30 per cent in the control and TTT groups, respectively. Pairwise differences are again significant at the two middle bid levels (table A2).

Finally, we account for respondent uncertainty in the control and ballot-box subsamples by recoding all 'yes' answers where the respondent was 'probably sure' to 'no' answers. (Again, because of an implementation oversight we did not collect certainty information on TTT respondents.) Table 3 reports these adjusted 'percent yes' numbers. Both remain monotonically declining with the bids, but the difference between the ballot box and control groups is now much smaller, especially at the highest price. Because no-one in the 'control' group reported being 'probably sure' about their vote, this correction affects only the ballot box group.

4.2. Multivariate analysis
We estimate multivariate probit models to analyze the decision to vote for the program. Because we do not observe protest responses in the ballot box treatment, for consistency we do not drop protesters/scenario rejecters in these models but treat them all as simple 'no' responses, possibly underestimating population WTP. The coefficients for the bid (the electricity surcharge) and income are both highly statistically significant and of the expected sign. Consistent with previous studies, the coefficient on time to think is negative and statistically significant (table 4). Respondents who completed 8–14 years of schooling are less likely to vote for the air quality program than those with a university degree. When we drop protest responses in the control and TTT subsamples, the coefficients on time to

and estimated a probit model of whether a response was a protest with a TTT dummy variable and other observable characteristics as explanatory variables. Respondents who had time to think were 11 per cent less likely to protest than the control respondents, an effect significant at the 5 per cent level (see table A3 in the online appendix). The difference is not statistically significant if we include 'the government must search for another policy' as another protest answer code.

think, bid, and income all remain statistically significant and have similar magnitudes (table A7).[10] After controlling for socio-economic characteristics, ballot box respondents are *more* likely to vote yes in the SBDC exercise than control respondents, although their responses are no different after adjusting for uncertainty.

4.3. Willingness-to-pay estimates

The non-parametric Turnbull 'lower bound' mean WTP from the SBDC exercise are reported in CFA francs in table 5; for international comparison these are US$0.48, US$0.37 and US$0.85 per household for the control, TTT and ballot box groups, respectively. Kristrom mid-point mean WTP is US$0.72, US$0.61, and US$1.22 for the control, TTT and ballot box groups, respectively. Time to think reduced the lower bound WTP by 24 per cent compared to the control group, and reduced the Kristrom 'mid-point' mean by 16 per cent. A *t*-test of differences in mean WTP is significant at the 1 per cent level, and 95 per cent confidence intervals calculated following Vaughan and Rodriguez (2001) do not overlap. The ballot box treatment increased Turnbull WTP by 74 per cent, although the difference is much smaller after correcting for uncertainty (by recoding 'probably sure' votes as 'no' votes). These WTP figures are underestimated since we treat all 'no' responses the same, even though some of the respondents answering 'no' might have in fact been willing to pay something to reduce air pollution with a different program. Excluding protest responses, the Turnbull mean WTP is US$0.75 and US$0.48 for the control and TTT groups (table A9).

   We calculate the parametric mean WTP of the SBDC following Cameron and James (1987). Mean WTP is US$0.70, US$0.61 and US$0.93 per household for the control, TTT and ballot box subgroups, respectively (table 5). Correcting for uncertainty, the mean WTP drops to US$0.54 for the ballot box group. We also compare the mean WTP across the three treatments using a bootstrap technique (Efron and Tibshirani, 1993). We simulate mean WTP using 1,000 replications, saving the results of each simulation. We then load and merge the two data sets and calculate the differences in

---

[10] This result is robust to assumptions about how we treat possible protest responses. If TTT affected the probability of a respondent protesting the scenario, one modeling approach would be to treat the first stage (protest vs. not protest) as a selection process. We use a Heckman selection model (following Whitehead *et al.*, 1993) with two equations: (1) a participation equation where the respondent does not protest (1 if the respondent has not protested and 0 otherwise); (2) the valuation equation where the respondent decides to pay for air quality program at the offered price (1 if the respondent agrees to pay and 0 otherwise). We then use a univariate probit and bivariate probit with sample selection. The results suggest that TTT is not inducing a selection bias in the mean WTP estimates. The likelihood ratio test indicates that the null hypothesis of independence cannot be rejected at conventional levels for the univariate probit (p-value = 0.96) and the bivariate probit (p-value = 0.67) for protest responses 1 and 4. Thus, deleting protest responses may not bias the estimates of mean WTP. The same results (p-value = 0.98 and 0.48, for the univariate probit and bivariate probit, respectively) are found if responses 1, 4 and 5 are considered protests.

Table 4. *Probit model of SBDC response*

Variables	Pooling control, TTT and ballot box subgroups (S.E.)	Pooling control and ballot box subgroups based on certainty (S.E.)
TTT	−0.48***	
	(0.17)	
BALLOT	0.41***	0.18
	(0.16)	(0.14)
Bid	−0.002***	−0.002***
	(0.003)	(0.003)
Male	−0.12	−0.14
	(0.15)	(0.15)
Age	−0.002	0.002
	(0.01)	(0.01)
Income	0.01***	0.01***
	(0.001)	(0.001)
Educationlow	−0.43*	−0.39
	(0.25)	(0.25)
Educationmid	−0.61***	−0.57***
	(0.15)	(0.15)
HHnumber	0.01	0.01
	(0.03)	(0.03)
ConcernAirpoll	−0.24	0.02
	(0.21)	(0.23)
Respiratory	0.06	0.25
	(0.19)	(0.18)
Intercept	−0.01	−0.30
	(0.42)	(0.41)

Notes: *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$. TTT takes 1 and 0 for the control. Ballot box takes 1 and 0 for the control.

mean WTP. Once a difference in mean WTP is calculated, the p-value is also calculated to test the corresponding null hypothesis of equality of the mean WTP of control and TTT. The lower the p-value relative to conventional significance levels, the more likely the mean WTP in the control group is higher than the mean WTP in the TTT group. Compared to the control group, mean WTP of the TTT group is significantly lower (p-value = 0.07), and mean WTP among the ballot box group is significantly higher (p-value = 0.00). After we adjust for uncertainty in the control and ballot box samples, however, there is no statistically significant difference (p-value = 0.49) in mean WTP between the two groups.

4.4. Debriefing questions on the control and time to think study
The most common reason respondents gave for agreeing to the plan was 'good health'. Seventy-three per cent of respondents in the TTT subsample said they used the opportunity to discuss their answers with their spouses, family members, friends or neighbors. The average time respondents reported that they spent thinking about the task is 30 minutes (see

Table 5. *Mean WTP (CFA francs) for SBDC, by experimental group*

| | Non-parametric | | | | Parametric (CI[a]) | |
	Turnbull (CI)		Kriström (CI)			
Control	260	(256, 264)	387	(382, 392)	375	(324, 426)
Time to think	198	(193, 203)	327	(322, 332)	326	(275, 379)
Ballot box – original	452	(446, 458)	652	(645, 660)	496	(444, 547)
Ballot box – adjusted	255	(251, 259)	415	(411, 420)	289	(176, 403)

Notes: CI means 95% confidence intervals. US$1 = 534 CFA francs.
[a]The 95% confidence interval of the mean WTP, obtained by bootstrap on 1,000 draws.

online appendix figure A3 for the distribution). We used these debriefing questions to investigate how the elements of time to think might influence WTP responses. A continuous variable of minutes spent thinking (zero for the control group) was not statistically significant (p-value = 0.6), nor was a dummy variable for whether the respondent discussed the decision (p-value = 0.24).

4.5. Retests among control and ballot box subgroups
When given the opportunity to revise their answers after a day of reflection, only three of 169 respondents in the control group and four in the ballot box group changed their answers to the SBDC question. All seven changed their responses from 'no' to 'yes'. Not surprisingly, the multivariate results do not change (table A10). This remarkable degree of preference stability contrasts with results in Cook *et al.* (2007), where a sizeable number of control respondents revised their answers downwards and mean WTP with and without time to think became indistinguishable. Not surprisingly, WTP estimates are also very similar (see tables A16 and A17).

5. Discussions
We begin with a discussion of the methodological results before turning to how our results might be used for policy in Douala. Consistent with previous studies, we find that time to think reduces average WTP estimates compared to a control group who do the valuation exercise in the 'conventional' way. Mean parametric WTP from the SBDC responses is 13 per cent lower with time to think; non-parametric measures are 24 per cent lower (Turnbull) and 16 per cent lower (Kristrom). This result remains after controlling for other important determinants of WTP and after adjusting for possible protest responses, although we are unable to test whether adjusting the respondents' certainty in their answers would affect this result. Consistent with a number of test-retest studies but inconsistent with Cook *et al.* (2007), however, we find that responses and mean WTP are stable when control (and ballot box) respondents are given a chance to reconsider their responses. This result is somewhat puzzling, since these respondents would have had the same opportunity as TTT respondents to reflect on

their budget constraints and discuss the decision with their spouses. They also had one day to reflect, and the enumerator primed them to reconsider their answers.[11]

Why did they not revise their answers, and in particular why did they not revise them downwards? One obvious explanation is that control and ballot box respondents anchored on their initial decisions. Unlike Cook *et al.* (2007), our respondents did a stoplight exercise in the first interview that asked them to think carefully about the range of prices they would pay, which may have further anchored responses. Even though they were given the opportunity to reflect, they may not have used it because they felt they had already completed the task; the cognitive costs were not worth the benefit. Unfortunately, we did not ask control or ballot box respondents if they discussed the decision or how long they spent thinking about the task. Unlike both Leggett *et al.* (2003) and Subade (2005), we find *higher* WTP with a ballot box approach, although the difference disappears if we correct for uncertainty by recoding 'probably sure' yes votes to no votes.

What is our best estimate of household WTP for policy purposes? We believe that time to think provides respondents with an opportunity to carefully consider their votes and their budget constraints, similar to that which would occur in a real citywide referendum on the program. While not all residents would take advantage of the opportunity to think about their votes, many would. We therefore feel most confident in these time to think estimates.

With regard to estimation technique, the non-parametric approach yields a lower bound and a midpoint WTP estimate of 225 CFA francs (US$0.42) and 360 CFA francs (US$0.67), while the WTP estimate from the parametric approach is 353 CFA francs (US$0.66). These results exclude protest responses (see tables A8 and A11) but also do not adjust for certainty. In the benefit-cost calculations below, we use the more conservative, less bias-prone estimate of 225 CFA francs (US$0.42), representing 0.2 per cent of average household annual income in our sample. Using this estimate, the total annual citywide benefit for a 25 per cent reduction in the health effects of air quality is 400 million CFA francs (US$749,064).[12]

We do not have the ability to identify exactly what policies would lead to a 25 per cent reduction in air quality related morbidity in Douala. Nevertheless, a rough comparison with some of the costs of the program

---

[11] Enumerators finished the first interview by saying: 'I'd like to give you another opportunity to answer the question I asked about whether you would vote for the one-time surcharge of FCFA ⸺ to improve air quality by 25 per cent after you have had more time to think about it. Feel free to discuss the environmental program, what you would have to pay for the air quality program, and other things that your household needs to spend money on with your family member, neighbor and colleagues. It is very important for our research that we have this second interview tomorrow.'

[12] The total benefit is calculated by multiplying mean WTP by the total number of households in Douala by the fraction of respondents who were 'in the market'. Thus, the total benefit is 2 million households * 0.89 * 225 = 400.5 million CFA francs.

described in the scenario may be illustrative. Using figures from the Douala Urban Council, we estimate a program to plant 1,000 seedlings might cost 7.6 million CFA francs (US$14,217).[13] The Cameroon transportation bureau estimated that there were approximately 2,300 minibuses on Douala roads in 2000 (the most recent estimate available) that were 15 years old or older. We assume there are now 3,000 minibuses of that vintage. The current market price for an older minibus is between 3 million and 4 million CFA francs (between US$5,618 and US$7,491). The cost to remove all 3,000 minibuses from the road would thus be 9–12 billion CFA francs (US$16.9 million–US$22.5 million). This investment in trees and minibus buybacks would have effects for more than just one year, however. To make them more commensurate with our annual estimate of benefits, we convert the costs to annual values by assuming that the tree planting and buybacks will have effects over 10 years and that the social discount rate is 10 per cent. The annualized cost of the tree planting program is thus approximately 1.24 million CFA francs, and the minibus buyback is 1.46–1.95 billion CFA francs (US$2.7 million–US$3.7 million). Our conservative estimate of citywide WTP for the air quality program (400 million CFA francs or US$749,064) is clearly insufficient to fund a complete buyback program. It would be sufficient to purchase approximately one-quarter of old minibuses at a price of 3 million CFA francs per bus,[14] although again we do not know whether this program would be sufficient to cause a 25 per cent drop in air pollution morbidity. Using our least conservative WTP estimate of 652 CFA francs (US$1.22) per household (from the Kristrom midpoint estimate in the ballot box sample, before correcting for uncertainty) still does not yield sufficient benefits to cover the entire cost of the program, though at US$2.17 million it is much closer.[15]

Supplementary materials and methods

The supplementary material referred to in this paper can be found online at journals.cambridge.org/EDE.

---

[13] The price of a sapling is 7,500 CFA francs. At 15 minutes per tree, and with a seven-hour working day, it would take 36 person-days to plant the trees, valued at the municipal wage rate of 2,000 CFA francs per day. We estimate total costs to transport trees to planting sites (within 15 km of the city center) to be 20,000 CFA francs.

[14] This calculation again uses an annualized cost estimate, however. If the 3 million purchase price is spread over 10 years at a 10 per cent discount rate, the annual cost is 490,000 CFA francs. 400 million/490, 000 = 816 buses. Household WTP estimates from the control group are in the range of 260–387 CFA francs (see table 5), implying citywide benefits from 462 million to 689 million CFA francs. This lower bound estimate (based on the Kristrom midpoint from the SBDC responses) would allow the purchase of almost half of old minibuses (689 million/490, 000 = 1, 406 buses).

[15] Using a similar calculation to that above, total benefits would be 2 million households * 0.89 in the market * 652 CFA francs = 1,161 billion CFA francs, or US$2.173 million.

References

Afroz, R., M.N. Hassan, M. Awang, and N.A. Ibrahim (2005), 'Willingness to pay for air quality improvements in Klang Valley Malaysia', *American Journal of Environmental Sciences* **1**(3): 194–201.

Arrow, K., P.R. Solow, E.E. Leamer, R. Radner, and H. Shuman (1993), 'Report of NOAA panel on contingent valuation method', *Federal Register* **58**(10): 4601–4614.

Bedate, A.M., L.C. Herrero, and J.A. Sanz (2010), 'Ex ante and ex post valuations of a cultural good. Are preferences or expectations changing?', *Journal of Environmental Planning and Management* **55**(1): 127–140.

Blumenschein, K., G.C. Blomquist, M. Johannesson, N. Horn, and P. Freeman (2008), 'Eliciting willingness to pay without bias: evidence from a field experiment', *Economic Journal* **118**(525): 114–137.

Brouwer, R. and I.J. Bateman (2005), 'Temporal stability and transferability of willingness to pay for flood control, and wetland conservation', *Water Resources Research* **41**(3): 1–6.

Bulte, E., S. Gerking, J.A. List, and A. de Zeeuw (2005), 'The effect of varying the causes of environmental problems on stated values: evidence from a field study', *Journal of Environmental Economics and Management* **49**(2): 330–342.

Cameron, T.A. (1988), 'A new paradigm for valuing non-market goods using referendum data: maximum likelihood estimation by censored logistic regression', *Journal of Environmental Economics and Management* **15**(3): 355–379.

Cameron, T.A. and D.D. Huppert (1989), 'OLS versus ML estimation of non-market resource values with payment card interval data', *Journal of Environmental Economics and Management* **17**(3): 230–246.

Cameron, T.A. and A. James (1987), 'Efficient estimation methods for "closed-ended" contingent valuation surveys', *Review of Economics and Statistics* **69**(2): 269–276.

Carson, R.T., M. Hanemann, R.J. Kopp, J.A. Krosnick, R.C. Mitchell, S. Presser, P.A. Ruud, and V.K. Smith (1994), 'Prospective interim lost use value due to DDT and PCB contamination in the southern California bight', Report to the National Oceanic and Atmospheric Administration pursuant to Task Order 56-DGNC-2-50070: Natural Resource Damage Assessment, La Jolla, CA: Natural Resource Damage Assessment.

Carson, R.T, W.M. Hanemann, R.J. Kopp, *et al.* (1998), 'Referendum design and contingent valuation: the NOAA panel's no-vote recommendation', *Review of Economics and Statistics* **80**(3): 484–487.

Carson, R.T., N.E. Flores and N.F. Meade (2001), 'Contingent valuation: controversies and evidence', *Environmental and Resources Economics* **19**(2): 173–210.

Champ, P.A., R.C. Bishop, T.C. Brown, and D.W. McCollum (1997), 'Using donation mechanisms to value nonuse benefits from public goods', *Journal of Environmental Economics and Management* **33**(2): 151–162.

Cook, J., D. Whittington, D.G. Canh, F.R. Johnson, and A. Nyamete (2007), 'The reliability of stated preferences for cholera and typhoid vaccines with time to think in Hue, Vietnam', *Economic Inquiry* **45**(1): 100–114.

Cook, J., M. Jeuland, B. Maskery, and D. Whittington (2012), 'Giving stated preference respondents "time to think": results from four countries', *Environmental and Resource Economics* **51**(4): 473–496.

Cummings, R. and L.O. Taylor (1999), 'Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method', *American Economic Review* **89**(3): 649–665.

Efron, B. and R.J Tibshirani (1993), *An Introduction to the Bootstrap*, New York: Chapman and Hall.

Gbinlo, E. (2006), 'Evaluation du coût social de la pollution de l'air par les taxis motos à Cotonou, Bénin', Institut de l'énergie et de l'environnement de la Francophonie (IEPF), Quebec.

Groothuis, P.A. and J.C. Whitehead (2002), 'Does don't know mean no? Analysis of 'don't know' responses in contingent valuation questions', *Applied Economics* **34**(15): 1935–1940.

Haab, T.C. and K.E. McConnell (1997), 'Alternative methods for handling negative willingness to pay in referendum models', *Journal of Environmental Economics and Management* **32**(2): 251–270.

Haab, T.C. and K.E. McConnell (2002), *Valuing Environmental and Natural Resources: The Econometrics of Non Market Valuation*, Cheltenham: Edward Elgar.

Harrison, G.W. and E.E. Rutström (2002), 'Experimental evidence on the existence of hypothetical bias in value elicitation methods', in V.L. Smith (ed.), *Handbook of Results in Experimental Economics*, New York: Elsevier Science.

Johansson, P., B. Kristrom, and K.G. Maler (1989), 'Welfare evaluations in contingent valuation experiments with discrete response data: comments', *American Journal of Agricultural Economics* **71**(4): 1054–1046.

Kealy, M.J., M. Montgomery, and J.F. Dovidio (1990), 'Reliability and predictive validity of contingent valuation: does the nature of the good matter?', *Journal of Environmental Economics and Management* **19**(3): 244–263.

Kristrom, B. (1990), 'A nonparametric approach to the estimation of welfare measures in discrete response valuation studies', *Land Economics* **66**(2): 135–139.

Leggett, C.G., N.S. Kleckner, K.J. Boyle, J.W. Dufield, and R.C. Mitchell (2003), 'Social desirability bias in contingent valuation surveys administered through in-person interviews', *Land Economics* **79**(4): 561–575.

List, J.A. and C.A. Gallet (2001), 'What experimental protocol influence disparities between actual and hypothetical stated values?', *Environmental and Resource Economics* **20**: 241–254.

Loomis, J.B. (1989), 'Test-retest reliability of the contingent valuation method: a comparison of general population and visitor responses', *American Journal of Agricultural Economics* **71**(1): 76–84.

McConnell, K.E., I.E. Strand, and S. Valdes (1998), 'Testing temporal reliability and carry-over effect: the role of correlated responses in test-retest reliability studies', *Environmental and Resource Economics* **12**(3): 357–374.

Meyerhoff, J. and L. Liebe (2008), 'Do protest responses to a contingent valuation question and a choice experiment differ?', *Environmental and Resource Economics* **39**(4): 433–446.

Robinson, L.A. and J.K. Hammitt (2009), *Final Report of the Sub-saharan Africa Refinery Project Health Study: Volume I-A*, Washington, DC: World Bank.

Strazzera, E., M. Genius, R. Scarpa, and G. Hutchinson (2003), 'The effect of protest votes on the estimates of WTP for use values of recreational sites', *Environmental and Resource Economics* **25**(4): 461–476.

Subade, R.F. (2005), *Valuing Biodiversity Conservation in a World Heritage Site: Citizens' Non-use Values for Tubbataha Reefs National Marine Park, Philippines*, Singapore: Economy and Environment Program for Southeast Asia.

Svedsater, H. (2007), 'Ambivalent statements in contingent valuation studies: inclusive response formats and giving respondents time to think', *Australian Journal of Agricultural and Resource Economics* **51**(1): 91–107.

Turnbull, B.W. (1976), 'The empirical distribution function with arbitrarily grouped, censored, and truncated data', *Journal of the Royal Statistical Society* **38**(B): 290–295.

Vaughan, W.J. and D.J. Rodriguez (2001), 'Obtaining welfare bounds in discrete-response valuation studies: comment', *Land Economics* **77**(3): 457–465.

Wang, H. and D. Whittington (1999), *Willingness to Pay for Air Quality Improvements in Sofia, Bulgaria*, Washington, DC: World Bank.

Wang, X.J., W. Zhang, Y. Li , K.Z. Yang, and M. Bai (2006), 'Air quality improvement estimation and assessment using contingent valuation method, a case study in Beijing', *Environmental Monitoring Assessment* **120**(1–3): 153–168.

Wang, Y., Y. Zhang, Q. Wang, and W. Wang (2007), 'Residents' willingness to pay for improving air quality in Jinan, China', *Chinese Journal of Population, Resources and Environment* **5**(2): 12–19.

Whitehead, J.C., P.A. Groothuis, and G.C. Blomquist (1993), 'Testing for non-response and sample selection in contingent valuation', *Economics Letters* **41**(2): 215–220.

Whittington, D. (1998), 'Administering contingent valuation surveys in developing countries', *World Development* **26**(1): 21–30.

Whittington, D. (2002), 'Improving the performance of contingent valuation studies in developing countries', *Environmental and Resource Economics* **22**(1): 323–367.

Whittington, D., S.V. Kerry, A. Okorafor, A. Okore, L.J. Long, and A. McPhail (1992), 'Giving respondents time to think in contingent valuation studies: a developing country application', *Journal of Environmental Economics and Management* **22**(2): 205–225.

Whittington, D., D.T. Lauria, A.M. Wright, K. Choe, J.A. Hughes, and V. Swarna (1993), 'Household demand for improved sanitation services in Kumasi, Ghana: a contingent valuation study', *Water Resource Research* **29**(6): 1539–1560.

Whittington, D., D. Sur, J. Cook, *et al.* (2008), 'Rethinking cholera and typhoid vaccination policies for the poor: private demand in Kolkata, India', *World Development* **37**(2): 399–409.

WHO (2012), *World: Exposure to Particulate Matter with an Aerodynamic Diameter of 10 µm or Less (PM10) in 1100 Urban Areas, 2003–2010*, Geneva: World Health Organisation.

World Bank (2004), *Etude sur la qualité de l'air en milieu urbain: Le cas de Douala*, Washington, DC: World Bank.

Yaduma, N., M. Kortelainen, and A. Wossink (2013), 'Estimating mortality and economic costs of particulate air pollution in developing countries: the case of Nigeria', *Environmental and Resource Economics* **54**(3): 361–387.