nature sustainability

Analysis

https://doi.org/10.1038/s41893-023-01259-6

Pervasive over-crediting from cookstove offset methodologies

Received: 2 July 2023

Accepted: 12 December 2023

Check for updates

Annelise Gill-Wiehl[1], Daniel M. Kammen[1,2,3,4] & Barbara K. Haya[2,5]

Cookstove carbon offset projects can progress multiple Sustainable Development Goals (SDGs), including climate, energy, health, gender, poverty and deforestation. However, project emission reductions must be accurately or conservatively estimated to avoid undermining climate action and long-term SDG financing. Here we conduct a comprehensive, quantitative, quality assessment of offsets by comparing five cookstove methodologies with published literature and our own analysis. We find misalignment, in order of importance, with fraction of non-renewable biomass, firewood–charcoal conversion, stove adoption, stove usage, fuel consumption, stacking (using multiple stoves), rebound and emission factors. Additionality, leakage, permanence and overlapping claims require more research. We estimate that our project sample is over-credited 9.2 times. Gold Standard's metered methodology, which directly monitors fuel use, is most aligned with our estimates (1.5 times over-credited) and has the largest potential for emission abatement and health benefit. We provide recommendations to align methodologies with current science and SDG progress.

Roughly 2.4 billion people globally cook with smoky solid fuels or kerosene, contributing to 2–3 million premature deaths annually[1] and roughly 2% of global greenhouse gas (GHG) emissions[2]. Efficient cookstoves can support multiple Sustainable Development Goals (SDGs) including climate, energy, health, gender, poverty and deforestation. Monetizing the GHG emission reductions from efficient cookstove projects through the voluntary carbon market (VCM) has the potential to provide substantial financing for these projects.

Efficient stoves can reduce emissions by (1) using less fuel or switching to a less GHG-intensive fuel and/or (2) reducing the release of methane and other pollutants through more complete fuel combustion. While improved stoves are often touted for their health benefits, only solar, electric, gas, ethanol and, currently, two forced-draft pellet stoves reduce smoke enough to meaningfully reduce disease risk and meet the World Health Organization (WHO) definition of 'clean'[3] (Supplementary Information).

Cookstove projects with credits on the VCM are registered under the Gold Standard (GS) and Verified Carbon Standard offset registries, and estimate carbon emission reductions using methodologies primarily developed by GS and the Clean Development Mechanism (CDM). Cookstoves, one of the fastest growing project types on the VCM, represented 1,213 out of the 7,933 project activities (individually registered or included in a programme of activities) on the VCM[4] and generated ~78.9 million total issued credits (as of 10 May 2023). Most VCM cookstove project activities replace three stone fires or inefficient biomass stoves with improved firewood stoves, while 43 project activities distribute only WHO-defined clean stoves/fuels (Fig. 1).

Studies of offset project quality have documented substantial excess crediting (as much as 13 times from single factors) from improved forest management[5,6], avoided deforestation[7,8] and the United Nations system[9,10]. Over-crediting is harmful to effective climate action, the buyer and the cookstove sector. Poor-quality credits can undermine climate action by justifying ongoing emissions and replacing direct emission reduction and other more effective climate mitigation activities, even if some reduction is achieved. Excess crediting obscures the overall effectiveness of climate efforts and progress

[1]Energy and Resources Group, University of California, Berkeley, CA, USA. [2]Goldman School of Public Policy, University of California, Berkeley, CA, USA. [3]Department of Nuclear Engineering, University of California, Berkeley, CA, USA. [4]United States Agency for International Development, Washington, DC, USA. [5]California Institute for Energy and Environment, University of California, Berkeley, CA, USA. e-mail: agillwiehl@berkeley.edu

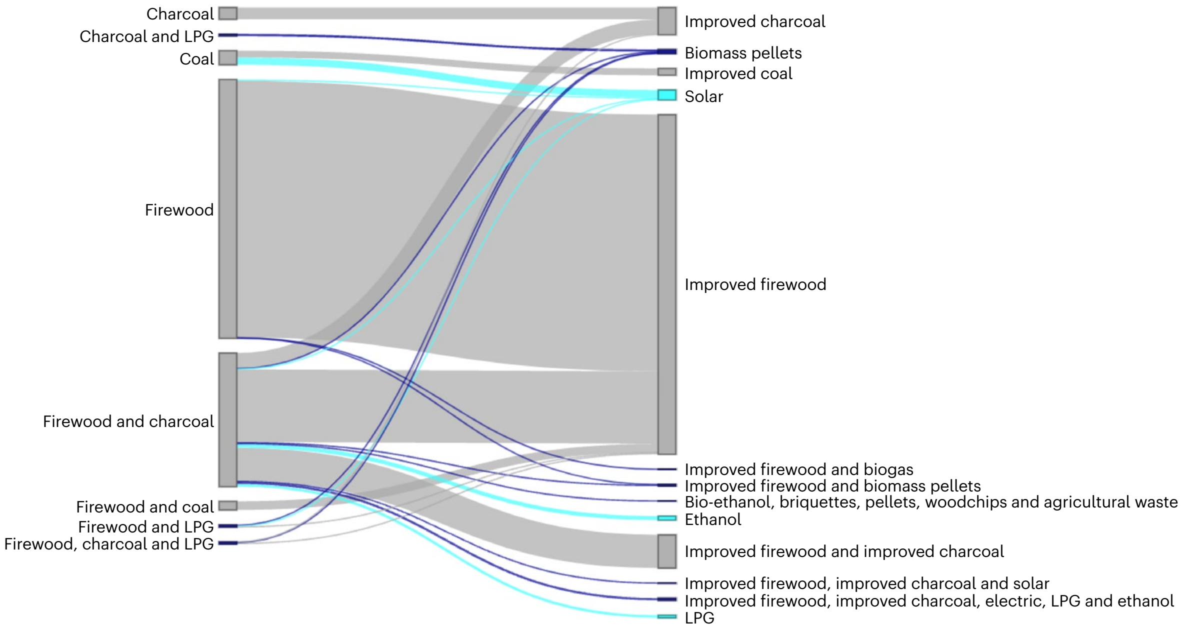

Fig. 1 | Transitions from baseline to project fuels by cookstove carbon offset projects. The left side of the diagram indicates the majority baseline fuels before intervention, and the right side represents the project fuel/stove that the VCM-funded project implemented. The width of the link indicates the relative number of projects. Grey indicates WHO polluting or transitional fuels or stoves (tiers 0–3). Dark blue indicates a mix of WHO clean, transitional or polluting fuels and stoves, while cyan indicates only WHO clean fuels or stoves. We exclude six projects that do not change the stove, but only replaced firewood with agricultural waste. As of 9 November 2022, 4% of cookstove project activities (43 out of 992 projects) registered on the VCM distribute only cooking fuels or stoves that meet the WHO's definition of clean, that is, they meet tier 5 for carbon monoxide and tier 4 for particulate matter.

towards ambitious climate targets. Over-crediting also creates confusion and reputational/legal risk for buyers. Lack of trust that a credit actually represents one metric ton less carbon dioxide equivalent weakens the market and its ability to support efficient cookstoves and all of their SDG benefits.

Studies of cookstove offset projects, covering single or a few factors, found over-crediting from the choice of fraction of non-renewable biomass (fNRB)[11] and methods for track adoption/usage rates[12], and under-crediting from emission factors (EFs)[13,14]. Qualitative studies have discussed quantification challenges and uncertainty[15,16]. This study fills multiple research gaps by performing a comprehensive quantitative assessment of offset credit quality, taking into account interactions in over/under-crediting across all methodology factors for all major cookstoves methodologies, and demonstrating how such a quality assessment can be performed on an offset methodology.

In this Analysis, we (1) discuss the accuracy of all estimation factors used (or not addressed) by the four most prominent cookstoves offset methodologies (GS-technologies and practices to displace decentralized thermal energy consumption (TPDDTEC)[17], GS-simplified methodology for clean and efficient cookstoves (simplified)[18], CDM-energy efficiency measures in thermal applications of non-renewable biomass (AMS-II-G)[19] and CDM-switch from non-renewable biomass for thermal applications by the user (AMS-I-E)[20]) and the recent GS-methodology for metered and measured energy cooking devices (metered)[21] methodology (Table 1; past and current versions), drawing from published literature and our own analysis (Methods). (2) We then recalculate the carbon emission reductions of a purposive sample of 51 cookstove projects, addressing ranges of uncertainty using the Monte Carlo method (MCM), and compare those results with actual credit issuance across eight methodology/stove type categories. (3) We suggest a specific set of methodological reforms to generate high-quality credits. In doing so, (4) we develop and demonstrate an over/under-crediting analysis that can be used to systematically assess quality and inform methodology improvements across all offset project types.

Data and sampling

We identified the 15 countries with the most credits from cookstove projects on the market, and for each country selected the largest projects for each methodology. In addition, we randomly included small- and medium-size projects globally, and covered all types of fuel transition, except electric (Methods). This approach resulted in a sample of 51 projects spanning 25 countries, and accounts for 40% of all issued credits from these cookstove methodologies on the VCM (as of 10 May 2023; Fig. 2).

Results

Here, we first summarize the major factors affecting offset quality assessed in our quantitative analysis and the accuracy of their treatment by each of the methodologies compared with the published literature. For a more detailed discussion of each factor, as well as discussion of additionality, leakage, permanence and overlapping claim, see Supplementary Information.

Adoption, usage and stacking rates

Efficient cookstove projects reduce emissions to the extent that users (1) 'adopt' a more efficient project stove defined as the percentage of distributed stoves actually in use; (2) use the project stove, where 'usage' is defined as the percentage of meals cooked using the project stove; and (3) stop or reduce 'stacking', defined as the percentage of meals cooked using the baseline stove(s) in concert with the project stove. These rates are used to determine the change between pre- and post-project fuel use.

Methods for monitoring adoption, usage and stacking fall into three categories: the AMS-I-E, GS-simplified and AMS-II-G, which track them through short cross-sectional surveys. GS-TPDDTEC requires in-field multi-day kitchen performance tests (KPTs) for a sample of households, capturing both usage and stacking rates by directly measuring daily fuel usage. The results are then applied to the full set of project households through surveyed adoption rates. GS-metered

https://doi.org/10.1038/s41893-023-01259-6

Table 1 | Outlining the quantification equations, approaches and data sources of five cookstove methodologies on the VCM

Parameter (typical units)								
Methodology	**Adoption × (stove-days)**	**[Baseline fuel use × (tons per stove per day)**	**Baseline EFs[a] (CO_2–equivalent per ton of fuel)**	**– Project fuel use × (tons per stove per day)**	**Project EFs (CO_2–equivalent per ton of fuel)]**	**× Adjustment for stacking (%)**	**×/– Leakage**	**Scope/applicability**
GS-TPDDTEC	Surveys	Default value, historical data, literature or KPT (rarely chosen)	Methodology defaults, IPCC values, or literature non-CO_2 gases and upstream emissions are optional	KPT	Same EF as the Baseline EF	N/A captured in KPT	Choice between discounting emission reductions by 5%, evaluating leakage through surveys or ignore entirely if deemed very low. This methodology can include stacking in the leakage figure	Credits for less GHG-intensive stoves (for example, improved biomass, heat retention, solar, LPG, electric). This is the most versatile methodology of the five. As of 2020, clean stoves must be registered under GS-metered.
CDM-AMS-I-E	Surveys	Ethanol projects determine the equivalent amount of biomass equal to the amount of energy from ethanol used in the project scenario from a KPT, surveys or literature.	Uses EF of a projected fossil fuel constructed per region from a weighted average of fossil fuel types	KPT or surveys	Differs by fuel type Ethanol projects typically include ethanol production, electricity consumption and transport emissions here	Surveys	Choice between discounting emission reductions by 5%, evaluating leakage through surveys or ignore entirely if deemed very low	Replaces non-renewable biomass with renewable energy (for example, renewable biomass, biogas, bioethanol, electric stoves)
GS-simplified	Surveys	Default value, historical data or sample surveys	Methodology defaults, IPCC values, or literature non-CO_2 gases and upstream emissions are optional	Multiply baseline consumption by the ratio baseline and project stove efficiencies derived from a water boil test or other efficiency assessment	N/A emission reductions are calculated by stove efficiency improvement	Surveys	Choice between discounting emission reductions by 5%, evaluating leakage through surveys or ignore entirely if deemed very low	Replaces traditional cookstoves with improved wood or charcoal stoves. Designed for smaller projects, capped at 10,000 CO_2e per year
CDM-AMS-II-G	Surveys	Default value, historical or survey data, or country specific default	Uses EF of a projected fossil fuel constructed per region from a weighted average of fossil fuel types	Multiply baseline consumption by the difference between baseline and project stove efficiencies derived from a water boil test or other efficiency assessment	N/A emission reductions are calculated by stove efficiency improvement	Surveys	Discount emission reductions by 5% This methodology can include upstream emissions for charcoal or processed biomass production in leakage.	Replaces traditional cookstoves with improved biomass (wood, charcoal, pellet and so on) cookstoves, ovens or dryers Designed for smaller projects, capped at 60 GWh
	Project energy (TJ) ×		**Baseline EFs (–equivalent per TJ)**		**– Project emissions (–equivalent per TJ)**		**×/– Leakage**	
GS-metered[b]	Meters or sales data		Constructed from a weighted average of baseline fuel types displaced in the project activity from historical, literature or survey data		Differs for fuel type (electric or fossil fuel); however, for pellets, which is the GS-metered stove type our sample covers, project emissions are included as either additional leakage (that is, 5–10%) or a standard tons CO_2 per year value for pellet production.		Choice between discounting emission reductions by 5%, evaluating leakage through surveys or ignore entirely if deemed very low. This methodology can include upstream emissions and stacking in leakage figure	Designed for cookstoves with metered or other direct fuel monitoring (for example, purchase records) such as electric, LPG, biogas, bioethanol or biomass pellet

All methodologies allow projects to use fNRB default values or calculate their own with a CDM tool. Nominal caloric value values are from IPCC estimates or project specific testing. We include the exact equations in Supplementary Information, but here outline the approaches using the language defined in the main text (for example, our definition of adoption rate). In the column heads, the '×' and '–' are multiplication and subtraction, respectively, to indicate the following formula: adoption × (baseline fuel use × baseline EF – project fuel use × project EF) ×/– leakage. N/A, not applicable. [a]EFs are calculated as fNRB (%)×(CO_2 emissions (ton of CO_2 per terajoule (TJ) of fuel)+non-CO_2 emissions (ton of CO_2–equivalent per TJ))× nominal caloric value of fuel (TJ per ton). [b]GS-metered also has an option based on specific energy consumption, where thermal energy efficiency is complicated by other factors such as pressure.

https://doi.org/10.1038/s41893-023-01259-6

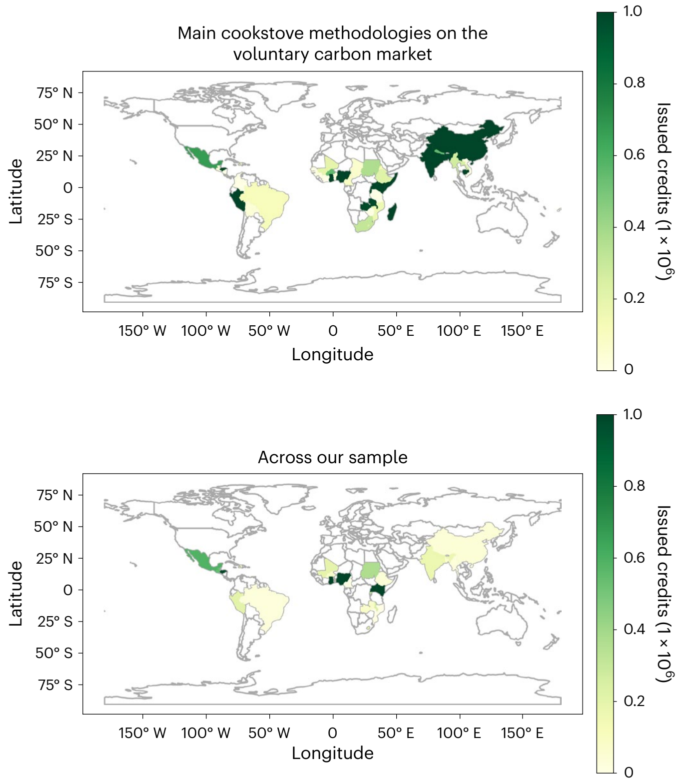

Fig. 2 | Issued credits across the VCM and our sample. Credits issued so far on the VCM across the five methodologies covered as of 9 November 2022 (top panel) and from the 51 cookstove project activities in our sample (bottom panel). We cover the GS-TPDDTEC, GS-simplified, CDM-AMS-II-G, CDM-AMS-I-E and GS-metered.

uses the most robust approach, directly tracking project stove and fuel use in all participating households through meters or fuel sales data.

The methodologies' default surveys range in quality, but all are infrequent and vulnerable to social desirability[22,23] and recall[23,24] biases. For example, AMS-II-G's default survey simply asks households if they used the improved stove in the last week or month. Credits are generated for all households that reply 'yes' as if they used the stove 100% of the time for the entire 1–2 year crediting period, with a discount if they also reported using the baseline stove in the last week or month. In 2017, GS updated their methodologies to provide projects different monitoring options, varying in rigour and capping the survey-derived adoption rate according to the rigour of the option; however, none of the surveys is designed to avoid social desirability bias, which has been well documented in survey methods across disciplines[22] as well as specifically and systematically in cookstove projects[23]. Social desirability bias occurs when participants provide responses (for example, inflating adoption/usage up to two times[24]), which they believe the surveyors (hired by the cookstove project developer) want to hear. Survey-based methods are further complicated as households may suffer from recall bias in remembering stove use over the past year[23].

KPTs, if done well, are reasonably robust, yet still have weaknesses. As a form of social desirability bias, called the Hawthorne effect, households may change their behaviour in the presence of project staff who can observe their stove choices while weighing the fuel[25]. Due to cost, KPTs are only required biennially on a sample of households; however, stove usage, stacking, and fuel quality and availability can be seasonal and highly variable[26]. Thus, KPTs might not accurately represent stove use across the participant pool over the 2 year crediting period.

Our sampled projects use surveys, and report adoption and usage rates much higher than rates documented in the literature (86% adoption rate and 98% usage rate compared with 58% and 52% from our literature reviews[27–37]), and stacking rates that are much lower (2% stacking rate compared with 68% in the literature[26–32,38,39]). These empirical studies, performed on cookstove projects very similar to those participating in the studied offset programmes (Supplementary Information), are designed to avoid bias with frequent, comprehensive, longitudinal surveys, triangulated with photos, field tests and/or stove monitors, and conducted by trained enumerators, unaffiliated with the project.

Since the offset project surveys have known biases, to estimate project carbon emission reductions, we replace all survey-derived adoption and usage rates with literature values as the best data available (Supplementary Information). We use empirical ranges in the MCM using a triangle distribution: adoption 58% (40%, 92%)[27–35], usage 52% (16%, 85%)[36] and stacking 68% (19.3%, 100%)[26–32,38,39]. We discount KPT-derived (that is, GS-TPDDTEC) usage and stacking rates with the MCM using a uniform distribution with the maximum based on an empirical study estimating the Hawthorne effect (−53% in usage and 29% in stacking)[25]. We do not correct GS-metered.

Fuel consumption

Methodologies use three approaches to estimate the difference between baseline and project fuel consumption. AMS-II-G and GS-simplified start by estimating baseline fuel use, and then use differences in the baseline and project stove efficiencies to estimate fuel use savings on the basis of surveyed adoption and usage rates. GS-TPDDTEC determines baseline and project fuel consumption separately and calculates emission reductions as the difference between the two. GS-metered/AMS-I-E start with measured/surveyed project fuel use and back-calculate baseline fuel consumption, assuming the equivalent energy would have been used in the baseline by the less-efficient baseline stove.

Methodologies give projects several options to determine most inputs. AMS-II-G, GS-simplified and GS-TPDDTEC allow projects to determine total baseline fuel use using a default value (0.4–0.5 tons of firewood per capita per year[40]), literature, national/project survey data or a KPT (rarely chosen)[16] (Table 1). AMS-II-G and GS-simplified use default values for the baseline stove efficiency and determine the project stove efficiency with a laboratory test. GS-metered and AMS-I-E determine baseline fuel consumption with default values, literature or surveys. GS-TPDDTEC and GS-metered require KPTs and metered or sales data, respectively.

CDM's previous default baseline stove efficiencies are lower than those found in the literature[41], while laboratory-derived project stove efficiencies are higher than actual performance in the field[42]. For projects that use default efficiencies, we update them to the CDM Methodology Panel's 2022 recommendations, which reflect current literature[40] (for example, from 10% and 20% to 15% and 25%, respectively, for firewood and charcoal).

Baselines constructed with project-led and national[43,44] fuel consumption surveys are vulnerable to social desirability[22,23] and recall[23,24] biases as households may want to present affluence and struggle to estimate kilograms of fuel used[23]. These biases can result in abnormally high baseline and/or low consumption values, especially when used together. Without a way to ground truth fuel consumption, we simply confine fuel consumption values to a reasonable literature-derived range of 2–4 MJ per capita per day[45,46] energy delivered to the pot (Supplementary Information).

fNRB

Projects that reduce biomass use should only be credited for the proportion of CO_2 emissions reduced from non-renewable sources. Previously, all methodologies relied on inaccurate CDM fNRB default values. As these defaults have now expired, projects may calculate fNRB values from a CDM tool[47] or assume a 30% default (rarely chosen). Both the earlier defaults and the tool overstate forest degradation compared

Table 2 | Outlining the factors and adjustments to each methodology based on published literature and our own analysis (Methods and Supplementary Information) and then the amount of over- or under-crediting from each individual factor across the issued credits from our sample of projects

Total amount of over-crediting across issued credits of studied projects from the average in our Monte Carlo Method (95% confidence interval)								
All factors	**Adoption rates[a]**	**Usage rates[a]**	**Stacking rates[a,b]**	**Fuel consumption**	**fNRB**	**EFs**	**Firewood–charcoal conversion**	**Rebound**
Definition	Percentage of distributed stoves actually in use	Percentage of meals cooked using the project stove	Percentage of meals cooked using the baseline stove in concert with the project stove	Amount of cooking fuel used by project households before and after obtaining the project stove	fNRB	The carbon dioxide equivalent emissions of fuel used, including upstream and non-carbon dioxide gases	Amount of firewood (on a wet basis) needed to produce the equivalent weight of charcoal (on a dry basis)	Increase in a household's overall cooking energy consumption with access to an improved stove
9.2 (7.0, 11.5)	1.4 (1.0, 1.7)	1.4 (1.1, 1.8)	1.1 (0.8, 1.4)	1.4 (1.1, 1.7)	1.7 (1.3, 2.1)	0.6 (0.5, 0.8)	1.5 (1.1, 2.0)	1.0 (0.8, 1.3)
Adjusted with	MCM using a triangle distribution: 58% (40%, 92%)	MCM using a triangle distribution: 52% (16%, 85%)	MCM using a triangle distribution: 68% (19.3%, 100%)	CDM's updated default baseline stove efficiencies if used and contained values within 2–4 MJ per capita per day delivered energy.	MCM using a triangle distribution from 'Scenario B–low yield' of Bailis et al. [2]	EFs for each cooking fuel from Floess et al. [49]	Charcoal upstream and point-of-use emissions factors from Floess et al. [49]	Literature-derived rebound effect: 22%
GS-TPDDTEC	✓	Discounted with an MCM using a uniform distribution with a maximum of a 53% decrease in usage	Discounted with MCM using a uniform distribution with a maxiimum of a 29% increase in stacking	✓	✓	✓	✓	
CDM-AMS-I-E (specific to ethanol projects)	✓	✓		✓	✓	✓	✓	✓
GS-simplified	✓	✓	✓	✓	✓	✓	✓	✓
CDM-AMS-II-G	✓	✓	✓	✓	✓	✓	✓	✓
GS-metered				✓	✓	✓	✓	✓

A check mark means that the approach outlined in the 'Adjusted with' row was applied; a blank cell means no adjustment was made and the text describes our approach. [a]One GS-TPDDTEC requires the removal of the baseline stove, and one AMS-II-G builds the improved stove in the exact spot of the baseline stove. We use slightly different Monte Carlo method distributions for these projects (see Supplementary Information). [b]Projects typically report a percentage of baseline stove use, which is then incorporated into the fuel consumption calculation. Using the project's documentation, we separate these two parameters.

with published literature[2]. WISDOM model of Bailis et al. [2] estimates fNRB, accounting for biomass regrowth and geographical, ecological and land use heterogeneity at the subnational level[2]. The most robust fNRB approach so far is a dynamic landscape model, Modelling Fuelwood Sustainability Scenarios[48]. When our study was conducted, few national values were available. Using the MCM, we replace project fNRB values with the 'Scenario B–low yield' 'minimum value' of Bailis et al. as the low boundary, 'expected value' as the mode and 10% over the expected value as the high boundary. On average, the projects chosen fNRBs are 3.0 (minimum 1.1, maximum 16.4) times the values of Bailis et al. [2] (Table 2 and Supplementary Information).

EFs

To translate fuel use into GHG emissions, GS uses 2006 Intergovernmental Panel on Climate Change (IPCC) default EFs and allows, but does not require, the inclusion of upstream emissions. Counterintuitively, to work around an early agreement prohibiting the crediting of reduced deforestation, CDM cookstoves methodologies apply a baseline EF assuming future fossil fuel use rather than biomass. This is a source of under-crediting[9]. We replace each approach with cooking fuel-specific EFs, including upstream emissions, from Floess et al. [49], the most comprehensive, up-to-date cooking fuel EF database. We also update all global warming potentials to the most recent IPCC values, accounting for distinctions for renewable/non-renewable biomass[50]. Due to high uncertainty around the climate impacts of black carbon emissions from cookstove projects[51], we, like the current methodologies, exclude black carbon.

Firewood–charcoal conversion

All methodologies allow projects replacing charcoal to use a firewood–charcoal conversion factor to estimate the amount of firewood (on a wet basis) needed to produce the equivalent weight of charcoal (on a dry basis). All used a default of six, which a CDM methodology panel updated to four in 2022 after our sample selection, based on literature[40]. Alternatively, methodologies allow projects to use literature to establish this conversion factor. All projects using

https://doi.org/10.1038/s41893-023-01259-6

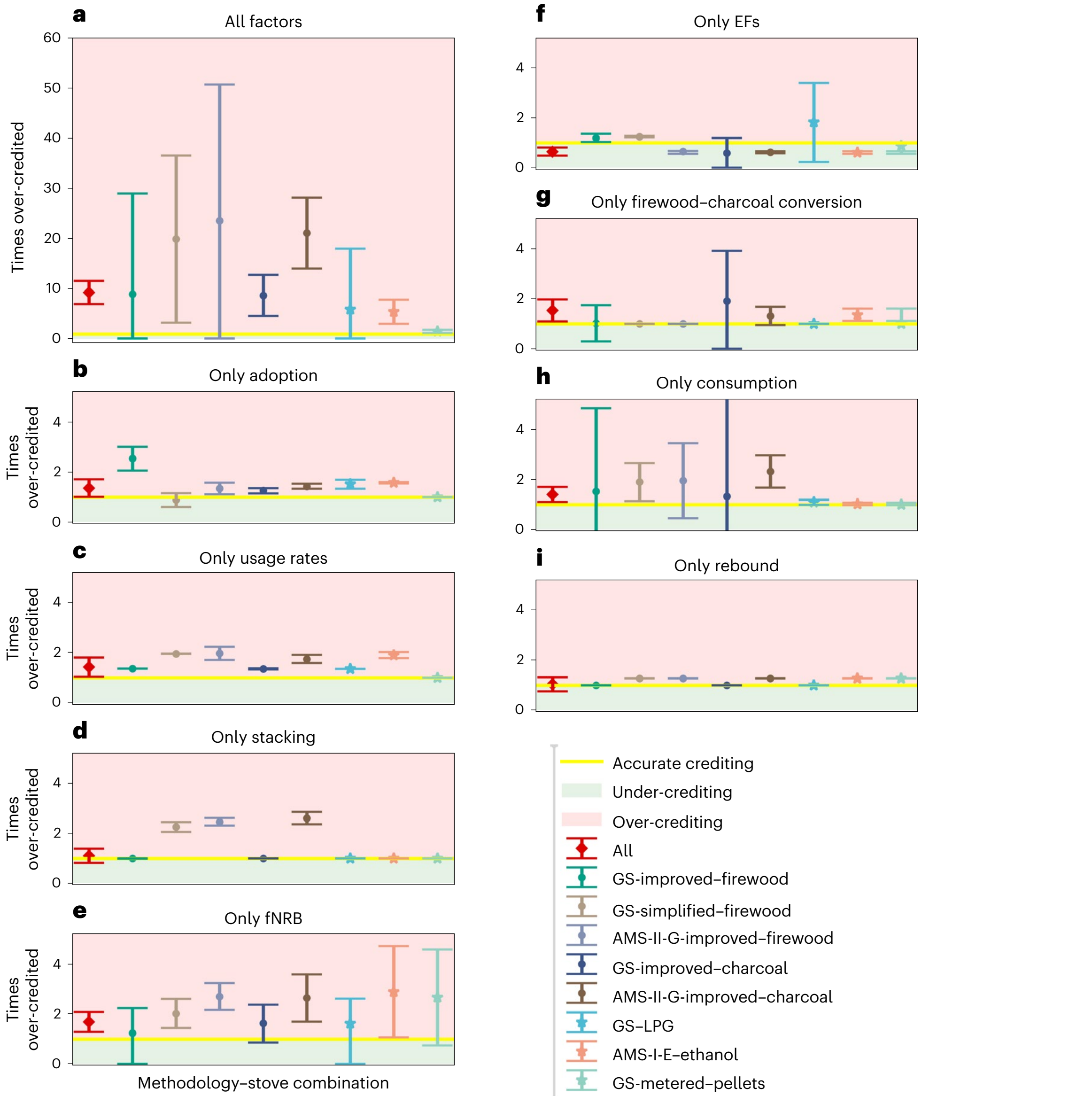

Fig. 3 | Over/under-crediting across factors. a–i, The mean amount of total over/under-crediting after quantifying all factors (*n* = 51 projects) (**a**) and individual factors by methodology–stove combinations for adoption (**b**), usage rates (**c**), stacking (**d**), fNRB (**e**), EFs (**f**), firewood–charcoal conversion (**g**), consumption (**h**) and rebound (**i**) methodologies only. GS–firewood (*n* = 9 projects), GS-simplified–firewood (*n* = 9 projects), GS–charcoal (*n* = 7 projects), GS–LPG (*n* = 4 projects), CDM-AMS-II-G–firewood (*n* = 13 projects), CDM-AMS-II-G–charcoal (*n* = 3 projects), CDM-AMS-II-E–ethanol (*n* = 4 projects) and GS-metered–pellet (*n* = 3 projects). The points indicate the total over- or under-crediting, while the error bars refer to the 95% CI for the total over-crediting across our sample of projects and the categories we delineate. We limit the CI's lower bounds to 0 (Methods and Supplementary Information). EFs include point-of-use emissions including non-CO_2 emissions and upstream emissions. Less than 1 (green shading) indicates under-crediting. Red shading indicates over-crediting, and yellow indicates accurate crediting.

this conversion used a value of 4.8 or higher. However, conversion efficiency is highly dependent on the specific location and charcoal production practices[40]. We do not use a firewood–charcoal conversion factor but instead use charcoal upstream and point-of-use EFs from Floess et al. [49].

Rebound effect

Households commonly increase their overall cooking energy consumption with access to an improved stove (for example, ref. 52). The improved stove lowers the 'cost' of cooking and provides another burner, allowing the household to increase their fuel consumption. Only projects that utilize KPTs capture this increase, which we confirm within our sample. We reduced our emission reduction estimation by 22% for projects that do not utilize KPTs, drawing on published literature that models or tracks the time stoves were used before and after the acquisition of an improved/clean stove through temperature sensors (Supplementary Information)[29,52–55].

Over/under-crediting analysis results

To find the total amount of over-crediting across our sampled portfolio, we estimate each project's over-crediting across analysed monitoring reports, then apply that to their total issued credits and compare our total ER estimates with their total issued credits. We estimate that our sample of cookstove projects are 9.2 times over-credited ((95% confidence interval (CI) 7.0, 11.5); Table 2 and Fig. 3). That is, the sample generated 26.7 million offset credits (as of May 2023), which is

https://doi.org/10.1038/s41893-023-01259-6

Table 3 | Recommendation for cookstove methodology reforms

To avoid over-crediting, new and current cookstoves methodologies should require, and until then, project developers should choose:	
Factor	**Recommendation**
fNRB	The 'Scenario B–low yield' value of Bailis et al. [2] at the lowest subnational level. Update to the Modelling Fuelwood Sustainability Scenarios value at the lowest subnational level as new research emerges.
Adoption, usage, stacking and rebound	One of the following options: 1. Meters or collect fuel purchase data for adoption, usage and stacking; a longitudinal survey or a conservative, literature-derived default for rebound; if a project has metered or fuel purchase data, this option is required 2. KPTs for usage and stacking, adjusted for the Hawthorne effect with a literature-derived default; robust longitudinal survey or conservative literature-derived default for adoption 3. Robust longitudinal surveys 4. Conservative literature-derived default values
Fuel consumption	Initial and update baseline KPTs and/or robust project-led surveys; enforce a reasonable range of 2–4 MJ-delivered per capita per day.
EFs	Upstream, point-of-use and non-CO_2 EFs for each cooking fuel from Floess et al., removing the need for a firewood–charcoal conversion factor IPCC's separate renewable/non-renewable global warming potentials for methane and nitrous oxide emissions, but continue to exclude black carbon pending future research

For full details on how to implement these recommendations, see our accompanying website[63].

over nine times our estimated carbon emission reductions of roughly 2.9 million tCO_2e.

Using the same approach, we extrapolate our estimates of over-crediting to the entire credit pool by methodology–stove combination. We find a total impact of roughly 5.2 million tCO_2e compared with the total 55.3 million VCM-issued credits.

We find that the average project in our sample is over-credited 27.6× (see Supplementary Information Section 6).

Respectively, fNRB, firewood–charcoal conversion, fuel consumption, adoption and usage produce the most over-crediting: 1.7, 1.5, 1.4, 1.4 and 1.4 times (Table 2). On average, only correcting the EFs resulted in under-crediting (0.6 times), while stove stacking and rebound minimally affects crediting amounts (1.1 and 1.0 times, respectively).

We find that all methodology–stove combinations over-credit (Fig. 3). AMS-II-G–firewood is the most over-credited project type from our sample (23.5 (0, 49.3)), stemming from specific project values (fNRB ~2.7× and consumption ~2.4×) and the methodology's approach (stacking ~2.5×, usage ~2.0×, adoption ~1.4× and rebound ~1.3×) that together have a multiplier effect. AMS-II-G–charcoal is the second most over-credited project type (21.0 (12.7, 29.4)) from the same sources, except their usage rates were closer to literature-derived values, while they had an additional source of over-crediting from the firewood–charcoal conversion (~1.3×). The CDM methodologies' weak monitoring approach overcomes the under-crediting from their use of the EF from a projected fossil fuel (~0.6–0.7×). GS-simplified–firewood (19.8 (2.5, 37.2)) is more over-credited than GS–firewood (8.9 (0, 26.9)) and GS–charcoal (8.6 (4.5, 12.8)), under GS-TPDDTEC, due to their less robust monitoring approach (that is, GS-simplified does not require KPTs). Compared with GS–firewood, GS–charcoal projects over-credited less from adoption, but over-credited from the firewood–charcoal conversion (~1.9×). GS–liquefied petroleum gas (LPG) over-credits by 5.9 (0, 16.3) times, from fNRB, adoption, usage and EFs. AMS-I-E–ethanol over-credits 5.4 (3.2, 7.6) times from adoption (1.6×), usage (1.9×), fNRB (2.9×) and rebound (1.3×), but under-credits from CDM's use of fossil fuel EFs (0.6×). GS-metered–pellets have the least over-crediting (1.5 (0.6, 2.4)), stemming only from fNRB and rebound, with slight under-crediting from EFs.

Over-crediting from fNRB stems from location-specific differences in the values of Bailis et al. [2] (Supplementary Fig. 4). Adoption, usage and stacking rates affect methodology–stove combinations based on the methodology's requirements (for example, meters, KPTs and surveys). GS–LPG, AMS-I-E–ethanol and GS-metered–pellets, on average, did not report fuel consumption values outside of a reasonable range, probably due to the use of KPTs, meters or sales data.

EF choices result in overall under-crediting (0.6×) from five methodology–stove combinations: CDM methodologies use the low EF of a projected fossil fuel as the baseline, GS–charcoal projects do not always include upstream emissions and GS-metered–pellets projects construct a weighted average baseline EF, which ultimately is lower than those in Floess et al. [49]. The EFs used by GS–firewood and GS–LPG for the baseline fuels are slightly higher than Floess et al. [49], leading to slight over-crediting, stemming from project-chosen values, not the switch to LPG.

Per stove-day, GS-metered–pellets and AMS-I-E projects reduce emissions by roughly 0.007 and 0.006 tCO_2e due to their renewable feedstocks, and thus minimal project emissions. They are followed, on a per stove-day basis, by GS–charcoal (0.003 tCO_2e), AMS-II-G–firewood, GS–firewood and LPG (0.001 tCO_2e), AMS-II-G–charcoal (0.0004 tCO_2e) and GS-simplified–firewood (0.0002 tCO_2e).

Discussion

We conservatively estimate that the total amount of over-crediting across our sample's issued credits is 9.2 (7.0, 11.5), stemming from misalignment across numerous, compounded factors.

The majority of over-crediting stems from lack of rigour and flexibility in how methodologies determine fNRB, adoption, usage, stacking and fuel consumption, despite periodic methodological updates. We provide recommendations for aligning methodologies with current science (Table 3). Regular updates will be needed to reflect future research advancements. Currently, project developers, who benefit financially from more credits, hire verifiers directly, possibly conflicting with the International Organization for Standardization (17029) that requires the verifier to be impartial (C5.3)[56]. The developers' incentives are evident, as robust fNRB values have been published for 8 years, yet all projects have opted to use higher CDM tool-derived or default values, and some projects track purchase data, yet fail to use it in reduction estimation. Eliminating the flexibility and requiring robust or conservative methods could reduce over-crediting easily, universally by 1.4–1.7 times for each factor.

Developers can apply these recommendations without incurring extra cost. For adoption, usage and stacking, while meters, longitudinal surveys and KPTs are the most accurate, they also can be costly depending on project infrastructure and size. For these factors, we include in our recommendations the option of literature-derived values that have no cost, and despite being less accurate, are likely to avoid over-crediting.

Additionally, increases in offset prices could make these needed reforms more affordable. There is a feedback loop—poor quality keeps offset prices too low to support accurately credited projects. Higher prices for accurately estimated reduction could incentivize and fund projects to promote behaviour change, increase awareness and address other market and behavioural barriers to cooking energy transformation[57].

In the current landscape, buyers are left confused about what constitutes quality, and often turn to rating companies. Similarly, for project co-benefits, some buyers are willing to pay more for projects with more co-benefits, but have been reported to care more about the number of SGDs than the quality of that contribution[58]. Project's claimed co-benefits are measured, unfortunately, alongside the adoption, usage and stacking rates, through single cross-sectional surveys, which are subject to the same biases our analysis outlines[58]. Low-quality tracking of both the carbon abatement and co-benefits leads to surface level, performative action, rather than meaningful, sustainable impact.

Our results are a call to action to overhaul offset programme design and the dominance of improved but not WHO-defined clean stoves. Prioritizing metered fuel switch projects and accurately quantifying their emission reductions would progress climate, energy and health SDGs. Our analysis indicates that these stoves currently offer the least over-credited credits and have the greatest abatement potential and health benefit. Further, they are often the most challenging for users to sustainably use, given the need for continuous fuel purchases, and thus are the cookstove project types that could most benefit from carbon finance. Our results further support Gill-Wiehl and Kammen's call for the VCM to exclusively fund WHO-defined clean stoves[59], and highlight the lost opportunity to use cookstove offsets to accelerate access to the cleanest stoves/fuels. Quality cookstove offsets could sustainably, instead of performatively, improve the health of people and the planet.

Methods

Due to the nature of this analysis, the results of our study of carbon accounting methods for cookstove projects are also the methods we used in our over/under-crediting analysis and inform our recommendations. Thus, our methods are summarized in the main text. Here, in the methods section, we include further explanation of how we adjusted factors, performed the MCM and estimated over/under-crediting, and discuss the limitations of our work. Further explanation and justification of our methods for each factor is provided in Supplementary Information.

Sample selection

We evaluate the quality of offset credits from the methodologies with the largest number of cookstove offset project activities on the VCM: GS-TPDDTEC, GS-simplified, CDM-AMS-II-G and CDM-AM-I-E. We also review the new GS-metered, released October 2021.

The methodologies deploy different project stoves. GS-TPDDTEC (previously GS's Methodology for Improved Cook-Stoves and Kitchen Regimes) is the most versatile methodology covering any thermal domestic technology switch that is less GHG intensive, including but not limited to improved biomass, heat retention, solar, LPG and electric stoves. CDM-AMS-I-E replaces non-renewable biomass with renewable energy (for example, renewable biomass, biogas, bioethanol and electric stoves). Designed for smaller projects, GS-simplified and CDM-AMS-II-G have limited scopes, only allowing for biomass efficiency projects (for example, traditional fuelwood stove to an improved fuelwood stove). GS-metered is designed for cookstoves with metered or other direct fuel monitoring (for example, purchase records) such as electric, LPG, biogas, bioethanol or advanced biomass pellet stoves.

Most cookstove projects are structured as programme of activities, in which multiple similar project activities (called voluntary project activities (VPAs) on the VCM and component project activities on the CDM) are bundled together to allow for rapid replication, only requiring a quick check from a validator and not a full registration procedure[60]. To reflect the diversity of projects on the VCM, we evaluated VPAs separately. CDM methodologies are used on both the CDM and the VCM, but we limited our scope to only VCM-registered projects (that is, those certified by GS or Verra).

In March 2021, we identified the 15 countries with the most credits from cookstove projects on the market and, for each country, selected the projects with the most credits for each methodology. For the GS-TPDDTEC, GS-simplified and CDM-AMS-II-G projects, we chose projects that posted at least one monitoring report and provided their exact calculations and the stove-days. There were very few projects under AMS-I-E and GS-metered and the only one that had been issued credits was also credited under AMS-I-I and so was not included in our sample. For these two methodologies, we selected all registered projects that provided enough information to recreate offset credit calculations on a stove-day basis for individual stove types. We included these methodologies because they offered different methods for monitoring stove usage and fuel consumption, and because of the greater potential emission reductions and health benefits from fuel switch projects that these protocols accommodate. We added additional projects as needed to ensure that our sample covered all types of fuel transitions, with the exception of electric stoves. There were no issued projects actively deploying an electric stove, and the only listed electric project under GS-metered had no files available. We do not include GS-metered's most recent methodology update, which allows for the participation of more complex cooking devices such as pressure cookers, in a new option called 'specific consumption' (Supplementary Information).

Additionally, we randomly selected ten small/medium-sized projects from GS-TPDDTEC (four), AMS-II-G (four) and GS-simplified (two) to ensure that our sample was representative of both large and small projects. We investigate the relationship between the amount of over/under-crediting and project size, and find a slight negative relationship between amount of over-crediting and total verified credits (evaluated on the log scale; Supplementary Fig. 1). This trend is not statistically significant and the R^2 is very low, but it indicates that our approach of focusing on large projects may have led to lower estimates of over-crediting.

This approach resulted in a sample of 51 projects, spanning 25 countries and 8 methodology–project type combinations: (1) GS–firewood, (2) GS-simplified–firewood, (3) GS–charcoal, (4) GS–LPG, (5) CDM-AMS-II-G–firewood, (6) CDM-AMS-II-G–charcoal, (7) CDM-AMS-II-E–ethanol and (8) GS-metered–pellet (WHO tier 4+ biomass pellet stove). Our sample covers 40% of all issued credits on the VCM from these methodologies (as of 10 May 2023). We have no reason to believe that these projects are not representative of the entire pool of cookstove credits on the VCM. The 31 GS projects in our sample represent 46% of the covered GS methodologies credits on the VCM. The 16 AMS-II-G contain 25% of that methodologies' credits.

Our sample of 51 projects tangentially represents 478 projects and 64% of total credits issued under the five studied methodologies as many projects are structured as largely identical VPAs under programme of activities.

Uncertainty

Quantification of emission reductions from offset programmes is inherently uncertain. Emission reductions must be estimated against an immeasurable counterfactual scenario. Other factors, notably fNRB, upstream emissions and leakage are also difficult to estimate, and with limited research so far, involve substantial uncertainty. Since offset credits often are used to 'offset' or trade with direct emission reductions, to maintain the integrity of an emission reduction claim, offset programmes are tasked with estimating programme impacts conservatively when there is uncertainty. Here, conservative means more likely to under-credit than to over-credit. Our analysis uses the most rigorous and up-to-date values from the literature when available (for example, fNRB). Instead of choosing conservative methods for all factors, we do not or minimally correct factors with little published research, notably additionality, leakage, non-permanence and overlapping claims, and instead recommend more research. In this way, we make methodological choices that probably underestimate the amount of over-crediting.

Methodology updates

All methodologies, except for recently released GS-metered, have undergone considerable updates over the years of credit generation that affect the methodological factors we study. Our recommendations and discussion below focus on the most recent version of each methodology and any updates proposed by the registry. However, most credits on the VCM, including those still available for purchase, are issued under previous methodology versions. Therefore, our quantitative over/under-crediting analysis assesses the credits generated regardless of the methodology version used. We note in the main text and detail in Supplementary Information where updated methodologies address over-crediting.

Adjusting factors

Using the values listed in the latest verified monitoring report or project documents of these 51 projects, we calculated the number of VERs on a per stove-day basis. We only included projects (or monitoring reports from projects) in our sample if we were exactly able to replicate the number of VERs either in total or on a per stove-time basis. Once we replicated the credits generated under the methodologies, we then adjusted all the identified factors contributing to over/under-crediting as described above. Then, we conducted analyses isolating each factor.

To make the factor analysis of EFs, firewood–charcoal conversion factor and consumption for GS-metered–pellet and AMS-I-E–ethanol comparable to all other methodology–stove combinations, we remove GS-metered and AMS-I-E's calculation approach and calculate the baseline emissions and project emissions separately. For example, we use the baseline and project consumption reported in their project documents to calculate the difference between baseline and project emissions instead of using their baseline conversion factor approach (see the 'Fuel consumption' section).

Finally, we conduct one analysis excluding adoption, usage and stacking rates, which are the only factors that are always monitored ex post. We do this for fair comparison with GS-metered–pellet and AMS-I-E–ethanol projects, which, as of the time of sampling, had generated no credits. In our main analysis, we use their ex ante values for adoption and stacking rates from the project documents rather than ex post values from monitoring reports as with all other projects.

In total, we have analyses in which (1) all factors are adjusted, (2) only adoption rates are adjusted, (3) only usage rates are adjusted, (4) only stacking rates are adjusted, (5) only fNRB values are adjusted, (6) only EFs (including upstream emissions) are adjusted, (7) only the firewood–charcoal conversion is adjusted, (8) only consumption (baseline and project) values are adjusted, (9) only rebound consumption is adjusted and (10) all factors are adjusted, except adoption, usage and stacking (Supplementary Information).

MCM

The MCM is a statistical framework that calculates possible outcomes when input parameters are randomly varied within a specified range using a given distribution[61]. When used for fNRB, adoption, usage and stacking rates, the MCM generates values within our defined limits, following the distribution defined in each factor's section, assuming independence (see 'Limitations' section). We specified the simulation to run 10,000 times, randomly generating new values for each of these factors and calculating an associated emission reduction. We acknowledge the inherent uncertainty within our factors and bound each one within a literature-derived range. We take this approach over other methods of error propagation given the inherent uncertainty and imprecision in the ranges within the literature. Johnson et al., for example, propagated error as they had direct field measurements for their study site for fNRB, EFs and fuel consumption. Without this level of precision for each carbon offset location, we take a higher level, although less precise approach. However, as discussed, we make methodological decisions that result in likely underestimation of the amount of over-crediting.

Estimating over/under-crediting

We estimate the over-crediting across our sampled portfolio in three ways. To estimate the total over-crediting of our sample, we estimate each project's over-crediting across analysed monitoring reports, apply that value to each project's total issued credits and compare our total ER estimates with their total issued credits (Fig. 3 and Supplementary Table 5). For the projects in our sample that have not generated credits (see the section on sampling), we use their estimated annual emission reductions from their project design documents. We then splice the results by methodology–project type combination (Fig. 3) and then by country (Supplementary Fig. 3).

Second, we average over-crediting by project across our sample (Supplementary Table 6). Finally, we take an average of our data points at the highest level of granularity, that is, at the level of the monitoring report or stove type within a monitoring report (Supplementary Table 7).

We construct CIs around the total amount of over-crediting by finding the standard deviation across the total over-crediting by project based off the average MCM for all and for the specific factor analysis. These CIs become larger within the subanalyses due to smaller sample sizes. Negative lower bounds of the CI are a function of large standard deviations due to specific project values and smaller sample sizes. Note that, within this over-crediting reporting framework, under-crediting is indicated by a value between 0 and 1, not negative. We thus limit the lower limit of CIs to zero.

To extrapolate to the entire cookstove market, we take the total rate of over-crediting for each methodology–stove combination found above, and then apply these rates to the total amount of credits issued for each methodology–stove combination. Thus, we find that the whole market is over-credited 10.6 times weighting by methodology–stove combination.

Commercial credits

A few of our sample projects included some stoves used for commercial purposes (restaurants, schools and so on), representing a small fraction of these projects' total credits. We do not adjust commercial stoves' adoption, usage or stacking rates, or baseline/project fuel consumption. There are still barriers to adoption, usage and ending stacking for commercial institutions; however, the literature on these rates is limited[62], and thus an area for future research.

Limitations

Our study has some limitations that must moderate our conclusions. This analysis does not cover 100% of projects under the five studied methodologies. We cover 40% of the market, and projects in 25 countries; however, we attempted to have a fully representative sample across methodologies, location and project type. We were limited to projects that were transparent enough to provide their exact calculations or stove-days within their monitoring or validation reports. All factors involve some amount of uncertainty, which we address with the MCM for some factors. We were limited by the details provided by the projects and the standards. For example, numerous projects did not specify the rural or urban setting or more specific administrative units, which is important for fNRB.

Finally, a key limitation in our work is that we assume that all factors are independent. This is an appropriate assumption for all factors, potentially except for adoption, usage and stacking. For example, there is no evidence in the literature that fNRB or EF is correlated with stove adoption; however, there could be correlation with stove adoption and usage. This correlation, however, would be highly context dependent and probably time variant (that is, a household's relationship with and use of an intervention has been shown to change over time). In creating the distributions for adoption, usage and stacking, we create ranges of uncertainty, since rates of adoption, usage and stacking have been reasonably well studied and there is an established literature that we draw from. Unfortunately, the correlation between these rates has not been well established and would require less grounded assumptions. This is also a reason that we pursue triangle distributions as we hesitate to make definitive claims on the underlying distributions, opting rather to simply present that the literature has established general ranges for these values as described above. Given this context, we therefore assume independence of all factors. This is a limitation of our work, but one that probably leads to more conservative findings. This is because incorporating the covariance between adoption, usage and stacking would further limit the input distribution of these factors and thus shrink our reported CIs. Thus, our reported ranges provide more coverage. We further feel comfortable with this methodological decision given the other areas that probably result in underestimation, as above.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data and code are publicly available online at https://github.com/agillwiehl/GillWiehl_et_al_Pervasive_over_crediting.

Code availability

All data and code are publicly available online at https://github.com/agillwiehl/GillWiehl_et_al_Pervasive_over_crediting.

References

1. Abbafati, C. et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1223–1249 (2020).
2. Bailis, R., Drigo, R., Ghilardi, A. & Masera, O. The carbon footprint of traditional woodfuels. *Nat. Clim. Change* **5**, 266–272 (2015).
3. Defining clean fuels and technologies. *World Health Organization* https://www.who.int/tools/clean-household-energy-solutions-toolkit/module-7-defining-clean (2021).
4. So I., Haya, B. & Elias, M. Voluntary registry offsets database v.8. *University of California, Berkeley* https://gspp.berkeley.edu/faculty-and-impact/centers/cepp/projects/berkeley-carbon-trading-project/offsets-database (2023).
5. Stapp, J. et al. Little evidence of management change in California's forest offset program. *Commun. Earth Environ.* **4**, 331 (2023).
6. Haya, B. *Policy Brief: the California Air Resources Board's US Forest Offset Protocol Underestimates Leakage* (Goldman School of Public Policy, University of California, Berkeley, 2019); https://gspp.berkeley.edu/assets/uploads/research/pdf/Policy_Brief-US_Forest_Projects-Leakage-Haya_4.pdf
7. Haya, B. et al. Quality assessment of REDD+ carbon credit projects. Berkeley carbon trading project. *Goldman School of Public Policy, University of California, Berkeley* https://gspp.berkeley.edu/research-and-impact/centers/cepp/projects/berkeley-carbon-trading-project/REDD (2023).
8. West, T. A. P. et al. Action needed to make carbon offsets from forest conservation work for climate change mitigation. *Science* **381**, 873–877 (2023).
9. Cames, M. et al. How additional is the clean development mechanism? Analysis of the application of current tools and proposed alternatives. *DG CLIMA* https://doi.org/10.13140/RG.2.2.23258.54728 (2016).
10. Haya, B. K. *Carbon Offsetting: An Efficient Way to Reduce Emissions or to Avoid Reducing Emissions? An Investigation and Analysis of Offsetting Design and Practice in India and China*. PhD thesis, University of California, Berkeley (2010).
11. Bailis, R., Wang, Y., Drigo, R., Ghilardi, A. & Masera, O. Getting the numbers right: revisiting woodfuel sustainability in the developing world. *Environ. Res. Lett.* **12**, 115002 (2017).
12. Ramanathan, T. et al. Wireless sensors linked to climate financing for globally affordable clean cooking. *Nat. Clim. Change* **7**, 44–47 (2017).
13. Freeman, O. E. & Zerriffi, H. How you count carbon matters: implications of differing cookstove carbon credit methodologies for climate and development cobenefits. *Environ. Sci. Technol.* **48**, 14112–14120 (2014).
14. Sanford, L. & Burney, J. Cookstoves illustrate the need for a comprehensive carbon market. *Environ. Res. Lett.* **10**, 084026 (2015).
15. Simon, G. L., Bumpus, A. G. & Mann, P. Win–win scenarios at the climate–development interface: challenges and opportunities for stove replacement programs through carbon finance. *Glob. Environ. Change* **22**, 275–287 (2012).
16. Lee, C. M., Chandler, C., Lazarus, M. & Johnson, F. X. Assessing the climate impacts of cookstove projects: issues in emissions accounting. *Chall. Sustain.* **1**, 53–71 (2013).
17. Reduced emissions from cooking and heating—technologies and practices to displace decentralized thermal energy consumption (TPDDTEC). *The Gold Standard Foundation* https://globalgoals.goldstandard.org/407-ee-ics-technologies-and-practices-to-displace-decentrilized-thermal-energy-tpddtec-consumption/ (2021).
18. The Gold Standard simplified methodology for clean and efficient cookstoves. *The Gold Standard Foundation* https://globalgoals.goldstandard.org/408-ee-ics-simplified-methodology-for-efficient-cookstoves/ (2022).
19. AMS-II.G.: energy efficiency measures in thermal applications of non-renewable biomass version 12.0. *Clean Development Mechanism* https://cdm.unfccc.int/methodologies/DB/10PELMPDW951SVSW1B2NRCQEBAX96C (2022).
20. AMS-I.E.: switch from non-renewable biomass for thermal applications by the user version 12.0. *Clean Development Mechanism* https://cdm.unfccc.int/methodologies/DB/BLVN9ULDP1FRUVS2LYWW6WPYN9W78E (2021).
21. Methodology for metered and measured energy cooking devices. *The Gold Standard Foundation* https://globalgoals.goldstandard.org/news-methodology-for-metered-measured-energy-cooking-devices/ (2022).
22. Krumpal, I. Determinants of social desirability bias in sensitive surveys: a literature review. *Qual. Quant.* **47**, 2025–2047 (2013).
23. Kar, A., Brauer, M., Bailis, R. & Zerriffi, H. The risk of survey bias in self-reports vs. actual consumption of clean cooking fuels. *World Dev. Perspect.* **18**, 100199 (2020).
24. Wilson, D. L. et al. in *Technologies for Development* (eds Hostettler S. et al.) 211–221 (Springer, 2015).
25. Simons, A. M., Beltramo, T., Blalock, G. & Levine, D. I. Using unobtrusive sensors to measure and minimize Hawthorne effects: evidence from cookstoves. *J. Environ. Econ. Manage.* **86**, 68–80 (2017).
26. Shankar, A. V. et al. Everybody stacks: lessons from household energy case studies to inform design principles for clean energy transitions. *Energy Policy* **141**, 111468 (2020).
27. Hanna, R., Duflo, E. & Greenstone, M. Up in smoke: the influence of household behavior on the long-run impact of improved cooking stoves. *Am. Econ. J. Econ. Policy* **8**, 80–114 (2016).
28. Burwen, J. & Levine, D. I. A rapid assessment randomized-controlled trial of improved cookstoves in rural Ghana. *Energy Sustain Dev.* **16**, 328–338 (2012).
29. Beltramo, T., Blalock, G., Harrell, S., Levine, D. & Simons, A. M. The effects of fuel-efficient cookstoves on fuel use, particulate matter, and cooking practices: results from a randomized trial in rural Uganda. *UC Berkeley Center for Effective Global Action* https://escholarship.org/uc/item/365778pn (2019).
30. Rosa, G. et al. Assessing the impact of water filters and improved cook stoves on drinking water quality and household air pollution: a randomised controlled trial in Rwanda. *PLoS ONE* **9**, e91011 (2014).

31. Bensch, G. & Peters, J. The intensive margin of technology adoption—experimental evidence on improved cooking stoves in rural Senegal. *J. Health Econ.* **42**, 44–63 (2015).
32. Ruiz-Mercado, I., Masera, O., Zamora, H. & Smith, K. R. Adoption and sustained use of improved cookstoves. *Energy Policy* **39**, 7557–7566 (2011).
33. Islam, M. M. et al. Assessing the effects of stove use patterns and kitchen chimneys on indoor air quality during a multiyear cookstove randomized control trial in rural India. *Environ. Sci. Technol.* **56**, 8326–8337 (2022).
34. García-Frapolli, E. et al. Beyond fuelwood savings: valuing the economic benefits of introducing improved biomass cookstoves in the Purépecha Region of Mexico. *Ecol. Econ.* **69**, 2598–2605 (2010).
35. Agurto Adrianzén, M. Social capital and improved stoves usage decisions in the Northern Peruvian Andes. *World Dev.* **54**, 1–17 (2014).
36. Jeuland, M., Soo, J. S. T. & Shindell, D. The need for policies to reduce the costs of cleaner cooking in low income settings: implications from systematic analysis of costs and benefits. *Energy Policy* **121**, 275–285 (2018).
37. Ruiz-Mercado, I., Canuz, E., Walker, J. L. & Smith, K. R. Quantitative metrics of stove adoption using stove use monitors (SUMs). *Biomass Bioenergy* **57**, 136–148 (2013).
38. Pine, K. et al. Adoption and use of improved biomass stoves in rural Mexico. *Energy Sustain. Dev.* **15**, 176–183 (2011).
39. Pattanayak, S. K. et al. Experimental evidence on promotion of electric and improved biomass cookstoves. *Proc. Natl Acad. Sci. USA* **116**, 13282–13287 (2019).
40. MP88: meeting report/recommendations to the executive board. *CDM Methodologies Panel* https://cdm.unfccc.int/Panels/meth/index.html (2022).
41. Life cycle assessment of cooking fuel systems in India, China, Kenya, and Ghana. *US Environmental Protection Agency* https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=339679&Lab=NRMRL&simplesearch=0&showcriteria=2&sortby=pubDate&timstype=Published+Report&datebeginpublishedpresented (2021).
42. Wathore, R., Mortimer, K. & Grieshop, A. P. In-use emissions and estimated impacts of traditional, natural- and forced-draft cookstoves in rural Malawi. *Environ. Sci. Technol.* **51**, 1929–1938 (2017).
43. Stockwell, T. et al. Estimating under- and over-reporting of drinking in national surveys of alcohol consumption: identification of consistent biases across four English-speaking countries. *Addiction* **111**, 1203–1213 (2016).
44. Ezzati, M., Martin, H., Skjold, S., Hoorn, S. V. & Murray, C. J. L. Trends in national and state-level obesity in the USA after correction for self-report bias: analysis of health surveys. *J. R. Soc. Med.* **99**, 250–257 (2006).
45. *Concept Note CDM-MP85-A07. Analysis and Options Regarding Caps Used in AMS-I.E, AMS-II.G and TOOL30 Version 01.0* (Clean Development Mechanism, 2013); https://cdm.unfccc.int/sunsetcms/storage/contents/stored-file-20210708220535947/MP85_EA07_Concept%20Note%20-%20Caps.pdf
46. Daioglou, V., van Ruijven, B. J. & van Vuuren, D. P. Model projections for household energy use in developing countries. *Energy* **37**, 601–615 (2012).
47. Tool 30: calculation of the fraction of non-renewable biomass (version 3). *Clean Development Mechanism* https://cdm.unfccc.int/methodologies/PAmethodologies/tools/am-tool-30-v1.pdf/history_view (2020).
48. Ghilardi, A. et al. Spatiotemporal modeling of fuelwood environmental impacts: towards improved accounting for non-renewable biomass. *Environ. Model Softw.* **82**, 241–254 (2016).
49. Floess, E. et al. Scaling up gas and electric cooking in low- and middle-income countries: climate threat or mitigation strategy with co-benefits? *Environ. Res. Lett.* **18**, 034010 (2023).
50. Whitman, T. L. & Lehmann, C. J. Systematic under and overestimation of GHG reductions in renewable biomass systems. *Clim. Change* **104**, 415–422 (2011).
51. Huang, Y. et al. Global radiative effects of solid fuel cookstove aerosol emissions. *Atmos. Chem. Phys.* **18**, 5219–5233 (2018).
52. Bailis R., et al. Enhancing clean cooking options in peri-urban Kenya: a pilot study of advanced gasifier stove adoption. *Environ. Res. Lett.* **15**, 084017 (2020).
53. Forum on natural capital accounting for better policy decisions: taking stock and moving forward. *World Bank Group* http://documents.worldbank.org/curated/en/904211580129561872/Forum-on-Natural-Capital-Accounting-for-Better-Policy-Decisions-Taking-Stock-and-Moving-Forward (2017).
54. Dufournaud, C. M., Quinn, J. T. & Harrington, J. J. A partial equilibrium analysis of the impact of introducing more-efficient wood-burning stoves into households in the Sahelian Region. *Environ. Plan. Econ. Space* **26**, 407–414 (1994).
55. Lambe, F. et al. Opening the black pot: a service design-driven approach to understanding the use of cleaner cookstoves in peri-urban Kenya. *Energy Res. Soc. Sci.* **70**, 101754 (2020).
56. ISO/IEC 17029:2019 general principles and requirements for validation and verification bodies. *ISO* https://www.iso.org/standard/29352.html (2019).
57. Khavari, B., Ramirez, C., Jeuland, M. & Fuso Nerini, F. A geospatial approach to understanding clean cooking challenges in sub-Saharan Africa. *Nat. Sustain.* **6**, 447–457 (2023).
58. Bakhtary, H., Tierney, M., Galt, H. & Gill-Wiehl, A. *More than just a Carbon Project: How Clean Cooking Projects Certified Under the Gold Standard Approach SDG Claims* (Climate Focus, 2023); https://mecs.org.uk/wp-content/uploads/2023/05/FINAL-SDG-Briefing-More-than-just-a-carbon-project.pdf
59. Gill-Wiehl, A. & Kammen, D. M. A pro-health cookstove strategy to advance energy, social and ecological justice. *Nat. Energy* **7**, 999–1002 (2022).
60. The handbook for programme of activities: practical guidance to successful implementation. *Climate Focus* https://climatefocus.com/publications/handbook-programmes-activities-practical-guidance-successful-implementation/ (2011).
61. Fitzpatrick, D. in *Analog Design and Simulation Using OrCAD Capture and PSpice* 2nd edn (ed Fitzpatrick, D.) 151–164 (Newnes, 2018).
62. Robinson, B. L., Clifford, M. J., Hewitt, J. & Jewitt, S. Cooking for communities, children and cows: lessons learned from institutional cookstoves in Nepal. *Energy Sustain. Dev.* **66**, 1–11 (2022).
63. Gill-Wiehl, A., Hogan, M. & Haya, B. A comprehensive quality assessment of cookstoves carbon credits. *Golman School of Public Policy, University of California, Berkeley* https://gspp.berkeley.edu/research-and-impact/centers/cepp/projects/berkeley-carbon-trading-project/cookstoves (2023).

Acknowledgements

The research that led to these results received funding from the Center for African Studies, University of California, Berkeley (A.G.-W.), Carbon Direct (B.K.H.), Google.org (D.K.), Katherine Lau Family Foundation (A.G.-W. and D.K.), the National Science Foundation's Graduate Researcher Fellowship Program (A.G.-W.), the Mulago Foundation/Better Cooking Company (A.G.-W.) and the Zaffaroni Family Foundation (D.K.). We thank A. Haya for help with initial data collection. We also thank Calyx Global and D. Lee. A.G.-W. thanks I. Ray and A. Hubbard for additional advising support.

Author contributions

A.G.-W. and B.K.H. co-led the research design. A.G.-W. compiled the data, conducted the analysis and co-led the write up of the paper. B.K.H. originated the idea and co-led the write up of the paper. D.K. contributed to the research design and write up, as well as funding the work.

Competing interests

A.G.-W. has received research support from Better Cooking Company Limited, whose leadership also provided comments on a draft of the manuscript. B.K.H. has received research support from Carbon Direct. D.K. declares no competeing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41893-023-01259-6.

Correspondence and requests for materials should be addressed to Annelise Gill-Wiehl.

Peer review information *Nature Sustainability* thanks Philippe Delacote, Ben Groom and Bruno Marino for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Annelise Gill-Wiehl

Last updated by author(s): Dec 1, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used.
Data analysis	We used Python 3.0 to analyze the data. Figure 1 was created in R 4.2.1. All data and code are available at https://github.com/agillwiehl/GillWiehl_et_al_Pervasive_over_crediting

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data and code are available at https://github.com/agillwiehl/GillWiehl_et_al_Pervasive_over_crediting

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

We conduct a comprehensive, quantitative quality assessment of carbon offsets comparing cookstove offset methodologies and projects to published literature and our own analysis. We (1) discuss the accuracy of all estimation factors used (or not addressed) by the four most prominent cookstoves offset methodologies (GS-TPDDTEC17 , GS-Simplified18, CDM-AMS-II-G19, and CDM-AMS-I-E20) and the recent GS-Metered21 methodology (Table 1) (past and current versions) drawing from published literature and our own analysis (see methods). We then (2) recalculate the carbon emission reductions of a purposive sample of 51 cookstoves projects, addressing ranges of uncertainty using the Monte Carlo Method (MCM), and compare those results with actual credit issuance across eight methodology/stove type categories. We (3) suggest a specific set of methodological reforms to generate high-quality credits. In doing so, we (4) develop and demonstrate an over/under crediting analysis that can be used to systematically assess quality and inform methodology improvements across all offset project types

Research sample

Our sampling approach resulted in a sample of 51 projects spanning 25 countries and eight methodology-project type combinations: (1) GS-Firewood, (2) GS-Simplified-Firewood, (3) GS-Charcoal, (4) GS-LPG, (5) CDM AMS-II-G-Firewood, (6) CDM-AMS-II-G-Charcoal, (7) CDM-AMS-II-E-Ethanol, and (8) GS-Metered-Pellet (WHO Tier 4+ Biomass Pellet Stove). Our sample covers 40% of all issued credits on the VCM from these methodologies (as of May 10th, 2023). We have no reason to believe that these projects are not representative of the entire pool of cookstove credits on the VCM. The 31 GS projects in our sample represent 46% of the covered GS methodologies credits on the VCM. The 16 AMS-II-G contain 25% of that methodologies' credits.

Sampling strategy

In March 2021, we identified the 15 countries with the most credits from cookstove projects on the market and for each country selected the projects with the most credits for each methodology. For the GS-TPDDTEC, GS-Simplified, and CDM-AMS-II-G projects, we chose projects that posted at least one monitoring report and provided their exact calculations and the stove days. There were very few projects under AMS-I-E and GS-Metered and the only one that had been issued credits was also credited under AMS-I-I and so was not included in our sample. For these two methodologies, we selected all registered projects that provided enough information to recreate offset credit calculations on a stove-day basis for individual stove types. We included these methodologies because they offered different methods for monitoring stove usage and fuel consumption, and because of the greater potential emission reductions and health benefits from fuel switch projects that these protocols accommodate. We also added additional projects as needed to ensure that our sample covered all types of fuel transitions, with the exception of electric stoves. There were no issued projects actively deploying an electric stove, and the only listed electric project under GS Metered had no files available. We also do not include GS Metered's most recent methodology update which allows for the participation of more complex cooking devices such as pressure cookers, in a new option called "specific consumption". See supplemental methodology equation information.

Additionally, we randomly selected 10 small/medium sized projects from GS-TPDDTEC (4), AMS-II-G (4), and GS-Simplified (2) to ensure that our sample was representative of both large and small projects. We investigate the relationship between the amount of over/under crediting and project size and find a slight negative relationship between amount of over-crediting and total verified credits (evaluated on the log scale; see supplemental Figure S1). This trend is not statistically significant and the R-squared is very low, but it indicates that our approach of focusing on large projects may have led to lower estimates of over-crediting.

Data collection

The first author obtained all data from the publicly available databases from Gold Standard (SustainCert) and Verra's registry.

Timing and spatial scale	We selected the initial sample of projects and the respective documents from SustainCert in March 2021. We added 5 GS Simplified projects in March of 2023 after receiving feedback from Gold Standard. We added 10 projects after feedback from the review process.
Data exclusions	No data were excluded from the analysis once the sample was established.
Reproducibility	All attempts to repeat the analysis were successful.
Randomization	Randomization is not applicable to our study design.
Blinding	Blinding was not applicable to our study as we did not implement an experimental design.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging