

https://dspace.org/

Fuel stacking housing quality and health disparities in rural South Africa A double machine learning approach

Nouran Zenelabden; Adeola Oyenubi; Johane Dikgang

ELSEVIER SCIENCE BV
https://hdl.handle.net/10539/47054
WOS

Downloaded from a DSpace repository. https://dspace.org/

Energy Economics 151 (2025) 108926



Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneeco





Fuel stacking, housing quality, and health disparities in rural South Africa: A double machine learning approach

Nouran Zenelabden [a,*], Adeola Oyenubi [a], Johane Dikgang [b,c]

[a] School of Economics and Finance, University of the Witwatersrand, Johannesburg, Gauteng, South Africa
[b] Department of Economics and Finance, Florida Gulf Coast University, Fort Myers, FL, USA
[c] The Water School, Florida Gulf Coast University, Fort Myers, FL, USA

ARTICLE INFO

JEL codes:
Q42
I15
R20
C55
O18
I32

Keywords:
Fuel stacking
Health disparities
Housing materials
Indoor air pollution
Double machine learning, sorted effects.

ABSTRACT

The relationship between fuel type, housing quality, and health outcomes is examined in rural South Africa using the 2019 General Household Survey dataset from Statistics South Africa (StatsSA). We apply double machine learning within a multivalued treatment effects framework to distinguish between clean, unclean, and mixed fuel types as treatment regimes, allowing us to infer health impacts from fuel-stacking behavior. Our findings show that using mixed fuels does not lead to health benefits from partially switching away from unclean fuels, which challenges common beliefs about "energy stacking" in low-income households. We also use Sorted Effects Analysis (SEA) to explore differences within the population and identify vulnerable groups most affected by the health risks of unclean fuel use. The SEA results reveal that the health effects of polluting fuels are not uniform, and there is some evidence that housing quality—such as roofing and walls—moderates the relationship between polluting fuels and health outcomes. These findings suggest that effective policies should consider the roles of fuel type, housing quality, and socioeconomic vulnerabilities to promote a fair and equitable transition to clean energy in rural South Africa.

1. Introduction

Indoor air pollution (IAP) is the most significant environmental health risk globally, and this is particularly true for individuals with low incomes. This study makes three contributions to the existing body of literature on the health risks associated with energy poverty in developing settings. The first innovative aspect of our analysis is that we conceptualize fuel stacking as a multivalued treatment instead of relying on the binary energy-choice framework prevalent in the current literature. While the simultaneous combustion of different fuels is said to help users move towards cleaner energy (Masera et al., 2000; Ruiz-Mercado and Masera, 2015), we instead question the hypothesized benefit of such behavior by asking whether the use of all the available fuels is beneficial for health rather than using the unclean fuel solely.

Secondly, we focus on the housing quality as a significant moderator of the relationship between fuel use and health. Through the interactions of fuel use and housing characteristics, especially low-quality building materials, we expand the nascent literature that considers housing quality in the energy-health link (Kowsari and Zerriffi, 2011; Checkley et al., 2016).

Third, we apply Sorted Effects Analysis (SEA) to discover heterogeneity in health effects and sub-groups more susceptible to specific polluting fuels, which can refine the policy opportunities in low-data settings, such as developing countries.

While the fuel-health relationship has been a topic of numerous studies, the body of literature primarily centers around China and reveals a dearth of understanding for other geographical areas such as Sub-Saharan Africa (Liu et al., 2020; Tian et al., 2021). In addition, past studies also have methodological shortcomings as they adopted a binary treatment method that does not appropriately reflect the existing situation where fuel stacking, which is a common practice (Tatham and Muchapondwa, 2021; Twumasi et al., 2020), is ignored in the modelling exercise.

While Hou et al. (2022) studied fuel-stacking behavior, the generalizability of their findings is limited by the sample size and study area. To fill these gaps, our study applies a multivalued fuel index (clean, mixed, unclean) and analyzes nationally representative data from the 2019 Statistics South Africa (StatsSA) General Household Survey. While the primary analysis leverages the causal machine learning approach (Knaus, 2021) to control for confounding based on observed factors,

* Corresponding author.
E-mail addresses: nouran.zen@gmail.com (N. Zenelabden), adeola.oyenubi1@wits.ac.za (A. Oyenubi), jdikgang@fgcu.edu (J. Dikgang).

https://doi.org/10.1016/j.eneco.2025.108926
Received 6 November 2024; Received in revised form 21 September 2025; Accepted 21 September 2025
Available online 22 September 2025
0140-9883/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

additional analysis confirms the robustness to unobserved factors. Lastly, we use the innovative Sorted Effects Analysis to investigate the role of housing quality in the energy-health nexus. The study holds practical significance for energy policymaking, particularly in the context of transitioning to clean fuels and monitoring trends towards sustainability targets.

The remaining structure of the paper consists of background information, theoretical framework, methodology, results analysis, and policy recommendations, along with future research avenues.

2. Background

South Africa is one of the world's most unequal nations; this stems from the legacy of apartheid, which deprived non-white communities of resources and rights (Sulla et al., 2022; World Bank, 2020). Post-apartheid policies sought to address this imbalance through various programmes such as the National Electrification Programme and the Integrated Resource Plan, which promoted renewable energy and prioritized rural areas (Rathi and Vermaak, 2018). Despite these efforts, while general access to power has improved, rural regions continue to lag (Stats SA Statistics South Africa, 2019).

Household energy use patterns in South Africa vary substantially, depending on location, cultural norms, income, and fuel supply. Wood is frequently the primary fuel in rural areas with scarce or expensive electricity, whereas low-income urban households rely on kerosene. In the Mpumalanga Province, residents near coal mines often use coal for cooking and heating their homes (Roomaney et al., 2022). Despite a projected electrification rate of 89 % in 2021, financial limitations prevent many low-income households from relying solely on electricity. Although electricity is the primary cooking fuel for 77.7 % of households, others continue to utilize wood (7.7 %), kerosene (3.5 %), and coal (0.4 %). These trends underscore the persistent challenges of providing universal access to clean fuels, despite substantial efforts to electrify.

IAP remains a significant public health concern in South Africa, particularly in low-income communities, where ambient particulate matter (PM) levels are higher than in industrialized areas (Adesina et al., 2020; Piketh et al., 2016). In 2012, 17.6 % of South Africans were exposed to IAP, resulting in an estimated 8862 fatalities, mainly from acute respiratory infection and cardiovascular disorders (Roomaney et al., 2022). The provinces with the highest IAP exposure from unclean fuels were KwaZulu-Natal (26.4 %), Mpumalanga (27.4 %), and Limpopo (50.0 %). Given these health hazards, it is crucial to examine household energy choices and their consequences to develop policies that promote the transition to cleaner energy while addressing energy poverty and gender inequality.

3. Theoretical frameworks and determinants of fuel type choices

3.1. Theoretical review

Understanding the complex dynamics of household fuel choices and their health impacts is crucial for developing effective interventions and policies that address these issues. This review aims to examine the current state of knowledge in this field critically. By synthesizing and critically analyzing the existing literature, this review seeks to identify critical research gaps.

The study of household fuel choices has been dominated by two primary theoretical frameworks: the energy ladder and fuel stacking models. Hosier and Dowd (1987) introduced the energy-ladder model, which posits a linear progression from traditional biomass fuels to modern, cleaner energy sources as household income increases. However, this model has faced criticism for oversimplifying the fuel-transition process and neglecting non-economic factors (Van der Kroon et al., 2014).

In response to these limitations, Masera et al. (2000) proposed the fuel-stacking model, recognizing that households often adopt new cooking technologies while retaining traditional methods. This model offers a better understanding of household energy-use patterns, acknowledging the influence of factors beyond income, such as cultural preferences and supply constraints (Batchelor and Brown, 2021).

While the fuel-stacking model represents an advancement in conceptualizing household energy transitions, it, too, has limitations. Notably, it needs more quantitative predictive power and may need to adequately address supply-side factors and market dynamics (Ruiz-Mercado and Masera, 2015). Recent studies have further recommended a more advanced overarching theory that would merge aspects of both these paradigms, considering perspectives from behavioral economics and sociotechnical transition theory (Kowsari and Zerriffi, 2011; Van der Kroon et al., 2014).

Another work that contributed to this discussion is Kowsari and Zerriffi (2011), specifically their concept of a "three-dimensional energy profile." This construct studies energy carriers and devices' physical properties, energy service peculiarities, and the socioeconomic and cultural factors that affect fuel selection. This model can describe why some households may engage in unhealthy fuel-stacking practices despite clear adverse health impacts. The authors claim that fuel selection is part of the overall household decision-making process concerning life and housing conditions.

The theoretical connections between fuel use and health effects were found to have empirical corroboration. Ruiz-Mercado and Masera (2015) argued that health impacts associated with fuel selection are mediated by housing, behavioral, and technical factors, providing some insights to comprehend why mixed-fuel employments used turned out to have worse health impacts. This framework was upgraded by Checkley et al. (2016), who held that housing conditions formed "microenvironments" in which health impacts related to fuel selection might be amplified or muted. Poor housing conditions might also trigger "feedback loops" that heighten pollutant exposure related to insufficient ventilation, decrease combustion fuel efficiency related to excessive dampness, and absorption and re-emission of pollutants due to specific building materials.

3.2. Understanding household fuel choice: determinants and transition factors

3.2.1. Diverse factors influencing fuel selection

Empirical studies have identified various factors influencing household fuel choices, broadly categorized into economic, socio-demographic, and geographic/infrastructural determinants. Income and wealth consistently emerge as significant predictors of fuel choices, with higher-income households more likely to use cleaner fuels (Wassie et al., 2021; Zhu et al., 2022). However, the relationship between income and fuel choice is not always linear, and the income threshold for transitioning to cleaner fuels varies across contexts (Muller and Yan, 2018). Although credit accessibility also plays a crucial role in enabling rural households to switch to cleaner energy sources (Twumasi et al., 2020), the long-term sustainability of credit-enabled fuel transitions remains understudied.

Education levels are consistently associated with preferences for cleaner energy sources (Dongzagla and Adams, 2022; Qiu et al., 2023). However, the mechanisms through which education influences fuel choices, beyond associated income effects, require further investigation. Gender dynamics are substantial in energy poverty, particularly in rural regions (Hou et al., 2022; Liu et al., 2020). Women and girls often bear the burden of collecting and using traditional biomass fuels, highlighting the need for gender-sensitive energy policies and interventions.

3.2.2. Fuel transition intricacies

Fuel switch trends in developing contexts highlight a highly non-linear path. The fuel choice made by households is highly complex and involves adopting multiple types of fuel due to volatility in supply

and economic prospects. The cultural and behavioral dimensions are also significant, as households often adhere to traditional fuel usage despite the availability of modern fuels (Ruiz-Mercado and Masera, 2015).

The second step involves establishing incentives for parents to overcome the switch barrier and adopt cleaner fuel technologies. Transition towards cleaner fuel technologies implies not only access to technologies but also overcoming ingrained habits. According to research by Carter et al. (2019), households would need to abandon old solid fuel technologies to adopt cleaner ones. Unless the new cleaner fuel options have more attractive prospects, the established sociotechnical systems would continue to prevail.

Demographic factors play an important role in fuel switching. Based on research conducted by Elasu et al., 2024, households headed by elderly individuals are more likely to use traditional fuels due to financial constraints and the high costs associated with switching from conventional to clean cooking fuels. According to Sharma (2019), even though households intend to switch to clean fuels such as LPG and electricity, they remain stuck at the bottom tier of the energy ladder due to economic reasons and values.

Another important determinant is the availability of infrastructure. As stated by Rosenthal et al. (2018), the overlap between old and new technologies is more frequently found in areas with high clean fuel prices and underdeveloped distribution systems; thus, the reason interventions neglecting the local infrastructure usually do not work.

"Fuel stacking" introduces more complexity in potential transition pathways. Malakar and Day (2020) argued that households with access to cooking with LPG still chose to rely on firewood as they believed that the food taste and the cooking steps associated with certain foods require the use of biomass fuels. This implies that specific households may never fully transition from the traditional to the modern fuels, and policies should be designed to accommodate this outcome instead of attempting to impose a one-dimensional transition pathway.

The study by Hanna et al. (2016) shows that the health benefits of new fuel alternatives can be negated by behaviors associated with them. This means that providing cleaner fuels alone does not necessarily lead to improvements in health, as the practice of using harmful fuels continues, often due to socio-cultural dynamics or a lack of alternatives.

Housing conditions play a crucial role in mediating the effect of fuel use on health. Structural housing characteristics, including ventilation systems, room configuration, and building design, are correlated with the emission of harmful pollutants produced when fuels are combusted. The influence of housing characteristics on fuel use also highlights the limitations of policies that promote fuel switching but employ a generalized approach to achieve expected health impacts (Bruce et al., 2015).

3.3. *Health impacts of indoor air pollution from fuel use*

Numerous studies have linked indoor air pollution from solid fuel use to various health problems, including respiratory infections, cardiovascular diseases, and adverse birth outcomes (Edwards and Langpap, 2012; Wang et al., 2016). The health impacts appear particularly pronounced among vulnerable groups such as children, women, and older people (Krishnamoorthy et al., 2022; Liang et al., 2013; Xu et al., 2022).

In the South African context, several studies have focused on the link between indoor air pollution and acute respiratory infection in children. Barnes et al. (2009) argue that indoor air pollution increases young children's risk of acute respiratory infections (ARI) due to their immature lungs and immune systems. Albers et al. (2015) found a notably higher rate of respiratory diseases in schoolchildren in Mpumalanga who resided in homes that used non-electric cooking fuels. Similar patterns have been observed in other regions of South Africa (Buthelezi et al., 2019; Shirinde et al., 2014).

Despite the substantial body of evidence linking the use of solid fuels to adverse health outcomes, several methodological challenges complicate the interpretation of these findings. Many studies rely on proxy exposure measures (e.g., fuel type) rather than direct measurements of pollutant concentrations, potentially leading to exposure misclassification. Socioeconomic status, nutrition, and access to healthcare can confound the relationship between fuel use and health outcomes, making it difficult to isolate the specific effects of indoor air pollution. Most studies have focused on short-term health impacts, neglecting the potential long-term consequences of chronic exposure to indoor air pollution.

These methodological limitations partially explain some of the conflicting evidence in the literature. For instance, Tatham and Muchapondwa (2021) found only a slight increase in acute respiratory infections and general acute respiratory infection associated with solid fuel use in rural South Africa, with the effect not being statistically significant for specific age and gender subgroups. This contrasts with other studies showing more pronounced health impacts, particularly among women and young children.

This critical review highlights the complex interplay between household fuel choices and health impacts in developing countries, focusing on South Africa. While substantial progress has been made in understanding the determinants of fuel choices and the associated health risks, several critical research gaps and methodological challenges persist.

4. Methodology

Machine learning (ML) is gaining popularity in energy economics owing to its superior performance in processing, categorizing, and predicting large-scale and complicated data (Ghoddusi et al., 2019). Since ML models do not make any pre-specified assumptions about the equation's functional form, the interaction between variables, and the statistical distribution of parameters, they can manage large amounts of structured and unstructured data and make quick decisions or forecasts (Athey, 2017; Mullainathan and Spiess, 2017). This gives machine learning (ML) methods an advantage over methods proposed by classical econometrics.

A recent development in the application of ML models is in causal inference, whereas traditional ML focuses on predicting outcomes. Causal ML, on the other hand, leverages this predictive power to quantify changes in outcomes due to treatment (Feuerriegel et al., 2024). Specifically, ML models can be applied in the context of the well-known Propensity Score Matching (PSM) to account for selection on observables. This offers an improvement on PSM because it can flexibly account for higher-order and interaction terms, leveraging the variable selection attributes of ML models.

The implementation of causal machine learning used in our analysis is called Double Machine Learning (DML). DML is a data-driven approach that improves the plausibility of the ignorability assumption[1] by offering more flexibility in defining the propensity score equation. This is achieved by controlling complex interactions and higher-order terms, which can increase the number of variables and potentially impact the common support assumption (Bodory et al., 2022; Chernozhukov et al., 2018; Knaus, 2021). However, mitigates this by relying on the spacing assumption, which improves the plausibility of the standard support assumption. Furthermore, unlike the single-equation approach that uses either the outcome or the propensity score model to estimate treatment effect under the ignorability assumption (e.g., propensity score matching), DML is considered "doubly robust" as it remains consistent even if the outcome or treatment equation is misspecified (Knaus, 2021).

In this analysis, we apply Knaus's (2021) DML approach to investigate the health impacts of fuel choices in a binary and multivalued treatment setting with different treatments from a random set $D_i \in \{0, 1,$

[1] This entails the conditional-independence and common-support assumption.

…, T}, $T = 3$ (i.e., treatment levels clean, mixed, and polluting fuels). Thus, we estimate the average potential outcomes (APO) $\mu_t = E[Y_i^t]$ and their differences. Following Knaus (2021), we estimate the average treatment effect (ATE) for the difference in possible outcomes $\mu_1 - \mu_0 = E[Y_i^1 - Y_i^0]$. We also consider various pairwise comparisons that could be made for the multivalued treatment, such as $\mu_m - \mu_k = \mathrm{E}[Y_i^m - Y_i^k]$ for m ≠ k $D_i \in \{0, 1, \ldots.\mathrm{T}\}$. The average potential outcomes conditional on covariates are obtained as follows:

$$\mu_t = E\left[\mu_t(X_i) + \frac{d_i^t(Y_i - \mu_t(X_i))}{p_t(X_i)}\right] \quad (1)$$

Where $d_i^t = \{D_i = t\}$ is the treatment indicator, $\mu_t(x)$ the conditional outcome predictions, and the propensity score predictions:

$$\sigma^2_{\mu,t} = E\left[\left(\frac{d_i^t(Y_i - \mu_t(X_i))}{p_t(X_i)} + \mu_t(X_i) - \mu_t\right)^2\right] \quad (2)$$

The post-model-selection problem does not affect the validity of the statistical inference on the possible outcome.[2] To implement the DML, the lasso is used as a variable selector to get the predictions for the outcome $\mu_t(x)$ and the propensity scores $p_t(x)$, following which both projections are inserted into the sample analogs of eqs. (1) and (2) to estimate the average potential health outcomes (e.g., self-reported poor health and ARI). We then estimate the difference between the three categories of fuel choices to get the multivalued treatment effect. The parameters $\mu_t(x)$ and $p_t(x)$ are calculated using the post-lasso method with cross-validation. The post-lasso is based on the lasso estimator proposed by Tibshirani (1996). The lasso solves the following optimization problem:

$$\min_{\beta}\left[\sum_{i=1}^{n}(Y_i - X_i\beta)^2\right] + \lambda\sum_{j=1}^{p}|\beta_j| \quad (3)$$

Where "Y_i" is the observed outcome, "X_i" is a vector of covariates, and "λ" is a penalty factor on the total of the absolute coefficients. In our analysis, we control for 16 variables and a model with second-order terms encompassing more than 301[3] covariates. Since some coefficients may be too small to have significance, the Lasso acts as a variable selector. It reduces to zero any variables with little or no predictive power, depending on the value λ selected by cross-validation. To evaluate the robustness of our results to unobserved factors, we employ a two-stage least squares (2SLS) estimation approach to account for model uncertainty related to variable selection and endogeneity bias. We use the proportion of people at the primary sampling unit (PSU) level who use unclean fuels at home as an instrument (IV) for fuel choices, as the acceptance of modern technology is frequently affected by other residents in the community (Conley and Udry, 2010).

The fuel preferences of other households may have little to no effect on the indoor air quality of a particular household, but they should not impact the outcome variable (Liu et al., 2020). The literature has demonstrated that using the average perception of other people as an instrumental variable is a reliable IV for determining fuel preferences (Liu et al., 2018; Liu et al., 2020). We also use Lewbel's (2012) heteroskedasticity-based technique to create additional instrumental variables. This approach, along with the IVs, has been employed in earlier empirical studies (see Liu et al., 2020). The generation of the IV is discussed in the next section.

Lastly, we conduct a sorted effects analysis to examine the heterogeneity in the relationship between different categories of fuel and the health outcomes examined in this paper. This approach departs from the usual approach of estimating a single quantity called the average effect to assessing the full distribution of effects. Introduced by Chernozhukov et al. (2018), this method uses nonlinear models (in our case, a logit model) to estimate and sort individual-level effects from the smallest to the largest, thereby identifying the full range of effects (or heterogeneity). The method also allows for classification analysis, i.e., a study that compares the mean of covariates among those who are at the top of the sorted partial effects (SPE) distribution (most affected) to the distribution of covariates among those who are at the bottom of the SPE distribution (least affected).

By comparing extreme groups, the analysis identifies policy-relevant covariates. This approach offers an alternative way to examine effect heterogeneity without splitting the sample into predefined subgroups—a useful property in data-constrained developing-country settings. We focus on housing variables that capture how household structure moderates the relationship between polluting fuels and health. Similar designs have been used to study heterogeneity in the gender gap in earnings and in mental health (Posel et al., 2024).

5. Data and descriptive statistics

5.1. Data

This study utilizes data from StatsSA's 2019 GHS, which assesses living conditions and service delivery nationwide.[4] We focused exclusively on rural households, excluding urban samples. The original sample consisted of 27,056 observations. Due to missing or invalid responses, 2087 observations were excluded for the variable FBE (Free Basic Electricity), and 25 observations were excluded for the General health status variable. The final sample size used for most analyses is therefore 24,944. Since nearly all households used electricity for lighting, we concentrated on energy sources for cooking and heating. Households reported their primary energy source for each, and we created dummy variables for unclean fuel use (e.g., paraffin, wood, coal) versus clean fuel (e.g., electricity, gas).

Unlike previous studies that focused solely on solid fuels as unclean (Alem et al., 2016; Hou et al., 2022; Liu et al., 2020), we followed the WHO's classification, grouping paraffin/kerosene with unclean fuels like biomass and solid fuel. We created a score index combining cooking and heating fuel types to examine fuel-stacking behavior based on Zenelabden and Dikgang (2022). A score of 1 indicated only clean fuels, 2 for mixed fuels, and 3 for unclean fuels. We generated a "fuel binary" variable to represent a binary treatment effect, coded 1 for mixed or unclean fuels and 0 for clean fuels.

The outcome variables of interest were self-reported health and acute respiratory infection (ARI). For the former, respondents rated their general health from 1 (excellent) to 5 (poor). We created a dummy variable for poor health, coded as 1 for fair or poor health and 0 otherwise (see Oliveras et al., 2021; Quesnel Vallée, 2007). To measure ARI, respondents were asked if they had experienced ARI in the past three months, with responses coded as 1 for yes and 0 for no.

We chose fuel choice determinants based on the energy and public health literature (Alem et al., 2016; Bakhsh et al., 2020; Boch et al., 2020). Individual characteristics included age, education, gender, and marital status. Size, income, dwelling style, and kitchen availability were all determinants at the household level. We developed dummy variables for construction materials i.e. unconventional walls (walls made of plastic, cardboard, wattle and daub, grass and asbestos),

[2] Knaus (2021) provides formulas for deriving the pairwise ATEs by deducting the doubly robust scores of the possible outcomes.

[3] We have 16 core covariates, these are covariates that are not subject to Lasso selection (see Table 1), in addition we include province dummies (8 dummies) for a total of 24 covariates. Finally, we include interaction terms of all covariates ($C_2^{24} = 276$). The total of 301 is made up of 276 + 24 + 1 where 1 is the constant term.

[4] We use the following province abbreviations: EC (Eastern Cape), FS (Free State), GP (Gauteng), KZN (KwaZulu-Natal), LP (Limpopo), MP (Mpumalanga), NC (Northern Cape), NW (North-West), WC (Western Cape).

unconventional roofs (roofs made of wood, plastic, cardboard, mud and cement mix, wattle and daub and mud) and unconventional floors (i.e. planks, dung and earth/sand) to assess house quality. These factors and free basic electricity (FBE) availability were included because they are associated with physical functioning and vulnerability, particularly in older persons (Boch et al., 2020; Chen et al., 2024).

5.2. Descriptive statistics

The analyses were conducted on the final sample of 24,944 observations. We dropped cases with missing or out-of-range responses. Table 1 defines the variables and their descriptive statistics.

In our study, close to 11 % of the sampled households reported poor health status. This is consistent with national estimates of self-reported health in rural South Africa (Stats SA Statistics South Africa, 2019). About 10 % also reported having acute respiratory infections (ARIs) within the last three months. This prevalence is slightly lower than the 13.5 % ARIs reported by Albers et al. (2015) in similar rural settings.

Fuel index is another important measure that represents the complexity of energy decision-making in households. Approximately 44 % of the rural households in the sample use unclean fuels for heating and cooking. The average fuel index score is 1.6 (where 1 is the cleanest and 3 the dirtiest option), suggesting that the members of rural households often practice fuel stacking (use both clean and unclean fuels).

Furthermore, the comparatively large share of households relying on clean fuels (56 %) as primary energy may be attributed to relatively successful electrification policies in South Africa after the end of apartheid, notably the National Electrification Programme and the Integrated Resource Plan focused on rural areas (Rathi and Vermaak, 2018). Nevertheless, as illustrated by our fuel index, many of these allegedly "electrified" households are alternatively using traditional fuels with electricity, which is in line with affordability constraints and pre-existing cultural habits, thus backing the fuel-stacking theoretical model elaborated on above.

The sociodemographic characteristics of the sample indicate that the average age of household members is 28 years. 19 % were married, 47 % completed senior education, and 34 % completed junior education. The average household size is approximately 6 members, which is slightly lower than the national average of 7 % for rural households (Stats SA, 2019). This may be attributed to the prevalence of extended families in rural South Africa. 53 % of the household members were female, which is consistent with the national demographic profile.

Housing described in our sample is representative of the inequalities present in rural areas of South Africa. Despite 82 % of households residing in formal dwellings—an indicator of some level of development—quality issues persist. About 1 % of houses had an unconventional material used for the walls and roof, and 6 % had a non-standard flooring material. These characteristics linked to houses that can be classified as substandard are most informative for the analysis due to the theoretical association between housing quality and the health effects of certain fuels (Checkley et al., 2016).

Moreover, 84 % of the surveyed households were reported to have separate kitchens, which is relevant for the exposure assessment as separate kitchens can mitigate the household members' direct exposure to cooking fuel emissions. Approximately 17 % of the households in the sample had access to free basic electricity, as part of the government's policy to accelerate the clean-energy transition.

The subsequent analysis of relationships between fuel type, housing quality, and health effects will rely on these descriptive statistics. The observed data patterns, such as fuel stacking and housing quality variability, inform our multivalued treatment choices by indicating the relevance of housing characteristics as moderators in the relationships between fuel type and health.

Table 1
Description of variables of interest and summary statistics.

Variables	Definitions	Mean	S.D.
Health variables			
Poor health	1 = if reporting fair or poor health; 0 = otherwise	0.105	0.305
ARI	1 = if reporting acute respiratory illness; 0 = otherwise	0.098	0.298
Treatment variables			
Fuel index	1 = if households use only clean fuels; 2 = household uses mixed fuels; 3 = household use only unclean fuels	1.617	0.767
Fuel binary	1 = if households use unclean or mixed fuels; 0 = households use only clean fuels	0.441	0.496
Control variables			
Age	Age (years)	27.89	21.637
Junior education	1 = if the highest degree is junior or primary school; 0 = otherwise	0.339	0.473
Senior education	1 = if the highest degree is senior high school or higher; 0 = otherwise	0.468	0.499
Female	1 = female; 0 = male	0.53	0.498
Married	1 = married; 0 = otherwise	0.1925	0.394
HH size	Number of household members	5.765	3.074
log income	Log of total household income	8.095	1.493
Structure	1 = formal; 0 = informal /attached/ traditional	0.819	0.384
unconventional wall	1 = if wall material from plastic, cardboard, wattle and daub, grass, and asbestos, 0 = otherwise	0.009	0.098
unconventional roof	1 = if the roof material is plastic, cardboard, mud, wattle and daub 0 = otherwise	0.005	0.072
non-standard flooring	1 = if floor material from planks, dung, sand 0 = otherwise	0.062	0.242
Kitchen	1 = if having a kitchen 0 = otherwise	0.844	0.362
FBE	1 = if the household receives free basic electricity from the government; 0 = otherwise	0.172	0.377
Spring[a]	1 = if data was collected in September, October and November; 0 = otherwise	0.251	0.433
Summer	1 = if data was collected in December, January and February; 0 = otherwise	0.247	0.431
Autumn	1 = if data was collected in March, April and May; 0 = otherwise	0.252	0.434

[a] Winter is the omitted baseline for seasons (hence we list Spring/Summer/Autumn only).

6. Results

6.1. Multivalued treatment effects

Table 2 presents the results of the binary treatment (clean versus polluting or mixed) and the multivalued treatment analysis. The result for the binary treatment is consistent with expectation, i.e., using mixed/polluting fuels increases the likelihood of reporting poor self-assessed health and ARI by 2.5 and 1.8 percentage points, respectively.

The multivalued treatment effects shown in Table 2, however, make

Table 2
Treatment effect of fuel choices on health status.

	Binary treatment	Multivalued treatment		
Variable	polluting vs. clean fuels (1)	mixed vs. clean fuels (2)	polluting vs. clean fuels (3)	polluting vs. mixed fuels (4)
Poor health	0.025***	0.023***	0.021**	−0.002
	(0.004)	(0.005)	(0.009)	(0.011)
ARI	0.018***	0.014*	0.022***	0.008
	(0.004)	(0.007)	(0.006)	(0.008)

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

a significant point against the widely accepted belief that such energy transitions are beneficial. The exclusive use of clean fuel offers considerable health benefits, while fuel stacking, regarded as a step towards improving heating, does not statistically outperform the exclusive use of polluting fuel. Specifically, for both outcomes, using mixed fuels as opposed to exclusively using polluting fuels does not offer a statistically significant benefit in terms of health outcomes. However, for self-assessed health, using mixed and polluting fuels increases the likelihood of reporting poor health by 2.3 and 2.1 percentage points, respectively, relative to the exclusive use of clean fuels.

Further, the possibility of reporting ARI increases by 1.4 and 2.2 percentage points for mixed and polluting fuels (respectively) relative to exclusive use of clean fuels. The evidence suggests that stacking fuel may offer minimal benefits in terms of health outcomes. Relative to clean fuels, both mixed and polluting fuels raise the probability of poor self-assessed health; mixed vs polluting is not statistically distinguishable. The potential consequences for each fuel category are illustrated in Figs. 1 and 2 below.

Figure 1 presents self-assessed health across three fuel options. Again, it is the three separate bars that can be compared. Clearly, from the figure, users of mixed and polluting fuels are significantly more likely to report poor health than users of clean fuels, where the levels of outcomes are strikingly similar between users of mixed fuels and those of polluting fuels.

The potential outcomes, showing the likelihood of reporting ARI for each fuel category, are presented in Fig. 2. The potential outcomes show that using mixed or polluting fuel increases the probability of reporting ARI compared to exclusive use of clean fuel.

6.2. Model specification and balance diagnostics

Note that to generate the results presented in Table 2, the core covariates shown in Table 1 are not subject to selection by the Lasso algorithm, because these variables are included based on their theoretical relationship with the treatment/outcome and their empirical importance as shown by extant literature. However, interaction and higher-order terms based on these variables are subject to selection by the Lasso algorithm. In the appendix, we present the second-order terms selected by the Lasso algorithm for each of the estimates shown in Table 2. Importantly, we note that in most cases, the chosen interaction terms include variables that describe the physical structure of the dwelling.

Further, the validity of our inference (for both the binary and multivalued treatment) hinges on balance in the distribution of covariates. As indicated in the extant literature (Knaus, 2021), we check balance across all covariates (even those that the Lasso model did not select).

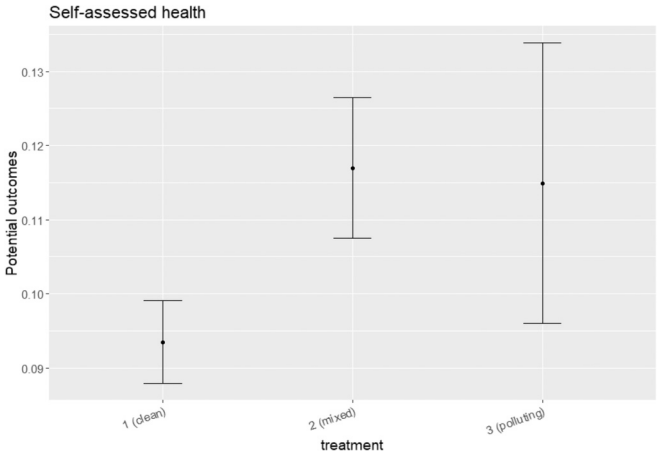

Fig. 1. Fuel choice and Self-assessed health.

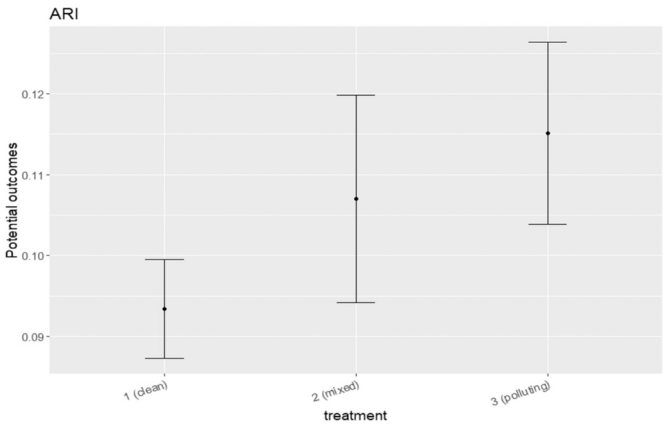

Fig. 2. Fuel choice and ARI.

Balance diagnostics are essential for validating the relationship claims of our estimates. Figs. 3 and 4 show the balance statistics for the multivalued treatment.

As depicted in Fig. 3, balance is also beneficial for the self-assessed health outcome. All standardized differences are significantly less than the standard cut-off of 0.2, indicating that the self-assessed health outcome estimates are valid. While in the case of self-assessed poor health, all covariates have standardized differences in means below the recommended cut-off of 0.2 (Fig. 3).

In Fig. 4, there are some noticeable imbalances in the distribution of covariates when the outcome is ARI (i.e., in the polluting versus clean "3 vs 1" and polluting versus mixed "3 vs 2" categories). The implication is that an imbalance may influence results for the ARI analysis in some covariates (or higher-order terms). Moreover, the results provided above rely on the arguably implausible conditional independence assumption; hence, we take the result from Table 2 as evidence of associations between the fuel choice and various health outcomes. To resolve these issues and validate our claims, we employ the instrumental variables methodology.

6.3. Robustness check: instrumental variables analysis

To check the robustness of the results, we use the instrumental variables approach under Lewbel's (2012) heteroskedasticity-based technique. As noted earlier, for this analysis, we use the proportion of people at the primary sampling unit (PSU) level who use unclean fuels at home as an instrument (IV) for fuel choices. While we conduct analysis based on this "external" instrument only (reported in the appendix), our primary analysis leverages the use of internal instruments. Our external instrument is peer effects, which is a familiar concept in the literature; however, we note that it is not immune to criticism. It has been pointed out that peer adoption may not be a valid instrument (Sacerdote, 2001).

For example, in our context, traditional dwellings are often densely packed together due to a lack of planning. Therefore, the use of polluting fuel by a neighbor can directly affect an individual's health through pollution. Manski (1993) also noted the problem of "common shocks" in what he referred to as the reflection problem. In our context, the observation that an individual's behavior is correlated with that of their neighbors may be due to standard shock, rather than neighbors influencing each other. For example, poor infrastructure in dwellings may constrain the adoption of non-polluting fuel, inducing correlation among neighbors.

Lewbel (2012) introduced an approach that can be used when there is reason to question the validity of the external instrument. This approach exploits heteroskedasticity in the errors of mismeasured or endogenous explanatory variables to construct instruments for those variables. Baum et al. (2019) discuss the assumptions under which this approach is valid, and tests that can be performed to validate these

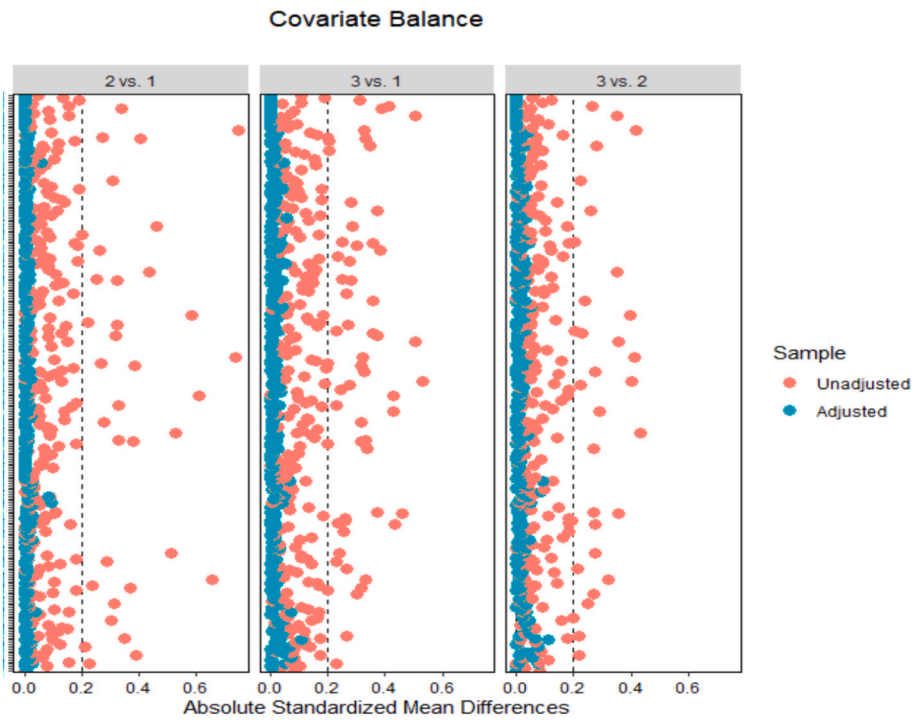

Fig. 3. Balance statistics across all covariates and interaction terms (Self-assessed health).

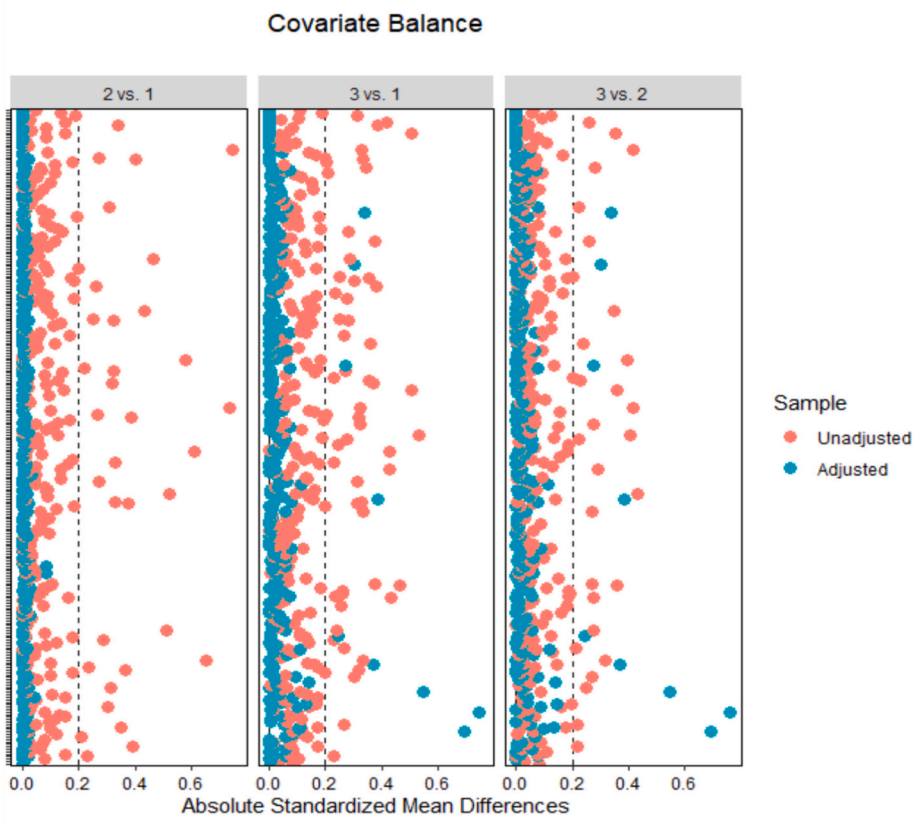

Fig. 4. Balance statistics across all covariates and interaction terms (ARI).

Table 3
Instrumental variable results.

VARIABLES	Poor health	ARI
Polluting fuels	0.038***	0.051***
	(0.006)	(0.007)
Married	−0.068***	0.006
	(0.006)	(0.005)
Junior education	−0.057***	−0.053***
	(0.005)	(0.006)
Senior education	−0.130***	−0.070***
	(0.005)	(0.006)
Age	0.007***	−0.000
	(0.000)	(0.000)
Log income	0.002*	0.004***
	(0.001)	(0.001)
HH size	−0.003***	−0.007***
	(0.001)	(0.001)
Structure	0.011**	−0.014**
	(0.006)	(0.006)
Roof unconventional	0.003	−0.014
	(0.026)	(0.026)
Walls unconventional	0.017	−0.013
	(0.020)	(0.019)
Floor non-standard	−0.025***	−0.042***
	(0.008)	(0.008)
Female	0.005	0.020***
	(0.003)	(0.004)
Kitchen	−0.020***	−0.005
	(0.005)	(0.005)
FBE	0.005	0.020***
	(0.005)	(0.005)
Spring	−0.001	−0.050***
	(0.005)	(0.006)
Summer	−0.008*	−0.067***
	(0.005)	(0.006)
Autumn	−0.001	−0.060***
	(0.005)	(0.006)
WC (Western Cape)	0.037**	0.040**
	(0.019)	(0.021)
EC (Eastern Cape)	0.067***	0.017***
	(0.006)	(0.006)
NC (Northern Cape)	0.044***	0.097***
	(0.011)	(0.014)
FS (Free State)	0.075***	0.103***
	(0.015)	(0.017)
KZN (KwaZulu-Natal)	0.044***	0.028***
	(0.005)	(0.005)
NW (North West)	0.032***	0.022***
	(0.007)	(0.007)
GAU (Gauteng)	0.010	0.002
	(0.014)	(0.013)
MP (Mpumalanga)	0.018***	0.045***
	(0.005)	(0.006)
Constant	−0.023*	0.165***
	(0.013)	(0.015)
Observations	24,944	24,944
R-squared	0.199	0.026
Weak identification test (Wald F statistic)	9234	13,708
Hansen J statistic	1.009	0.463
Chi-sq(2) P-val	0.604	0.496

Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10.

assumptions. We conduct these tests and discuss them in the footnote to Table 3. The instrumental variables result provides important validation of our main findings.

The last 3 rows of Table 3 show that the instruments are not weak (F-statistics greater than 10), satisfy the exclusion restriction (*p*-values 0.604 and 0.496), and the excluded external instrument is likely valid.[5] These diagnostic tests validate that our preferred specifications are correctly specified for the instrumental variable approach and that the outcomes (which further demonstrate even larger adverse health effects associated with polluting/mixed fuels) corroborate our conclusions.

6.4. *Heterogeneity analysis: sorted effects analysis*

Lastly, we examine heterogeneity between the use of polluting fuels and health using the Sorted Effects Analysis.[6] The idea is to quantify heterogeneity in the relationship of interest and compare those at the bottom of the Sorted partial Effects distribution with those at the top of the SPE distribution. Furthermore, the analysis allows for determining the effect of heterogeneity, i.e., which households are most affected by the health consequences of polluting fuels. Figs. 5 and 6 show the SPE (SPE) and the Average Partial Effect (APE) when the exclusive use of clean fuels is compared with the use of mixed fuels.

In Fig. 5, we observe considerable variation around the average health effect of using mixed fuels. In contrast, the average is negative; the sorted effects exhibit a wide range, suggesting that some households suffer notably more health detriment than others from fuel stacking. Consistent with our previous results, the APE shows that the use of mixed fuel increases the likelihood of reporting poor self-assessed health and ARI on average. However, the SPE reveals considerable heterogeneity, with effects at the extremes of the SPE being significantly different from those of the APE.

For example, while the APE is 0.023 in Fig. 5, the SPE ranges from a low of −0.06 to a high of 0.13. Specifically, about 20 % of respondents at the extremes of the SPE distribution in the case of poor self-assessed health have effect sizes that are significantly larger or lower than the APE. The ARI has approximately 10 % of respondents in both extremes, with effects that are substantially different from the APE.

For the ARI, approximately 10 % of respondents fell into both extremes, with effects that were significantly different from those of the APE. The presence of this heterogeneity has significant consequences for implementation, as it implies that specific characteristics of households expose some families to the health effects of polluting fuels more than others.

Tables 4 and 5 present the classification analysis for self-assessed health and ARI, respectively, when the exclusive use of clean fuels is

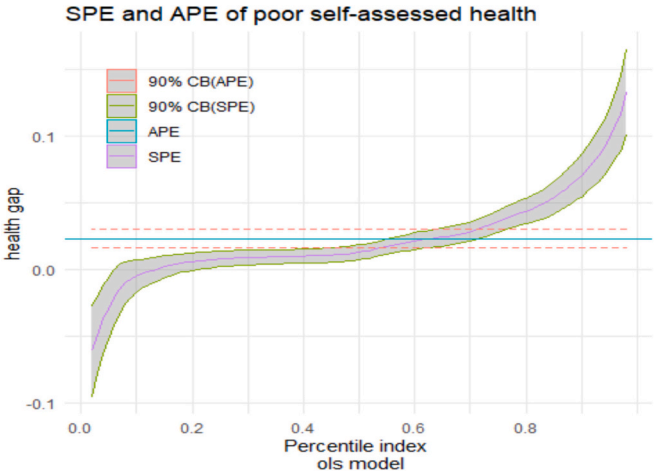

Fig. 5. SPE and APE for clean versus mixed fuels (self-assessed health).

[5] Note that this only gives an indication as the validity assumption is untestable.

[6] We study effect heterogeneity using **Sorted Effects Analysis (SEA).** SEA computes **sorted partial effects (SPE)** for each observation and summarizes their distribution (e.g., selected quantiles). Alongside the distribution of SPE, we also report the **average partial effect (APE).**

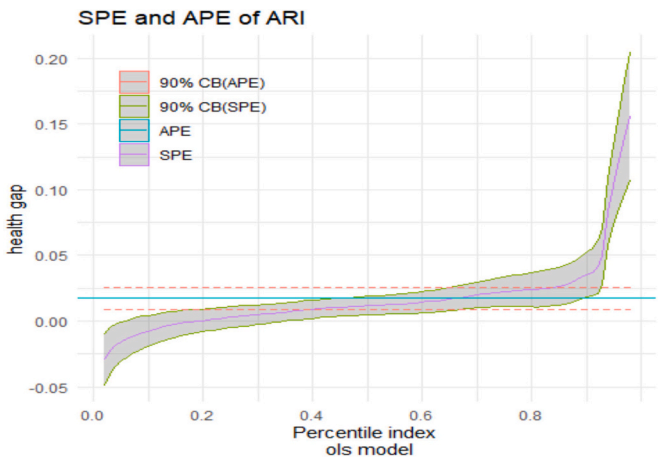

Fig. 6. SPE and APE for clean versus mixed fuels (ARI).

compared to mixed fuels. These tables highlight the attributes that distinguish households severely affected from those least affected.

The first column shows the mean among individuals in the most affected by the relationship (i.e., top 20 % in the case of self-assessed health and top 10 % in the case of ARI), the second column shows the standard error of this estimate. Columns 3 and 4 display the same statistics for the least affected group (i.e., the bottom 20 % for self-assessed health and the bottom 10 % for ARI). For example, in Table 4, the log of income is 7.68 in the most affected group and 8.39 in the least affected group, a statistically significant difference of −0.71 percentage points. The implication is that those who live in households with more income are less likely to be in the most affected group (when the outcome is self-assessed health).

Moreover, Table 4 identifies key moderators of the fuel-health relationship, with housing materials being significant. We focus on variables that describe the materials used for the roofs, walls, and floors of the households included in our study. Table 4 shows that those who use unconventional materials for roofs and walls are marginally more represented in the most affected group by 1 and 3 percentage points, respectively.

For Table 5, the use of unconventional material for walls is associated with being more represented in the most affected group by 8 percentage points. Indicates that households with unconventional wall materials are overrepresented in the most affected group for ARI by 8 percentage points. Other covariates show mixed results. For instance, while higher income and the presence of a kitchen are associated with being in the least affected group for poor health, income also appears to increase the probability of being in the most affected group for ARI. Caution should be exercised when comparing results from this analysis to those based on a sample splitting approach. The SPE approach does not only require the moderating effect of a variable to be significant, but it also requires the size of the moderating effect to be large relative to other moderating impacts, such that when partial effects are sorted, individuals with the characteristic of interest are placed in either extreme of the SPE distribution.

Another example from the existing literature is that the negative relationship between the polluting fuels and health outcomes is notably stronger for older individuals (age). Table 5 (the ARI results) provides evidence for this, indicating that the most affected group is, on average, about 13 years older than the least affected group (a significant 10 %). Additionally, the distribution by province also indicates considerable heterogeneity in the sample, with the most affected provinces (EC, KZN, NW, and MP) characterized by overrepresentation in the most affected groups for both health outcomes.

In the results presented in the appendix, we also conduct an analysis comparing the use of clean fuels with the exclusive use of polluting fuels. For this analysis, we did not find significant heterogeneity when the outcome is self-assessed health (see Fig. B1 in the appendix). There is considerable heterogeneity for ARI, and Table B presents the result of the classification analysis comparing the top 20 % with the bottom 20 %. While the results in Table 5 suggest that the use of unconventional materials for housing is associated with being in the most affected group in terms of ARI, the result comparing clean with exclusive use of polluting fuels did not show a significant result on the use of materials

Table 4
Sorted effects analysis, clean versus mixed fuels (Self-assessed health).

	Most (Top 20 %)	SE	Least (Bottom 20 %)	SE	Estimate	SE	PW P-vals
HH size	5.09	0.38	5.03	0.51	0.06	0.91	0.47
Age	38.49	2.72	42.12	3.93	−3.63	6.59	0.29
Female	0.53	0.05	0.60	0.08	−0.08	0.12	0.27
Married	0.27	0.04	0.31	0.07	−0.04	0.11	0.35
Senior education	0.50	0.06	0.47	0.08	0.02	0.14	0.44
Junior education	0.33	0.05	0.36	0.07	−0.03	0.12	0.42
FBE	0.32	0.05	0.06	0.03	0.26	0.08	0.00***
Log income	7.68	0.23	8.39	0.15	−0.71	0.37	0.03**
Spring	0.20	0.05	0.30	0.08	−0.11	0.12	0.20
Summer	0.26	0.05	0.21	0.06	0.05	0.10	0.31
Autumn	0.33	0.05	0.15	0.05	0.17	0.10	0.05**
Roof unconventional	0.01	0.01	0.00	0.00	0.01	0.01	0.08*
Walls unconventional	0.03	0.01	0.00	0.01	0.03	0.02	0.04***
Floor non-standard	0.05	0.02	0.13	0.05	−0.08	0.07	0.12
Structure	0.81	0.04	0.70	0.06	0.11	0.10	0.13
Kitchen	0.65	0.05	0.97	0.02	−0.33	0.07	0.00***
WC	0.00	0.00	0.04	0.01	−0.04	0.01	0.00***
EC	0.54	0.07	0.02	0.02	0.52	0.09	0.00***
NC	0.00	0.01	0.08	0.04	−0.08	0.05	0.07*
FS	0.08	0.02	0.00	0.00	0.08	0.03	0.00***
KZN	0.04	0.04	0.49	0.09	−0.45	0.12	0.00***
NW	0.21	0.08	0.00	0.02	0.20	0.10	0.02**
GAU	0.01	0.03	0.00	0.04	0.01	0.06	0.46
MP	0.01	0.02	0.24	0.09	−0.22	0.11	0.02**
LM	0.10	0.05	0.12	0.08	−0.02	0.11	0.42

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 5
Sorted effects analysis, clean versus mixed fuels (ARI).

	Most (Top 10 %)	SE	Least (Bottom 10 %)	SE	Estimate	SE	PW P-vals
HH size	4.89	0.39	4.48	0.51	0.41	0.83	0.31
Age	39.00	5.90	25.90	4.37	13.11	9.57	0.09*
Female	0.53	0.07	0.68	0.10	−0.14	0.17	0.19
married	0.26	0.07	0.22	0.08	0.04	0.15	0.39
Senior education	0.34	0.06	0.67	0.11	−0.33	0.14	0.01***
Junior education	0.36	0.09	0.10	0.08	0.26	0.16	0.05**
FBE	0.22	0.06	0.30	0.11	−0.08	0.16	0.31
Log income	8.28	0.20	7.07	0.52	1.20	0.71	0.05**
Spring	0.22	0.05	0.22	0.10	0.00	0.14	0.50
Summer	0.25	0.07	0.09	0.05	0.16	0.12	0.09*
Autumn	0.28	0.06	0.09	0.06	0.19	0.12	0.05**
Roof unconventional	0.00	0.01	0.00	0.02	0.00	0.03	0.49
Walls unconventional	0.08	0.04	0.00	0.00	0.08	0.04	0.02**
Floor non-standard	0.03	0.02	0.04	0.04	−0.01	0.06	0.45
Structure	0.82	0.06	0.81	0.11	0.01	0.15	0.47
Kitchen	0.82	0.06	0.83	0.08	−0.01	0.14	0.48
WC	0.00	0.00	0.08	0.01	−0.08	0.01	0.00***
EC	0.04	0.03	0.50	0.15	−0.47	0.16	0.00***
NC	0.28	0.02	0.00	0.00	0.28	0.02	0.00***
FS	0.19	0.03	0.00	0.00	0.19	0.03	0.00***
KZN	0.02	0.04	0.24	0.14	−0.22	0.17	0.10*
NW	0.01	0.03	0.13	0.14	−0.12	0.17	0.23
GAU	0.14	0.03	0.00	0.00	0.14	0.03	0.00***
MP	0.13	0.09	0.03	0.06	0.10	0.15	0.25
LM	0.19	0.10	0.01	0.02	0.18	0.11	0.06*

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

for roofing and flooring. However, it indicates that the use of unconventional materials for walls is associated with being in the least affected group (by 3 percentage points), although the relationship is only significant at 10 %.

Our construction materials are probably not necessarily related to being in the most affected group. In some cases, poor materials provide more ventilation, making the relationship between housing structure and health when using contaminant fuels more complex than the simple idea that poor or unconventional construction materials always worsen the relationship between contaminant fuels and health.

The sorted effects analysis suggests that fuel options present varying degrees of heterogeneity in health effects. The moderating function of housing material in these associations connects with significant policy implications. The clean energy transition does not necessarily require a standardized policy; it would be more effective to acknowledge housing conditions and target households that would benefit the most from adopting clean fuel. The findings also indicate that possibilities to enhance housing quality, particularly concerning conventional materials for walls and roof, may be an essential accompanying strategy to fuel-switching efforts in South Africa's rural homes.

7. Discussion

We contribute to the growing body of evidence on the complex relationship between fuel type, housing quality, and health in rural South Africa, debunking some of the conventional wisdom on energy transitions and demonstrating significant heterogeneity in fuel type effects across households.

7.1. Challenging the energy ladder paradigm

Our key outcome shows that the health impacts of fuel stacking are statistically not different than those of using polluting fuel. This result contradicts the dominant narrative on fuel stacking, which perceives this choice of households as a transition step in the energy ladder bearing positive health consequences (Masera et al., 2000; Ruiz-Mercado and Masera, 2015). As fuel stacking use is statistically not different from the polluting range, it carries all the positive health effects that come with the exclusive use of clean fuel. In this regard, exclusive clean use reduces the probability of self-assessed health and ARI being poor by 2.3 and 1.4 percentage points, respectively.

Furthermore, the findings of this research have significant implications for energy transition policies in developing countries. Current policy frameworks favor the use of clean fuels, partly to shift entirely towards greener fuels. Our results suggest that fueling processes may not be separated from transition-type frameworks, where they are expected to generate positive externalities even if the transition is not fully achieved. In such cases, the misuse of polluting fuels is likely a key reason for the lack of effects observed in our study, due to stacking behavior. During times when indoor air pollution exists, the improper use of polluting fuels probably outweighs the health benefits of occasional clean fuel use.

7.2. The critical role of housing quality

The heterogeneity results show that the type of materials used for housing moderates fuel effects. The preliminary analysis suggests that households with unconventional roof and wall materials are overrepresented in the extreme groups in terms of the health effects associated with polluting fuel usage. Regarding self-reported health, households with unconventional roof and wall materials are more likely to be in the highly affected category, with 1 percent and 3 percent of these households falling into this category. For ARI, the use of unconventional roof and wall materials yields an additional 8 percentage points to the probability.

This moderation effect likely functions along different mechanisms. The poor quality of raw materials may lead to a situation where pollutants accumulate in a poorly ventilated room, resulting in the formation of harmful compounds. However, room wall materials considered unconventional sometimes relate to more beneficial results (in the clean vs polluting fuel comparison), which means that there can be another mechanism - certain room materials can ensure more effective natural ventilation, although that may apply only to very few specific cases.

Housing quality is not only about the materials: Our estimates reveal that a separate kitchen reduces the likelihood of falling in the most vulnerable category (self-assessed health) by 33 percentage points,

which suggests that having this activity spatially isolated from residential places is an essential factor deterring indoor air pollution, irrespective of the fuel.

7.3. Demographic and geographic variations

Our data analysis results revealed the substantial heterogeneity of vulnerability of specific population groups to the adverse effects of polluting fuels. One of the heterogeneity moderators is age. Our findings have implications for public health and policy: older people are more vulnerable to the health impacts of polluting fuel consumption, which corresponds to previous studies proving age is a critical moderator of vulnerability to air pollution. About ARI effects, the most vulnerable group is, on average, 13 years older than the least vulnerable group.

The geographic differentials are also noticeable. Four provinces, namely the Eastern Cape, KwaZulu-Natal, North-West, and Mpumalanga, are overrepresented among the most susceptible. These geographic differentials may be related to climate factors, the nature of housing stock, the associated cooking practices, and the availability of substitutes. There are large provincial differentials, suggesting that national policies are inappropriate for these differentials, and that local circumstances and vulnerabilities should inform interventions.

Income appears to have a more complex moderating role. For self-assessed health, the moderating effect is positive, with the effect of log income being 0.71 points lower for the most impacted group, whereas it is negative in the case of ARI. This may reflect a potential trade-off, as wealthier households tend to enjoy better general health and nutrition levels. Still, they also have larger and better-insulated homes, where pollutants may accumulate due to inadequate ventilation.

7.4. Policy implications

Our results suggest that the energy transition policy in rural areas of South Africa and similar contexts should reconsider its policy direction, avoiding the strengthening of fuel stacking as an acceptable intermediate option, and instead encouraging clean fuel transitions. Although these may require higher upfront investments, they will yield a larger health payoff than transitional options.

With housing quality having such a strong moderating effect, energy programs may also be integrated with housing programs. Increasing ventilation, separate kitchens, and improved building materials may enhance the health benefits of households using cleaner fuels. For households that cannot fully transition to cleaner fuels, improving housing quality may provide at least some health benefits from polluting fuels, primarily through enhanced ventilation and improved design of spaces dedicated to cooking.

As the impacts are heterogeneous across the provinces, interventions should be developed on a regional basis. The provinces with higher vulnerability were identified (i.e., Eastern Cape, KwaZulu-Natal, North-West, and Mpumalanga), and these proved to require attention for the implementation of clean energy programs. Interventions should be tailored to regional characteristics, including common household forms, traditional cooking styles, and local climate conditions.

Results also highlight the importance of household-level determinants, enabling the targeting of interventions. Older households with inferior quality materials for their homes and those lacking a separate kitchen were identified as significant risk groups for using polluting fuels. The vulnerable households were also identified as needing to be targeted for clean energy subsidies and housing uplift interventions.

7.5. Methodological contributions

This study also has some methodological implications for further energy-health studies. First, we are performing multivalued treatment analysis, which allows treating fuel stacking as a separate treatment. This spares the researchers from the challenge of considering fuel stacking either as a clean fuel choice or a choice involving polluting fuel types. The findings of the present study also brought the original discovery that fuel stacking is not beneficial to health in comparison to choices involving the usage of only polluting fuels.

Secondly, the application of the SEA methodology in the analysis can provide more elucidating comparisons than average treatment effects. Analyzing averages and comparing the most and least affected groups can help highlight the features of households that correlate with the more problematic use of polluting fuels. This is particularly relevant for developing countries, where program heterogeneity is high, and average treatment effects can mask significant discrepancies in program influence.

Third, the application of double machine learning and Lewbel's heteroskedastic-based instrumental variables in the study is performed to address endogeneity issues and provide efficiency. The reliance on external instruments (peer effects) and internally constructed instruments allows one to conduct consistent causal identification while considering the possible problems with common exclusion restrictions.

8. Conclusion

These outcomes have presented some valuable insights into the synergistic relationship between fuel and health in rural South Africa housing and living conditions. Moreover, the findings are contrary to the energy ladder theory, which assumes a smooth transition of all households from using more polluting fuels to cleaner fuels. Instead of this paradigm, contrary to the expectations that fuel switching was generally praised as a step towards transition, there was no positive effect on health compared to the exclusive use of a single polluting fuel. Therefore, energy policymaking in developing nations should focus its efforts on accomplishing a complete transition to cleaner fuels rather than allowing an approach without health benefits.

Moreover, the heterogeneity revealed by the sorted effects analysis, concerning the second moderator — health risks from using the fuel — is not homogeneous across the sub-populations. The second moderator examined housing quality, building materials, and kitchen type, revealing that households that use unconventional materials for roofing and walls are more vulnerable to the effects of polluting fuels, suggesting synergies for joint interventions aimed at improving energy use and housing quality.

There are significant limitations to the study, which also point towards the future scope of work in this area. At the outset, even though we adopted instrumental variables and double machine learning techniques, causal inferences cannot be drawn from the cross-sectional nature of this design. Longitudinal studies tracing the energy transitions of households over time would lend more robustness to claims about the temporality of the relationship between health effects and changes in fuel use patterns.

Second, the health outcomes, self-assessed health status, and self-reported ARI may be susceptible to reporting bias. Future research that incorporates objective health measures, such as lung function or biomarkers of exposure to polluting agents from indoor air pollutants, shall be considered more conclusive. Furthermore, this study assessed individual-level health outcomes. It did not account for potential spillover health effects among household members, particularly children

who are more vulnerable to the impact of indoor air pollution.

Third, existing literature proposed that climatic variables (i.e., rainfall and temperature shocks) are key factors determining both energy selection and health impacts (Nawaz, 2021). Unfortunately, the analysis is unable to include climate data since the GHS dataset does not contain such particulars. A future study on the relationship between fuel consumption and health consequences, utilizing weather data, can significantly broaden our insight into the topic.

Fourth, although our theoretical framework and SEA results suggest significant potential interplays between fuel selection and housing situations, we were unable to empirically test the causal link between fuel stacking and housing quality due to limitations inherent in the cross-sectional nature of the GHS dataset. To obtain the directionality of the link, longitudinal data should be utilized.

Fifth, we acknowledge that housing materials may serve as significant potential moderators. However, we lack information on ventilation, cooking behavior, and time-activity distributions that could clarify some of the inconsistencies in health outcomes. Future efforts should focus on obtaining further information on housing characteristics, such as room volume, window position, and ventilation, to evaluate the potential mechanisms underlying these characteristics as moderators of health impacts linked to fuel usage.

It would be interesting to conduct further studies to gain insight into the disparity among provinces across different geographical regions. Qualitative studies may be conducted to investigate each province's regional traditional cooking practices, fuel availability, and other cultural factors that explain the apparent vulnerability of certain provinces to the detrimental impact of polluting fuels. Climatic variations among provinces may also contribute to differences in ventilation behavior or pollution levels, which are linked to temperature and humidity..

Our findings provide evidence that could facilitate a shift in the prevailing policy paradigm concerning energy transitions in developing economies. Policymakers should avoid viewing fuel stacking as an acceptable and tolerable outcome of energy transitions. Instead, they should recognize that it is only with complete energy transitions towards cleaner fuels that health gains can materialize. This would entail a major confrontation with current subsidy policies that promote the partial adoption of cleaner fuels and technologies. Instead of allocating resources to further this partial adoption, they should aim to encourage complete energy transitions for the most impacted and vulnerable households.

The existing research on the relationship between housing quality, fuel access, and related health impacts suggests that energy and housing policies working separately may not be optimal. There is potential for joint health benefits from energy access and housing quality improvements offered through a joined-up program. For families who, on a priority basis, may not be able to migrate to clean fuels, establishing ventilation and distinct cooking spaces for health professionals may be an acceptable interim measure.

Considering the diversity of the evidence based on the health impacts of clean fuel shift, strategies for implementation should be tailored for priority groups. These groups could include the older population, households dwelling in low-quality building materials, and households living in provinces with high vulnerability. A limited resource subsidy can be targeted through means-testing mechanisms to ensure that it is utilized most optimally by families in the greatest need of a clean fuel shift impact

Finally, we combine our methodological approach with the multi-valued treatment effect and majority effect patterns. It can serve as a model for further research in the development sphere. There is heterogeneity of treatment effects, and it can be tested. It is possible to suggest targeted advice instead of "one-size-fits-all" treatment effect-based advice. It is the compromise between the tailoring development policy and the circumstances of policy intervention in developing countries.

In summary the experiences of rural South Africa suggest that the availability of clean fuels is insufficient to achieve health benefits through energy transition. It is necessary to enable full transitions, attend to the relevant complementarities, and reach the most vulnerable with the interventions. These are essential lessons as countries worldwide are seeking to achieve sustainable development goals related to energy access and health, which are relevant for making energy policy more effective.

CRediT authorship contribution statement

Nouran Zenelabden: Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization, Investigation, Methodology, Validation, Visualization. **Adeola Oyenubi:** Writing – review & editing, Supervision, Formal analysis, Methodology. **Johane Dikgang:** Writing – review & editing, Supervision, Validation, Visualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used the AI language model Claude to assist in the preliminary organization of ideas and drafting (i.e., reasoning). After using this supportive tool/service, the author(s) thoroughly reviewed and edited the content as needed and took full responsibility for the content of the published article.

Declaration of competing interest

None.

Acknowledgements

The authors sincerely thank the African Economic Research Consortium (AERC) for the Doctoral Research Grant that supported the research from which this paper is derived.

Appendix A. Variables selected by Lasso algorithm

The following variables were selected by the Lasso algorithm for inclusion in the final models. The main covariates, which included marital status, education categories, age, income, household size, dwelling type, housing materials, gender, access to kitchen, whether they are from a female-headed household or not, free basic electricity, season, and province, have not gone through variable selection since they were included based on theory. The interaction and polynomial terms of these variables were included through Lasso selection.

A.1. Binary treatment model variable selection

Table A1: Binary active variables.

Binary interactions terms	
SAH	ARI
married:walls_others	married:walls_others
married:EC	married:floor_non-standard
married:NC	married:summer
married:FS	junior_education:age
senior_education:roof_others	junior_education:roof_others
senior_education:autumn	junior_education:EC
age:hh_size	junior_education:NC
age:KZN	junior_education:MP
age:NW	senior_education:age
age:LM	senior_education:FBE
hh_formal:EC	senior_education:NC
hh_formal:FS	senior_education:FS
roof_others:KZN	age:spring
floor_non-standard:summer	age:EC
floor_non-standard:autumn	age:NC
floor_non-standard:KZN	age:GAU
floor_non-standard:LM	logincome:MP
female:LM	hh_size:EC
Kitchen:MP	hh_size:FS
FBE:autumn	hh_size:KZN
FBE:EC	hh_formal:floor_non-standard
FBE:NC	hh_formal:autumn
FBE:LM	hh_formal:EC
spring:FS	hh_formal:NC
autumn:EC	hh_formal:FS
	hh_formal:MP
	roof_others:KZN
	walls_others:female
	walls_others:FBE
	walls_others:EC
	walls_others:NW
	walls_others:MP
	floor_non-standard:FBE
	floor_non-standard:MP
	female:FS
	female:KZN
	female:NW
	female:LM
	Kitchen:NW
	Kitchen:GAU
	FBE:spring
	FBE:EC
	FBE:NC
	FBE:MP
	FBE:LM
	spring:NC
	spring:LM
	summer:MP
	summer:LM
	autumn:FS
	autumn:MP

The table of binary active variables displays the key core covariates available in the SAH and ARI models.

Table A.2 Multivalued treatment model variable selection.

Multivalued treatment active variables	
PH	ARI
married	married
junior_education	junior_education
senior_education	senior_education
age	age
logincome	logincome
hh_size	hh_size
hh_formal	hh_formal
roof_others	roof_others
walls_others	walls_others

(*continued on next page*)

(*continued*)

Multivalued treatment active variables	
PH	ARI
floor_non-standard	floor_non-standard
female	female
Kitchen	Kitchen
FBE	FBE
spring	spring
summer	summer
autumn	autumn
	FS

Core covariates to the multivalued treatment models are presented in Table A2. They are akin to the binary models, with differences arising solely from the Lasso selection.

Table A3: Multivalued treatment interaction terms.

Multivalued treatment interaction terms	
PH	ARI
age:KZN	married:floor_non-standard
age:NW	married:FBE
	junior_education:age
	junior_education:roof_others
	junior_education:EC
	junior_education:NC
	junior_education:MP
	senior_education:FBE
	age:roof_others
	age:floor_non-standard
	age:female
	age:spring
	age:EC
	age:NC
	age:NW
	age:MP
	age:LM
	logincome:FS
	hh_size:KZN
	hh_size:MP
	hh_formal:roof_others
	hh_formal:floor_non-standard
	hh_formal:summer
	hh_formal:EC
	floor_non-standard:FBE
	floor_non-standard:summer
	floor_non-standard:KZN
	floor_non-standard:MP
	female:KZN
	female:LM
	Kitchen:NC
	Kitchen:NW
	FBE:EC
	spring:NC
	spring:MP
	summer:LM

Finally, the interaction terms from the multivalued treatment models reveal trends like those found in the binary models. Housing-related features again appear most often among the selected interactions, meaning that housing quality is a reasonable candidate as a moderator of the impact of fuel on health.

Appendix B. Additional sorted effects analysis

In this section, we conduct a sorted effects analysis for the contrasting case of using only clean fuels –exclusive use – versus using only polluting fuels – exclusive use, excluding mixed fuels.

B.1. Self-assessed health analysis

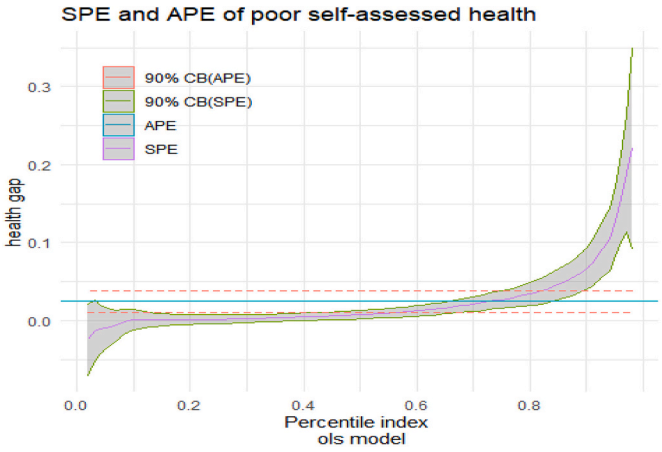

Fig. B1: SPE and APE for clean versus polluting fuels (self-rated health).

Figure B1 presents the sorted partial effects and average partial effects for self-assessed health. For users of clean fuels versus those using only polluting fuels, there is no heterogeneity in this comparison regarding self-assessed health, which differs from the analysis of mixed fuels. The SPE curve is relatively flat, which suggests that the exclusive use of polluting fuel compared to clean fuels is likely to have a similar impact on health across the population.

B.2. Acute respiratory infection analysis

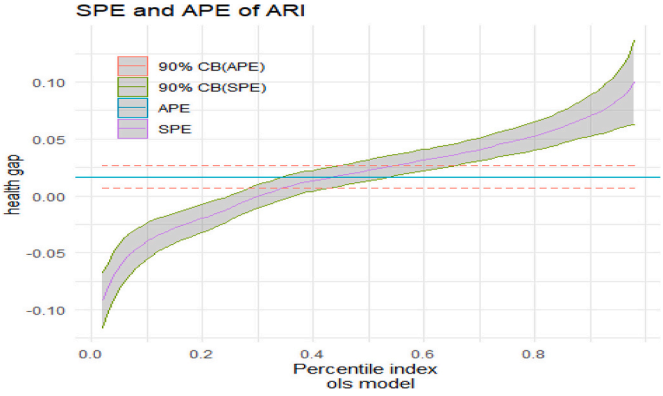

Fig. B2: SPE and APE comparisons of clean versus polluting fuels (ARI).

Figure B2 illustrates significant heterogeneity in the ARI outcomes of the comparison between users of clean and polluting fuels. Moreover, the spikes on both sides of the distribution show that the fuel type used in a household has a significant interaction with specific household characteristics related to ARI health outcomes.

Table B1: Sorted effects analysis clean versus polluting fuels (ARI).

	Most (Top 10 %)	SE	Least (Bottom 10 %)	SE	Estimate	SE	PW P-vals
hh_size	5.50	0.44	4.89	0.30	0.61	0.70	0.19
age	22.38	4.17	31.80	2.26	−9.42	6.31	0.07*
female	0.65	0.08	0.55	0.05	0.11	0.13	0.21
married	0.08	0.05	0.25	0.05	−0.17	0.09	0.03**
senior_education	0.12	0.09	0.55	0.06	−0.43	0.14	0.00***
junior_education	0.48	0.09	0.25	0.05	0.23	0.15	0.06**
FBE	0.06	0.04	0.32	0.05	−0.26	0.09	0.00***
logincome	8.37	0.18	7.77	0.22	0.60	0.37	0.06*
spring	0.23	0.07	0.24	0.05	−0.01	0.11	0.48
summer	0.07	0.06	0.27	0.06	−0.21	0.11	0.03**
autumn	0.40	0.08	0.15	0.04	0.25	0.11	0.01***
roof_others	0.02	0.01	0.00	0.00	0.01	0.01	0.13
walls_others	0.00	0.00	0.03	0.02	−0.03	0.02	0.10*
floor_non-standard	0.08	0.03	0.02	0.02	0.06	0.05	0.11

(continued on next page)

(*continued*)

	Most (Top 10 %)	SE	Least (Bottom 10 %)	SE	Estimate	SE	PW P-vals
hh_formal	0.80	0.05	0.74	0.05	0.06	0.10	0.26
Kitchen	0.83	0.06	0.82	0.04	0.01	0.10	0.45
WC	0.03	0.03	0.00	0.01	0.03	0.03	0.15
EC	0.00	0.01	0.48	0.09	−0.48	0.09	0.00***
NC	0.03	0.04	0.01	0.03	0.01	0.06	0.42
FS	0.00	0.00	0.10	0.00	−0.10	0.00	0.00***
KZN	0.51	0.10	0.01	0.02	0.50	0.11	0.00***
NW	0.00	0.00	0.36	0.08	−0.36	0.08	0.00***
GAU	0.01	0.01	0.00	0.00	0.01	0.01	0.05**
MP	0.09	0.07	0.03	0.04	0.05	0.10	0.30**
LM	0.33	0.08	0.01	0.02	0.32	0.09	0.00***

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B1 presents the results of the classification analysis method for the ARI outcome, comparing the most affected groups (top 20 %) and the least affected groups (bottom 20 %). The highlighted results are:

- Age: Individuals in the most-affected group were younger than those in the least-affected (22.38 vs 31.80 years; difference 9.42 years; $p < 0.10$).
- Education: Senior education was less common in the most-affected group than in the least-affected (12 % vs 55 %; difference 43 percentage points; $p < 0.01$).
- Housing materials: As for wall materials, unconventional wall materials were associated with the least affected group, with a 3 percentage-point difference ($p < 0.10$).
- Geographic location: There were strong provincial effects in the representation of the most and least affected groups in the Eastern Cape, KwaZulu-Natal, and Limpopo.

Appendix C. Robustness checks

C.1. External instrument only analysis

We reported in the main text the results of the analysis using the external instrument (the proportion of PSU-level unclean fuel use) only, without the Lewbel internal instruments. Results (omitted here for brevity) corroborated our main findings but yielded larger standard errors, as anticipated with fewer instruments.

C.2. Sensitivity to covariate selection

We assessed the robustness of our results to variations in the Lasso penalty parameter specifications. We apply a data-driven approach by Knaus (2021) that considers alternative values of λ along the penalty grid within one standard error. The analysis of results obtained using different penalties indicates that the method produces stable results for plausible penalty terms. The sensitivity analysis results for poor health are presented in Fig. C2, and the results for ARI are in Fig. C3.

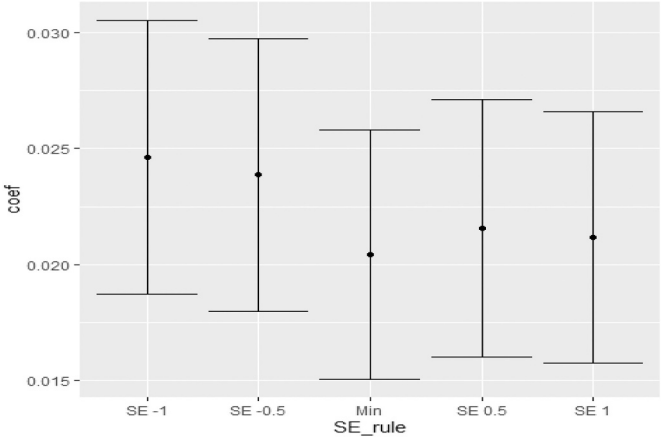

Fig. C2: Sensitivity analysis (poor health).

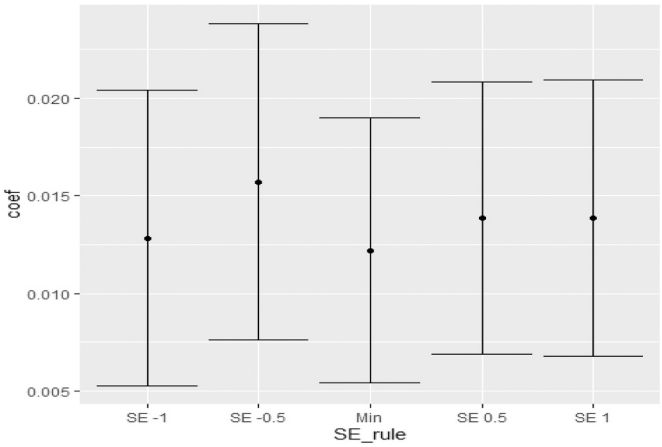

Fig. C3: Sensitivity analysis (ARI).

As model complexity increases, the estimated parameter for poor health status and ARI does not change significantly, which suggests that the ATE is unaffected by parameter selection.

Appendix D. Additional technical notes

D.1. Construction of housing quality variables

Roofing other than cement, tiles, or similar materials: corrugated iron/zinc, wood, plastic, cardboard, mud, and other unconventional materials. Unconventional walls: mud, wood, corrugated iron, plastic, and similar materials. Flooring other than cement, tiles, or similar materials: mud, wood, and other unconventional materials.

D.2. Fuel classification details

Fuel types are categorized into the following: clean fuels – main electricity, solar power, and gas; polluting fuels - wood, coal, paraffin, and dung; and mixed fuels - households that report use of both clean and polluting fuel types.

D.3. Diagnostic tests for the Lewbel IV approach

The following assumption is required from the Lewbel (2012) approach:

- Instruments generate heteroskedasticity in the treatment model but not the outcome model. Instruments are uncorrelated with the product of the disturbances of the treatment and outcome equations.
- To test such assumptions, we applied the Breusch-Pagan test. For self-assessed health, non-standard floors and roofs passed the test. For ARI, non-standard roofs passed. The overall result of the tests showed that the instrumental variables approach is valid.

Appendix E. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eneco.2025.108926.

References

Adesina, J.A., Piketh, S.J., Qhekwana, M., Burger, R., Language, B., Mkhatshwa, G., 2020. Contrasting indoor and ambient particulate matter concentrations and thermal comfort in coal and non-coal burning households at South Africa Highveld. Sci. Total Environ. 699, 134403.
Albers, P.N., Wright, C.Y., Voyi, K.V., Mathee, A., 2015. Household fuel use and child respiratory ill health in two Mpumalanga, South Africa, towns. S. Afr. Med. J. 105 (7), 573–577.
Alem, Y., Beyene, A.D., Köhlin, G., Mekonnen, A., 2016. Modeling household cooking fuel choice: a panel multinomial logit approach. Energy Econ. 59, 129–137.
Athey, S., 2017. Beyond prediction: using big data for policy problems. Science 355, 483–485.
Bakhsh, K., Sadiqa, A., Yasin, M.A., Haider, S., Ali, R., 2020. Exploring the nexus between households' choice of cooking fuels, sanitation facilities, and access to information in Pakistan. J. Clean. Prod. 257, 120621.
Barnes, B., Mathee, A., Thomas, E., Bruce, N., 2009. Household energy, indoor air pollution, and child respiratory health in South Africa. J. Energy South. Afr. 20 (1), 4–13.
Batchelor, S., Brown, E., 2021. Editorial for special issue: "Clean energy innovations: challenges and strategies for low- and middle-income countries,". Energies 14 (24), 8577.
Baum, C.F., Lewbel, A., Schaffer, M.E., Talavera, O., 2019. Advice on using Heteroskedasticity-based identification. The Stata Journal 19 (4), 757–767.
Boch, S.J., Taylor, D.M., Danielson, M.L., Chisolm, D.J., Kelleher, K.J., 2020. "Home is where the health is": housing quality and adult health outcomes in the survey of income and program participation. Prev. Med. 132, 105990.
Bodory, H., Huber, M., Laffers, L., 2022. Evaluating (weighted) dynamic treatment effects by double machine learning. Econ. J. 25 (3), 628–648.
Bruce, N., Pope, D., Rehfuess, E., Balakrishnan, K., Adair-Rohani, H., Dora, C., 2015. WHO indoor air quality guidelines on household fuel combustion: strategy implications of new evidence on interventions and exposure–risk functions. Atmos. Environ. 106, 451–457.
Buthelezi, S.A., Kapwata, T., Wernecke, B., Webster, C., Mathee, A., Wright, C.Y., 2019. Household fuel use for heating and cooking and respiratory health in a low-income, south African coastal community. Int. J. Environ. Res. Public Health 16 (4), 550.
Carter, E., et al., 2019. Household transitions to clean energy in a multiprovincial cohort study in China. Nat. Sustain. 2 (3), 224–233.

Checkley, W., Pollard, S.L., Siddharthan, T., Babu, G.R., Thakur, M., Miele, C.H., Van Schayck, O.C., 2016. Managing threats to respiratory health in urban slums. Lancet Respir. Med. 4 (11), 852–854.
Chen, K., Wang, W., Qiu, J., Guo, W., Du, J., Gao, B., Hu, Z., Qiao, H., 2024. Housing conditions, cooking fuels, and health-related quality of life among rural middle-aged and elderly in Northwest China: a ten-year balanced panel study. Prev. Med. Rep. 37, 102563.
Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J., 2018. Double/debiased machine learning for treatment and structural parameters. Econ. J. 21, C1–C68. https://doi.org/10.1111/ectj.12097.
Conley, T.G., Udry, C.R., 2010. Learning about a new technology: pineapple in Ghana. Am. Econ. Rev. 100 (1), 35–69. https://doi.org/10.1257/aer.100.1.35.
Dongzagla, A., Adams, A.M., 2022. Determinants of urban household choice of cooking fuel in Ghana: do socioeconomic and demographic factors matter? Energy 256, 124613.
Edwards, J.H., Langpap, C., 2012. Fuel choice, indoor air pollution, and children's health. Environ. Dev. Econ. 17 (4), 379–406.
Feuerriegel, S., Frauen, D., Melnychuk, V., Schweisthal, J., Hess, K., Curth, A., Bauer, S., Kilbertus, N., Kohane, I.S., van der Schaar, M., 2024. Causal machine learning for predicting treatment outcomes. Nature Medicine 30 (4), 958–968.
Ghoddusi, H., Creamer, G.G., Rafizadeh, N., 2019. Machine learning in energy economics and finance: a review. Energy Econ. 81, 709–727.
Hanna, R., Duflo, E., Greenstone, M., 2016. Up in smoke: the influence of household behavior on the long-run impact of improved cooking stoves. Am. Econ. J. Econ. Pol. 8 (1), 80–114.
Hosier, R.H., Dowd, J., 1987. Household fuel choice in Zimbabwe: an empirical test of the energy ladder hypothesis. Resour. Energy 9 (4), 347–361.
Hou, B., Wu, J., Mi, Z., Ma, C., Shi, X., Liao, H., 2022. Cooking fuel types and health effects: A field study in China. Energy Policy 167, 113012. https://doi.org/10.1016/j.enpol.2022.113012.
Knaus, M.C., 2021. A double machine learning approach to estimate the effects of musical practice on students' skills. J. Roy. Statist. Soc. Ser. A 184, 282–300. https://doi.org/10.1111/rssa.12623.
Kowsari, R., Zerriffi, H., 2011. Three-dimensional energy profile: a conceptual framework for assessing household energy use. Energy Policy 39 (12), 7505–7517.
Krishnamoorthy, Y., Rajaa, S., Ramasubramani, P., Saya, G.K., 2022. Association between indoor air pollution and cognitive function among a nationally representative sample of middle-aged and older adults in India—a multilevel modelling approach. Indoor Air 32 (1), e12929.
Lewbel, A., 2012. Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. J. Bus. Econ. Stat. 30 (1), 67–80.
Liang, L., Wu, W., Lal, R., Guo, Y., 2013. Structural change and carbon emission of rural household energy consumption in Huantai, northern China. Renew. Sust. Energ. Rev. 28, 767–776.
Liu, J., Hou, B., Ma, X.W., Liao, H., 2018. Solid fuel use for cooking and its health effects on the elderly in rural China. Environ. Sci. Pollut. Res. 25, 3669–3680.
Liu, Z., Li, J., Rommel, J., Feng, S., 2020. Health impacts of cooking fuel choice in rural China. Energy Econ. 89, 104811. https://doi.org/10.1016/j.eneco.2020.104811.
Malakar, S., Day, R., 2020. Differences in firewood users' and LPG users' perceived relationships between cooking fuels and women's multidimensional well-being in rural India. Nat. Energy 5 (6), 486–494.
Manski, C.F., 1993. Identification of endogenous social effects: the reflection problem. Rev. Econ. Stud. 60 (3), 531–542.
Masera, O.R., Saatkamp, B.D., Kammen, D.M., 2000. From linear fuel switching to multiple cooking strategies: a critique and alternative to the energy ladder model. World Dev. 28 (12), 2083–2103.
Mullainathan, S., Spiess, J., 2017. Machine learning: an applied econometric approach. J. Econ. Perspect. 31, 87–106. https://doi.org/10.1257/jep.31.2.87.
Muller, C., Yan, H., 2018. Household fuel use in developing countries: review of theory and evidence. Energy Econ. 70, 429–439.
Nawaz, S., 2021. Energy poverty, climate shocks, and health deprivations. Energy Economics 97, 105204.
Oliveras, L., Peralta, A., Palència, L., Gotsens, M., López, M.J., Artazcoz, L., Borrell, C., Marí-Dell'Olmo, M., 2021. Energy poverty and health: trends in the European Union before and during the economic crisis, 2007–2016. Health Place 67, 102294.
Piketh, S., Wernecke, B., Burger, R., et al., 2016. Household air pollution in south African low-income settlements: a case study. WIT Trans. Ecol. Environ. 207, 227–236.
Posel, D., Gelo, D., Casale, D., Oyenubi, A., 2024. Sorting the gender earnings gap: heterogeneity in the South African labor market. Feminist Economics 30 (1), 106–133.
Qiu, X., Jin, J., He, R., Zhang, C., 2023. Do solid fuels for cooking lead to an increased prevalence of respiratory disease? Empirical evidence from rural China. Energy Sustain. Dev. 74, 297–308.
Quesnel Vallée, A., 2007. Self-rated health: caught in the crossfire of the quest for 'true' health? Int. J. Epidemiol. 36 (6), 1161–1164.
Rathi, S.S., Vermaak, C., 2018. Rural electrification, gender, and the labor market: a cross-country study of India and South Africa. World Dev. 109, 346–359.
Roomaney, R.A., Wright, C.Y., Cairncross, E., Abdelatif, N., Cois, A., Turawa, E.B., Awotiwon, O.F., Neethling, I., Nojilana, B., Pacella, R., Bradshaw, D., 2022. Estimating the disease burden attributable to household air pollution from cooking with solid fuels in South Africa for 2000, 2006, and 2012. S. Afr. Med. J. 112 (8b), 718–728.
Rosenthal, J.K., et al., 2018. Clean cooking and the SDGs: integrated analytical approaches to guide energy interventions for health and environment goals. Energy Sustain. Dev. 42, 1–10.
Ruiz-Mercado, I., Masera, O., 2015. Patterns of stove use in the context of fuel–device stacking: rationale and implications. EcoHealth 12, 42–56.
Sacerdote, B., 2001. Peer effects with random assignment: results for Dartmouth roommates. Q. J. Econ. 116 (2), 681–704.
Sharma, A., 2019. Household fuel transition and determinants of firewood demand in Nepal. Econ. J. Develop. Iss. 25 (1–2), 1–10.
Shirinde, J., Wichmann, J., Voyi, K., 2014. Association between wheeze and selected air pollution sources in an air pollution priority area in South Africa: a cross-sectional study. Environ. Health 13, 1–12.
Stats SA (Statistics South Africa), 2019. General Household Survey. Statistics South Africa, Pretoria.
Sulla, V., Zikhali, P., Cuevas, P.F., 2022. Inequality in Southern Africa: An Assessment of the Southern African Customs Union. International Bank for Reconstruction and Development - The World Bank.
Tatham, J., Muchapondwa, E., 2021. The Impact of Solid Cooking Fuels on the Health of Rural South African Households. master's dissertation. University of Cape Town. https://www.gu.se/sites/default/files/2021-11/Thesis.pdf (accessed February 15, 2024).
Tian, Z., Tian, Y., Shen, L., Shao, S., 2021. The health effect of household cooking fuel choice in China: an urban-rural gap perspective. Technol. Forecast. Soc. Chang. 173, 121083.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Stat Methodol. 58, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
Twumasi, M.A., Jiang, Y., Ameyaw, B., Danquah, F.O., Acheampong, M.O., 2020. The impact of credit accessibility on rural households' clean cooking energy consumption: the case of Ghana. Energy Rep. 6, 974–983.
Van der Kroon, B., Brouwer, R., Van Beukering, P.J., 2014. The impact of the household decision environment on fuel choice behavior. Energy Econ. 44, 236–247.
Wang, B., Liu, Y., Li, Z., Li, Z., 2016. Association of indoor air pollution from coal combustion with influenza-like illness in housewives. Environ. Pollut. 216, 646–652.
Wassie, Y.T., Rannestad, M.M., Adaramola, M.S., 2021. Determinants of household energy choices in rural sub-Saharan Africa: an example from southern Ethiopia. Energy 221, 119785.
World Bank, 2020. Poverty and Equity Brief Sub-Saharan. South Africa, Africa.
Xu, M., Ke, P., Wang, C., Di, H., Meng, X., Xia, W., Gan, Y., He, Y., Tian, Q., Jiang, H., Lu, Z., 2022. Cooking with biomass fuels and mortality among Chinese older adults: a prospective cohort study. Indoor Air 32 (1), e12954.
Zenelabden, N., Dikgang, J., 2022. Satisfaction with water services delivery in South Africa: the effects of social comparison. World Dev. 157, 105861.
Zhu, X., Zhu, Z., Zhu, B., Wang, P., 2022. The determinants of energy choice for household cooking in China. Energy 260, 124987.
Elasu, A., Ntayi, J., Adaramola, M.S., Buyinza., F., Atukunda, R., Ngoma, M., 2024. Perceived satisfaction and household fuel transition: evidence from Ugandan households. Front. Sustain. 5, 1404806.