BRC Baekrokdam Research Commons
백록담 리서치 커먼스
Independent Open Research Journal

RESEARCH PROTOCOL BRC-2026-W103 Published 2026-08-15 Creative Commons Attribution 4.0 International

A prospective observation protocol for recurrent localized back heat

Yeonseung Choe [1] **Baekrokdam Research Commons (BRC)** [1]

[1] Baekrokdam Research Commons (BRC), Republic of Korea

Article type	Submitted	Published	Persistent ID
Research protocol	2026-08-15	2026-08-15	BRC-2026-W103
Last updated	**Review status**	**License**	**Correspondence**
2026-08-15	Awaiting public assessment	Creative Commons Attribution 4.0 International	Baekrokdam Research Commons (BRC)

Canonical URL https://research.baekrokdam.com/en/articles/localized-back-heat-observation-protocol

Suggested citation Choe Y, Baekrokdam Research Commons (BRC). A prospective observation protocol for recurrent localized back heat. Baekrokdam Research Commons. 2026; BRC-2026-W103.

Abstract

This protocol specifies a small prospective feasibility study of recurrent heat, warmth, burning, or adjacent thermal expressions localized to the back. Participants' original wording and drawn boundaries are preserved before investigators apply a prespecified eligibility dictionary. Event, symptom-free comparator, and standardized-visit clocks are separated; sensation maps, skin-surface temperature, evoked thermoception, functional impact, and context remain non-interchangeable observation lanes. The study does not validate a diagnostic tool, and it treats safety exits and missingness as reportable outcomes.

Keywords: observational protocol · event capture · sensory mapping · skin temperature · missing data

Why an observation protocol is needed

The evidence gap is not a missing diagnostic label. In the frozen and verified source set, no study prospectively recruited people because they experienced recurrent localized back heat and then aligned a spontaneous episode with the participant's mapped wording and depth, visible or palpable skin findings, controlled back-surface temperature, evoked thermoception, associated findings, and episode context. Existing sources address parts of that observation problem in itch-, pain-, healthy-volunteer, provocation, or safety-defined settings. Their methods cannot be joined after the fact into a participant-level episode record.

The proposed response is therefore a prospective, repeated within-person observation protocol. Its purpose is to learn whether separate observations can be captured at the same anatomical site and relevant time, how often each component is completed, and which within-person patterns are reproducible enough to justify later study. It does not make an objective measure the judge of a reported sensation, presume that the measures will agree, or convert a feasibility-stage observation set into a questionnaire, diagnostic test, mechanism study, referral tool, or care pathway. No recruitment, validation, registration, data collection, or ethics approval is represented by this note.

Who and what counts as an episode

The target population is adults who prospectively report recurrent heat, warmth, burning, or an adjacent thermal expression whose participant-drawn boundary includes the back and is distinguishable from a whole-body-only hot or cold episode. Before recruitment, investigators must version an eligibility dictionary and adjudication rule for adjacent expressions. It must name the deciding roles, restrict their decision to the participant's preserved original wording, localization, and other source text specified in advance, define how uncertain entries are retained or excluded, and route whole-body-only reports out of this phenotype. The decision is made before any surface measurement, QST result, or other objective outcome is viewed. The dictionary is investigator-selected and unvalidated: it preserves original wording and does not force equivalence across Korean, English, or other languages. Recruitment must use a named, ethics-approved route that does not draw from private patient records. Setting, invitations, eligibility, exclusions, enrollment, and retention require separate accounting. A prespecified run-in should confirm at least two prospectively recorded episodes and whether event capture is feasible; its duration and final event requirement are pilot parameters that must be fixed before recruitment rather than selected after outcomes are seen.

The primary unit is a participant-reported episode, not a diagnosis, temperature value, or researcher-assigned syndrome. At each qualifying report, the participant first records their own Korean, English, or other wording and marks a body map with boundary, spinal or back level, laterality, area, and perceived depth. Researchers may later map wording to prespecified analytic categories, but they must preserve the original wording and must not treat warmth, burning, itch, pain, measured skin temperature, or an evoked thermal response as equivalent constructs.

Eligibility does not require visible change, palpable warmth, elevated surface temperature, an abnormal evoked test, or a presumed diagnosis. Requiring one of those findings would build the desired association into the sample. Generalized thermal symptoms, posture, movement, clothing, ambient heat, exertion, sleep, touch, itch, pain, numbness, tingling, relief, skin change, and known diagnoses or medicines may be recorded as consented context; none is silently promoted to the cause of the episode. A participant who needs clinical assessment under the later safety procedure is deferred or paused and is not cleared by study measurements.

Align the clock, map, and measurement

The protocol uses three clocks that must not be collapsed: the participant-reported episode clock, the scheduled non-episode reference clock, and the end-of-observation accounting clock. At an episode, the participant first time-stamps perceived onset and records current wording, intensity, unpleasantness if used, mapped boundary, level, laterality, area, perceived depth, and context. The record must also state the delay from onset to the subjective entry and from that entry to any surface observation. These aligned times permit comparison; they do not presume that felt heat and skin-surface temperature are the same construct or will agree. [1]

A surface observation is attempted only when it can be made safely under the prespecified device, positioning, clothing-exposure, acclimation, anatomical-marker, region-of-interest, and environmental conditions. Each observation records the device and calibration status, relevant accuracy or sensitivity specifications, emissivity where applicable, distance, angle, assessor role, room temperature, humidity, airflow, recent sunlight or heat, exertion, food or drink, relevant medicines or substances, sweating, time of day, and elapsed time from onset. The raw native left-right orientation and participant-drawn region are retained. These requirements govern comparability and auditability; they are not a claim that the cited exercise-provocation procedure is a universal protocol for spontaneous episodes. [1]

Surface temperature is reported as surface temperature, not deep tissue temperature. For every valid observation, the protocol keeps three quantities separate: the absolute surface value at the mapped region; change from that participant's prespecified comparable reference; and a prespecified matched side or region contrast. A relative side or region difference may coexist with cooling from baseline, and the contralateral side may share bilateral or systemic features. No comparison is therefore renamed a healthy control, and no absolute, reference-change, or relative contrast may be substituted for another after the data are seen. [1]

The scheduled reference window repeats the same map, subjective fields, surface procedure, and metadata during a participant-defined non-episode state at a comparable clock time and under comparable controlled conditions where feasible. Before recruitment, the protocol must name whether each primary comparison is within-region episode versus reference, mapped versus contralateral, mapped versus another control region, or external reference. A non-episode record is not an unobserved episode peak, a contralateral site is not automatically unaffected, and external data are not internally recruited controls.

Prespecified short-interval repeats may describe the observed trajectory only when feasible and non-burdensome. Every repeat keeps its actual time relative to onset. A late observation is retained as late, never backfilled or presented as the episode peak; a device failure, unsafe capture, incomplete control condition, missed event, or absent map is reason-coded missing data rather than a normal or negative result. At the end of observation, participant reports, device records, timing, comparison labels, and failure reasons are reconciled without inferring an unreported sensation from a device trace or a surface result from the participant's wording. [1]

Evoked testing is a separate experiment

Quantitative sensory testing is optional and, if retained after ethics, safety, device, and burden review, is performed during a supervised research visit rather than as an instrumented replay of a spontaneous episode. Before data collection, the module must fix the tested modalities, device and safety limits, instructions, familiarization, stimulus and response rules, trial handling, site order, assessor role, environmental conditions, stopping criteria, and analysis outputs. These are investigator-selected protocol parameters requiring piloting; the verified sources do not supply a validated QST battery for recurrent localized back heat. [2, 3, 4]

Testing is anchored to the participant-drawn symptomatic site and to one prespecified anatomically matched comparison strategy. A mirror contralateral site offers a within-person side contrast but is not automatically unaffected or equivalent to a healthy control. A different control region asks a regional question, and published reference data ask a population-reference question whose age, sex, trunk site, modality, protocol, and sample composition must be compatible and reported. These comparisons may be secondary to one another, but they cannot be silently substituted or pooled after results are known. [2, 3, 4]

Every QST output remains modality- and site-specific and is labeled as an evoked detection, pain, or other prespecified response. It is not a measure of skin-surface temperature, perceived depth, or the occurrence, intensity, or truth of a spontaneous heat or burning episode. Whether an episode is present at the visit may be time-stamped as context, but the evoked result is analyzed in its own table before any exploratory cross-construct comparison. No threshold is called diagnostic, normative for this phenotype, or mechanistic. [2, 3, 4]

This draft does not prescribe a paradoxical-heat procedure. The main article for PMID 38835743 is verified, but its linked supplement was not lawfully acquired; the exact stimulus sequence, trial count, response definition, and scoring are therefore unavailable and must not be reconstructed from the reported results. Paradoxical responses may enter a later protocol only through a separately specified, lawfully verified and independently reviewed method or a versioned amendment. The published table's footnoted classification ambiguity is retained as reported and is not silently recalculated. [3]

The analysis must retain null, bilateral, and asymmetric patterns. In the verified 15-person itch-defined pilot, the symptomatic and contralateral sites did not differ on the reported sensory measures except mechanical-evoked itch, while several deviations from published reference data appeared on both sides; paradoxical responses were also reported on both sides. These findings justify preserving all comparison patterns, not predicting the outcome for this different target population. A null side difference does not prove equivalence or negate the report, and a bilateral or abnormal reference comparison does not identify a cause. [3]

Missingness and repeatability are outcomes

Event capture can fail at several distinct steps, and each failure is an observed feasibility outcome rather than a blank to erase. Every expected episode, reference window, repeat, surface procedure, and optional QST visit is assigned one prespecified status and, when incomplete, a reason: the episode was not noticed in time or was too brief; the participant was unavailable or asleep; measurement was unsafe or too burdensome; a safety stop or clinical redirect occurred; the device was unavailable or failed; environmental or acclimation conditions were outside protocol; a map or symptom field was incomplete; a surface measurement was absent; QST was not performed; follow-up ended; or the reason was unknown. Multiple reasons may be retained in a fixed hierarchy or as non-exclusive fields, but the choice must be frozen before analysis.

Completion is reported with component-specific numerators and denominators. At minimum, the flow distinguishes enrolled participants, participants whose recurrence was prospectively confirmed, reported qualifying episodes, episodes entered within each prespecified time window, episodes with complete subjective maps, attempted and valid surface observations, valid paired comparisons, completed QST visits if that module is used, and safety-paused observations. A participant or episode can therefore contribute to one component without being counted as complete for another. The report also shows the distribution of delays and reason codes by participant and episode context, without treating differential completion as biological evidence.

The primary feasibility and descriptive accounts use the components actually observed, with their explicit denominators. A missing diary does not mean that no sensation occurred; a missing, technically normal, or invalid surface record does not mean that the episode was absent or reassuring; and a later observation is not carried backward as an episode peak. The existence, wording, quality, map, or location of an unreported episode is never imputed. Any defensible model-based imputation proposed later is confined to a labeled sensitivity analysis under the prespecified analysis plan and cannot replace the observed-data account.

Procedural repeatability is assessed only in prespecified short-interval replicates made under nominally stable conditions: the same anatomical markers, positioning, device settings, environment, acclimation, assessor instructions, and region-of-interest rules. The protocol may add blinded duplicate region extraction for a prespecified subset and may repeat map or quality coding without displaying the prior analytic code. For continuous measures, reporting uses a within-participant measurement-error estimate and agreement interval appropriate to the scale rather than correlation alone. For categorical maps or codes, it reports raw agreement and a chance-adjusted measure only when that measure's assumptions are met. [1, 2]

Stable-condition repeatability is not episode recurrence. Change within an episode and heterogeneity between episodes remain biological and contextual descriptions unless a stable-condition design shows a procedural source of variation. Episode-to-episode variation is therefore reported separately and is not renamed poor instrument reliability. No smallest detectable, clinically meaningful, abnormal, or diagnostic difference is declared for this proposed phenotype because those measurement properties have not been established. [1, 2]

Burden is accounted for alongside completion: event prompts and attempts, time from report to each task, time spent on episode and reference procedures, short-interval repeats, interrupted sleep or activity where consented, procedures declined or stopped, device or travel demands, safety pauses, withdrawal, and free-text burden feedback are summarized without placing submission text or other sensitive content in analytics events. A schedule that produces selective late capture, repeated technical failure, excessive burden, or withdrawal is a feasibility result that can require redesign; it is not repaired by retaining only complete episodes.

Analysis without a hidden diagnosis

The analysis has nested units. Participants are the independent sampling units; prospectively reported episodes are repeated observations within participants; and time points, regions, sides, map elements, and procedures are observations within an episode or reference period. Every summary states which unit supplies its numerator and denominator. Models or interval calculations must account for within-participant dependence, and an episode is never analyzed as though it were a different person. If the pilot is too small or sparse for a defensible repeated-measures model, participant-level distributions and trajectories remain the primary report rather than a falsely precise pooled estimate.

The feasibility pilot estimates event rate, prospective recurrence-confirmation yield, capture delay, complete-map yield, valid controlled surface-observation yield, valid-pair yield, within-person variation, burden, retention, safety pauses, and attrition with component-specific denominators and interval estimates where the data support them. These parameters answer whether a later study is practicable and how it should be sized; they do not establish prevalence, a normal range, a clinically meaningful threshold, or a diagnostic property. No success threshold or sample size is invented in this note. A later study must justify its size from one frozen primary estimand, its precision target, anticipated within-participant dependence, missingness, and expected number of usable episodes.

Before outcome analysis, the protocol identifies one primary anatomical comparison and one primary episode time point or window. Skin-surface observations are then reported in separate estimands: the absolute surface value at that time, change from the participant's prespecified non-episode reference, and the paired difference from the prespecified side or region comparator. Their directions need not agree, and none substitutes for deep temperature. Contralateral, regional, and non-episode references answer different questions; the pilot may compare their availability, but a confirmatory analysis must select rather than retrospectively promote whichever comparison gives the largest contrast.

Subjective episode records, mapped boundaries, visible or palpable findings, surface temperature, optional QST, and safety variables are first summarized in separate tables. Their matched-time joint account uses neutral prespecified observation categories: both the subjective map and a technically valid surface observation available within the frozen time window; both available but outside that window; subjective record available with the surface observation unavailable or technically invalid; or surface observation available with the subjective map incomplete. Within the first two categories, the underlying participant-drawn map, spatial relation, timing, and continuous absolute, reference-change, and side-or-region surface values are shown rather than collapsed into concordant or discordant labels. Every temporal, spatial, direction, and temperature rule is frozen before outcomes

are viewed, identified as investigator-selected and unvalidated, and never derived from pilot results. Not-evaluable and technically invalid observations remain visible. These categories do not validate, invalidate, diagnose, or adjudicate the participant report. The participant report remains the observation of the subjective construct: objective warmth is not required to define the episode, and neither a null surface comparison nor a null or bilateral QST pattern proves equivalence, benignity, or absence of an unmeasured process. [3, 1]

Multiplicity is limited by freezing the primary outcome family, anatomical contrast, time point, direction where justified, and analysis version before outcomes are examined. All other time points, regions, sides, qualities, contexts, QST modalities, and cross-construct associations are labeled secondary or exploratory. The report gives estimates with a prespecified interval level and states any multiplicity procedure used for a confirmatory family; exploratory intervals remain visible and are not converted into a search for a single significant subtype. Null, imprecise, inconsistent, and directionally opposed estimates are retained.

Time from onset, anatomical region, laterality, environmental conditions, exertion, posture, skin findings, itch or pain, and whole-body heat may be described, stratified, or entered into a prespecified repeated-measures model only when sample size, cell counts, and missingness support that analysis. These variables are context or effect-modification candidates, not adjudicated causes. A cross-construct model, if justified after the separate accounts are complete, must name its estimand and temporal assumptions; it cannot relabel an association as mediation, mechanism, a causal subtype, treatment indication, or safety clearance.

The prohibited analyses are explicit: no data-derived diagnostic or referral cutoff; no cluster-derived etiologic subtype; no composite thermal abnormality score that pools spontaneous sensation, skin-surface temperature, and QST; no independence assumption for repeated episodes; and no phenotype restricted to episodes with objective warmth or complete measurements. Measurement results cannot diagnose, exclude, clear, or rank a cause. Code, variable definitions, transformations, exclusions, estimands, and the analysis version are frozen before outcome analysis, and every deviation is reported with its timing and reason.

Safety stops without inventing a pathway

Before recruitment, each Korean study site must have an ethics-reviewed response procedure that names the person responsible for receiving an alert, the lawful local route by which a participant can obtain clinical assessment, the hand-off and documentation steps, adverse-event oversight, capacity and consent safeguards, and privacy protections. That local procedure must define the broader applicable screening and response responsibilities; this publication object neither reproduces nor invents them. NICE NG127 and NG234 supply only selected, non-exhaustive, question-relevant UK safety boundaries. They are not a comprehensive assessment, universal red-flag list, Korean pathway, or patient-facing triage tool, and their referral intervals, coordinators, and service routes are not Korean implementation instructions. [5, 6]

For an adult participant, stop the research procedure and activate the approved assessment hand-off when rapidly progressive symmetrical numbness over hours to days occurs together with weakness or imbalance. Do the same when severe low-back pain radiating into a leg occurs together with new bladder, bowel, or sexual dysfunction or new perineal numbness. Every qualifier is required: these UK neurological recommendations do not make stable, recurrent, isolated back heat an immediate-referral criterion, and a thermal adjective does not substitute for the named tempo, distribution, pain, radiation, or associated findings. [5]

Also stop the research procedure and activate the approved assessment hand-off when a participant with past or current cancer has a named symptom or sign of cord compression: bladder or bowel dysfunction, gait difficulty, limb weakness, neurological signs of cord or cauda-equina compression, numbness, paraesthesia or sensory loss, or radicular pain. Separately, pause and use the site procedure when past or current cancer accompanies a specified NG234 back-pain pattern, or when cancer is independently suspected and such a pattern is present. NG234's UK emergency, 24-hour, coordinator, and urgent-oncology routes remain jurisdiction-bound; neither heat wording alone nor unqualified back pain establishes this constellation. [6]

At participant and procedure level, stop or pause whenever continuing measurement could delay needed assessment or increase risk, and for a protocol-defined device contraindication, intolerable pain, skin injury, distress, loss of capacity, withdrawal of consent, or a privacy breach. QST, thermography, imaging, or completion of a study visit must never be a gatekeeper to the clinical hand-off. A research measurement cannot diagnose, exclude, reassure, clear, or determine the urgency of care. The local procedure, not an investigator improvising from the publication, governs the response.

At study level, suspend the affected procedure after an unexpected serious adverse event plausibly related to it until the prespecified independent safety review is complete. Pause recruitment if the alert route is unavailable, responsibility is ambiguous, or event monitoring indicates that the protocol is delaying care. Apply a prespecified modification or stopping rule when technical failure, burden, missingness, or safety-pause frequency makes a measurement component's primary estimand uninterpretable, and record the amendment, timing, reason, affected participants and estimands, and restart decision. These are proposed governance rules, not observed safety results.

Targeted searches of the verified NG127 and NG234 documents found no recognition or referral criterion stated as back heat, localized warmth, burning, skin temperature, or temperature. That silence is bounded: it neither proves the complaint benign nor excludes neurological, malignant, dermatological, vascular, autonomic, or other disease. It cannot be used as a futility rule, a no-referral rule, or reassurance. This protocol therefore records the participant's thermal language while safety decisions remain tied to the approved local process and to the complete associated constellations that prompted concern. [5, 6]

Pilot decisions and redesign thresholds

This note does not establish an optimal run-in, observation duration, recurrence threshold beyond the provisional minimum of two prospectively reported episodes, event window, sampling frequency, sample size, temperature device, QST procedure, or local clinical pathway. Before recruitment, investigators must select and version each of those parameters, give its feasibility or governance rationale, and identify the estimand it serves. Prespecification makes a choice auditable; it does not make the choice validated or evidence-mandated.

The feasibility pilot estimates the inputs needed for a later design: prospective recurrence yield, event rate, record-to-measurement delay, complete-map and valid-pair yields, usable episodes per participant, within-person variance, stable-condition measurement error, burden, retention, attrition, technical failure, and safety pauses. Numerical continuation, modification, or futility thresholds must be justified and frozen in the pilot protocol before outcomes are viewed; this draft supplies none. Pilot estimates are reported with their denominators and uncertainty and cannot become prevalence, normative, meaningful-change, diagnostic, referral, or safety cutoffs. [1, 2, 4]

Redesign the capture module before confirmatory use when the prespecified event window, reference schedule, environmental controls, anatomical registration, device procedure, or participant burden fails its frozen feasibility criterion and therefore cannot yield interpretable paired observations for the named estimand. A redesign must state what failed, preserve the original result, version the protocol and analysis materials, and pilot the changed component again. Missing or technically invalid observations remain missing; neither late capture nor a substituted comparator may be used to manufacture feasibility.

Comparator availability is itself a pilot question. If contralateral, within-region baseline, non-episode, and external-reference observations are systematically unavailable or answer materially different questions, the confirmatory design must select one primary comparison and demote, redesign, or remove the others. It must not retrospectively choose the comparison with the largest contrast. Likewise, a validated measurement-property study may justify updating a questionnaire, map, surface-temperature procedure, or QST module, but it does not retroactively validate the version used here. [1, 2, 4]

Split the work into a separately registered design when new evidence supports a reference standard, diagnostic-accuracy question, prognostic estimand, intervention effect, or etiologic hypothesis requiring its own comparator and bias controls. Also split or reject any component that requires a diagnostic cutoff, causal subtype, generic red-flag list, or patient-facing advice merely to make this methods note appear complete. A direct prospective dataset should be linked as original research or motivate a versioned protocol update; it must not be written back into this proposal as though it had been available when the design was formed.

Lawful access to a previously unavailable method source may trigger a bounded module review, not a silent rewrite: the PMID 38835743 supplement or full text for PMIDs 24525274 or 33077130 must materially change a proposed detail before revision is warranted. Ethics review and local Korean governance control any required change to consent, privacy, device, safety, or response procedures.

Until approvals, feasibility piloting, applicable registration, and versioned analysis materials exist, this remains a protocol proposal. It establishes no repeatability, agreement, measurement error, responsiveness, meaningful change, concordance, diagnosis, clinical pathway, or observed safety result. [1, 2, 4]

Declarations

Author contributions (CRediT)	Yeonseung Choe: Conceptualization, Methodology, Clinical framing, Research-system design · Baekrokdam Research Commons (BRC): Investigation, Data curation, Writing – original draft, Evidence verification
Data availability	Only public literature was used. The evidence ledger linking citations to passages, permitted claims, and prohibited inferences is downloadable with this article.
Funding	No external funding.
Competing interests	The authors and publisher share research infrastructure with Baekrokdam Korean Medicine Clinic. This relationship does not support treatment-effect claims or causes not examined in this work.
AI use disclosure	AI tools assisted retrieval, structuring, bilingual drafting, and format conversion. Citations were verified at passage level, and claim boundaries, counter-hypotheses, bibliographic links, and form fit were rechecked by internal publication gates. BRC retains publication accountability.
Ethics	This public-literature work contains no individual patient data. Any human-participant implementation of the protocol requires separate ethics review and consent.

References

1. Effects of Unilateral Muscle Fatigue on Thermographic Skin Surface Temperature of Back and Abdominal Muscles—A Pilot Study PMID 35324650. https://doi.org/10.3390/sports10030041
2. Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): reference data for the trunk and application in patients with chronic postherpetic neuralgia PMID 24525274. https://doi.org/10.1016/j.pain.2014.02.004
3. Quantitative sensory testing in notalgia paresthetica reveals small fiber-type-specific differences in non-pruritic sensitivity: a pilot study. PMID 38835743. https://doi.org/10.1097/PR9.0000000000001162
4. Thermal sensitivity mapping - warmth and cold detection thresholds of the human torso. PMID 33077130. https://doi.org/10.1016/j.jtherbio.2020.102718
5. National Institute for Health and Care Excellence. Suspected neurological conditions: recognition and referral (NG127). Updated 2026. Source
6. National Institute for Health and Care Excellence. Spinal metastases and metastatic spinal cord compression (NG234). Updated 2026. Source