Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A global benchmark of urban heat adaptation

Nabig Chaudhry

nabigac@berkeley.edu

University of California, Berkeley https://orcid.org/0000-0002-3783-539X

David Anthoff

University of California, Berkeley https://orcid.org/0000-0001-9319-2109

Tarun Khanna

Harvard Business School

Spencer Glendon

Probable Futures

Marco Billi

University of Chile

Christa Brelsford

Los Alamos National Laboratory https://orcid.org/0000-0002-3490-8020

Kate Cullen

University of California, Berkeley

Michaela Deglon

University of Cape Town

Kristin Fabbe

European University Institute

Radley Horton

Lamont Doherty Earth Observatory, Columbia University https://orcid.org/0000-0002-5574-9962

Toshiaki Ichinose

National Institute for Environmental Studies

John S. Ji

Tsinghua University https://orcid.org/0000-0002-5002-118X

Cora Kingdon

UC Berkeley https://orcid.org/0000-0001-6207-1048

Stephen Livesley

The University of Melbourne

Heleen Mees

Utrecht University

Vuyisile Moyo

University of Cape Town

Olasunkanmi Okunola
United Nations University Institute for Environment and Human Security
Sushila Pandit
European University Institute
Sebastian Pfautsch
School of Social Science and Psychology, Western Sydney University https://orcid.org/0000-0002-4390-4195
Malini Roy
The University of Texas at Arlington
Himanshu Shekhar
United Nations University – Vice-Rectorate in Europe
Pamela Smith
University of Chile
Liyang Wang
University of California, Berkeley
Saskia Werners
Wageningen University & Research
Andrew Jones
University of California, Berkeley

Social Sciences - Article

Keywords:

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11731493/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

A global benchmark of urban heat adaptation

Nabig A. Chaudhry[1*], David Anthoff[1], Tarun Khanna[2], Spencer Glendon[3], Marco Billi[4], Christa Brelsford[5], Kate A. Cullen[1], Michaela Deglon[6], Kristin Fabbe[7], Radley M. Horton[8,9], Toshiaki Ichinose[10,11], John S. Ji[12], Cora Kingdon[1], Stephen Livesley[13], Heleen Mees[14], Vuyisile Precious Moyo[6], Olasunkanmi Habeeb Okunola[15], Sushila Pandit[7], Sebastian Pfautsch[16], Malini Roy[17], Himanshu Shekhar[18], Pamela Smith[19], Liyang Wang[1], Saskia Werners[15,20], and Andrew D. Jones[1]

[1] Energy and Resources Group, University of California, Berkeley, Berkeley, CA 94720, USA.
[2] Harvard Business School, Harvard University, Boston, MA 02163, USA.
[3] Probable Futures, Massachusetts, USA.
[4] Department of Rural Management and Innovation, Faculty of Agricultural Sciences, University of Chile, Santiago, Chile.
[5] Los Alamos National Laboratory, Los Alamos, NM 87545, USA.
[6] African Climate and Development Initiative, University of Cape Town, Cape Town 7701, South Africa.
[7] Florence School of Transnational Governance, European University Institute, Florence 50133, Italy.
[8] Columbia Climate School, Columbia University, New York, NY 10027, USA.
[9] Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA.
[10] Social Systems Division, National Institute for Environmental Studies, Tsukuba, Ibaraki 305-8506, Japan.
[11] Graduate School of Environmental Studies, Nagoya University, Nagoya, Aichi 464-8601, Japan.
[12] Vanke School of Public Health, Tsinghua University, Beijing 100084, China.
[13] School of Agriculture, Food and Ecosystem Sciences, The University of Melbourne, Parkville, VIC 3010, Australia.
[14] Copernicus Institute of Sustainable Development, Faculty of Geosciences, Utrecht University, Utrecht 3584 CB, Netherlands.
[15] United Nations University Institute for Environment and Human Security (UNU-EHS), Bonn 53113, Germany.
[16] School of Social Sciences, Western Sydney University, Penrith, NSW 2751, Australia.
[17] Department of Public Affairs and Planning, College of Architecture, Planning and Public Affairs, The University of Texas at Arlington, Arlington, TX 76019, USA.
[18] United Nations University - Vice-Rectorate in Europe (UNU-VIE), Bonn 53113, Germany.
[19] Department of Geography, Faculty of Architecture and Urbanism, University of Chile, Santiago, Chile.
[20] Earth Systems and Global Change Group, Wageningen University & Research, Wageningen 6708 PB, Netherlands.

* Corresponding author: Nabig A. Chaudhry (nabigac@berkeley.edu)

Summary

Climate change is increasing temperatures and worsening extreme heat[1–5]. Heat threatens human and natural systems[2,3,5–7]. In cities, urban heat islands and concentrations of people, infrastructure, and economic activity amplify risks[6,8–12]. Cities are adapting to heat, yet evidence remains fragmented and difficult to compare because adaptation is multidimensional, context-specific, and scale-dependent[3,6,7,12–18]. Here we introduce and apply the Heat Adaptation Benchmark, a multi-criteria framework for evaluating heat adaptation through structured expert judgment. Twenty-four expert evaluators assessed a purposive sample of 46 cities spanning six global regions and differing in three city characteristics: metropolitan area population, country-level economic resources, and historical climate. Across all cities, the mean total score suggested considerable scope to strengthen documented heat adaptation (36.1 out of 72). Of the six criteria, risk assessment had the highest mean score (8.4 out of 12) and monitoring and evaluation the lowest (3.4). In every city, the combined score for risk assessment, planning, and actions exceeded that for strategy, monitoring and evaluation, and governance. Higher-scoring cities had different combinations of strengths and gaps and were found in every region and city characteristic category. Our results highlight challenges in translating documented assessment, planning, and identified actions into sustained implementation and demonstrated heat risk reduction. We anticipate that the benchmark will support cross-city comparison and learning by illustrating a range of heat adaptation approaches and identifying practices that cities facing similar conditions or challenges could draw on. Across contexts, our findings point to the importance of durable institutions that can implement heat adaptation plans, sustain action through long-term strategies and reliable funding, and assess whether those actions reduce heat risk. Repeated assessments with the benchmark could help track progress and inform adjustments as heat risks evolve.

Main text

Anthropogenic climate change is increasing average temperatures and the frequency, intensity, and duration of extreme heat events[1–5]. These changes manifest as chronic heat stress (long-term rising temperatures) or acute heat shocks (short-term extreme events) and carry profound consequences for society[1,2,6,8].

Heat is a complex hazard that often interacts with and compounds other risks[3,6,8]. The effects of heat can cascade across human and natural systems, posing serious threats to human health, physical infrastructure, agricultural production, economic activity, and ecosystem functioning[2,3,5–7]. In particular, heat-related health impacts can be fatal, with risks especially high among populations with greater exposure or vulnerability[2,3,7,8].

Together, these growing impacts and risks underscore the urgency of adapting to heat ('heat adaptation')[1]. Heat adaptation broadly refers to efforts to reduce heat risk by limiting exposure and vulnerability, strengthening adaptive capacity, or increasing resilience[2,3,6–8,14]. These efforts can include behavioral, technological, infrastructural, institutional, and ecosystem-based strategies[2–6].

In urban and metropolitan areas (here called 'cities'), heat risks are further amplified by additional warming from the urban heat island effect and by the concentration of people, infrastructure, and economic activity exposed to heat[6,8–12]. Cities also play a central role in adaptation planning and implementation at local and regional scales[7,16]. Accordingly, cities are increasingly pursuing heat adaptation strategies, such as heat action plans, early warning and emergency response systems, tree canopy expansion, and governance arrangements, including chief heat officer positions[6,12–15].

As heat adaptation efforts have expanded, so has demand for systematic approaches to evaluate and benchmark them[17,19,20]. Systematic evaluation and benchmarking can document progress and outcomes, examine links between planning and implementation, identify needs and gaps, support comparison and learning, and inform future action[10,17,18,20,21].

However, evidence on how cities are adapting to heat remains fragmented, and evaluation is complicated because adaptation spans multiple dimensions, varies by context, and may be planned and implemented at different geographic or administrative scales[3,7,17,18,22]. Establishing a common basis for assessing and aggregating heat adaptation efforts is thus difficult[15,17,18,22]. These challenges hinder consistent comparisons across cities and the synthesis of local evidence in global assessments[3,10,17,18,20]. Transparent evaluation frameworks and comparative assessments are therefore needed to strengthen the evidence base for heat adaptation[3,10,16,17,19,20,22,23].

Here we present a systematic approach to evaluating and benchmarking heat adaptation across cities. Our aim is to characterize the current state of documented heat adaptation efforts, analyze broad patterns across cities, highlight strengths and gaps, and inform cross-city learning. Using a bottom-up approach that aggregates city-level assessments, we examine local heat adaptation efforts in 46 cities ('city case studies') spanning six global regions and differing in three city characteristics: metropolitan area population, country-level economic resources, and historical

climate (see Methods for city selection details). We operationalize this approach through a multi-criteria framework for assessing heat adaptation efforts ('Heat Adaptation Benchmark') and a structured expert judgment process involving 24 researchers ('expert evaluators') from around the globe.

This approach provides a comparative baseline for understanding documented heat adaptation across varied city contexts. Findings reflect the evidence available to expert evaluators and their judgments, drawing on a purposively selected sample that is not statistically representative of cities globally. Although the benchmark considers evidence of outcomes, higher total scores do not establish greater effectiveness in reducing heat risk.

We first describe the benchmark, present results for the full sample of 46 cities, and compare scores across global regions and city characteristic categories. We then examine patterns, strengths, and gaps in heat adaptation efforts through assessment question scores, evaluator rationales, and illustrative city profiles. Finally, we draw lessons from selected city case studies for implementing, sustaining, and evaluating heat adaptation efforts.

Heat Adaptation Benchmark

We designed the Heat Adaptation Benchmark to guide the city case study assessments (Table 1). The benchmark builds on the adaptation assessment approach of Baker et al. (2012)[23] and the GAP-Track coastal adaptation assessment framework developed by Magnan et al. (2023)[20]. We refined its structure and tailored it specifically for heat adaptation through further literature review and iterative feedback from the expert evaluators[2,3,7,8,14,16,19,24].

Table 1 | Heat Adaptation Benchmark

Benchmark criterion	Assessment question
1. Climate impact, vulnerability, and risk assessment	**1.1** Are current heat hazards identified?
	1.2 Are current drivers of heat exposure and vulnerability identified?
	1.3 Are future heat risks projected?
2. Adaptation plan	**2.1** Is there a dedicated heat adaptation plan?
	2.2 Is there a range of short- to long-term goals?
	2.3 Are stakeholder and community engagement mechanisms in place?
3. Adaptation actions	**3.1** Are actions addressing heat hazards identified?
	3.2 Are actions addressing drivers of heat exposure and vulnerability identified?
	3.3 Are identified actions being implemented?
4. Adaptation strategy	**4.1** Are costs and benefits of heat adaptation options analyzed?
	4.2 Are synergies and trade-offs between heat adaptation options assessed?
	4.3 Are heat adaptation options planned in a sequenced, pathway-based approach?
5. Monitoring and evaluation	**5.1** Are adaptation actions monitored and evaluated?
	5.2 Is there evidence of reducing heat risk?
	5.3 Is there evidence of reducing maladaptation risk?

6. Heat governance	**6.1** Are heat governance structures established?
	6.2 Is there technical expertise in place and accessible?
	6.3 Is there dedicated and sustainable heat adaptation funding?

Note: Detailed scoring guidance for each assessment question is provided in the Heat Adaptation Scoring Rubric (see Supplementary Information).

The Heat Adaptation Benchmark was designed as a practical framework for assessing documented heat adaptation across cities with varied contexts[20]. It comprises six benchmark criteria, each containing three assessment questions, for a total of 18 assessment questions (Table 1). The criteria represent interconnected dimensions of heat adaptation, rather than sequential stages through which all cities progress.

For each city case study, two expert evaluators independently scored each of the 18 assessment questions from 0 (none) to 4 (very high), using the detailed Heat Adaptation Scoring Rubric (see Supplementary Information). Evidence sources included publicly available documents (such as adaptation plans and reports), stakeholder survey responses, and expert evaluators' firsthand knowledge where available. Evaluators also assigned confidence scores for each benchmark criterion from 0 (no confidence) to 4 (very high confidence).

Differences of three or more points between paired assessment or confidence scores prompted a discussion of the evidence and its interpretation. Evaluators could revise or retain their scores during this reconciliation process, and consensus was not required. Following reconciliation, 80.1% of paired assessment score comparisons and 84.4% of paired confidence score comparisons differed by no more than one point on the 0-4 point scale (see Supplementary Information). These percentages do not describe agreement between the initial independent evaluations.

Assessment scores were then averaged by question, and the three mean assessment scores within each criterion were summed into a benchmark criterion score (0-12). The six benchmark criterion scores were then summed to produce a total heat adaptation benchmark score (0-72). Total benchmark scores were classified into four equal ranges: low (0-18), moderate (>18-36), high (>36-54), and very high (>54-72) (see Methods for scoring and aggregation details). Confidence scores were averaged and analyzed separately and were not used to weight assessment scores. Scores should be interpreted alongside confidence scores and written rationales.

Global results

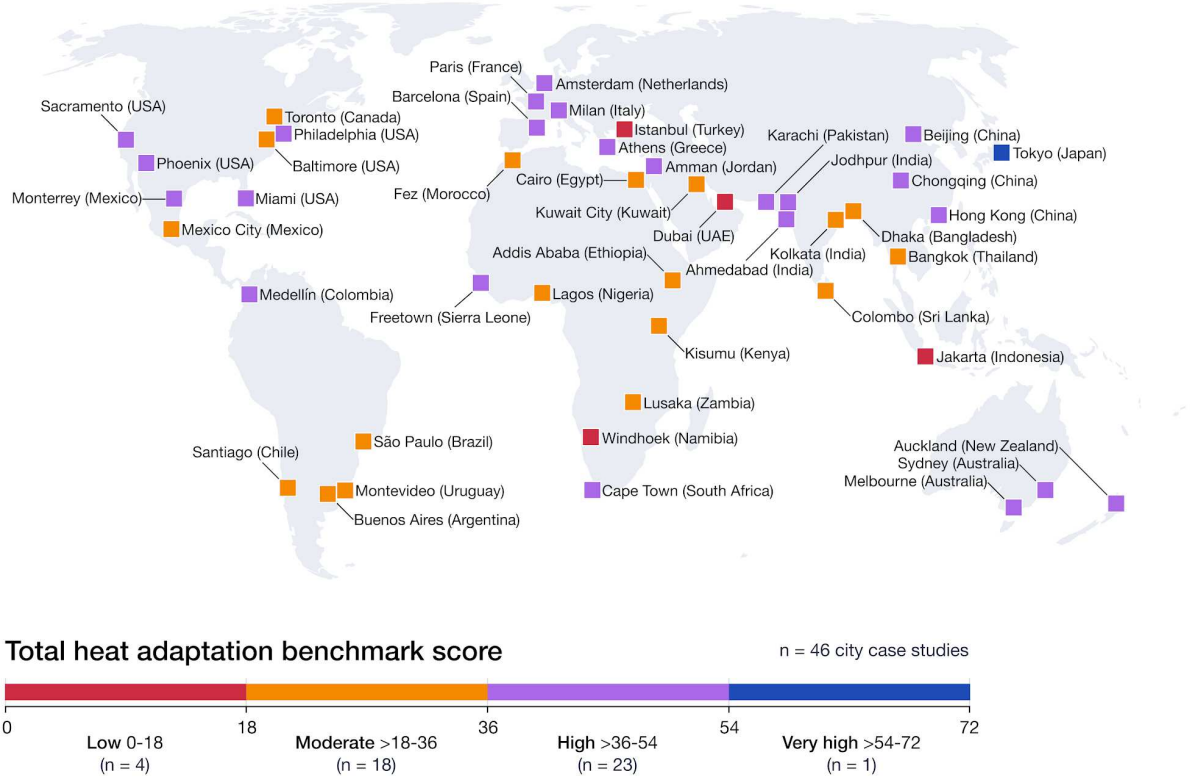


Fig. 1 | Global city case studies map. Each square marks the approximate location of one of the 46 city case studies. Square colors indicate the total heat adaptation benchmark score category: low (0-18, *n = 4*), moderate (>18 and ≤36, *n = 18*), high (>36 and ≤54, *n = 23*) or very high (>54 and ≤72, *n = 1*). Underlying city-level scores and classifications are provided in Supplementary Data 1.

Among the 46 city case studies, 23 had high total heat adaptation benchmark scores, 18 had moderate scores, four had low scores, and one had a very high score (Fig. 1). Tokyo, Japan, had the highest score (57.0 out of 72), followed by Jodhpur, India (52.5), whereas Jakarta, Indonesia (13.5) and Dubai, UAE (14.5) had the lowest. Across our global sample of 46 cities, the mean total heat adaptation benchmark score was 36.1 out of 72, approximately half the maximum and just above the threshold separating moderate and high scores (Fig. 2). This suggests considerable scope to strengthen heat adaptation in the areas assessed.

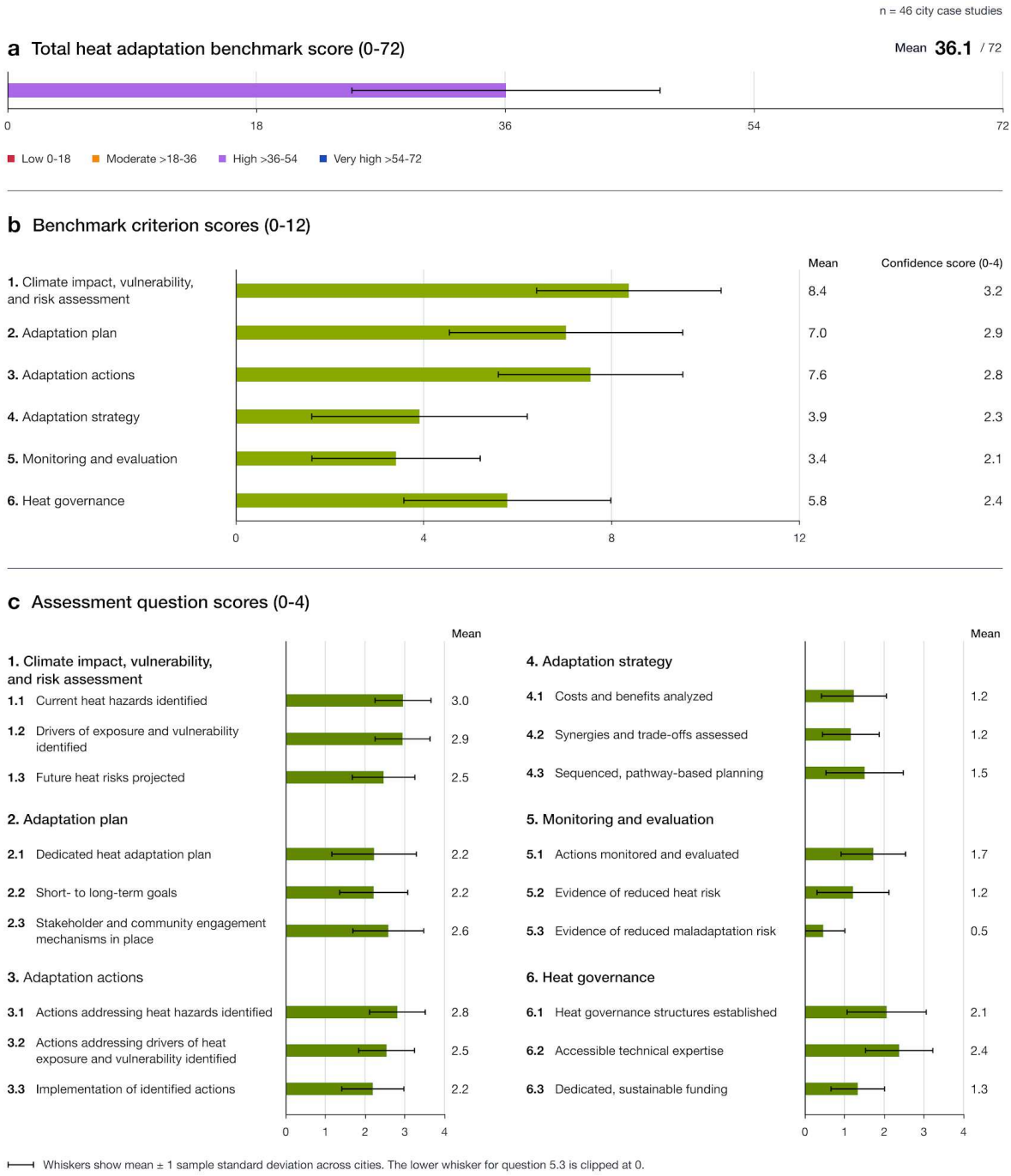


Fig. 2 | Global score summary. Bars show mean scores across 46 city case studies. **a**, Total heat adaptation benchmark score (0-72). Purple indicates that the sample mean falls in the high score category. **b**, Scores for the six benchmark criteria (0-12), with the corresponding mean confidence scores (0-4) shown at right. **c**, Scores for the 18 assessment questions (0-4), grouped by benchmark criterion. Whiskers show the mean ± 1 sample standard deviation across city case studies. The lower whisker for question 5.3 is truncated at 0, the lower bound of the scoring scale. Displayed means are rounded to one decimal place. Underlying scores and bar chart statistics are provided in Supplementary Data 1 and 2, respectively.

At the benchmark criterion level, mean benchmark criterion scores were highest for climate impact, vulnerability, and risk assessment (8.4 out of 12), followed by adaptation actions (7.6) and adaptation plan (7.0) (Fig. 2). Monitoring and evaluation had the lowest mean benchmark criterion score (3.4), followed by adaptation strategy (3.9) and heat governance (5.8). Mean confidence scores followed a similar pattern, with the highest mean for climate impact, vulnerability, and risk assessment (3.2 out of 4) and the lowest for monitoring and evaluation (2.1).

For all 46 city case studies, the combined score across the first three benchmark criteria (risk assessment, adaptation plan, and adaptation actions) exceeded the combined score across the latter three (adaptation strategy, monitoring and evaluation, and heat governance).

At the assessment question level, mean scores were highest for identifying current heat hazards (3.0 out of 4), drivers of exposure and vulnerability (2.9), and actions addressing heat hazards (2.8) (Fig. 2). Evidence of reduced maladaptation risk scored lowest (0.5), while other comparatively low-scoring questions concerned costs and benefits, synergies and trade-offs, sequenced planning, evidence of reduced heat risk, and sustainable funding.

These global results summarize documented adaptation efforts but do not establish whether those efforts are sufficient to address local heat risks. Such risks depend on local climatic conditions, exposure, and vulnerability[5,14].

Regional comparisons

When city case studies were grouped by global region, the mean total heat adaptation benchmark score was highest in Australia and New Zealand (41.5 out of 72), closely followed by North America (41.0), and lowest in Africa (29.9) (Supplementary Data 1). The 11.6-point range in regional means exceeded the corresponding ranges across city characteristic categories (Supplementary Data 1). Regional total score distributions overlapped, and scores varied substantially within regions (Fig. 4).

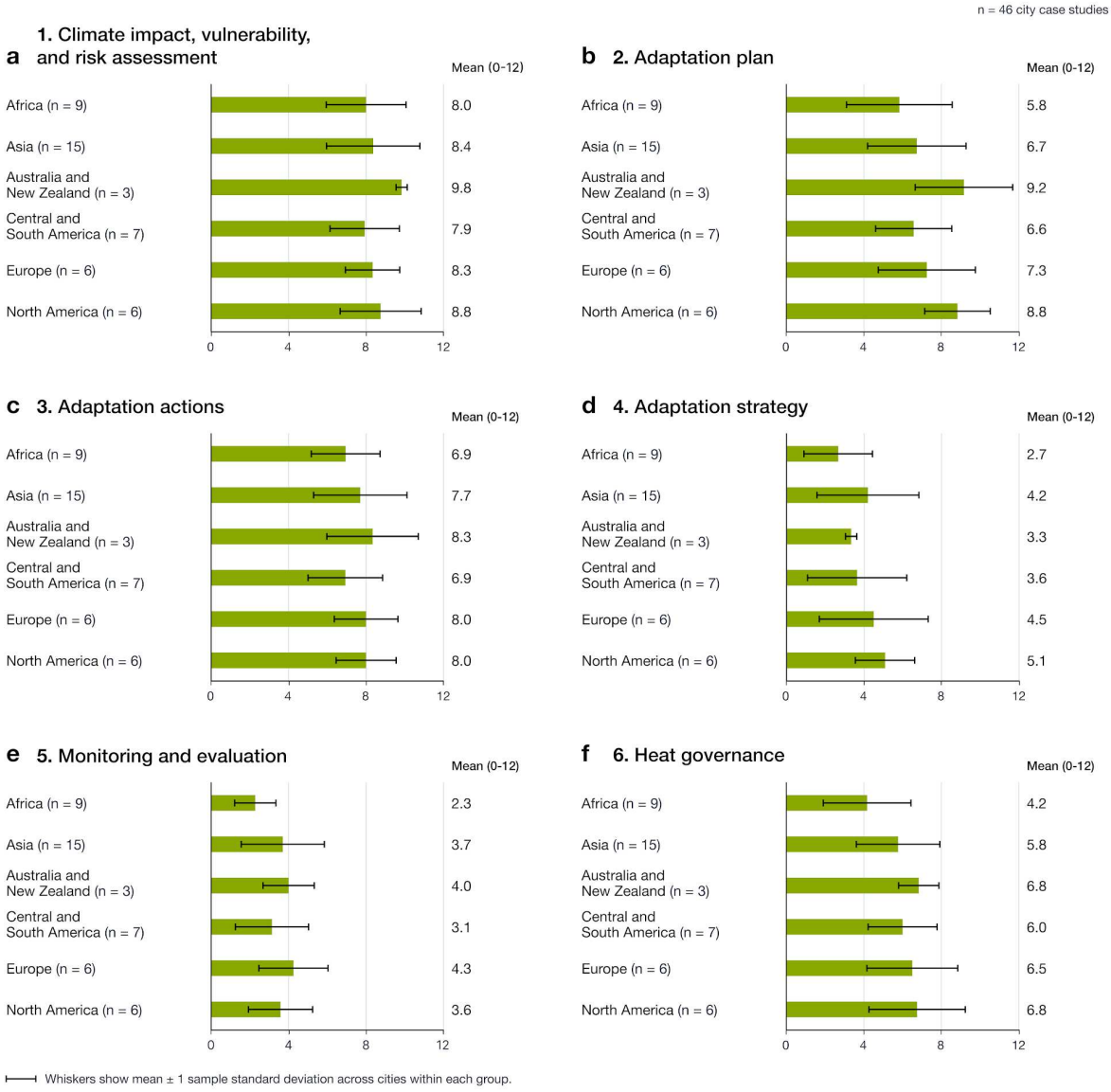


Fig. 3 | Mean benchmark criterion scores by global region. Each panel shows scores for one benchmark criterion across six global regions. **a**, Climate impact, vulnerability, and risk assessment. **b**, Adaptation plan. **c**, Adaptation actions. **d**, Adaptation strategy. **e**, Monitoring and evaluation. **f**, Heat governance. Bars show mean criterion scores (0-12), and whiskers show the mean ± 1 sample standard deviation across city case studies within each region. The sample included Africa (*n = 9*), Asia (*n = 15*), Australia and New Zealand (*n = 3*), Central and South America (*n = 7*), Europe (*n = 6*) and North America (*n = 6*). Underlying scores and bar chart statistics are provided in Supplementary Data 1 and 2, respectively.

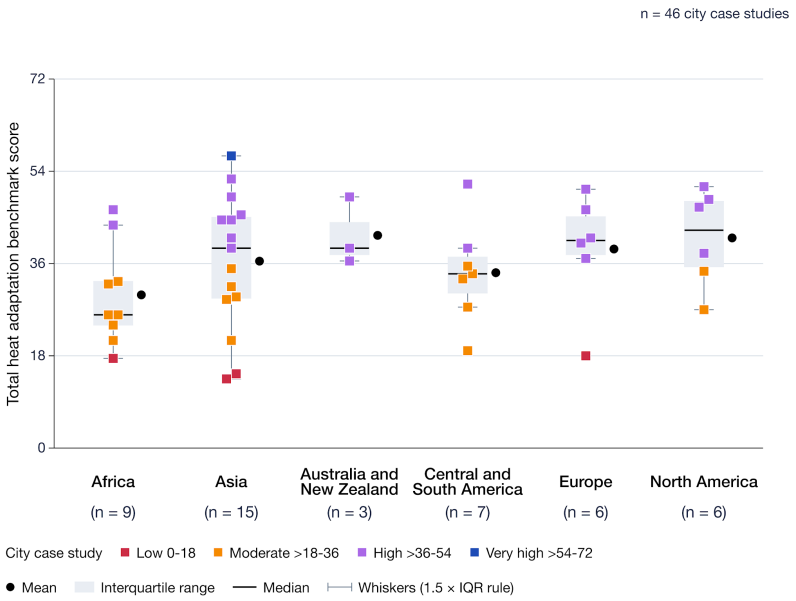


Fig. 4 | Distribution of total heat adaptation benchmark scores by global region. Squares show individual city case study scores, colored by score category: low (0-18), moderate (>18 and ≤36), high (>36 and ≤54) or very high (>54 and ≤72). Black circles indicate means, horizontal lines within boxes indicate medians, and boxes span the first to third quartiles (interquartile range, IQR). Whiskers extend to the most extreme observed scores within 1.5 × IQR below the first quartile and above the third quartile. The sample included Africa (*n = 9*), Asia (*n = 15*), Australia and New Zealand (*n = 3*), Central and South America (*n = 7*), Europe (*n = 6*) and North America (*n = 6*). Underlying scores and boxplot statistics are provided in Supplementary Data 1 and 2, respectively.

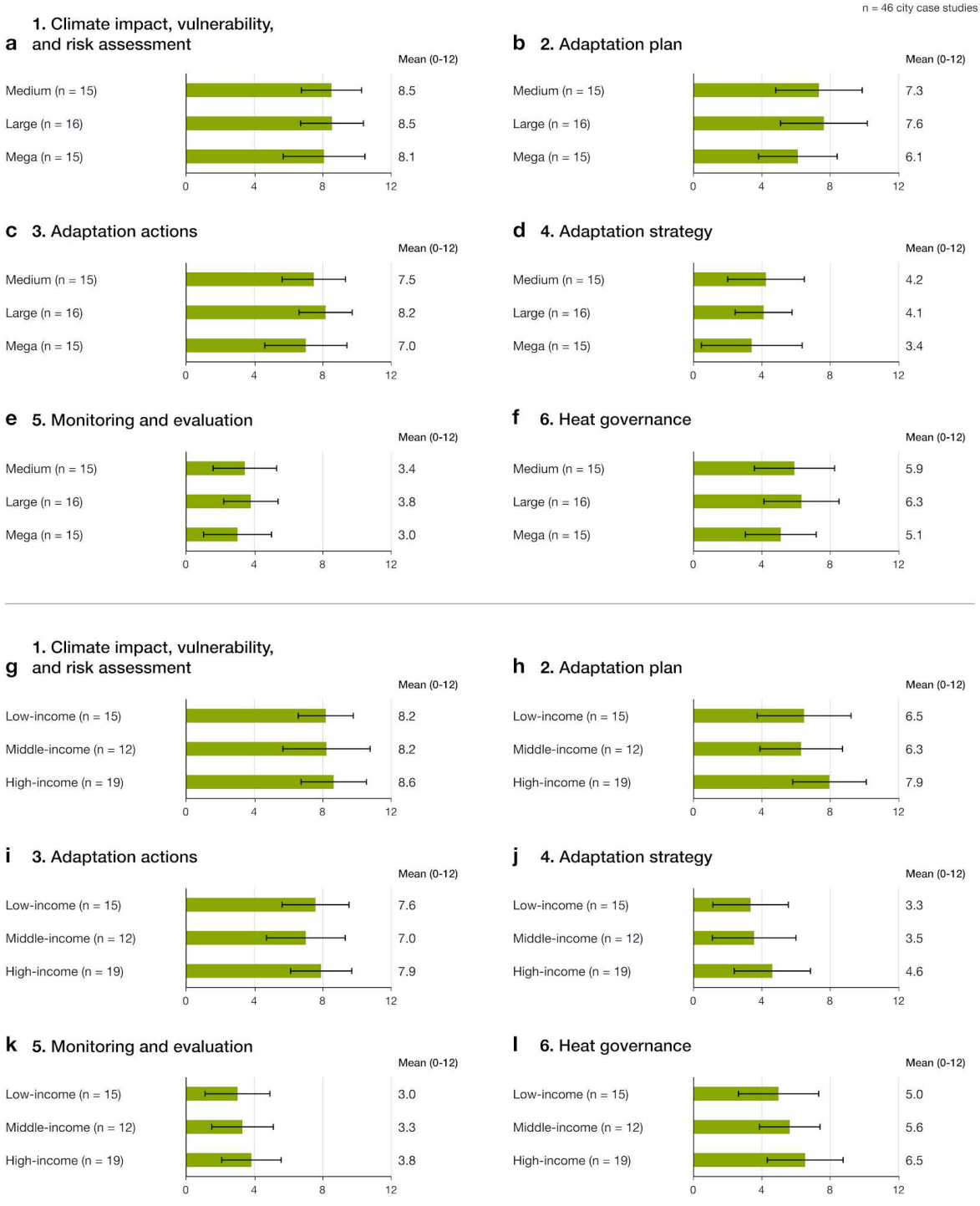
At the benchmark criterion level, Africa had the lowest mean scores for four of the six criteria. Central and South America scored lowest for risk assessment and adaptation actions (its adaptation actions mean and Africa's both round to 6.9 out of 12) but had a comparatively higher heat governance score (6.0 out of 12) (Fig. 3). Australia and New Zealand scored highest for four criteria, North America for adaptation strategy, and Europe for monitoring and evaluation. At the assessment question level, evidence of reducing maladaptation risk was the lowest-scoring question in every region (ranging from 0.2 to 0.8 out of 4) (Supplementary Data 1). Confidence patterns broadly aligned with assessment results, with Africa recording the lowest mean confidence score for each of the six benchmark criteria (Supplementary Data 1).

Substantial within-region variation was evident in Africa, Asia, Central and South America, and Europe (Fig. 4 and Supplementary Data 1). Freetown, Sierra Leone (46.5 out of 72) and Cape Town, South Africa (43.5) were among Africa's higher-scoring cases. Asia had the widest range, from Jakarta, Indonesia (13.5) to Tokyo, Japan (57.0). Central and South America ranged from São Paulo, Brazil (19.0) to Medellín, Colombia (51.5), while Europe ranged from Istanbul, Turkey (18.0) to Paris, France (50.5).

Regional comparisons should be interpreted cautiously because group sizes and national diversity differed, including only three cities in Australia and New Zealand and five US cities among the six North American cases.

City characteristic comparisons

For descriptive comparisons, cities were grouped by metropolitan area population (medium: <5 million; large: 5 to <10 million; mega: ≥10 million), country-level economic resources (low-, middle-, and high-income groups), and historical climate (hot and temperate/cool) (Figs. 5 and 6). These characteristics also informed the second phase of city selection (see Methods for category definitions and selection procedures).



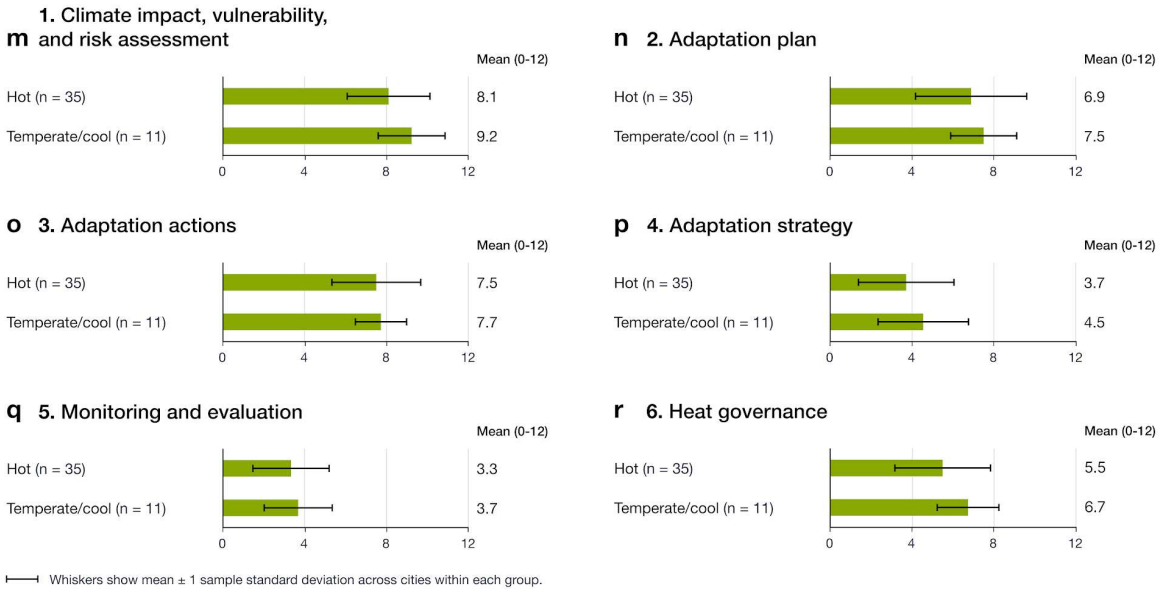


Fig. 5 | Mean benchmark criterion scores by city characteristic. **a-f**, Metropolitan area population, with 15 medium, 16 large and 15 mega city case studies. **g-l**, Country-level economic resources, with 15, 12 and 19 city case studies in the low-, middle- and high-income groups, respectively. **m-r,** Historical climate, with 35 hot and 11 temperate/cool city case studies. Each set shows the six criteria in the same order: climate impact, vulnerability, and risk assessment, followed by adaptation plan, adaptation actions, adaptation strategy, monitoring and evaluation, and heat governance. Bars show mean criterion scores (0-12), and whiskers show the mean ± 1 sample standard deviation across city case studies within each category. Categories are defined in Methods Table 2. Underlying scores and bar chart statistics are provided in Supplementary Data 1 and 2, respectively.

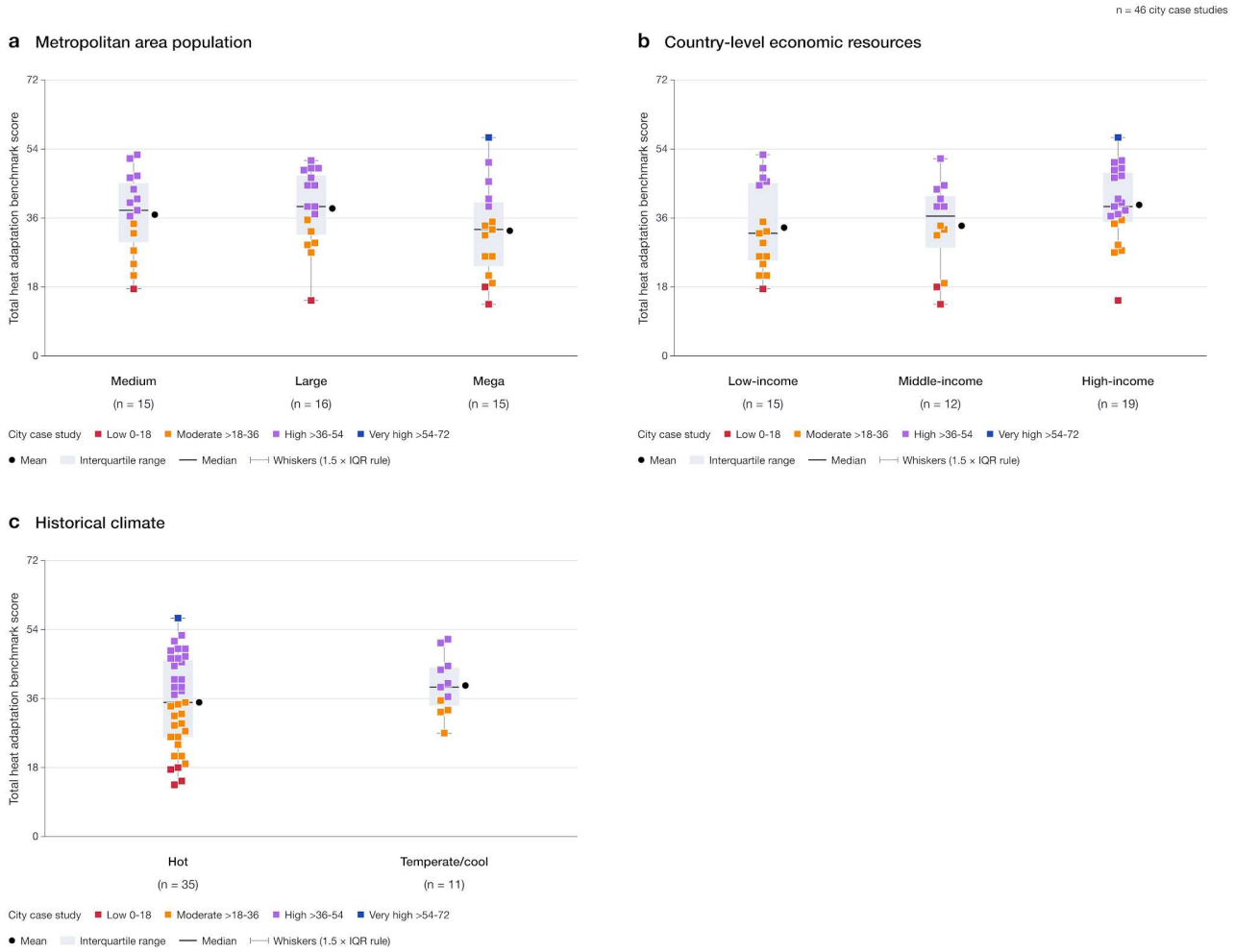


Fig. 6 | Distribution of total heat adaptation benchmark scores by city characteristic. **a**, Metropolitan area population (medium, *n = 15*, large, *n = 16*, and mega, *n = 15*). **b**, Country-level economic resources (low-income group, *n = 15*, middle-income group, *n = 12*, and high-income group, *n = 19*). **c**, Historical climate (hot, *n = 35*, and temperate/cool, *n = 11*). Categories are defined in Methods Table 2. Squares show individual city case study scores, colored by score category: low (0-18), moderate (>18 and ≤36), high (>36 and ≤54) or very high (>54 and ≤72). Black circles indicate means, horizontal lines within boxes indicate medians, and boxes span the first to third quartiles (interquartile range, IQR). Whiskers extend to the most extreme observed scores within 1.5 × IQR below the first quartile and above the third quartile. Underlying scores and boxplot statistics are provided in Supplementary Data 1 and 2, respectively.

For metropolitan area population, medium and large cities had similar mean total heat adaptation benchmark scores (36.9 and 38.5 out of 72, respectively), whereas mega cities scored lower (32.7) (Supplementary Data 1). Mega cities also had the lowest mean scores across all six criteria, with the largest differences in adaptation planning (Fig. 5). Nevertheless, total score distributions overlapped substantially across the three population categories (Fig. 6).

For country-level economic resources, mean total heat adaptation benchmark scores increased from cities in the low-income (33.5 out of 72) and middle-income (34.0) groups to those in the

high-income group (39.4) (Supplementary Data 1). Mean benchmark criterion scores for climate impact, vulnerability, and risk assessment varied little across economic categories (ranging from 8.2 to 8.6 out of 12), while larger differences occurred in adaptation plan, heat governance, and adaptation strategy (Fig. 5). Cities in the high-income group had higher mean scores for dedicated plans, governance structures, and sustainable funding (Supplementary Data 1). Distributions nevertheless overlapped, including high-scoring cities in the low-income group and lower-scoring cities in the high-income group (Fig. 6).

For historical climate, the mean total heat adaptation benchmark score was higher for historically temperate/cool cities (39.4 out of 72) than historically hot cities (35.0) (Supplementary Data 1). Differences in mean benchmark criterion scores were largest for risk assessment and heat governance, while mean scores for adaptation actions were similar (Fig. 5). Historically hot cities had a wider score range (13.5-57.0 out of 72) than temperate/cool cities (27.0-51.5), although the groups differed greatly in sample size (35 hot and 11 temperate/cool cities) (Fig. 6).

In general, mean scores for city characteristic categories provide only a partial picture of documented heat adaptation, making individual city cases important for interpreting these comparisons.

City characteristic patterns

Two contextual patterns regarding city characteristics warrant closer consideration. The differences in scores for mega cities raise questions about coordination across metropolitan areas, while the economic comparisons show how individual cities can depart from group averages.

For mega cities, comparatively low scores for adaptation planning, adaptation strategy, monitoring and evaluation, and heat governance may partly reflect the difficulty of coordinating heat adaptation across large, institutionally fragmented metropolitan systems. Coordination across neighborhoods, authorities, departments, and jurisdictions may therefore be especially important in mega cities for aligning heat adaptation plans, clarifying responsibilities, and tracking implementation[25,26].

Among individual cities, higher country-level economic resources did not consistently correspond to higher benchmark scores. Cities such as Jodhpur, India; Ahmedabad, India; Amman, Jordan; Freetown, Sierra Leone; and Karachi, Pakistan exceeded the mean scores of cities in the high-income group for at least five of the six benchmark criteria (Supplementary Data 1). These comparisons do not imply that economic resources are unimportant, especially because the sample was not statistically representative and economic resources were measured using a country-level proxy. Instead, these cases may warrant closer examination of how institutional arrangements, political priorities, or lower-cost approaches may support heat adaptation under resource constraints.

In general, higher-scoring cases were found in every category of metropolitan area population, country-level economic resources, and historical climate. In each comparison, some cities in groups with lower means matched or outscored cities in groups with higher means. Examining the documented efforts behind these cases can show how cities tailor and implement heat adaptation

under different local contexts and constraints. Cities in similar circumstances can learn from these approaches.

Global strengths and gaps

Across these varied city contexts, the benchmark results suggest a common challenge in sustaining implementation and demonstrating reductions in heat risk. Documented strengths in risk assessment, planning, and identifying actions were accompanied by gaps in the longer-term strategy, institutional capacity, and evaluation systems needed to support these efforts.

Mean assessment scores were higher for identifying actions addressing heat hazards (2.8 out of 4) than for implementing identified actions (2.2). This contrast indicates that comparatively strong scores within the adaptation actions criterion did not necessarily correspond to equally strong evidence of implementation. Longer-term decision-making, sustainable funding, and evidence of reduced heat or maladaptation risk also scored comparatively low, reinforcing gaps in institutional follow-through and outcomes.

Monitoring and evaluation was the lowest or tied-lowest benchmark criterion in 31 of the 46 city case studies (67.4%) (Supplementary Data 1). By contrast, no city had risk assessment, adaptation plan, or adaptation actions as its lowest criterion. Evidence of reduced maladaptation risk was especially limited, with no city scoring above 2.0 out of 4. The evaluated cases therefore provided limited documented evidence that adaptation efforts avoided or reduced unintended harms. Lower confidence scores for adaptation strategy and monitoring and evaluation indicate greater uncertainty in these assessments, which may partly reflect limitations in the availability, accessibility, or quality of supporting evidence.

Accordingly, confidence scores provide essential context for interpreting assessment results. A low assessment score with high confidence indicates that expert evaluators judged efforts related to that question to be limited. When confidence is low, it is less clear whether a low score reflects limited efforts or gaps in the available evidence. Considering both scores helps distinguish documented gaps from uncertainty in the evidence base. This distinction may be especially relevant for the city case studies in Africa, which had the lowest mean confidence score for each of the six benchmark criteria. This confidence gap may partly reflect more limited availability or accessibility of public evidence and planning documents for those cities.

Evaluator rationale themes

Expert evaluator scoring rationales provide additional context for the gaps identified across the benchmark criteria (see Supplementary Information for city case study scoring sheets). Because rationales varied in length and detail, we explore three themes and do not estimate their prevalence across the sample.

First, rationales highlighted a gap between the approaches to uncertainty used in assessing future heat risks and developing adaptation strategies. Some heat risk assessments used multiple scenarios to explore a range of possible futures. Yet the associated longer-term strategies did not

explain how actions would be selected, sequenced, or adjusted to account for those different futures. For example, some strategies assigned actions to target years, such as 2050, without specifying the thresholds or decision points that would prompt cities to accelerate, redesign, or replace an intervention. Target dates therefore provided a timetable for action without necessarily establishing a flexible adaptation pathway. Examples included Amman, Jordan; Kuwait City, Kuwait; and Istanbul, Turkey.

Second, rationales highlighted two challenges in demonstrating reduced heat risk: measuring outcomes and attributing them to specific adaptation actions. Expert evaluators described cities tracking activities such as trees planted, shelters established, or emergency protocols activated. These indicators documented implementation but did not, by themselves, demonstrate outcomes such as reductions in heat exposure, vulnerability, or risk. Even when temperature, heat-related illness, or mortality was monitored, rationales noted the difficulty of attributing observed changes to particular heat adaptation actions because outcomes could also reflect broader changes in healthcare, technology, or individual behavior. Examples included Lusaka, Zambia; Barcelona, Spain; and Tokyo, Japan.

Third, rationales distinguished establishing heat-related roles, offices, or partnerships from building lasting municipal capacity for heat adaptation. Heat-specific roles and offices, such as Chief Heat Officers, can provide focal points for coordination, while external partnerships can provide technical expertise or financial support. For roles and offices, rationales questioned whether responsibilities and capabilities could outlast a particular officeholder or project. Some established roles or offices were described as discontinued, no longer funded, or having unclear authority. In some cases, broader governance arrangements persisted after heat-specific roles or offices ended or changed. For external partnerships, rationales also questioned whether the expertise provided had become embedded in routine municipal capacity and could be sustained beyond the partnership itself. Examples included Miami, USA; Dhaka, Bangladesh; Lagos, Nigeria; and Monterrey, Mexico.

Heat adaptation profiles

To examine how these strengths and gaps combined within individual cities, we considered contrasting score-based heat adaptation profiles. We examine the 24 city case studies with “high” or “very high” total heat adaptation benchmark scores, alongside the four with “low” scores. These profiles show different configurations of documented heat adaptation across cities.

Among the 24 high- and very high-scoring cases, we illustrate four descriptive, nonexclusive profiles using selected city examples. These profiles were not derived through formal statistical grouping methods. For the examples in the first three profiles, the mean score across the highlighted criteria exceeded the mean across the remaining criteria on the 0-12 point scale. The fourth illustrates comparatively strong scores across all six criteria.

- **Assessment- and planning-led adaptation.** These cities had comparatively high scores for risk assessment and adaptation plan, which could suggest strong foundations for characterizing heat risk and establishing formal direction. However, these strengths did not

necessarily coincide with high scores for strategy, monitoring and evaluation, or governance. Examples include Sacramento, USA (mean score of 9.8 for the highlighted criteria versus 4.6 for the remaining criteria) and Cape Town, South Africa (10.5 versus 5.6).

- **Action-led adaptation.** These cities had comparatively high scores for adaptation actions, reflecting strong documented action portfolios. These strengths did not necessarily coincide with high scores for longer-term strategy or monitoring and evaluation. Strong action scores also often accompanied high risk assessment and planning scores. Examples include Sydney, Australia (11.0 versus 7.6), and Monterrey, Mexico (9.0 versus 6.0).
- **Planning- and governance-led adaptation.** These cities combined comparatively high adaptation plan and heat governance scores, highlighting how formal plans may be paired with documented institutional arrangements. However, these strengths did not necessarily coincide with sequenced decision-making or demonstrated outcomes. Examples include Freetown, Sierra Leone (9.8 versus 6.8), and Phoenix, USA (9.8 versus 6.9).
- **Balanced adaptation.** These cities had comparatively high scores across all six benchmark criteria, with every criterion score exceeding 6.0 out of 12. Examples include Tokyo, Japan; Jodhpur, India; Medellín, Colombia; and Paris, France. All four were among the five highest-scoring cities in the sample. Nevertheless, monitoring and evaluation remained the lowest-scoring criterion for all four.

Among the four city case studies with "low" total heat adaptation benchmark scores, relative strengths varied: risk assessment scored highest in Windhoek, Namibia, and Istanbul, Turkey; adaptation actions in Dubai, UAE; and heat governance in Jakarta, Indonesia. Adaptation strategy and monitoring and evaluation were the two lowest-scoring criteria in all four cases. Compared with these four cases, the 24 cases with "high" or "very high" total scores had both higher maximum criterion scores (ranges of 8.0-12.0 versus 4.5-6.0) and higher minimum criterion scores (ranges of 1.5-7.0 versus 0.0-0.5).

These profiles should not be interpreted as adaptation pathways or evidence that one configuration produces stronger adaptation. They show that high total scores can reflect different combinations of documented strengths and gaps. These differences provide a starting point for examining how cities put specific heat adaptation practices into effect and what arrangements support them.

Heat adaptation lessons

These contrasting heat adaptation profiles offer opportunities for cross-city learning on implementing, sustaining, and evaluating heat adaptation. Drawing on the benchmark scores and evaluator rationales, we highlight three complementary lessons from selected city case studies (see Supplementary Information). These examples illustrate some of the learning opportunities across the sample. Other lessons may be more relevant depending on a city's heat risks, institutions, and existing adaptation strengths and gaps.

First, plans can support sustained implementation by establishing clear roles and coordination structures, alongside regular review of progress and evaluation of outcomes. In Ahmedabad, India, the heat action plan assigned an Ahmedabad Municipal Corporation nodal officer to coordinate heat

response, defined departmental responsibilities, and specified annual post-season review[27]. A published pilot evaluation compared all-cause mortality before and after implementation, although the plan's contribution to observed changes remained uncertain[28].

Second, engagement with affected communities can inform measures tailored to local exposure and vulnerability. In Freetown, Sierra Leone, shade structures in open-air markets were co-designed with market women and community stakeholders[29]. Their design combined protection from sun and heavy rain with sufficient daylight for trading, addressing heat exposure alongside vendors' practical working needs. Reflective roof pilots also targeted informal settlements, although securing funding to sustain and expand both of these efforts remained a challenge[29].

Third, broader urban policies can connect immediate heat protection with longer-term adaptation. In Paris, France, the climate action plan combined access to cool spaces and outreach to vulnerable residents during heatwaves with greening and measures to reduce overheating in buildings[30]. This linked responses to current heat risks with changes intended to reduce exposure as the climate warms. Many measures were planned or ongoing, and their effects over time remained uncertain.

Other opportunities for learning include Phoenix's coordination across departments and jurisdictions and Medellín's evidence of local cooling associated with green corridors. Across these examples, the wider challenge remains connecting documented plans and identified actions with the capacity to sustain implementation and demonstrate reductions in heat risk.

Conclusion

We designed and applied the Heat Adaptation Benchmark to assess documented heat adaptation efforts across 46 cities using structured expert judgment. We found that documented strengths were concentrated in risk assessment, planning, and identifying actions, whereas evidence of sustained implementation and reductions in heat risk was more limited. Higher-scoring cities were found in every region and city characteristic category, and they showed different combinations of strengths and gaps. Within our sample, a city's region, metropolitan area population, country-level economic resources, and historical climate provided only a partial guide to its documented adaptation efforts. The case studies illustrate how cities can shape their heat adaptation efforts within local constraints through choices about what to prioritize, who leads implementation, and where to direct available resources. Cities in comparable circumstances can draw on these examples to identify practices they could adapt to address their own gaps.

The evaluator rationales highlighted challenges in incorporating uncertainty into long-term strategy, establishing whether measures reduced heat risk, and sustaining municipal capacity beyond individual roles or projects. Together, the benchmark scores, evaluator rationales, and selected city case studies point to several potential lessons for pursuing heat adaptation. These include assigning clear responsibilities and regularly reviewing implementation and outcomes, working with affected communities to tailor measures to local needs, and combining immediate heat protection with longer-term measures to reduce exposure. Making these practices part of

urban heat adaptation efforts can help cities sustain implementation across heat seasons and use evidence on outcomes to decide which measures to maintain, expand, or adjust as risks change.

Several limitations should guide interpretation of these findings (see Methods). Structured expert judgment remains influenced by evaluator interpretation despite standardized materials, paired evaluations, and score reconciliation. Scores also depended on the availability, accessibility, and quality of evidence and may have favored formal, well-documented, and English-language sources. The results therefore represent efforts documented or known to expert evaluators rather than complete inventories of heat adaptation. The city case studies were selected for geographic and contextual diversity and are not statistically representative of cities globally. Finally, equal weighting and aggregation simplify an otherwise complex adaptation process and may obscure differences among individual dimensions of adaptation.

The benchmark results are best understood as comparative indicators of documented strengths, gaps, and uncertainty rather than complete measures or definitive rankings of heat adaptation performance. Together with the accompanying scoring sheets, the benchmark can help cities examine how heat adaptation efforts are planned, organized, implemented, and assessed. Cities can use the results to decide which gaps to address first, particularly in implementation, strategy, monitoring and evaluation, and governance. They can also look to other cities for practices and institutional arrangements they could adapt to their own contexts. Over time, repeated evaluations and benchmarking could track changes in these strengths and gaps as heat risks evolve.

Main text references

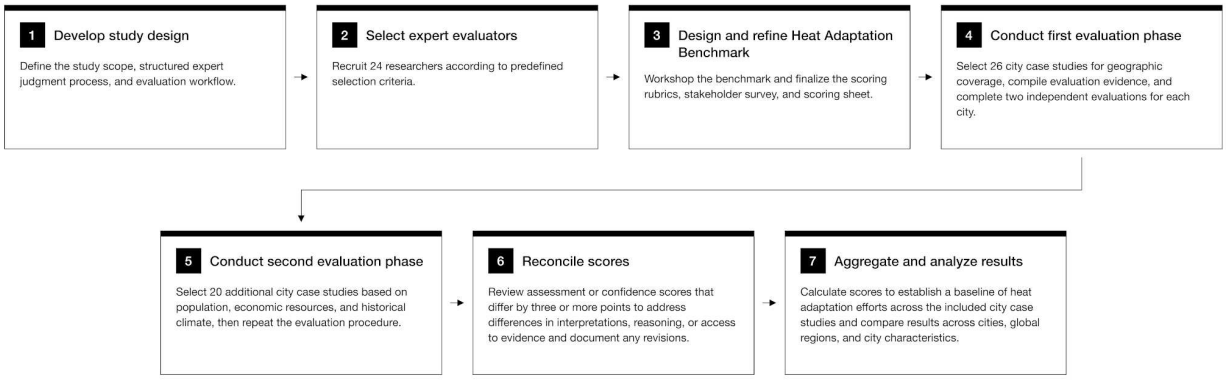
1. Fischer, E. M. *et al.* Record-breaking extremes in a warming climate. *Nat. Rev. Earth Environ.* **6**, 456–470 (2025).
2. Singh, C. *et al.* Applying recent advances in climate adaptation research to urban heat risk management. *WIREs Clim. Change* **15**, e901 (2024).
3. Turek-Hankins, L. L. *et al.* Climate change adaptation to extreme heat: a global systematic review of implemented action. *Oxf. Open Clim. Change* **1**, kgab005 (2021).
4. Ebi, K. L. *et al.* Hot weather and heat extremes: health risks. *The Lancet* **398**, 698–708 (2021).
5. Coffel, E. D. *et al.* The Science of Adaptation to Extreme Heat. in *Resilience* 89–103 (Elsevier, 2018). doi:10.1016/B978-0-12-811891-7.00007-4.
6. Keith, L., Meerow, S., Hondula, D. M., Turner, V. K. & Arnott, J. C. Deploy heat officers, policies and metrics. *Nature* **598**, 29–31 (2021).
7. Gabbe, C. J., Pierce, G., Petermann, E. & Marecek, A. Why and How Do Cities Plan for Extreme Heat? *J. Plan. Educ. Res.* **44**, 1316–1330 (2024).
8. Keith, L., Meerow, S. & Wagner, T. Planning for Extreme Heat: A Review. *J. Extreme Events* **06**, 2050003 (2019).
9. Mahmuda, S. & Webb, R. Climate adaptation and urban planning for heat islands: a case study of the Australian Capital Territory. *Aust. Plan.* **53**, 127–142 (2016).
10. Araos, M. *et al.* Climate change adaptation planning in large cities: A systematic global assessment. *Environ. Sci. Policy* **66**, 375–382 (2016).
11. Morales-Burnett, J. & Marx, R. *The Rise of the Chief Resilience Officer*. https://www.urban.org/sites/default/files/2022-09/Rise%20of%20CROs%20Brief.pdf (2022).
12. Rizwan, A. M., Dennis, L. Y. C. & Liu, C. A review on the generation, determination and mitigation of Urban Heat Island. *J. Environ. Sci.* **20**, 120–128 (2008).
13. Meerow, S. & Keith, L. Cities at the forefront of emerging US heat governance. *One Earth* **7**, 1330–1334 (2024).
14. Stone, B., Lanza, K., Mallen, E., Vargo, J. & Russell, A. Urban Heat Management in Louisville, Kentucky: A Framework for Climate Adaptation Planning. *J. Plan. Educ. Res.* **43**, 346–358 (2023).
15. Aylett, A. Institutionalizing the urban governance of climate change adaptation: Results of an international survey. *Urban Clim.* **14**, 4–16 (2015).
16. United Nations Environment Programme *et al. Adaptation Gap Report 2024: Come Hell and High Water - As Fires and Floods Hit the Poor Hardest, It Is Time for the World to Step up Adaptation Actions*. (United Nations Environment Programme, 2024). doi:10.59117/20.500.11822/46497.
17. Berrang-Ford, L. *et al.* Tracking global climate change adaptation among governments. *Nat. Clim. Change* **9**, 440–449 (2019).
18. United Nations Framework Convention on Climate Change. *Approaches to Reviewing the Overall Progress Made in Achieving the Global Goal on Adaptation: Technical Paper by the Adaptation Committee*. (United Nations, 2024). doi:10.18356/9789210014540.
19. Albers, R. A. W. *et al.* Overview of challenges and achievements in the climate adaptation of cities and in the Climate Proof Cities program. *Build. Environ.* **83**, 1–10 (2015).
20. Magnan, A. K. *et al.* Status of global coastal adaptation. *Nat. Clim. Change* **13**, 1213–1221 (2023).
21. OECD. *Measuring Progress in Adapting to a Changing Climate: Insights from OECD Countries*. (OECD, 2024). doi:10.1787/8cfe45af-en.

22. Magnan, A. K., Li, J., Tanguy, A., Hallegatte, S. & Buffet, C. The value of structured expert judgment to help assess climate adaptation. *Clim. Risk Manag.* **47**, 100692 (2025).
23. Baker, I., Peterson, A., Brown, G. & McAlpine, C. Local government response to the impacts of climate change: An evaluation of local climate adaptation plans. *Landsc. Urban Plan.* **107**, 127–136 (2012).
24. Luber, G. & McGeehin, M. Climate Change and Extreme Heat Events. *Am. J. Prev. Med.* **35**, 429–435 (2008).
25. Araos, M., Ford, J., Berrang-Ford, L., Biesbroek, R. & Moser, S. Climate change adaptation planning for Global South megacities: the case of Dhaka. *J. Environ. Policy Plan.* **19**, 682–696 (2017).
26. Meerow, S. Double exposure, infrastructure planning, and urban climate resilience in coastal megacities: A case study of Manila. *Environ. Plan. Econ. Space* **49**, 2649–2672 (2017).
27. Ahmedabad Municipal Corporation & Natural Resources Defense Council. Ahmedabad Heat Action Plan. (2019).
28. Hess, J. J. *et al.* Building Resilience to Climate Change: Pilot Evaluation of the Impact of India's First Heat Action Plan on All-Cause Mortality. *J. Environ. Public Health* **2018**, 1–8 (2018).
29. Kargbo, E. Freetown Heat Action Plan. (2025).
30. Ville de Paris. Paris Climate Action Plan 2024–2030. (2024).
31. Majszak, M. & Jebeile, J. Expert judgment in climate science: How it is used and how it can be justified. *Stud. Hist. Philos. Sci.* **100**, 32–38 (2023).
32. World Bank. World Bank Country and Lending Groups. *World Bank Data Help Desk* https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.
33. Metreau, E., Young, K. E. & Eapen, S. G. Understanding country income: World Bank Group income classifications for FY26 (July 1, 2025-June, 2026). *World Bank Data Blog* https://blogs.worldbank.org/en/opendata/understanding-country-income--world-bank-group-income-classifica (2025).
34. United Nations, Department of Economic and Social Affairs, Population Division. *World Urbanization Prospects: The 2018 Revision*. https://population.un.org/wup/assets/Publications/WUP2018-Report.pdf (2019).
35. Probable Futures. Climate Zones. https://probablefutures.org/maps/?selected_map=climate_zones&version=latest&view=mercator&scenario=0.5#2.2/0/0.

Methods

Here we describe our methods for systematically evaluating and benchmarking global heat adaptation in cities. Our study design draws on established approaches to measuring and evaluating adaptation[3,10,15,20,22,23,31].

The study proceeded through seven stages: (1) developing the initial study design, (2) selecting a group of expert evaluators ('expert evaluators'), (3) designing and refining the Heat Adaptation Benchmark, (4) conducting the first phase of city case study evaluations, (5) conducting the second phase of city case study evaluations, (6) reconciling scores, and (7) aggregating and analyzing the results (Methods Fig. 1). Following their selection, the expert evaluators participated as members of the broader research team, providing input and feedback that shaped the study's design and implementation.



Methods Fig. 1 | Study design and evaluation workflow. The seven stages cover study design (1), selection of 24 expert evaluators (2), development and refinement of the Heat Adaptation Benchmark (3), two phases of city case study evaluation (4-5), score reconciliation (6), and aggregation and analysis (7). The first evaluation phase included 26 city case studies selected for geographic coverage. The second added 20 city case studies to broaden variation in metropolitan area population, country-level economic resources and historical climate. The analysis compared scores across cities, global regions and city characteristic categories.

Scope

The 'city case study' served as the primary unit of evaluation. Each city case study comprised a specific metropolitan area and the available evidence on its efforts to address chronic heat stress and acute heat shocks. The evaluation of each city case study produced two independently completed scoring sheets, one from each expert evaluator. Each scoring sheet included an assessment score and corresponding rationale for every Heat Adaptation Benchmark assessment question, as well as a confidence score for each benchmark criterion.

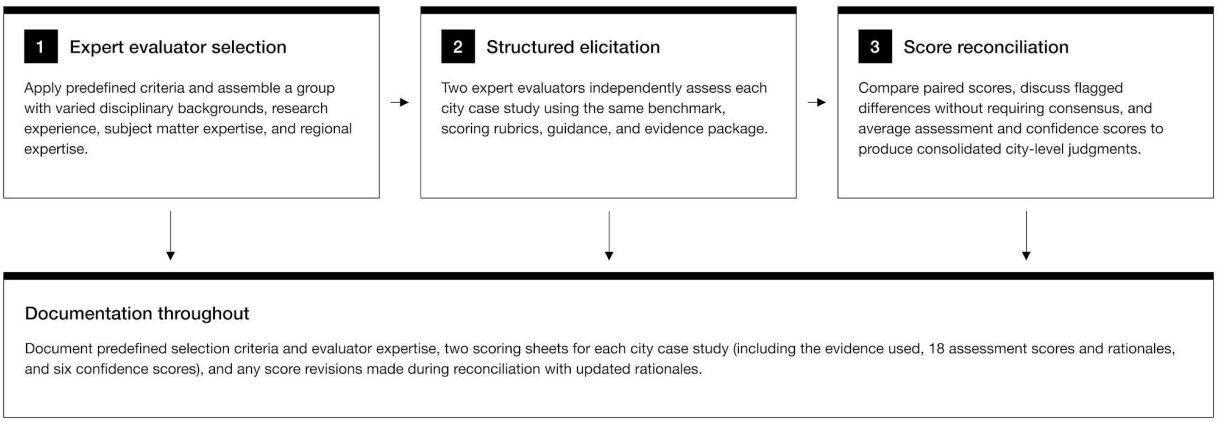
We evaluated deliberate human efforts intended to reduce observed heat impacts or anticipated heat risks to human health, human activity, infrastructure, and other urban systems[20]. These actions or strategies could take any form, such as behavioral or technological adaptations. Evidence of heat adaptation efforts could be drawn from publicly available documents, stakeholder surveys, and expert evaluators' firsthand knowledge of the local context[10].

We included only adaptation efforts explicitly framed as heat adaptation or heat risk reduction. We excluded those that might reduce heat risk indirectly through broader objectives, such as economic development[20]. The assessment focused on actions and strategies undertaken at the local scale, although these could be informed or supported by action at regional, national, or global scales[20].

Structured expert judgment

To evaluate heat adaptation efforts within this scope, we used a structured expert judgment process. Structured expert judgment is a systematic approach to eliciting and synthesizing expert assessments when empirical evidence is limited or uncertain[20,22,31]. Our process was informed by the structured expert judgment approach used by Magnan et al. (2023)[20] to assess coastal adaptation. In this study, structured expert judgment allowed us to integrate multiple forms of evidence, account for contextual diversity, and apply the Heat Adaptation Benchmark where empirical evidence varied in availability and completeness[20,22,31].

The process comprised three components: (1) selecting the expert evaluators, (2) eliciting and documenting their evaluations and scores, and (3) reconciling and aggregating the scores (Methods Fig. 2). Each component was documented, with the two scoring sheets for each city case study capturing the evidence used, an assessment score and rationale for each assessment question, and a confidence score for each benchmark criterion.



Methods Fig. 2 | Structured expert judgment process. Expert evaluators were selected using predefined criteria (1). Two expert evaluators independently assessed each city case study using standardized materials (2). Paired assessment or confidence scores that differed by three or more points were discussed, after which evaluators could revise or retain their scores. Assessment and confidence scores were then averaged separately across the two evaluators (3). Documentation included evaluator selection criteria and expertise, the evidence used, assessment scores and rationales, confidence scores, and any revisions and updated rationales from reconciliation.

Expert evaluators

Expert evaluators played a central role throughout the study and were integral to the structured expert judgment process. In addition to conducting the two phases of city case study evaluations, they contributed to developing and refining the study design, particularly the Heat Adaptation Benchmark. We recruited researchers with expertise in climate risk, adaptation and resilience, and urban systems. To support evaluation across diverse city contexts, we selected researchers with varied disciplinary backgrounds, research experience, subject-matter expertise, and regional expertise.

To be selected, candidates were required to have:

- Doctoral-level research training, either completed or in progress
- Experience applying interdisciplinary perspectives
- Foundational knowledge of heat hazards, exposure, vulnerability, and risk
- Research or professional experience relating to urban systems and environments

We identified potential expert evaluators through professional referrals and searches of academic publications, researcher profiles, and institutional websites. This process resulted in a group of 24 expert evaluators who participated in the structured expert judgment process (Methods Table 1).

Methods Table 1 | Expert evaluators

Expert evaluator	Subject matter expertise	Regional expertise
Marco Billi	Climate governance, climate risk, climate resilience and adaptation, water-food-energy security, metrics, and the science-policy interface	Central and South America, Europe
Christa Brelsford	Cities, urban infrastructure, and geography	North America
Nabig Chaudhry	Climate adaptation, climate risk, and climate services	Asia, Europe, North America
Kate Cullen	Climate governance, glaciology, hydroclimatology, and political economy	Central and South America, North America
Michaela Deglon	Heat adaptation intervention trials and environmental health	Africa
Kristin Fabbe	Comparative politics, migration, and geopolitics	Europe
Spencer Glendon	Climate science communication, climate risk, and climate adaptation	North America
Radley Horton	Climate extremes, climate impacts, and climate adaptation	North America
Toshiaki Ichinose	Urban climatology, energy systems, and urban planning	Asia
John S. Ji	Climate adaptation and health impacts of air pollution, greenness, and temperature extremes	Asia
Andrew Jones	Human and environmental systems	North America
Tarun Khanna	Heat risk and social and economic development	Asia, North America
Cora Kingdon	Climate-related health impacts and urban adaptation planning	North America

Stephen Livesley	Urban climatology, climate impacts, and climate adaptation	Australia and New Zealand
Heleen Mees	Climate adaptation governance, urban sustainable development, and climate justice	Europe
Vuyisile Precious Moyo	Climate adaptation and transformation, heat adaptation, community mobilization, stakeholder engagement, and policy implementation	Africa
Olasunkanmi Habeeb Okunola	Urban planning, disaster risk reduction, and climate adaptation	Africa, Asia, Europe
Sushila Pandit	Climate resilience, anticipatory action, vulnerability reduction, people-centered and locally led climate adaptation, and Indigenous and local knowledge	Asia, Europe
Sebastian Pfautsch	Climate change impacts on cities, heat mitigation, and urban cooling	Australia and New Zealand
Malini Roy	Socially equitable urban planning and extreme and compound climate impacts	Asia, North America
Himanshu Shekhar	Urban well-being, urban planning, and climate adaptation	Africa, Asia, Europe
Pamela Smith	Urban climate, climate change, resilience, and climate justice	Central and South America
Liyang Wang	Climate science, infrastructure planning, uncertainty, and equity	Asia, North America
Saskia Werners	Climate adaptation pathways and urban heat	Africa, Asia, Europe

Notes: Subject matter expertise summarizes each expert evaluator's primary areas of research and professional knowledge relevant to the study. Regional expertise indicates global regions in which evaluators had relevant research, professional, or contextual experience. Expert evaluators were not restricted to assessing cities only within their regional expertise.

City case studies

Two expert evaluators assessed and scored each city case study using the Heat Adaptation Benchmark and accompanying Heat Adaptation Scoring Rubric. We selected the 46 city case studies in two phases. City case studies were intentionally selected to maximize geographic and contextual diversity rather than to produce a statistically representative sample of cities.

In the first phase, we selected 26 city case studies to achieve geographic coverage, with at least three cities from each of six global regions: Africa, Asia, Australia and New Zealand, Europe, Central and South America, and North America. We refined the final sample for the first phase based on input from the expert evaluators.

In the second phase, we selected 20 additional city case studies to broaden variation across three city characteristics: metropolitan area population, country-level economic resources, and historical climate (Methods Table 2). These characteristics represent contextual attributes that may shape how metropolitan areas approach heat adaptation. Across the complete sample of 46 city case studies, the final sample included at least ten cities in each category of the three characteristics.

Methods Table 2 | City characteristics

City characteristic	Definition	Characteristic categories
Metropolitan area population	The population of the metropolitan area represented by the city case study, based on the most recent available estimate. When metropolitan area population data were unavailable, we used data for the metropolitan region, metropolis, municipality, or city proper (in that order of preference). We used this measure as a proxy for the demographic scale of the city.	**Medium:** Metropolitan area population < 5,000,000
		Large: Metropolitan area population ≥ 5,000,000 and < 10,000,000
		Mega: Metropolitan area population ≥ 10,000,000
Country-level economic resources	The country's 2024 gross national income (GNI) per capita, calculated using the World Bank Atlas method (in current US dollars)[32,33]. We used this measure as a proxy for the financial and economic capacity of the city.	**Low-income:** GNI per capita < $4,496 (merges World Bank low- and lower middle-income groups)
		Middle-income: GNI per capita ≥ $4,496 and < $13,936 (World Bank upper middle-income)
		High-income: GNI per capita ≥ $13,936 (World Bank high-income group)
Historical climate	The historical Köppen-Geiger climate classification of the city at approximately 0.5 °C of global warming relative to 1850-1900. We used this classification as a broad indicator of the city's historical climatic conditions.	**Hot:** Köppen-Geiger class at ~0.5 °C baseline in A (Tropical), hot-dry B (BWh, BSh), hot-summer Temperate C (Csa, Cfa, Cwa), or hot-summer Continental D (Dsa, Dfa, Dwa).
		Temperate/Cool: Köppen-Geiger class at ~0.5 °C baseline in cold-dry B (BWk, BSk), Temperate C (non-hot-summer) (Csb, Csc, Cfb, Cfc, Cwb, Cwc), Continental D (non-hot-summer) (Dsb, Dsc, Dsd, Dfb, Dfc, Dfd, Dwb, Dwc, Dwd), or E (Polar).

Note: Metropolitan area population thresholds of 5 million and 10 million are based on the United Nations Department of Economic and Social Affairs classification of large cities and megacities, respectively[34]. We grouped all city case studies below 5 million into a single medium category. Country-level economic resource thresholds are based on the World Bank FY2026 income thresholds, which used 2024 GNI per capita information[32,33]. Historical climate classification was based on Probable Futures data representing 1971-2000, when global mean surface temperature was approximately 0.5 °C above the 1850-1900 baseline[35].

Together, the two phases produced a sample of 46 city case studies spanning diverse geographic, demographic, economic, and climatic contexts (Methods Table 3).

Methods Table 3 | City case studies

City	Study phase	Global region	Metropolitan area population category	Country-level economic resources category	Historical climate category
Addis Ababa, Ethiopia	Phase 2	Africa	Large	Low-income	Temperate/cool
Cairo, Egypt	Phase 1	Africa	Mega	Low-income	Hot
Cape Town, South Africa	Phase 1	Africa	Medium	Middle-income	Temperate/cool
Fez, Morocco	Phase 2	Africa	Medium	Low-income	Hot
Freetown, Sierra Leone	Phase 1	Africa	Medium	Low-income	Hot
Kisumu, Kenya	Phase 2	Africa	Medium	Low-income	Hot
Lagos, Nigeria	Phase 2	Africa	Mega	Low-income	Hot
Lusaka, Zambia	Phase 1	Africa	Medium	Low-income	Hot
Windhoek, Namibia	Phase 1	Africa	Medium	Low-income	Hot
Ahmedabad, India	Phase 1	Asia	Large	Low-income	Hot
Amman, Jordan	Phase 2	Asia	Large	Low-income	Temperate/cool
Bangkok, Thailand	Phase 2	Asia	Mega	Middle-income	Hot
Beijing, China	Phase 1	Asia	Mega	Middle-income	Hot
Chongqing, China	Phase 1	Asia	Mega	Middle-income	Hot
Colombo, Sri Lanka	Phase 2	Asia	Large	Low-income	Hot
Dhaka, Bangladesh	Phase 1	Asia	Mega	Low-income	Hot
Dubai, UAE	Phase 2	Asia	Large	High-income	Hot
Hong Kong, China	Phase 2	Asia	Large	Middle-income	Hot
Jakarta, Indonesia	Phase 2	Asia	Mega	Middle-income	Hot
Jodhpur, India	Phase 2	Asia	Medium	Low-income	Hot
Karachi, Pakistan	Phase 2	Asia	Mega	Low-income	Hot
Kolkata, India	Phase 2	Asia	Mega	Low-income	Hot
Kuwait City, Kuwait	Phase 1	Asia	Large	High-income	Hot
Tokyo, Japan	Phase 1	Asia	Mega	High-income	Hot
Auckland, New Zealand	Phase 1	Australia and New Zealand	Medium	High-income	Temperate/cool
Melbourne, Australia	Phase 1	Australia and New Zealand	Large	High-income	Temperate/cool
Sydney, Australia	Phase 1	Australia and New Zealand	Large	High-income	Hot
Buenos Aires, Argentina	Phase 1	Central and South America	Mega	Middle-income	Hot

Medellín, Colombia	Phase 2	Central and South America	Medium	Middle-income	Temperate/cool
Mexico City, Mexico	Phase 2	Central and South America	Mega	Middle-income	Temperate/cool
Monterrey, Mexico	Phase 1	Central and South America	Large	Middle-income	Hot
Montevideo, Uruguay	Phase 1	Central and South America	Medium	High-income	Hot
Santiago, Chile	Phase 1	Central and South America	Large	High-income	Temperate/cool
São Paulo, Brazil	Phase 2	Central and South America	Mega	Middle-income	Hot
Amsterdam, Netherlands	Phase 1	Europe	Medium	High-income	Temperate/cool
Athens, Greece	Phase 1	Europe	Medium	High-income	Hot
Barcelona, Spain	Phase 1	Europe	Large	High-income	Hot
Istanbul, Turkey	Phase 2	Europe	Mega	Middle-income	Hot
Milan, Italy	Phase 2	Europe	Large	High-income	Hot
Paris, France	Phase 1	Europe	Mega	High-income	Temperate/cool
Baltimore, USA	Phase 1	North America	Medium	High-income	Hot
Miami, USA	Phase 1	North America	Large	High-income	Hot
Philadelphia, USA	Phase 2	North America	Large	High-income	Hot
Phoenix, USA	Phase 1	North America	Medium	High-income	Hot
Sacramento, USA	Phase 1	North America	Medium	High-income	Hot
Toronto, Canada	Phase 2	North America	Large	High-income	Temperate/cool

Note: City names refer to the metropolitan areas evaluated as city case studies. For the regional comparisons, the Central and South America group includes Mexico City and Monterrey. The North America group comprises the case studies in Canada and the United States. For the income groups, we retained Ethiopia's FY2025 low-income classification for Addis Ababa (Ethiopia was temporarily unclassified in FY2026) and used China's 2024 Atlas GNI per capita to classify Hong Kong. Metropolitan area population, country-level economic resources, and historical climate categories are defined in Methods Table 2.

Benchmark development and operationalization

We developed the Heat Adaptation Benchmark by adapting concepts from established approaches to evaluating adaptation and reviewing additional research on heat risk and adaptation. This process was used to identify dimensions of adaptation that could be assessed consistently across cities while capturing activities ranging from risk assessment and planning to implementation, governance, and outcomes.

The draft framework underwent iterative review and refinement with the expert evaluators. Evaluator feedback was used to clarify the scope and wording of assessment questions, reduce overlap between questions, standardize terminology, and improve applicability across different

geographic and institutional contexts. The framework was finalized before the city case study evaluations began.

We then further operationalized each assessment question into question-specific guidance in the Heat Adaptation Scoring Rubric (see Supplementary Information). For each question, the rubric specifies what was assessed, the geographic or administrative scale, the evidence relevant to scoring, and the criteria for each score level. We also developed a standardized scoring sheet that required evaluators to document the evidence used, their assessment score and rationale for each question, and their confidence score for each benchmark criterion. These materials were intended to support consistent independent evaluations and provide a transparent record of how each judgment was reached.

Evaluation evidence

Expert evaluators could draw on three sources of evidence when assessing city case studies using the Heat Adaptation Benchmark and Heat Adaptation Scoring Rubric: publicly available documents, stakeholder survey responses, and firsthand knowledge of the local context. Expert evaluators documented the evidence sources used in each independently completed scoring sheet.

Evaluation evidence identified, submitted, or updated between July 1, 2025, and August 21, 2026, including during reconciliation, was incorporated into the final assessments and scores. Evidence made available after this period was excluded.

We conducted targeted rather than systematic searches for publicly available evidence relevant to each city case study and the Heat Adaptation Benchmark. These searches were intended to assemble evidence for the evaluations rather than produce an exhaustive inventory of all available documents.

Publicly available documents included peer-reviewed and gray literature, reports, plans, policies, and other written materials related to adaptation. We defined publicly available documents as materials accessible to the general public, whether freely online, through a formal request (for example, public records or information access procedure), or through physical access at an archive or library. Materials that were private, confidential, or restricted in use were excluded.

Stakeholder surveys elicited information from local government representatives and practitioners (‘stakeholders’) with relevant knowledge or experience in heat adaptation within the city being evaluated. The survey instrument was designed to correspond to the evidence needed to apply the Heat Adaptation Benchmark and Heat Adaptation Scoring Rubric (Supplementary Table 1).

Stakeholders were identified through publicly available sources and professional networks and invited to participate by email. The stakeholder survey was reviewed and approved by the Committee for Protection of Human Subjects at the University of California, Berkeley (protocol no. 2025-04-18570). Participation was voluntary, and informed consent was obtained through the

recruitment email. Survey responses were de-identified and summarized before being provided to the expert evaluators assessing the relevant city case study.

We contacted 68 stakeholders, including at least one for each city case study. Seven stakeholders responded and completed the survey, yielding a response rate of 10.3%. The absence of a stakeholder survey response was treated as an unavailable source of evidence, rather than as evidence of limited heat adaptation. When the overall evidence available for a city case study was limited, however, evaluators could assign lower confidence scores to the affected benchmark criteria.

Expert evaluators could also draw on firsthand knowledge when they had relevant academic, research, or professional experience with a city and its heat adaptation efforts. The ability to incorporate such contextual knowledge is a benefit of structured expert judgment, which can synthesize multiple forms of information[22,31]. When evaluators used firsthand knowledge, they identified it explicitly as evidence in their assessment.

Degree of confidence

Based on the evidence available for each city case study, each expert evaluator assigned one confidence score to each Heat Adaptation Benchmark criterion. The confidence score represented the evaluator's confidence in the assessment scores assigned within that criterion.

The Degree of Confidence Scoring Rubric defines a five-point scale ranging from 0 (no confidence) to 4 (very high confidence) (Methods Table 4). The scores reflect the quantity and quality of available evidence and the remaining uncertainty in the expert evaluator's judgment. The Degree of Confidence Scoring Rubric was informed by Magnan et al. (2023)[20].

Methods Table 4 | Degree of Confidence Scoring Rubric

Confidence score	Confidence level	Scoring criteria
0	No confidence	No supporting evidence is available, or evidence is entirely unreliable
1	Low confidence	Limited or low-quality evidence, with significant gaps or uncertainties
2	Moderate confidence	A moderate amount of evidence of varying quality, with some gaps or uncertainties
3	High confidence	A substantial body of reliable evidence supports the judgment, with minor uncertainties
4	Very high confidence	A comprehensive and high-quality evidence base strongly supports the judgment, with minimal uncertainty

Note: Confidence scores were assigned at the benchmark criterion level and recorded separately from assessment scores. They indicate confidence in the assessment scores within the associated criterion and were not used to weight or adjust assessment scores.

Evaluation procedure

After the expert evaluator group had been assembled and the evaluation materials finalized, we conducted the city case study evaluations in two phases. For each city case study, we assigned two expert evaluators to independently assess the city using the available evaluation evidence. This paired evaluation design supported consistency between expert evaluators, enabled comparison and aggregation of independently assigned scores, and reduced the influence of any single expert evaluator's judgment on the results. The first phase of city case study evaluations began on July 8, 2025, and the second phase began on November 14, 2025. Score reconciliation began on July 12, 2026, and all evaluations and reconciliation were completed by August 21, 2026.

Assignments for city case study assessment prioritized expert evaluators who had indicated an interest in or experience with the city, followed by those with expertise in the global region where the city was located. When a sufficient number of evaluators meeting these considerations was unavailable, we assigned evaluators outside their areas of regional expertise. Expert evaluators assessed an average of about four city case studies each, with individual totals ranging from two to six, reflecting differences in availability and participation. The standardized benchmark, scoring rubrics, and supporting materials were designed to enable any expert evaluator meeting the study's selection criteria to complete the assessment.

We provided expert evaluators with detailed guidance and responded to questions about the evaluation procedure and terminology. For each city case study, we compiled an evaluation package containing the Heat Adaptation Benchmark, Heat Adaptation Scoring Rubric, Degree of Confidence Scoring Rubric, and available initial evaluation evidence (publicly available documents and any stakeholder survey responses). Evaluators could also consider additional publicly available evidence and their firsthand knowledge of the local context.

Each expert evaluator completed a scoring sheet for every city case study assigned to them. The scoring sheet documented the evaluation evidence used, the assessment score and corresponding rationale for each assessment question, and the confidence score for each benchmark criterion. The 46 city case studies therefore produced 92 independently completed scoring sheets. Each scoring sheet contained 18 assessment scores and six confidence scores, with both types of scores assigned using their respective 0-4 point scales. In the scoring sheets made available with the publication, evaluator names and other direct identifiers were removed, and evaluators were identified only as 'Expert Evaluator 1' or 'Expert Evaluator 2.'

Score reconciliation

After the two assigned expert evaluators had independently completed their scoring sheets for a city case study, we compared the two sheets to determine whether score reconciliation was required. A city case study was flagged for reconciliation if either of the following conditions was met:

1. The expert evaluators' assessment scores differed by three or more points on the 0-4 point scale for any assessment question.

2. The expert evaluators' confidence scores differed by three or more points on the 0-4 point scale for any benchmark criterion.

For each flagged city case study, we notified the evaluators of the specific assessment or confidence scores that triggered reconciliation and provided an opportunity for them to discuss their scoring. This process allowed the evaluators to clarify their interpretations of the available evaluation evidence, identify overlooked evidence or information, and confirm that both had access to the same evidence.

Following the discussion, each evaluator could revise or retain their original scores. Reconciliation did not require the evaluators to reach consensus, and scores could remain different when they reflected reasonable differences in interpretation rather than overlooked reasoning or evidence. Scores were revised only when the discussion produced new information or insights that changed an expert evaluator's judgment. Following reconciliation, evaluators recorded any changes directly in the corresponding scoring sheet by updating the relevant scores and written judgments. Evaluators could also update their written judgments to clarify their assessments without changing their scores.

Score aggregation and analysis

The Heat Adaptation Benchmark contains 18 assessment questions, each scored on a 0-4 point scale, allowing for a maximum possible score of 72. Each of the six benchmark criteria contains three assessment questions and thus has a maximum possible score of 12.

Following score reconciliation, we averaged the two expert evaluators' assessment scores for each question within each city case study. This value represented the consolidated judgment of the two expert evaluators.

For each city case study, we summed the three mean assessment scores within each benchmark criterion to calculate a benchmark criterion score reported on a 0-12 point scale. We then summed all 18 mean assessment scores to calculate a total heat adaptation benchmark score reported on a 0-72 point scale. We also averaged the two expert evaluators' confidence scores for each benchmark criterion. For grouped results, we averaged these city-level confidence scores across the included city case studies.

For grouped results, we calculated the mean assessment score for each of the 18 questions across the city case studies in the group. We then summed these 18 mean assessment scores to calculate a group-level total heat adaptation benchmark score reported on a 0-72 point scale.

We divided the possible range of total heat adaptation benchmark scores into four equal score categories:

- **Low:** 0 ≤ total benchmark score ≤ 18
- **Moderate:** 18 < total benchmark score ≤ 36

- **High:** 36 < total benchmark score ≤ 54
- **Very high:** 54 < total benchmark score ≤ 72

These categories were intended to describe and compare scores, not to define established levels of adaptation effectiveness.

We weighted expert evaluators, assessment questions, benchmark criteria, and city case studies equally. Expert evaluator scores therefore contributed equally to mean assessment scores, assessment questions contributed equally to benchmark criterion scores and total heat adaptation benchmark scores, and city case studies contributed equally to grouped results. Confidence scores were not included in these calculations and were not used to weight or adjust assessment scores or total benchmark scores.

To complement the quantitative results, we grouped recurring observations from evaluator rationales into themes and identified illustrative, nonexclusive profiles among high- and very high-scoring city case studies based on relative criterion strengths. We did not estimate theme prevalence, test group differences, or use formal statistical grouping methods. Selected city case studies were also used to illustrate practices relevant to the observed strengths and gaps.

Group-level comparisons across global regions and city characteristic categories were treated as descriptive given the intentional selection of city case studies and unequal group sizes. We did not conduct inferential statistical tests.

Limitations

Several methodological limitations should be considered when interpreting the results. Structured expert judgment incorporates the perspectives, experience, and interpretations of participating expert evaluators[22,31]. Standardized materials, paired independent evaluations, documented rationales, and score reconciliation supported consistency but did not eliminate subjectivity[20,22,31]. Different expert evaluator groups or city assignments could produce different scores, potentially limiting reproducibility[22].

The evaluations also depended on the availability, accessibility, and quality of evidence. Public documentation varied across cities and may have favored formal, well-documented initiatives over informal or community-led efforts. Public-facing materials may also have simplified technical details or complex heat adaptation considerations to improve accessibility, potentially affecting how evaluators interpreted the evidence and assigned assessment and confidence scores.

Although documents were reviewed in English or another language accessible to the expert evaluator and translated when needed, evidence searches generally favored English-language materials, potentially advantaging cities with more extensive English-language documentation in both assessment and confidence scores. The low stakeholder survey response rate also limited the representation of stakeholder perspectives. The absence of a survey response was not treated as evidence of limited heat adaptation, but it reduced the information available to evaluators.

Evidence also varied across geographic and administrative scales, with some city case studies drawing on a mixture of local, regional, and national plans. Determining how directly higher-scale plans informed or supported local heat adaptation could be difficult and may have introduced inconsistency across assessments. Uneven evidence availability and scale may therefore have affected what evaluators identified and their confidence in the assessments. The results thus represent efforts documented or known at the time of evaluation rather than a complete inventory of heat adaptation efforts.

The city case studies were intentionally selected for geographic and contextual diversity and are not statistically representative of cities globally. Grouped results should be interpreted as descriptive comparisons within the study sample. In addition, the three city characteristics provide simplified proxies for complex urban conditions, and their broad categories may obscure variation among cities.

Finally, the Heat Adaptation Benchmark is broad but not exhaustive. Equal weighting assumes that all assessment questions and benchmark criteria contribute equally to the total heat adaptation benchmark score, although their relative importance may vary across city contexts. Scores were assigned against question-specific rubric requirements and were not adjusted for differences in implementation difficulty, resource requirements, or the time needed to establish outcomes.

Aggregation may also obscure differences across individual dimensions of adaptation. Confidence scores indicated the strength of evidence supporting the assessments but were not used to weight assessment scores, so scores associated with lower and higher confidence contributed equally to aggregated results. The resulting scores should therefore be interpreted as comparative summaries of the documented evidence, alongside their confidence scores, rather than as definitive measures of city performance.

Data availability

Final assessment and confidence scores, city- and group-level summaries, and post-reconciliation score differences are provided in Supplementary Data 1. Plotting data and summary statistics for the bar charts and boxplots are provided in Supplementary Data 2. The final de-identified expert evaluator scoring sheets, including scoring rationales and evidence considered, are provided in the Supplementary Information.

Code availability

No custom code was used in this study.

Acknowledgements

We thank the local government representatives and practitioners who contributed information through the stakeholder survey.

Author contributions

N.A.C. conceived and led the study, including the study design, research coordination, compilation of initial evaluation evidence, and development of the Heat Adaptation Benchmark. All authors other than D.A. served as expert evaluators and conducted city case study evaluations. All authors provided feedback on the study methods and benchmark design. D.A. and A.D.J. reviewed the initial results. N.A.C. drafted the manuscript. All authors reviewed, commented on, and revised the manuscript.

Additional information

The authors declare no competing interests. Correspondence should be addressed to N. A. Chaudhry.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementarydata1globalbenchmarkonheatadapt.xlsx
- supplementarydata2globalbenchmarkonheatadapt.xlsx
- supplementaryinformationglobalbenchmarkonheatadapt.pdf