

Article

Indoor, Perceived, and Municipal Seismic Intensity in the Greater Tokyo Metropolitan Area: Persistent Divergence and No Detectable Perceptual Calibration

Takashi Furuya ^{1,2,*}, Hiroshi Tsuruoka ³, Danijel Schorlemmer ^{4,5} and Naoshi Hirata ^{2,3}

¹ International Research Institute of Disaster Science (IRIDeS), Tohoku University, 468-1 Aramaki Aza-Aoba, Aoba-ku, Sendai 980-8572, Japan

² Interfaculty Initiative in Information Studies, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan; hirata@eri.u-tokyo.ac.jp

³ Earthquake Research Institute, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan; tsuru@eri.u-tokyo.ac.jp

⁴ Swiss Seismological Service, ETH Zurich, Sonneggstrasse 5, 8092 Zurich, Switzerland; danijel.schorlemmer@sed.ethz.ch

⁵ GFZ Helmholtz Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany

* Correspondence: takashi.furuya.c6@tohoku.ac.jp

Abstract

Discrepancies between announced regional seismic intensity and shaking experienced indoors may affect household mitigation and risk communication. We report a three-year field experiment (2021–2024) in the Greater Tokyo Metropolitan Area combining citizen-hosted low-cost sensors with subjective reports. Up to 92 volunteers answered questionnaires after 96 target earthquakes; sensor hosts kept plug-in micro-electromechanical systems (MEMS) seismometers at home. Delivery logs fixed the denominator: 5724 responses to 8054 attempts. First, indoor seismic intensity exceeded municipal Japan Meteorological Agency (JMA) seismic intensity by a median of +0.34 classes against unrounded instrumental values; site-year means of the excess spread over 1.4 classes against the municipal class, and the excess varied between events within a home (median standard deviation 0.41 classes). Perceived seismic intensity fell between the two, partly by definition of the JMA scale. The excess of perceived over municipal intensity was confined to weak shaking: +0.31 classes at municipal intensities of 1.0–1.5, and not resolved above 1.5. Second, local intensity was associated with feeling an event independently of magnitude and distance (odds ratio 2.42 per class), whereas floor level was not resolved. Third, in this highly prepared volunteer panel, repeated feedback over up to 96 mostly weak events, with receipt unverified, produced no detectable calibration of the individual bias.

Keywords: seismic intensity; felt reports; citizen seismology; macroseismic observation; low-cost MEMS seismometers; earthquake early warning; indoor seismic hazard; risk communication; site amplification; Japan

Academic Editor: Ioanna Triantafyllou

Received: 30 August 2026

Revised: 15 October 2026

Accepted: 15 October 2026

Published: 17 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

A macroseismic scale measures the severity of earthquake shaking from its observed effects on people, objects, buildings, and the natural environment and has long played an

indispensable role in seismology. It serves four important purposes. First, it quantifies actual damage, which physical metrics recorded by seismometers, such as peak ground acceleration (PGA), cannot directly express. Second, it allows the size of historical earthquakes that predate instrumental networks to be reconstructed from documentary records. Third, it yields shaking distributions dense enough to fill spatial gaps between instruments. Fourth, it provides a common yardstick across countries with different instrumental standards and setups.

Citizen participation in seismology has historically developed along two distinct lineages: sensor hosting and felt reporting [1]. In the sensor-hosting lineage, volunteers host low-cost accelerometers, ranging from the global Quake-Catcher Network [2,3] to smartphone-based networks [4] and Raspberry Pi-based community networks in Japan [5], to densify instrumental observations; the participant's role is primarily passive in housing a device. In the felt-reporting lineage, participants actively report their own experiences to systems such as the United States Geological Survey (USGS) "Did You Feel It?" (DYFI) framework, which aggregates internet questionnaires into community intensities, and to the European-Mediterranean Seismological Centre (EMSC) LastQuake system [6]. Over five million responses accumulated over two decades have demonstrated that DYFI provides a surprisingly robust measure of earthquake ground motions [7].

However, each lineage inherently lacks what the other measures: sensor networks do not record human perception or response, whereas felt-reporting systems lack instrumental measurements at the respondent's exact location. A 20-year retrospective review of the DYFI system [8] reports two limitations arising from this gap: self-selection bias (fewer than 3% of entries report "not felt", leaving the true denominator unknown); and observer conditions that contribute a variance of about 0.6 intensity units [9], which is assumed to cancel out when many reports are averaged, although the settings themselves are not recorded. A third limitation follows from the geometry of the observation rather than from the system itself: validating a felt report against ground motion requires the record of the nearest network station, which is often far from the respondent. The review identified physical co-location of human and instrumental measurements as an important future direction for addressing these structural limitations.

Japan provides exceptional conditions for addressing this challenge. The Japan Meteorological Agency (JMA) operates one of the world's densest instrumental intensity networks and publishes municipality-level intensities within minutes of an earthquake, embedding the intensity scale itself into everyday language. In this setting, the critical question is not whether instruments can substitute for felt reports, but how well officially announced municipal JMA seismic intensity represents the shaking that residents actually experience inside their homes and whether that perceptual gap narrows as they accumulate experience alongside immediate, objective feedback.

This question bears directly on seismic mitigation. Earthquakes are experienced inside dwellings, and protective actions are taken or not. Toppling of furniture and injuries depend on indoor shaking rather than free-field intensity on open ground outside the building. Intensity-based alerting, earthquake early warning (EEW) systems, and community preparedness programs all implicitly target indoor environments. Determining how far published municipal intensity departs from indoor conditions is therefore important for assessing indoor seismic hazard. Indoor shaking has been examined by questionnaire in Japan, but in high-rise buildings [10] and in buildings already equipped with seismographs [11], rather than in ordinary dwellings.

Here, we report on a three-year panel experiment in the Greater Tokyo Metropolitan Area that combined repeated felt reports with co-located home-seismometer measurements and personalized feedback. Panel here denotes a fixed group of residents who reported repeatedly across successive earthquakes, so that the same individuals are

observed under many different levels of shaking. Sensor-hosting residents kept a plug-in micro-electromechanical systems (MEMS) seismometer in their homes and reported after target earthquakes how strongly they felt the shaking and where they were located. The system also provided personalized information, including measured indoor intensity, nearest-station intensity, and EEW-predicted intensity, as applicable. Human and instrumental observations were collected across 96 earthquakes.

The advance over existing citizen seismology lies in three respects. First, relative to DYFI and LastQuake [6,7], for participants hosting a sensor, the panel records the felt report of a person together with an instrumental value measured inside that person's dwelling for the same earthquake. Second, the delivery log fixes the denominator, so the earthquakes a participant did not feel are counted as well as those they did. Third, relative to the Japanese studies of indoor shaking [10,11], the observations are made in ordinary dwellings, follow the same individuals across many earthquakes, and are accompanied by feedback on the measured values. For participants hosting a sensor, co-location, repeated prompting, the salience of being a volunteer and exposure to feedback always occur together, and the comparison with non-hosts is observational (Section 4.6), so we do not claim an effect of co-location itself; the contribution lies in the quantities these observations yield, and Section 2.7 sets out what the design can and cannot establish. We test three core hypotheses:

H1 (local representativeness). *Officially announced municipal JMA seismic intensity under-represents indoor shaking in residences, which exhibits significant home-specific amplification.*

H2 (perception follows local input). *The probability of feeling an event is associated primarily with local intensity rather than magnitude or distance alone. The association with the floor level on which the participant mainly stays is examined as a covariate. The data support an association independent of magnitude and distance, but not its primacy (Section 4.3).*

H3 (feedback calibration). *Repeated exposure to immediate, objective feedback reduces the systematic over-reporting error of felt reports over time. This hypothesis motivated the experimental design; the data from this panel do not support it, within the bound stated in Section 2.7.*

The behavioral outcomes of the same experiment, focusing on household preparedness and protective actions taken during these earthquakes, are examined in a companion paper (in preparation).

2. The Experiment

2.1. Plug-In Home Seismometer and Automated Survey System

Each participant assigned to host a sensor received a palm-sized MEMS accelerometer unit equipped with a Raspberry Pi-class processor and a mobile Wi-Fi router. The unit was installed by plugging it into a standard wall outlet [12]. This unit functions as the QuakeSaver sensor, and its technical deployment for community earthquake risk communication has been validated in previous operations [13]. The units continuously stream waveform data to a cloud server in miniSEED format. Waveforms are retrieved from the cloud every 30 min and, after each target earthquake, upon the earthquake early warning trigger. An analysis server at the Earthquake Research Institute (ERI), The University of Tokyo, automatically detects seismic events and computes a JMA-equivalent measured instrumental intensity (I_{inst}) for each event, referred to here as “indoor seismic intensity.” Figure 1 shows a unit installed in a participant's home.

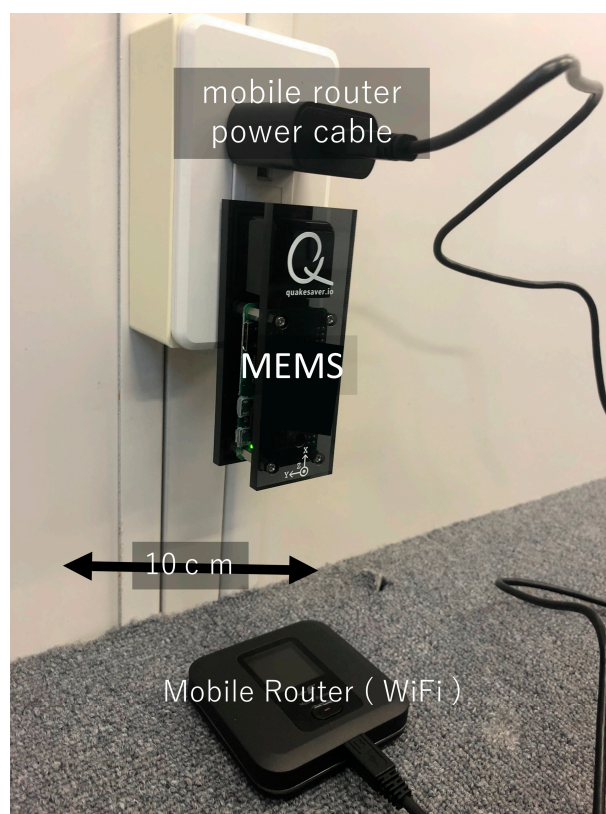


Figure 1. The plug-in home seismometer installed in a participant’s home: the micro-electromechanical systems (MEMS) sensor unit plugged directly into a wall outlet and the mobile Wi-Fi router that transmits data to the cloud. The arrow indicates a 10 cm scale.

Triggered by high-grade JMA Earthquake Early Warning (EEW) messages, the automated survey pipeline compiled and dispatched personalized messages to participants within approximately 8 to 10 min after each target earthquake via email (starting in 2022) or the LINE messaging application (starting in 2023). The pipeline waits for the JMA municipal intensity bulletin, which is published about 7 min after origin time, and this wait sets the delay. Consistent with that figure, the earliest felt report in the panel was returned 12.4 min after origin time, and no response arrived before it. These notifications reported (a) the participant’s own measured indoor seismic intensity, (b) the observed instrumental intensity at the nearest JMA station, and (c) the site-specific EEW-predicted intensity at their home location. Participants who did not host a sensor received the same messages without item (a) or the waveform link. The EEW-predicted intensity was calculated from the final real-time EEW telegram using the standard ground-motion attenuation relationship developed by Si and Midorikawa [14], accounting for local site amplification factors. Each message also contained a link to a secure smartphone/PC questionnaire and a link to the seismic waveform recorded at the sensor. Figure 2 summarizes the complete pipeline from the sensor in the home, through the cloud, to the analysis and delivery server at ERI, and back to the participant.

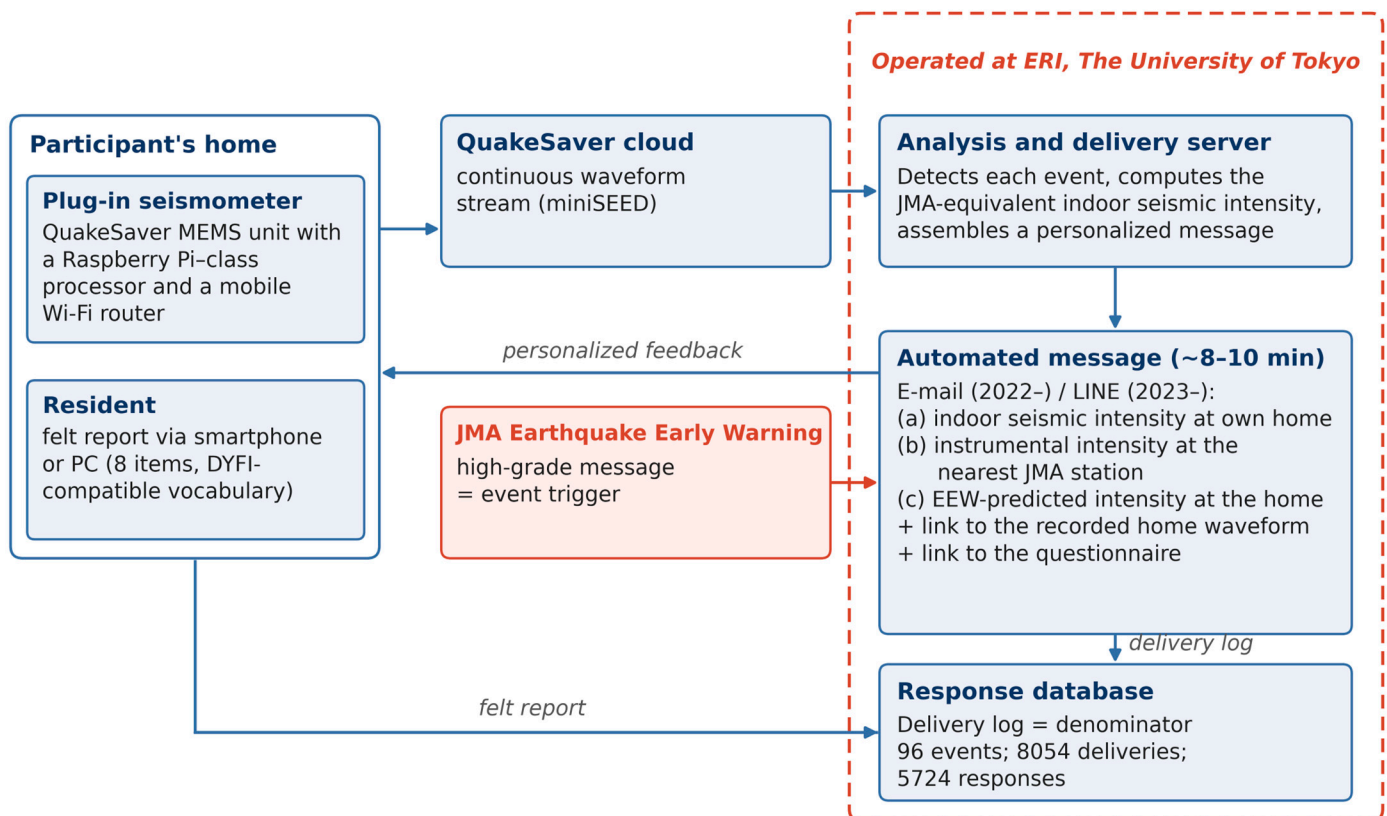


Figure 2. The plug-in home seismometer and automated per-event survey pipeline. High-grade Japan Meteorological Agency (JMA) Earthquake Early Warning messages trigger the pipeline. The analysis server computes indoor seismic intensity from continuously streamed waveforms and assembles a personalized message that reaches the participant within about 8 to 10 min. The resident's felt report is then returned to the response database, whose delivery log provides the denominator for the response rates (96 events; 8054 deliveries; 5724 responses). The dashed frame marks the components operated on the analysis and delivery server at the Earthquake Research Institute (ERI), the University of Tokyo.

2.2. Configuration of Seismic Intensity Stations and Indoor Seismometers

High-density seismic observation networks are essential for rapid post-event emergency response after a destructive earthquake. Japan maintains one of the world's densest instrumental intensity networks, with 4424 active stations nationwide as of October 2023. The Kanto region contains approximately 720 stations, including 708 active stations within our immediate study area (34.9–36.9° N, 138.9–140.9° E). This configuration provides a mean station spacing of 3.38 km (standard deviation (SD) 2.11 km; median 3.06 km; range 0.0–14.7 km).

Station coordinates were taken from the official catalogs published by the Japan Meteorological Agency, the National Research Institute for Earth Science and Disaster Resilience (NIED), and the Headquarters for Earthquake Research Promotion. The JMA synthesizes real-time data from this combined network to publish a single representative maximum intensity value for each administrative municipality immediately after an event. We use these official data as our baseline regional metric, defined as “municipal JMA seismic intensity.”

Despite the density of the public network, a mean station spacing of ~3.4 km remains insufficient to capture the microscale shaking variations experienced within an individual's living space. Figure 3 overlays the geographic distribution of the citizen-hosted indoor seismometers on the public station network. Hosting was counted separately in each

study period, so the 106 installations correspond to 76 distinct hosts: 50 residents hosted a sensor in 2022 and 56 in 2023, and 30 of them hosted in both periods. The public stations are spread uniformly across the Kanto plain, whereas the participant-hosted sensors are concentrated within a densely populated urban belt extending from the Tokyo wards through southern Saitama, northwestern Chiba, and eastern Kanagawa, reflecting our recruitment channels. Figure 3b places the target earthquakes in their regional context: their epicenters extend well beyond the study frame, from the 2024 M_j 7.6 Noto Peninsula sequence in the west to events off Aomori in the north and in the Izu–Ogasawara region in the south.

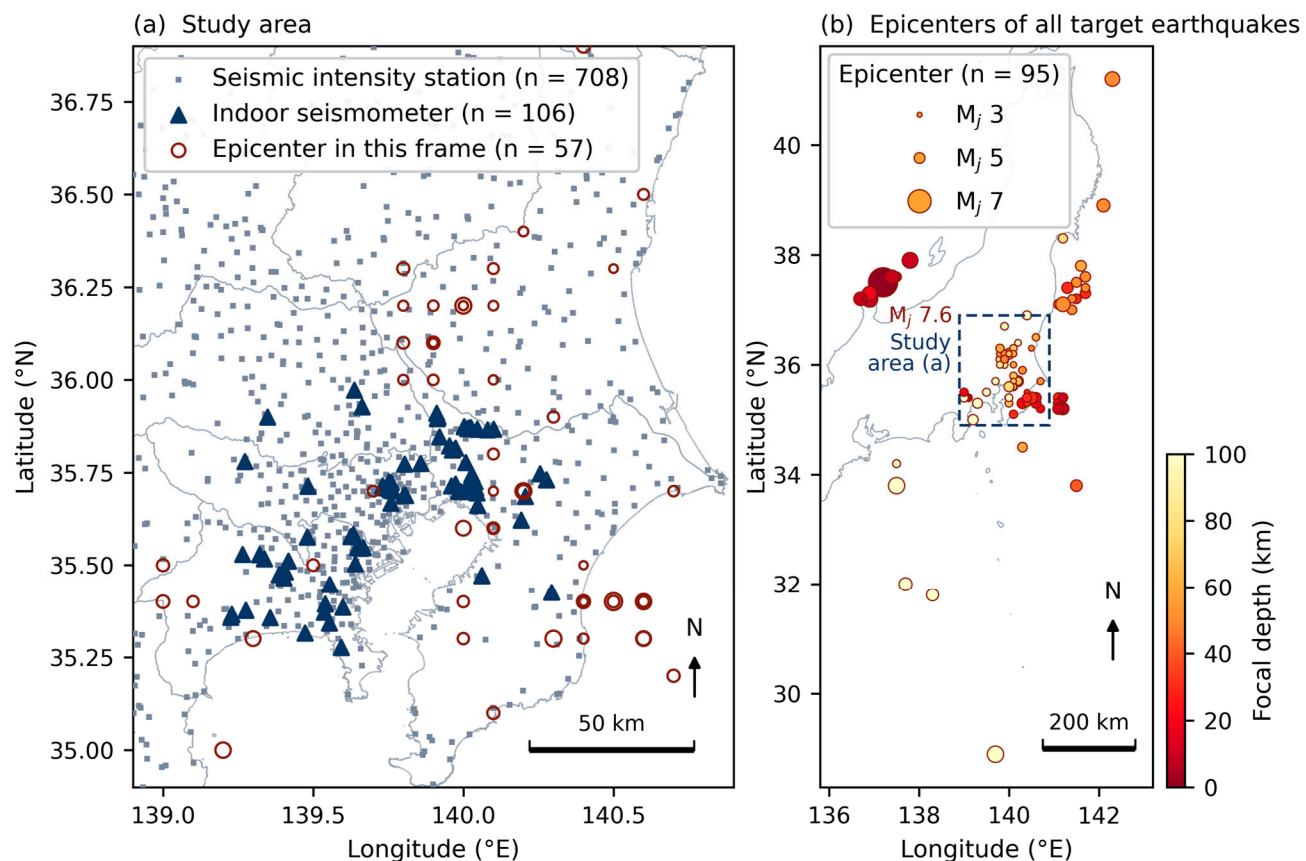


Figure 3. Seismic intensity stations, indoor seismometers, and target earthquakes. (a) The 708 seismic intensity stations (gray squares) and 106 indoor seismometer installations by 76 participants (navy triangles) within the study area (34.9–36.9° N, 138.9–140.9° E), together with the epicenters of the 57 target earthquakes located inside this frame (open circles). (b) Epicenters of 95 of the 96 target earthquakes; no hypocenter record could be assigned to the remaining event, for which JMA issued no seismic intensity bulletin (Section 3). Symbol area scales with M_j , and fill color indicates focal depth. The dashed rectangle marks the frame of Panel (a). Epicenters are plotted at the 0.1° resolution of the JMA catalog, so nearby events can share a plotting position. Prefectural boundaries in (a) are from the National Land Numerical Information administrative-area data (Ministry of Land, Infrastructure, Transport and Tourism (MLIT), 1 January 2024; dataset N03; <https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-N03-2024.html>, accessed on 13 August 2026), processed by the authors to dissolve the municipal polygons into prefectural outlines. These data are used under the MLIT terms of use, which conform to the Government of Japan Standard Terms of Use (Version 2.0) and are compatible with CC BY 4.0. The coastline in (b) is from Natural Earth (1:10 m), which is in the public domain. The aspect ratio is set to true distance at the mean latitude of each panel.

The linear distance from each participant's residence to the single nearest active public station averaged 2.08 km (SD 1.37 km; median 1.78 km; range 0.01–8.15 km). Thus, even in densely built metropolitan environments where the public network is most concentrated, an average spatial gap of 2 km separates a resident's home from the nearest ground-surface station. This observational gap motivates the use of co-located indoor instruments and forms the basis for the nearest-station robustness checks detailed in Section 4.1.

2.3. Three-Year Cohort Design

The longitudinal panel experiment spanned three distinct study periods between 2021 and 2024 (Figure 4; Table 1). A preliminary pilot study in Zama City in 2021 (enrolling 10 sensor hosts and 10 questionnaire-only participants from a local disaster-volunteer network) established the basic questionnaire wording and highlighted the operational need for an automated, real-time survey delivery pipeline.

Three-year co-location experiment: cohorts, responses, and telemetry

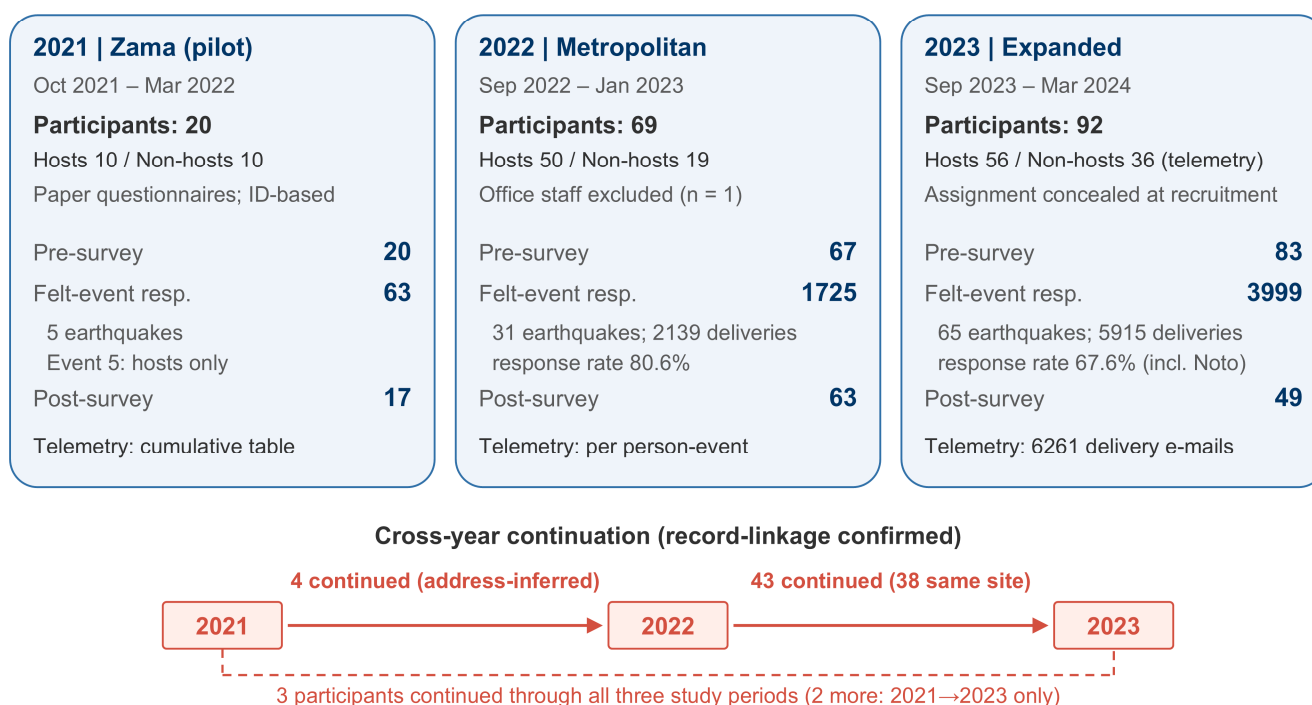


Figure 4. Overview of the three-year experiment: cohorts, installed/non-installed groups, per-phase response counts, and cross-year continuation (43 residents continued from 2022 to 2023, 38 of them at the same address, and 30 hosted a sensor in both periods). Response rates are delivery-based and cover residents only, with project staff excluded; pre- and post-survey counts are unique respondents, with resubmissions removed; host status in 2023 was determined from telemetry, using a recording rate of 50% or more; and the 2021 linkage was inferred from the installation address at room level.

Table 1. Cohorts, events, and responses (residents only). EEW, earthquake early warning.

	2021 Zama (Pilot)	2022 Metropolitan	2023 Expanded
Survey period	October 2021–March 2022	September 2022–January 2023	September 2023–March 2024
Participants (hosts/non-hosts)	20 (10/10)	69 (50/19)	92 (56/36)
Target earthquakes	5	31	65

Felt-survey deliveries	—	2139	5915
Felt-survey responses (rate)	63	1725 (80.6%)	3999 (67.6%)
Not-felt share of responses	—	77%	85%
At-home felt reports with municipal intensity	—	330	509
Pre-survey respondents (unique)	20	67	83
Post-survey respondents (unique)	17	63	49
Delivery channel	paper	e-mail	e-mail + LINE
Feedback per event	none (survey only)	indoor + station + EEW	indoor + station + EEW

The 2022 metropolitan study period enrolled 69 residents (50 sensor hosts, 19 non-hosts) across 31 target earthquakes occurring between September 2022 and January 2023, generating 2139 survey deliveries. The expanded 2023 study period enrolled 92 residents across 65 target earthquakes between September 2023 and March 2024, generating 5915 deliveries; this period captured the far-field seismic sequence of the 1 January 2024 M_j 7.6 Noto Peninsula earthquake. To mitigate potential selection bias, applicants were recruited without being informed of their sensor-hosting assignment at the time of recruitment, and active recording status was continuously verified via server telemetry.

Our participants are active members of safety-leader volunteer networks and university-affiliated community groups. The panel therefore represents a highly disaster-prepared stratum, with a mean baseline score of 7.4 to 7.6 on the ten household preparedness measures of the Tokyo Fire Department's public checklist, Ten Preparations against Earthquakes (https://www.tfd.metro.tokyo.lg.jp/lfe/bou_topic/jisin/sonae10.html, accessed on 16 October 2026). The preparedness score and the building attributes come from the pre-survey, which this paper uses only to describe the participants; change between the pre- and post-surveys is not analyzed here and is reported in a companion paper. Recruitment through these networks means that the panel is not representative of the general population, and our data cannot determine whether the over-reporting bias of ordinary residents would be larger or smaller than the values reported here (Sections 2.7 and 5.5).

2.4. Felt-Report Questionnaire

The per-event digital questionnaire was limited to eight items to optimize response rates and data quality. It captured (i) the respondent's exact indoor location and posture at the time of the shock; (ii) a descriptive shaking assessment using vocabulary compatible with international macroseismic frameworks such as DYFI (e.g., not felt, weak, moderate, strong, violent) [15,16]; (iii) a directly self-assigned perceived seismic intensity using the JMA scale; (iv) any situational information known to the respondent before making their judgment; (v) the specific media channel through which they learned the official JMA intensity; and (vi) whether any pre-planned protective actions were executed, which actions were taken, and which could not be taken (three items).

To address the self-selection and non-reporting biases affecting anonymous, cross-sectional systems [8,17,18], an explicit “did not notice the shaking” pathway was embedded in the interface starting in 2023. This feature raised the “not-felt” capture rate to 85% of total responses in 2023 (Table 1). Metadata analysis of response timestamps further shows that participants referenced the delivered measured indoor intensity before submitting their subjective judgment in only 6% to 10% of felt responses. Thus, in most felt responses, participants submitted their subjective judgments without first consulting the delivered indoor seismic intensity, limiting the potential for immediate feedback to bias their perceived seismic intensity reports.

2.5. The JMA Intensity Scale in International Context

Throughout this study, “intensity” refers strictly to the seismic intensity scale of the Japan Meteorological Agency (JMA), which has been in continuous operational use since the late nineteenth century and is deeply embedded in public language and understanding in Japan. The modern discrete JMA scale consists of ten categories: 0, 1, 2, 3, 4, 5 Lower (5−), 5 Upper (5+), 6 Lower (6−), 6 Upper (6+), and 7. Since 1996, official JMA bulletins have used automated instrumental calculation rather than human observational reporting: a continuous instrumental intensity value is derived from band-pass-filtered, three-component acceleration waveforms using a 0.3 s duration threshold criterion [19]. The discrete JMA class is then assigned using fixed mathematical boundaries (e.g., class 1 spans 0.5 to 1.5; class 2 spans 1.5 to 2.5). The relation between the instrumental JMA value and conventional strong-motion parameters has been established empirically [20]. Indoor seismic intensity, as the term is used in this paper, is the JMA instrumental intensity computed by this same procedure from acceleration recorded by a sensor inside a dwelling [12]. It is not a quantity that JMA publishes: the agency’s stations are sited in the free field or at the base of a building, and the published municipal value is the maximum over such stations. It is also not a macroseismic assessment because it records no effect on people or on objects, and it is not a measure of structural response such as a floor response spectrum. We introduce it because residents experience shaking inside the dwelling, so an instrumental value recorded there is the measurement that corresponds to what a resident is asked to report. Indoor seismometer readings are available as continuous instrumental intensity values and discrete JMA classes, whereas participants directly self-assign discrete perceived seismic intensity values.

For international comparison, Table 2 provides an approximate macroseismic correspondence with the European Macroseismic Scale (EMS-98) [21], following the guidelines established by Musson et al. [22]. Across the weak-to-moderate motion domain analyzed in this study (JMA classes 0 to 4), the scale corresponds approximately to EMS-98 (and Modified Mercalli Intensity, MMI) degrees I to V.

Table 2. Approximate correspondence between the JMA intensity scale and the European Macroseismic Scale (EMS-98) (approximately equal to the Modified Mercalli Intensity (MMI)).

JMA class	0	1	2	3	4	5 Lower	5 Upper	6 Lower	6 Upper	7
EMS-98 (approximately MMI)	I	II–III	IV	IV–V	V	VI	VII	VIII	IX–X	XI

Two methodological caveats are important. First, such conversions compare the linguistic descriptors of scale definitions rather than empirical co-assignments [22,23]. Second, a modern JMA value represents automated physical ground motion at a point, whereas EMS-98 or MMI traditionally evaluate aggregated human and environmental effects; work toward an international macroseismic scale addresses this divergence directly [24]. Our survey design also differs from the USGS DYFI system [15]. Because Japanese citizens are fluent in the JMA scale, they directly self-assign a discrete intensity value, whereas DYFI computes a Community Decimal Intensity (CDI) through a weighted matrix of reported effects [25]. This direct self-assignment is a deliberate feature of our panel design, allowing us to compare the class that residents themselves use and communicate during an emergency with the measured values.

2.6. Validity of the Plug-In Units at Low Intensity

Because our core statistical comparisons rely on indoor instrumental intensity data, we first need to establish that low-cost plug-in MEMS sensors provide scientifically meaningful data during weak ground shaking. A well-documented limitation of low-cost

MEMS accelerometers is their relatively high electronic self-noise floor. For example, the smartphone-based i-jishin network exhibits a continuous internal noise floor of approximately ± 5 gal, which routinely obscures earthquake waveforms at or below JMA intensity 2 [26]. Laboratory testing of low-cost seismographs provides the reference values for such noise floors [27], and in citizen-hosted networks the ambient noise at the installation site is dominated by human activity [28].

We evaluated the data validity of our QuakeSaver sensors in two ways. First, we isolated the 3659 person-events for which the JMA published no municipal intensity bulletin, indicating that regional ground motion fell below the instrumental threshold of 0.5. Within this sub-threshold dataset, the median intensity reported by the home sensors was 0.47, with a 5th percentile of 0.06; 80% of all idle readings remained below 1.0. For the 77 sensor units that recorded at least ten events, every sensor reported a value of 0.44 or lower during its least-active interval, with a median per-unit minimum of 0.06. Because any active reading combines ambient ground motion with electronic noise, these empirical minima bound the true instrumental noise floor of the devices well below intensity 1.0. We also examined a quiet interval directly: in a two-minute window of one record that contained no earthquake signal, the amplitude was about ± 0.5 gal, which corresponds to an equivalent instrumental seismic intensity of 0.2. This level is the sum of instrument self-noise and ambient ground motion, which we cannot separate, but it bounds both, and it is an order of magnitude below the ± 5 gal floor reported for the smartphone-based network and well below the value of 0.5 at which JMA class 1 begins.

Second, indoor intensities were compared with municipal instrumental intensities for the 1209 person-events for which municipal values were available (Figure 5). The two metrics were strongly correlated (Pearson's $r = 0.76$, Spearman's $\rho = 0.75$), with indoor values remaining systematically higher because of structural amplification. The lower edge of the distribution is particularly informative. Because indoor shaking is mechanically expected to be at least as intense as free-field ground-surface motion, indoor values for the least-amplified structures should coincide with the municipal baseline. If electronic noise were artificially inflating low-intensity readings, the lower bound of the scatter plot would instead flatten horizontally above the municipal values.

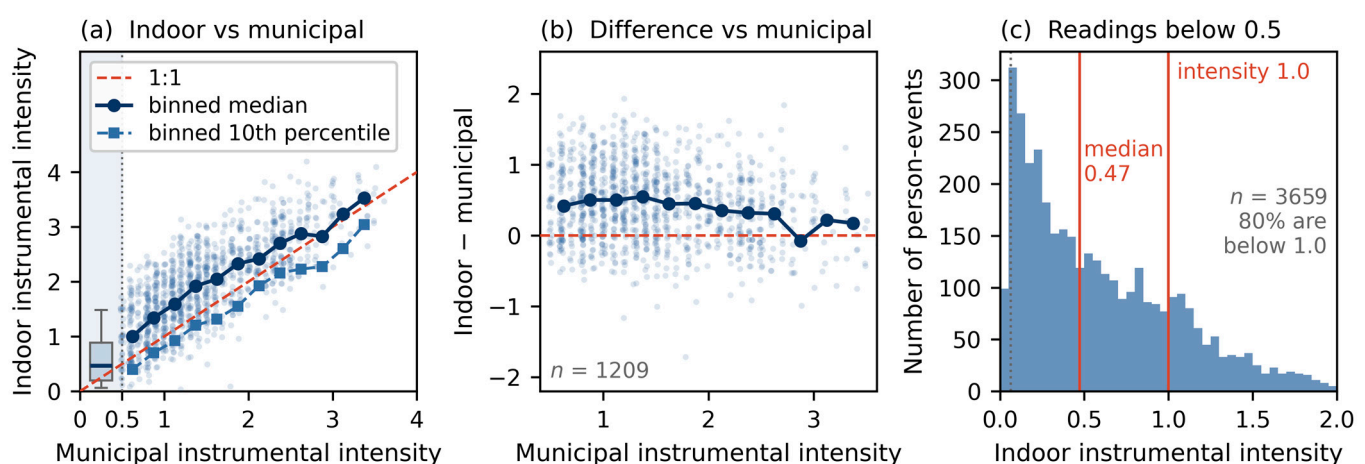


Figure 5. Indoor (device) versus municipal instrumental intensity for 1209 person-events from 77 sensor units: (a) indoor against municipal intensity, (b) their difference against municipal intensity, and (c) indoor readings for events with no published municipal intensity. Blue points represent individual person-events (municipal values are published to 0.1 resolution; a horizontal jitter of ± 0.02 is applied for legibility). In (a), the solid line joins binned medians and the dashed blue line binned 10th percentiles, computed in 0.25-wide bins of the horizontal axis, and the dashed vermilion line is the 1:1 line. In (a), the shaded band on the left covers events for which no municipal intensity

was published (below 0.5); there, device values alone are shown as a box plot (median, quartiles, 5th–95th percentiles; $n = 3659$), and the band carries no horizontal information. Between municipal intensities of 0.5 and 1.5, the binned 10th percentile runs parallel to the 1:1 line with an offset of -0.14 to -0.20 ; in panel (c), the dotted line marks the 5th percentile of the device readings (0.06), and the solid lines mark their median (0.47) and intensity 1.0.

For municipal intensities between 0.5 and 1.5, the binned 10th percentile of indoor intensity runs parallel to the 1:1 line with an offset of -0.14 to -0.20 and tracks the municipal baseline linearly down to 0.40 without visible flattening. These results indicate that electronic noise does not systematically inflate the readings at low intensity. They do not establish per-event accuracy at the threshold of JMA intensity class 1: our data cannot separate electronic self-noise from ambient ground motion, and we return to this limitation in Section 5.7.

2.7. What the Design Can and Cannot Establish

Before presenting the data, we set out what this design can and cannot decide. One distinction should be made at the outset. The quantity we examine is perceptual calibration, that is, whether the intensity class a respondent assigns moves closer to the measured value as experience accumulates. It is not risk awareness. A resident may become more aware of earthquake risk without assigning intensity classes any more accurately, and the converse is equally possible; we measure only the former.

Four measures, four definitions. This study compares four quantities that are often used interchangeably but are not equivalent (Table 3). Municipal JMA seismic intensity is the maximum over a municipality; nearest-station intensity refers to one instrument at a known distance from the respondent; indoor seismic intensity refers to one instrument inside the respondent's dwelling; and perceived seismic intensity is a class assigned by the respondent. They differ in spatial support, in scale type and in observational definition. Three of the four are continuous; the fourth is an ordinal class assigned by a person. The JMA scale is itself described partly in terms of what people indoors feel and how objects inside a room respond [29]. When respondents assign a class themselves, some convergence between perceived and indoor measured intensity therefore follows from the definition rather than from the accuracy of perception, and we return to this in Sections 4.1 and 5.2. The share of the ordering that follows from the definition cannot be identified with this design: a self-assigned class carries the definition of the scale and the accuracy of perception in a single value, and separating them would require a comparison group answering on a scale without indoor descriptors, which we did not run. We therefore do not count the position of perceived intensity between the municipal and indoor values among the established results (Section 5.1). Perception also depends on the circumstances of the observer, which has been documented independently of the present design [9,30].

Table 3. The four intensity measures compared in this study. They differ in spatial support, in scale type, and in observational definition.

Measure	Spatial Support	Scale Type	Observational Definition
Municipal JMA seismic intensity	Maximum over a municipality	Continuous (published to 0.1)	Instrumental, free-field
Nearest-station intensity	One station, median 1.8 km (range 0.01 to 8.2 km) from the dwelling	Continuous	Instrumental, free-field
Indoor seismic intensity	One instrument inside the dwelling	Continuous	Instrumental, indoor

Perceived seismic intensity	One person inside the dwelling	Ordinal class assigned by the respondent	Self-assignment against the JMA class descriptions
-----------------------------	--------------------------------	--	--

The design cannot isolate co-location. Participants who hosted a seismometer were also prompted after every target earthquake and received personalized feedback. For them, these three conditions always occur together, so the effect of co-location cannot be separated from that of repeated prompting, of the salience of being a volunteer, or of exposure to feedback. Separating them would require arms that receive prompting only, or no feedback; we did not run such arms.

The panel is not a sample of the general population. Participants were recruited through safety-leader volunteer networks and university-affiliated community groups, and their baseline preparedness scores lie close to the upper end of the scale used (Section 4.5). Our data cannot determine whether the over-reporting bias of ordinary residents would be larger or smaller than the values reported here. We therefore do not argue from this panel to the general population, either for the size of the bias or for its stability under feedback (Section 5.5).

Exposure is measured indirectly. We use the cumulative number of felt reports as a proxy for exposure to feedback. The delivery log records that a message was sent, not that it was opened or read, so the proxy is an upper bound on exposure. Section 4.4 reports the same analysis with the cumulative number of messages delivered in place of felt reports. Throughout this paper, the calibration result carries one bound: no calibration was detected in a highly prepared volunteer panel, over shaking mostly at JMA classes 0 to 3 (368 felt reports at municipal class 2 or above), with exposure measured by cumulative felt reports or by messages delivered rather than by verified receipt of feedback.

3. Data

This study uses three distinct measures of intensity: perceived seismic intensity, indoor seismic intensity, and municipal JMA seismic intensity. The latter represents regional shaking at the administrative level. For each municipality, JMA determines this baseline by selecting the maximum instrumental intensity (I_{inst}) recorded among all active stations (operated by JMA, local governments, and NIED) within the municipal boundary. This value is then mapped onto the 10-level discrete JMA seismic intensity scale ranging from 0 to 7. We use these three terms consistently throughout the text.

The analysis compares these three intensity measures for the same participant and earthquake event, with municipal JMA seismic intensity serving as the regional reference for the participant's municipality. The agreement rates reported below are not variables used in the analysis; rather, they verify the accuracy of the compiled data.

The delivery logs form the raw data pool for statistical analysis, with 5724 responses obtained from 8054 delivery attempts to residents (2139 in 2022; 5915 in 2023), representing response rates of 80.6% and 67.6%, respectively. The 96 target earthquakes comprise 31 events in the 2022 study period and 65 in 2023; the 2021 pilot used paper questionnaires and is not included in the event count. Response delays were generally brief (median: 3.3 h; 90% within 24 h; 96% within 72 h). Delayed responses were predominantly “not felt” reports, supporting the use of a 72 h (or until the subsequent event) analytical time window. All 96 events were cross-checked against official JMA epicenter/hypocenter records and municipal JMA seismic intensity data retrieved via the P2P Earthquake Information Archive. One event (21 December 2023, 23:17 Japan Standard Time, off the east coast of Chiba Prefecture) reached no participating municipality at JMA intensity class 1 or above, so JMA issued no seismic intensity bulletin for it. That event is retained in the delivery and response counts, but no municipal intensity or hypocenter record could be assigned to it, and it is therefore absent from the analyses that use the municipal class or the

hypocenter; in the comparison against unrounded municipal values by level of shaking (Table 4), its six at-home felt reports are counted in the censored band below 0.5.

Table 4. Excess of perceived over municipal seismic intensity by level of shaking, for at-home felt responses.

Municipal Instrumental Intensity	Felt Reports	Participants	Mean Person-Level Excess	95% CI	<i>p</i>	Share with a Negative Gap
Below 0.5 (not published)	245	67	+1.100	+0.969 to +1.231	<0.001	0%
0.5 to 1.0	79	48	+0.554	+0.412 to +0.696	<0.001	0%
1.0 to 1.5	153	71	+0.313	+0.171 to +0.455	<0.001	46%
1.5 to 2.0	148	77	+0.093	−0.032 to +0.218	0.141	46%
2.0 and above	220	85	−0.070	−0.174 to +0.034	0.186	37%

Note: Only responses in which the respondent was at home and reported feeling the earthquake are included, so the perceived class is at least 1 and the gap cannot be negative where municipal intensity is below 1.0. All 324 gaps in the first two rows are non-negative. The constraint no longer binds at 1.0 and above, where 37% to 46% of gaps are negative, so those rows carry the interpretable estimates. The excess is averaged within each participant and then across participants; the means over responses are given in Table A1. The first row includes the six at-home felt reports for the event without a JMA bulletin (Section 3), whose municipal ground motion is below 0.5. CI: confidence interval.

The assigned municipal intensities aligned closely with the nearest-station values delivered to individual users by our system (71% exact match; 98% within ± 1 intensity class; Spearman's $\rho = 0.70/0.77$). This comparison serves as a compilation check because the two values represent different quantities: the maximum within a municipality and the value at an individual's nearest station. Exact agreement is therefore not expected. Participant records across the three study periods were linked using anonymous individual identifiers and passed a 45-item data integrity audit. Research team members and administrative personnel were excluded from the analysis. A total of 991 felt reports were obtained from residents; 839 of these, submitted from home for events with defined municipal JMA seismic intensities, formed the core sample for the felt-report comparisons of Section 4.1 and for testing H3.

For the 95 events with a JMA record, we additionally compiled the instrumental intensities (to one decimal place) of every station that observed intensity 1 or greater from the seismic-intensity files of the Monthly Seismological Bulletin and the appendices of the Monthly Report on Earthquakes and Volcanoes. A participant's municipal instrumental intensity is not the value at the nearest station but the maximum across all stations within the municipality where that participant lives, following the definition of the published municipal JMA seismic intensity. Reclassifying these values reproduces, for 99.4% of person-events, the class independently assigned to the same municipality in the archive of real-time JMA bulletins. This provides a second compilation check because the same quantity is derived from two sources: the definitive monthly bulletins and the real-time telegrams. It is not a comparison with indoor or nearest-station values.

The 0.6% disagreement (48 of 7651 person-events) corresponded to 38 independent cases (municipality $\times$ earthquake) because several participants shared a municipality and event; all cases were checked individually. In 29 cases, the archive had failed to match the municipality name and assigned no intensity, 25 of them in a single municipality. In the remaining nine, the archive value exceeded the bulletin value, and all occurred during periods when earthquakes happened in quick succession. For such periods, the real-time announcements can carry intensities observed from a different earthquake, and the JMA provides per-earthquake intensities in its periodical bulletins. We therefore treated the

bulletin as authoritative at the level of the individual earthquake and used its values here. Stations below intensity 1 are not published, so instrumental values <0.5 are censored; analyses using them bound this censoring (midpoint 0.25; sensitivity bounds 0.00/0.49).

For these data, the statistical analyses comprise (i) class-agreement tabulations; (ii) linear mixed-effects models for the signed gap (felt minus municipal class) with crossed random intercepts for participants and for earthquakes, with the event-demeaned specification retained as a robustness check; (iii) logistic models for the probability of feeling an event, with two-way cluster-robust standard errors clustered simultaneously on participants and on earthquakes [31]; and (iv) person-level nonparametric tests and instrumental-intensity robustness checks. All analyses were performed in Python 3.13.5 with statsmodels 0.14.6, SciPy 1.17.1, NumPy 2.4.4, and pandas 3.0.2.

The H3 model for the signed gap y_{ij} of participant i in earthquake j is

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 m_{ij} + \beta_3 p_{ij} + u_i + v_j + \varepsilon_{ij}$$

where x_{ij} is cumulative felt experience (the number of the participant's earlier felt reports, standardized); m_{ij} is municipal intensity (centered on its sample mean); p_{ij} indicates the 2023 study period; and the participant effects u_i , earthquake effects v_j and residuals ε_{ij} are independent and normally distributed with variances σ_u^2 , σ_v^2 and σ^2 . The model was estimated by restricted maximum likelihood with the MixedLM routine of statsmodels, the earthquake intercepts being entered as a variance component. The intra-class correlations reported in Section 4.4 are the shares of the total variance attributable to participants and to earthquakes. The H2 models are logistic regressions fitted by maximum likelihood; their two-way cluster-robust covariance is the sum of the covariance matrices clustered on participants and on earthquakes minus the heteroskedasticity-robust matrix [31]. Model B1 uses the 5462 responses for which municipal intensity, epicentral distance and magnitude are all available (the 51 responses to the event without a bulletin and 211 responses from eight participants without a recorded epicentral distance are excluded), and Model B2 uses the 5322 of these for which the main floor of residence was reported. All analyses use complete cases, and no values were imputed. For H3 the sample is derived as follows: 8054 deliveries yielded 5724 responses, of which 991 reported the earthquake as felt; 845 of these were made at home, and 839 of those refer to an event with a published municipal intensity, from 107 participants and 89 earthquakes.

4. Results

The analysis covers the two primary study periods, 2022 and 2023. As outlined in Section 2.3, the longitudinal experiment included three phases. However, the preliminary 2021 pilot in Zama City was an operational study designed to refine the questionnaire, and its paper-based distribution left the exact response denominator undefined. Accordingly, the results below are based only on the systematic digital datasets from the second and third study periods.

4.1. Ordering of Municipal, Perceived, and Indoor Seismic Intensity and Its Dependence on Shaking Level (H1)

Figure 6 cross-tabulates at-home perceived seismic intensity against measured indoor seismic intensity and municipal JMA seismic intensity across both study periods. Compared with measured indoor seismic intensity, subjective assessments matched the instrumental classes in exactly 50% of cases in both 2022 and 2023, while under-reporting by exactly one class occurred in approximately one-third of cases (signed mean offset: −0.20 in 2022 and −0.26 in 2023). Perceived intensity was within one class of the indoor class in 96% (2022) and 94% (2023) of cases.

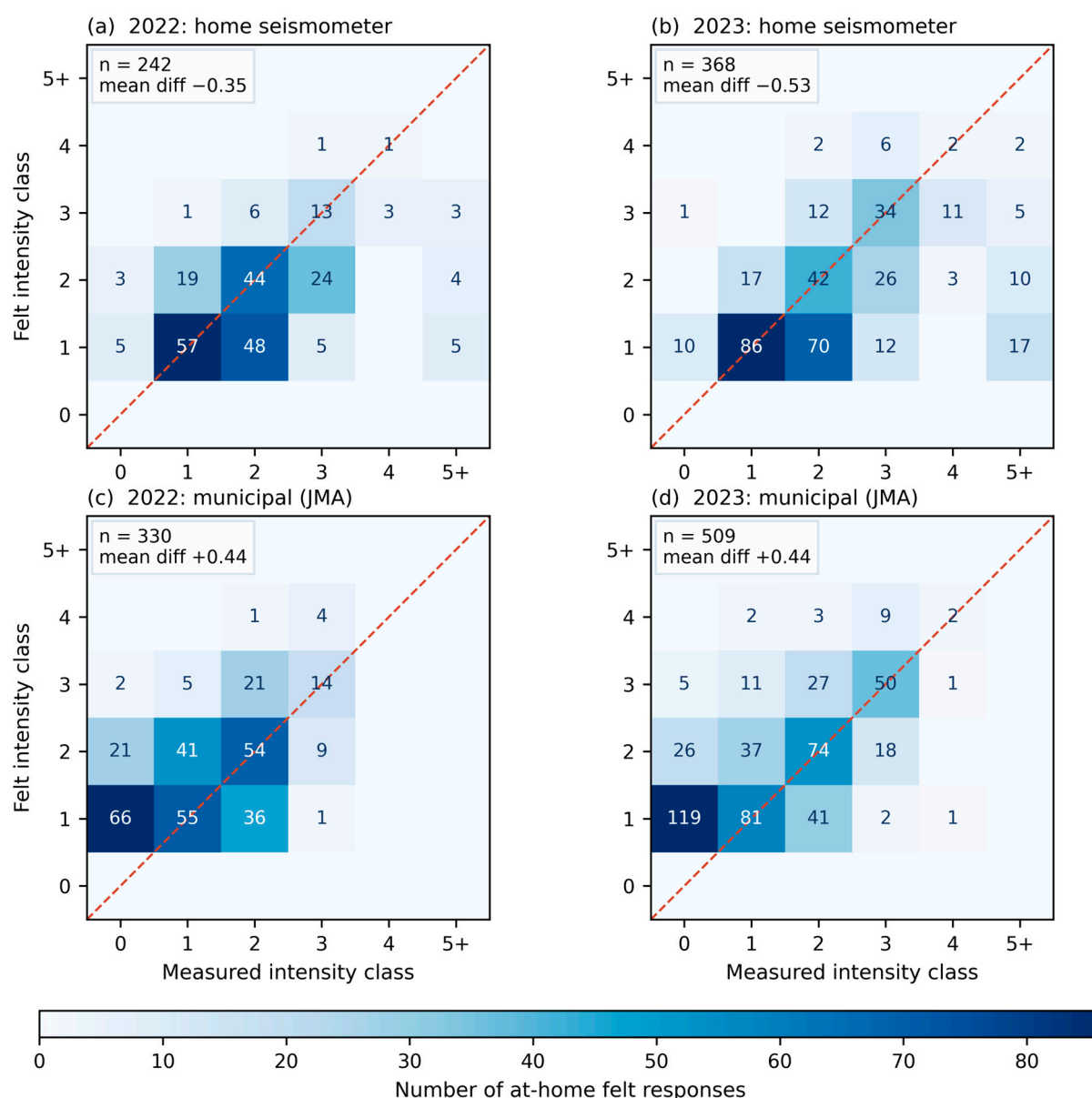


Figure 6. Perceived seismic intensity versus measured intensity classes for at-home felt responses: against the home seismometer (**top**) and municipal JMA seismic intensity (**bottom**), for 2022 (**left**) and 2023 (**right**). The dashed line marks exact agreement.

Conversely, relative to official municipal JMA seismic intensities, exact class matches were 37% and 41%, respectively. Perceived seismic intensity exceeded official municipal JMA seismic intensity by an average of +0.44 classes over the 839 at-home felt responses. This excess was reproduced across two annual cohorts and distinct seismic events, but, as shown below, it is not uniform across shaking levels. Table A1 (Appendix A) reconciles this and every other estimate of bias or excess reported in this paper.

These results show that the participants generally reported greater shaking than indicated by official municipal bulletins but less than their in-home instruments physically recorded. Perceived seismic intensity therefore lies between the free-field regional baseline and the measured indoor environment and is closer to the indoor measurement. This position is partly implied by the definition of the JMA scale, whose classes are described in terms of what people indoors feel (Section 2.7); it is reported here as an observation, not as evidence of perceptual accuracy. That perceived intensity is closer to the indoor measurement is supported by the mean absolute differences (0.56 classes against indoor

intensity versus 0.72 classes against municipal intensity). Among the 569 discrete felt responses, 219 were closer to the indoor measurement, whereas only 137 were closer to the municipal bulletin (binomial $p = 1.6 \times 10^{-5}$). Rank correlations with the two reference metrics were similar (Spearman's $\rho = 0.57$ against indoor intensity and $\rho = 0.58$ against municipal intensity).

Person-level analyses further support the average excess of perceived intensity over official published metrics. When aggregated within individual participants ($n = 107$), the mean individual perceived intensity exceeded the municipal JMA seismic intensity class by +0.38 (95% confidence interval (CI): +0.26 to +0.49; Cohen's $d = 0.63$; with Student's t , Wilcoxon signed-rank, and sign tests all yielding $p < 10^{-7}$). Using the unrounded municipal instrumental intensity derived to one decimal place from official JMA files does not change this conclusion (+0.36 class excess, 95% CI: +0.26 to +0.46, $p \approx 3 \times 10^{-10}$). The pattern also remains under highly conservative sub-threshold censoring bounds (+0.30 class excess, $p \approx 5 \times 10^{-9}$).

The over-reporting bias was similar among non-hosts (person-level mean excess: +0.33, $p \approx 2 \times 10^{-3}$) and active sensor hosts (+0.40, $p \approx 7 \times 10^{-8}$), with no detectable difference between the two groups ($p = 0.59$). Because non-hosts received the same nearest-station and early-warning intensity information as hosts, and the only difference in the information they received was the intensity and waveform recorded inside the hosts' own homes, this similarity gives no indication that feedback on indoor measurements is the source of the observed over-reporting bias. Both groups received the station and early-warning values, so the present design cannot separate any effect of that shared exposure from the bias.

The systematic over-reporting was concentrated at weak-shaking thresholds. Relative to the unrounded municipal instrumental intensity, the mean person-level excess declined from +1.10 classes for sub-threshold events (intensities below 0.5) through +0.55 at 0.5–1.0 to +0.31 and +0.09 classes within the 1.0–1.5 and 1.5–2.0 intensity ranges, and then to −0.07 classes for events exceeding intensity 2.0 (Table 4; the corresponding means over responses are given in Table A1). Below municipal intensity 1.0, this gap cannot be negative, because a respondent who felt an event reports at least class 1: none of the 324 felt reports in that range has a negative gap. Above 1.0, the constraint no longer binds, with 37% to 46% of gaps negative, so the estimates at 1.0 and above are the interpretable ones (Table 4). The difficulty of resolving low intensities from felt reports is recognized independently of the present data [32]. The three-way ordering was also reproduced using continuous unrounded reference values. Among 2023 sensor hosts, the mean nearest-station instrumental intensity (1.16) was lower than human perceived intensity (1.70), which was in turn lower than measured indoor seismic intensity (1.96; $n = 334$).

Although municipal JMA seismic intensity is the metric that citizens directly encounter, it is defined as the maximum observed value across an entire administrative boundary. This definition may make the regional baseline higher than the localized conditions at a participant's residence. To control for this spatial factor, we re-evaluated the three-way ordering using the single public observation station geographically closest to each resident's address instead of the municipal baseline.

We calculated linear distances between participant coordinates and official public station networks. The station identified geometrically was compared with the station recorded in our real-time delivery logs, showing strong baseline agreement (73% exact match in 2022 and 61% in 2023) and small differences in corresponding instrumental values (mean differences of +0.011 and +0.009; mean absolute errors of 0.06 and 0.09).

Audit logs showed that the station routing table used by the real-time telemetry system inadvertently contained several public stations that had been officially decommissioned before the study (affecting 10 out of 92 residents in 2023). Because published

instrumental datasets do not exist for inactive stations, using the system delivery logs directly would introduce a systematic downward bias for these participants. We therefore use the geometrically computed active station network across both cohorts. Restricting the validation model to system logs for active stations yields identical results.

Against this localized nearest-station reference, the mean individual perceived excess increased to +0.48 classes (95% CI: +0.38 to +0.58; Cohen's $d = 0.91$; $n = 107$ persons; split as +0.42 in 2022 and +0.48 in 2023). This increase is expected because the reference shifts from a municipality-wide maximum to a single localized surface instrument. The three-way ordering was retained under this localized control framework: nearest-station intensity (1.18) < human perceived intensity (1.66) < indoor seismic intensity (1.90; $n = 568$ person-events).

4.2. Site Amplification Is Large, Home-Specific, and Stochastically Variable (H1)

For the 106 site-years that accumulated at least 15 seismic events, measured indoor seismic intensity correlated strongly with municipal JMA seismic intensity within each fixed structure (median within-site $\rho = +0.72$). However, these models show a substantial site-specific mean excess of +0.55 classes (median) relative to the discrete municipal baseline, with site-year means spanning from -0.1 to $+1.3$ classes (Figure 7). This 1.4-class spread is consistent with variations in building materials, floor height, and localized geological site amplification factors.

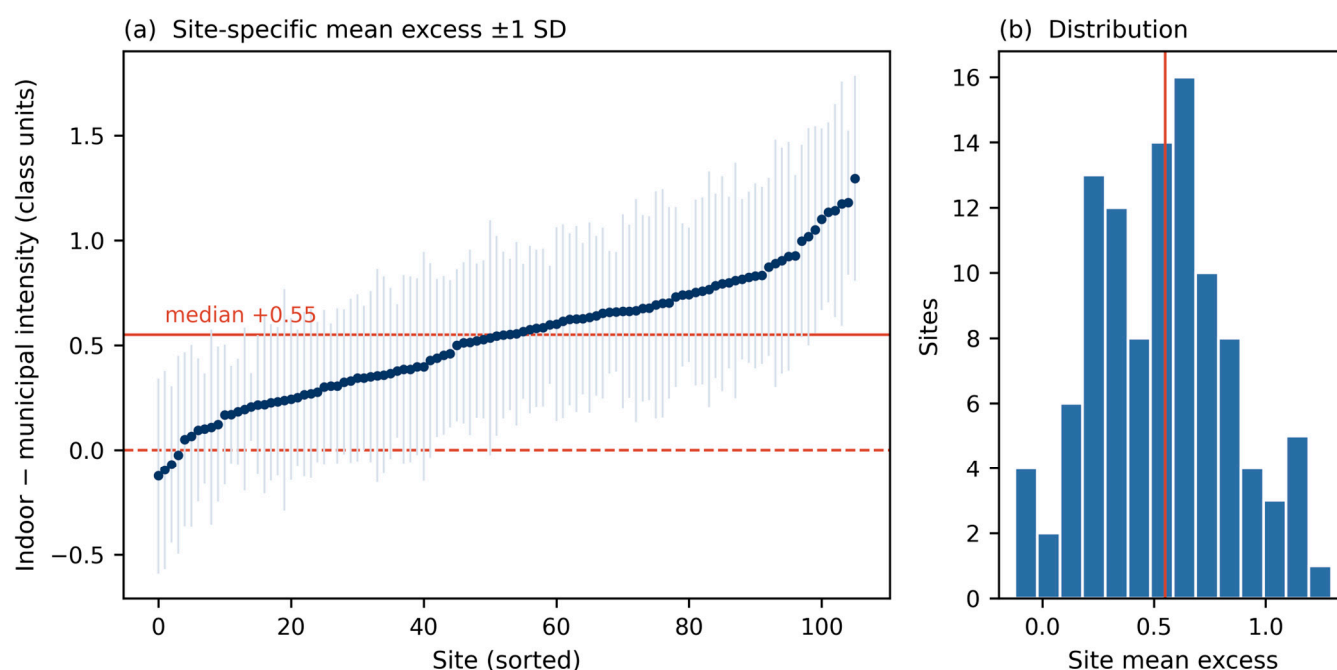


Figure 7. Site-specific mean excess of indoor over municipal JMA seismic intensity for 106 site-years (at least 15 events each): sorted means ± 1 standard deviation (SD) (a) and their distribution (b). The municipal reference here is the intensity class; against the observed unrounded instrumental reference, the median excess is +0.34 (see Section 4.2). The solid line marks the median of the site means (+0.55) in both panels, and the dashed line in (a) marks zero excess.

Because public stations registering below intensity class 1 (instrumental intensity 0.5) are omitted from standard JMA bulletins, this discrete class-referenced metric treats sub-threshold municipal ground motion as zero, representing an analytical upper bound. Re-estimating this amplification factor at the household level against continuous, observed unrounded municipal instrumental intensities yields a localized median excess of +0.34

classes (across 76 actively recording hosts; with strict sensitivity bounds evaluated at +0.15 to +0.53). We therefore use this +0.34 class value as the baseline estimate of structural amplification. This household-level estimate averages over all delivered events, most of which were not felt, whereas the person-level excess in Section 4.1 averages over at-home felt responses only. The two values are therefore not directly comparable. Across the 570 responses for which all three metrics are available, the three-way ordering is retained: municipal 1.24 < perceived 1.66 < indoor 1.91, the position of perceived intensity being partly implied by the definition of the scale (Section 2.7).

The event-to-event stochastic variability recorded within an individual structure (median within-site SD: 0.41) matches or exceeds the systematic variation measured between structures (between-site SD: 0.30). Restricting the comparison to events at municipal class 1 or above gives a median within-site SD of 0.43 (103 site-years with at least five such events). Structural amplification therefore cannot be modeled as a fixed multiplier. At a constant municipal JMA seismic intensity, the indoor seismic excess increases weakly with epicentral distance ($\rho = +0.14$ for class 2 events), a pattern consistent with the long-period resonant response of residential structures to distant deep-source earthquakes; perception data have been used elsewhere to recover the natural frequencies of buildings [33], and the perception of tremors in tall buildings has been modeled explicitly [34].

The structural amplification pattern remained similar when the nearest public station was substituted as the baseline reference (104 site-years; median excess: +0.37 classes; between-site SD: 0.27). Two of the 106 site-years were recorded at a participant's workplace rather than at home; excluding them leaves the median excess (+0.55), the median within-site SD (0.41) and the range unchanged, and changes the between-site SD from 0.30 to 0.31.

4.3. Probability of Feeling a Seismic Event Is Associated with Local Intensity Independently of Magnitude and Distance (H2)

To evaluate factors associated with noticing an earthquake, we fitted logistic regression models with two-way cluster-robust standard errors, clustered simultaneously on participants and on earthquakes, across 5462 distinct delivery responses; the model that adds floor height uses 5322 responses (Table 5). The odds of a resident noticing a seismic event increased 2.42-fold per municipal JMA seismic intensity class (95% CI: 2.00 to 2.91; Model B2), independently of magnitude and epicentral distance. Epicentral distance acted in the opposite direction (odds ratio (OR): 0.33 per log-unit), and magnitude contributed a further positive effect (OR: 2.60 per unit) after controlling for intensity. Clustering on earthquakes as well as on participants inflated the standard errors of the event-level predictors by factors of 1.3 to 1.4, but all of them remained significant at the 0.01 level. Per standard deviation of each predictor, the odds ratios are 2.00 for municipal intensity, 2.18 for magnitude and 0.38 for epicentral distance, so local intensity is associated with feeling an event alongside, not above, magnitude and distance.

Table 5. Logistic models for the probability of feeling an earthquake, with two-way cluster-robust standard errors. Model B1 excludes floor height; Model B2 adds it and uses the subset for which floor height was reported.

Predictor	Model B1 OR [95% CI]	<i>p</i>	Model B2 OR [95% CI]	<i>p</i>
Municipal JMA seismic intensity (+1 class)	2.41 [1.99–2.92]	<0.001	2.42 [2.00–2.91]	<0.001
In epicentral distance	0.35 [0.28–0.43]	<0.001	0.33 [0.27–0.42]	<0.001
Magnitude (M_j , +1)	2.55 [1.87–3.47]	<0.001	2.60 [1.91–3.52]	<0.001
Main floor ≥ 3 (Model B2 only)	—	—	1.46 [0.97–2.18]	0.067
Seismometer host	1.86 [1.31–2.62]	<0.001	1.81 [1.28–2.55]	<0.001
Study period 2023	0.62 [0.44–0.87]	0.006	0.57 [0.41–0.80]	0.001

Note: Model B1, $n = 5462$ delivery responses; Model B2, $n = 5322$. Standard errors are clustered simultaneously on participants and on earthquakes. OR: odds ratio.

The magnitude effect may reflect longer wave durations or long-period spectral content, which are more characteristic of larger ruptures and may affect human perception in ways not fully captured by automated instrumental intensity algorithms. Predicted probabilities rise from 8% at municipal JMA class 0 to 56% at class 3 for sensor hosts, and from 5% to 40% for non-hosts (Figure 8).

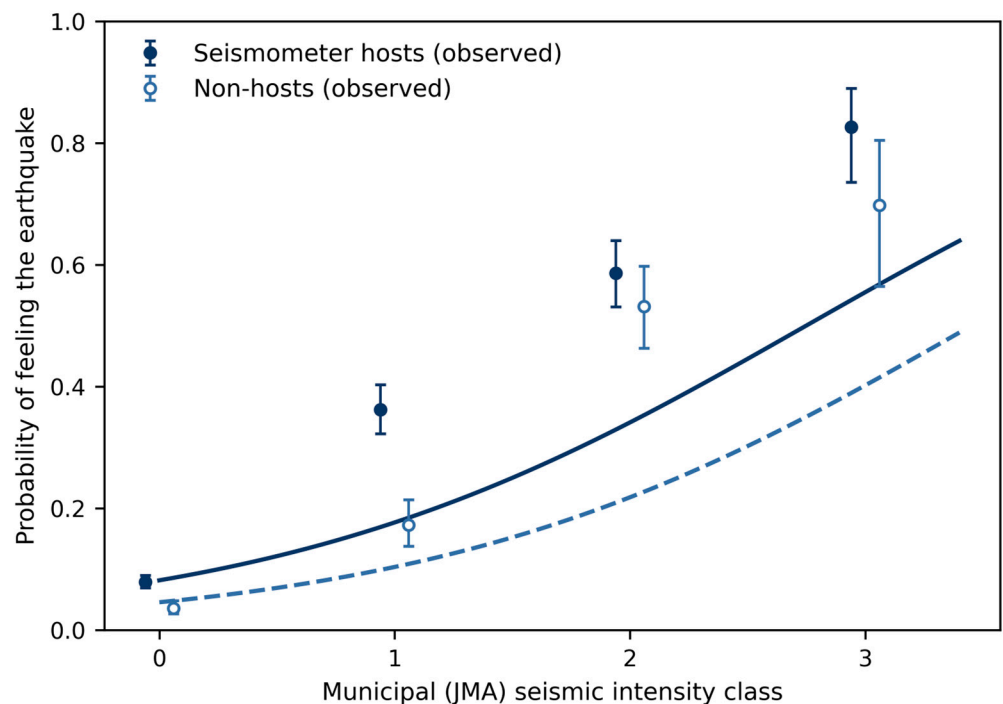


Figure 8. Probability of feeling an earthquake versus municipal JMA seismic intensity: observed proportions (95% Wilson confidence intervals) and model predictions at covariate means for hosts (solid curve) and non-hosts (dashed curve).

Living primarily on the third floor or higher of a residential building was associated with approximately 1.5-fold higher odds of feeling an event, but the estimate is not statistically resolved (OR: 1.46; 95% CI: 0.97 to 2.18; $p = 0.067$) and was directionally consistent with the physical instrumental amplification recorded by upper-floor sensors and with macroseismic surveys that report higher intensities on upper floors and in taller buildings [30,35]. Active sensor hosts also exhibited higher reporting odds (OR: 1.81; 95% CI: 1.28 to 2.55) than questionnaire-only participants, a pattern that may reflect greater personal salience or sustained civic motivation within the device-hosting panel.

4.4. No Detectable Perceptual Calibration with Iterative Personalized Feedback in This Panel (H3 Not Supported)

The core design of this field experiment hypothesized that exposing citizens to dozens of real-time feedback cycles, in which their home's exact instrumental intensity was reported alongside official benchmarks, would gradually reduce the gap between perceived seismic intensity and objective measurements. A naive time-series evaluation appears to support this hypothesis, as the raw signed gap (perceived minus municipal class) decreases linearly over the 2023 calendar timeline.

However, macroseismic analysis indicates that this apparent learning effect reflects changes in the earthquakes that occurred over the period rather than learning. When each participant's sequential felt reports are ordered by cumulative personal felt experience, the raw signed gap does not decline. It rises, from +0.33 classes over the first three reports to +0.66 over the thirteenth and later (Figure 9, panel a; Spearman's rank correlation +0.15), which is the opposite of calibration. After event demeaning to control for changes in the composition and source parameters of sequential earthquakes, the rise is about halved and is no longer detected (rank correlation +0.06, $p = 0.063$; Figure 9, panel b). In neither case does the gap move toward the measured values.

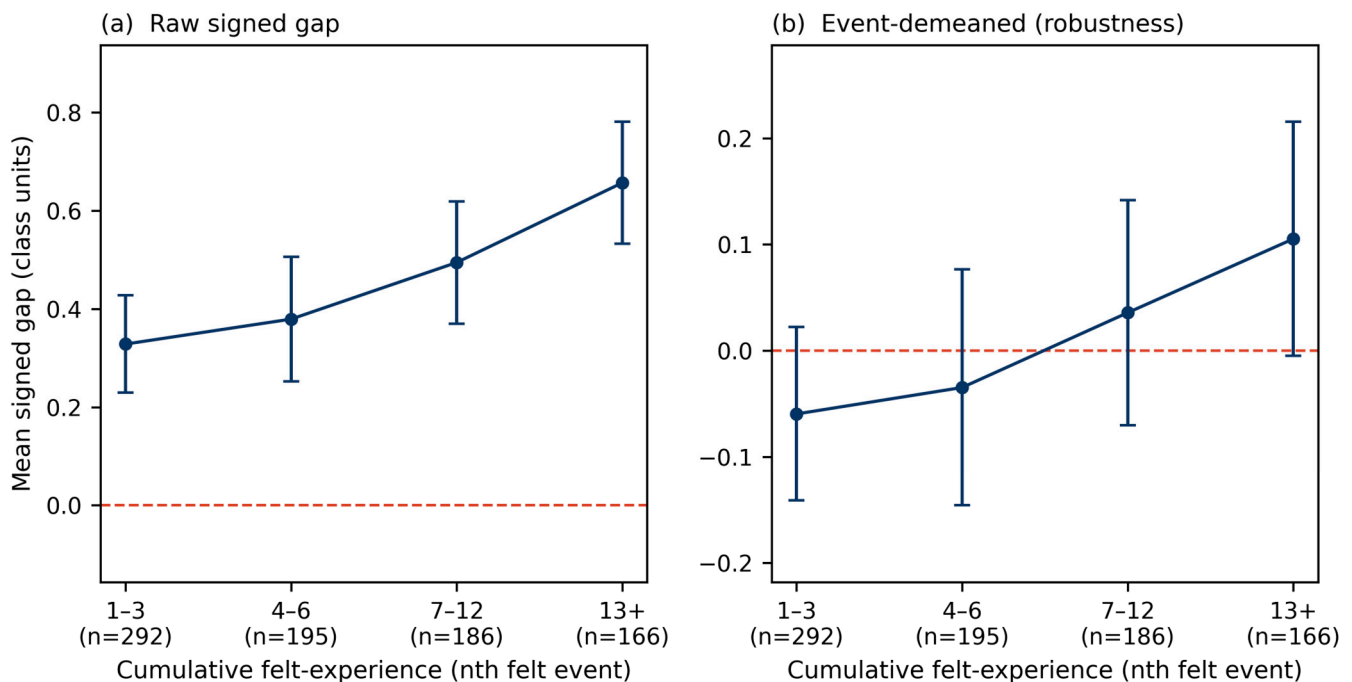


Figure 9. Test of feedback calibration: mean signed gap (perceived minus municipal JMA seismic intensity) by cumulative felt experience, raw (a) and event-demeaned (b); the mixed-model experience effect is $\beta = +0.048$ ($p = 0.135$).

To test this formally, we fitted linear mixed-effects models across 839 felt reports, specifying crossed random intercepts for the 107 participants and for the earthquakes (Table 6, Figure 10). Because the event random effect absorbs the composition of events, event demeaning is no longer required, and municipal intensity is retained as a covariate. The model indicated that the experience effect was not detected ($\beta = +0.048$ per standard deviation of cumulative events; $p = 0.135$), with no significant interaction at stronger shaking levels, whose sign was in any case opposite to calibration ($p = 0.093$). Random-slope models, and models in which exposure is measured by the cumulative number of messages delivered rather than by felt reports, produced the same null conclusion. At this sample size, the smallest effect detectable with 80% power is 0.09 classes per standard deviation of experience, so effects smaller than that cannot be excluded. Exposure to the feedback was itself limited: respondents had consulted the delivered intensity before submitting their own assessment in only 6% to 10% of felt responses (Section 2.4). The delivery log records that a message was sent, not that it was opened, so what is tested here is repeated availability of feedback rather than its verified use.

Table 6. Linear mixed-effects model for the signed gap between perceived and municipal JMA seismic intensity, with crossed random intercepts for participants and for earthquakes.

Fixed Effect	β	SE	p
Intercept	+0.438	0.068	<0.001
Cumulative felt experience (per SD)	+0.048	0.032	0.135
Municipal JMA seismic intensity (centered)	−0.592	0.031	<0.001
Study period 2023	−0.056	0.082	0.492
Variance: person/event/residual	0.043/0.029/0.371	ICC 0.098/0.066	—
Interaction model: experience $\times$ (intensity ≥ 2)	+0.079	0.047	0.093
Random-slope model: experience	+0.044	—	0.397

Note: $n = 839$ at-home felt reports from 107 participants and 89 earthquakes. Estimated by restricted maximum likelihood (statsmodels MixedLM); the model is given in Section 3. Cumulative felt experience is standardized and municipal intensity is centered on its sample mean, so the intercept is the expected gap in 2022 at mean experience and mean municipal intensity. ICC: share of the total variance attributable to participants and to earthquakes. SE: standard error.

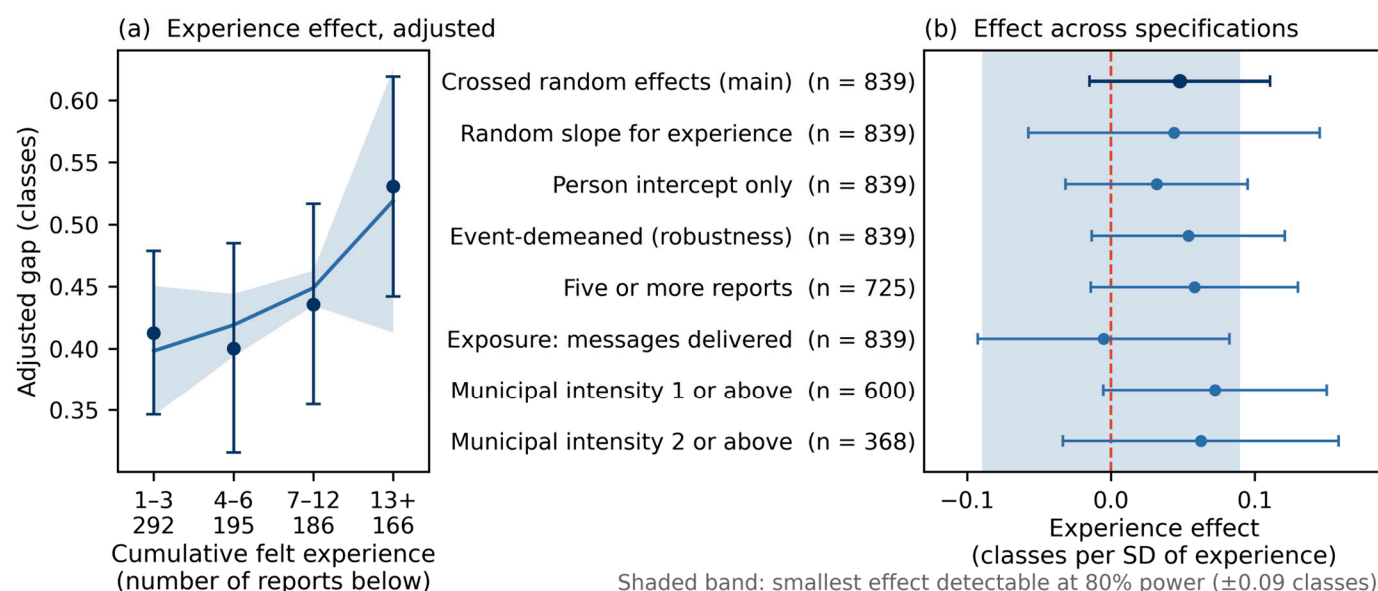


Figure 10. The experience effect under the main model (Table 6). (a) Signed gap between perceived and municipal JMA seismic intensity by cumulative felt experience, adjusted for municipal intensity, study period, participant and earthquake. Points are adjusted means with 95% confidence intervals, and the line is the model fit with its 95% band. (b) The experience effect under eight specifications, with 95% confidence intervals. The shaded band marks the smallest effect this sample can detect with 80% power (± 0.09 classes). Every interval includes zero and every estimate falls inside the band, so these data do not separate a small positive effect from no effect.

No association was detected between how often an individual reported feeling an earthquake and the size of that individual's over-reporting bias (rank correlation +0.09, $p = 0.37$), so the bias is not confined to the participants who notice minor shaking most often. Together with the longitudinal result, the person-level over-reporting bias of about +0.38 classes relative to official bulletins did not decline across the study periods in this panel, although personalized feedback was available after every target earthquake (Section 2.7).

4.5. Nonresponse, Attrition, and Selection

Because the panel is self-selected, we quantified the paths by which the analyzed sample could differ from the delivered sample (Table 7, Figure 11).

Response rates varied widely between individuals. Across 118 residents and 8054 deliveries, the overall response rate was 71%, but the median rate per participant-year was 81%, and the interquartile range spanned 62% to 94% (Figure 11a); the headline rates of 80.6% in 2022 and 67.6% in 2023 are therefore a mixture of participants who answered almost every time and participants who rarely answered.

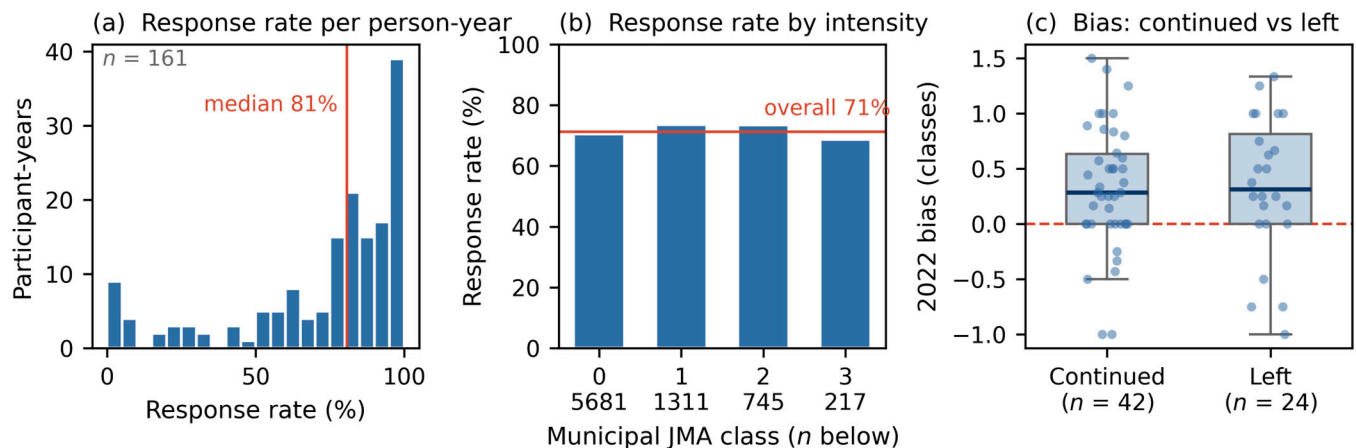


Figure 11. Nonresponse, attrition, and selection. (a) Distribution of the response rate for each participant-year ($n = 161$); the vertical line marks the median (81%). (b) Response rate by municipal JMA seismic intensity class, with the number of deliveries below each class; the horizontal line marks the overall rate (71%). The rank correlation between intensity and responding is $+0.023$ ($p = 0.038$). (c) Mean over-reporting bias in 2022 for the participants who continued into 2023 and for those who did not; boxes show the median and quartiles, and points show individual participants. The two groups do not differ detectably ($+0.324$ versus $+0.337$ classes, Cohen's $d = -0.02$, $p = 0.936$).

Nonresponse was almost independent of the strength of shaking. Response rates were 70.5% at municipal intensity class 0, 73.5% at class 1, 73.4% at class 2 and 68.7% at class 3. The rank correlation between intensity and responding is $+0.023$ ($p = 0.038$): statistically detectable at this sample size but negligible in magnitude and not monotonic (Figure 11b). Weak events are therefore not preferentially missing from the analyzed responses.

Of the 69 residents enrolled in 2022, 43 (62%) continued into 2023, and 26 (38%) did not. Those who continued and those who left did not differ detectably on any of the six attributes we could compare: 2022 response rate (83.9% versus 75.2%, $p = 0.125$), preparedness score (7.86 versus 7.08, $p = 0.177$), number of stories (2.81 versus 3.08, $p = 0.514$), main floor of residence (2.29 versus 2.80, $p = 0.415$), the proportion hosting a seismometer (79% versus 62%, $p = 0.165$), and, most directly relevant, the 2022 over-reporting bias itself ($+0.324$ versus $+0.337$ classes, Cohen's $d = -0.02$, $p = 0.936$; Figure 11c). Attrition was therefore not selective on the outcome.

Within respondents, the propensity to answer was unrelated to the size of the bias. Splitting person-years at the median response rate gives mean biases of $+0.418$ and $+0.291$ classes ($p = 0.204$), and the rank correlation between individual response rate and individual bias is $+0.033$ ($p = 0.730$).

We emphasize that the absence of detectable selection within the panel does not make the panel representative. Selection at recruitment is a separate matter and is stated in

Section 2.7. Global analyses of who contributes to crowdsourced intensity systems report that participation is far from uniform [17,18,36–38], so the absence of detectable selection here should be read as a property of this panel rather than of citizen-contributed intensity data in general.

Table 7. Paths by which the analyzed sample could differ from the delivered sample, and the quantity estimated for each.

Path	Quantity	Result
Nonresponse by level of shaking	Response rate at municipal class 0/1/2/3	70.5%/73.5%/73.4%/68.7% (rank correlation +0.023, $p = 0.038$)
Attrition	2022 participants continuing into 2023	43 of 69 (62%)
Attrition by attributes	Six attributes compared	No detectable difference (smallest $p = 0.125$)
Attrition by outcome	2022 over-reporting bias, continuers versus leavers	+0.324 versus +0.337 classes ($d = -0.02$, $p = 0.936$)
Selection within respondents	Individual response rate versus individual bias	rank correlation +0.033 ($p = 0.730$)
Ceiling of the preparedness scale	Preparedness score (maximum 10)	Mean 7.6 (2022) and 7.4 (2023); 55% to 58% scored 8 or more

Note: Nine deliveries at municipal class 4 are omitted from the first row. The comparison of bias between continuers and leavers uses the 42 continuers and 24 leavers who made at least one at-home felt report in 2022.

4.6. Sensor Hosts Versus Non-Hosts

Nineteen participants in 2022 and 36 in 2023 did not host a seismometer. They received the same nearest-station intensity and the same early-warning estimate as hosts, and the only difference in the information they received was the intensity and waveform recorded inside a host's own home. Because hosts were not randomly assigned and also hosted the device, the comparison does not isolate the effect of the in-home measurement (Section 2.7).

The magnitude of the over-reporting bias did not differ between the two groups: the person-level mean excess over municipal JMA seismic intensity was +0.33 classes among non-hosts and +0.40 among active hosts, with no detectable difference ($p = 0.59$). The probability of reporting an event as felt, however, was higher among hosts (odds ratio 1.81, 95% CI 1.28 to 2.55; Table 5).

Hosts were thus more likely to report an event as felt but did not differ detectably in how they scaled what they felt. We note that this contrast is observational: hosts were not randomly assigned, and the two groups may differ in ways we did not measure.

5. Discussion

This section sets out our interpretation of the results in Section 4. Except where we refer to a specific analysis, the statements here are our reading of the observations rather than findings established by the experiment.

5.1. What the Results Establish

Four observations are established by the analyses in Section 4. First, indoor seismic intensity exceeded municipal JMA seismic intensity by a median of about one-third of a class, and the site-year means of the excess against the municipal class spread over 1.4 classes. Second, within a single dwelling the excess varied from event to event with a median standard deviation of 0.41 classes, and this variability persisted when the weakest events were excluded (Section 4.2). Third, the probability of feeling an earthquake was

associated with local intensity (odds ratio 2.42 per class) independently of magnitude and epicentral distance. Fourth, the individual over-reporting bias relative to municipal intensity did not decline with cumulative experience of immediate feedback over up to 96 earthquakes, within the bound stated in Section 2.7. The position of perceived intensity between the municipal and indoor values is not among these observations because it is partly implied by the definition of the JMA scale (Section 2.7); the first observation compares two instrumental values and is not affected by that definition. These observations rest on three features of the data rather than on the novelty of the design: perceived, indoor measured and published intensity were recorded for the same person and the same dwelling; the delivery log fixed the denominator, so the earthquakes that were not felt are counted; and the amplification is resolved at the level of individual homes.

Two further statements hold only within a stated range. The excess of perceived over municipal intensity was confined to weak shaking: it is +0.31 classes at municipal intensities of 1.0 to 1.5 and is not statistically resolved above 1.5 (Table 4). And the association between the main floor of residence and the probability of feeling an event was in the expected direction but was not statistically resolved in this sample.

5.2. What Municipal Bulletins Miss and Where Perceived Intensity Lies

The comparison of the two instrumental values shows that municipality-level public intensity systematically under-represents measured indoor seismic intensity by a median of roughly one-third of an intensity class and by more than one class for the most amplified residential structures (site-year means up to +1.3 against the municipal class), and that within a single home the difference changes from event to event. Perceived seismic intensity lies between the two values, but because this position is partly implied by the definition of the scale (Section 2.7), these data do not show how much of the discrepancy perception recovers. From a risk-communication perspective, a resident's report that they "felt stronger shaking than the officially announced intensity" may therefore reflect a genuine difference between municipal and indoor conditions rather than only perceptual error. This interpretation is consistent with, though not established by, the finding that the participants' perceived seismic intensity was within a single discrete class of measured indoor seismic intensity in 94% to 96% of observed instances (Section 4.1), an agreement that is itself partly implied by the definition.

5.3. Significance of the Null Perceptual Calibration Result

Hypothesis H3 provided the analytical basis for our real-time feedback architecture and reflects an intuitive assumption in risk communication: continuous exposure to objective data may gradually refine subjective human judgment. However, no calibration was detected within this highly prepared volunteer panel, exposed to up to 96 feedback cycles over mostly weak shaking (Section 2.7); effects smaller than 0.09 classes per standard deviation of experience cannot be excluded (Section 4.4), and the result does not show that perception cannot be trained in other groups or at stronger shaking.

What this implies for presenting the measured indoor intensity is discussed in Section 5.8. We note that within the limited subsample of stronger shaking (municipal class ≥ 2), where the mean gap is close to zero, the experience effect is not resolved (Section 5.7); a conclusion for this range of shaking will require a larger longitudinal dataset containing high-intensity seismic events.

5.4. Bridging Measurement and Mitigation Paradigms

First, for indoor seismic-hazard assessment, the measured home-specific amplification (median +0.34 class against unrounded municipal values, with site-year means up to +1.3 against the municipal class) indicates how far indoor shaking can depart from free-

field hazard estimates, which operate at the ground-surface level; whether it can inform adjustments to them is discussed in Section 5.8.

Second, for the “last mile” of earthquake early warning (EEW), the source-based localized intensity predictions provided by our automated alert system correlated only moderately with the actual indoor seismic intensities measured post-event ($\rho = 0.34$; based on 4951 deliveries to residents for the target earthquakes of both study periods). When predicted instrumental intensity was below 0, the automated alert message stated “0 or less.” These entries were initially omitted as missing data; however, restricting the analysis to the numeric subset attenuated the correlation to $\rho = 0.08$, an artifact of dropping the weak-prediction cases. Treating these low-intensity data points as left-censored ranks increased the correlation to $\rho = 0.39$ in 2022 and 0.31 in 2023. This finding highlights the complementary roles of localized sensing and EEW: advance warning forecasts incoming shaking, whereas the in-home sensor records the shaking that actually reached the inside of the dwelling.

Third, for public risk communication, no perceptual calibration was detected in this panel (Sections 2.7 and 4.4), and whether ordinary residents would calibrate is not known (Section 5.5); disseminating observed indoor measurements, rather than relying on residents to adjust their own assessments, is therefore an option to be tested rather than a recommendation (Section 5.8). The extent to which this continuous information fosters long-term household earthquake preparedness and protective actions is investigated in an upcoming companion paper (in preparation).

5.5. Establishing a Baseline from Highly Prepared Individuals

Our longitudinal panel consists of members of safety-volunteer networks with an exceptionally high level of disaster preparedness (near-ceiling baseline). In this panel, participants neither showed detectable calibration of their subjective perception over time nor frequently consulted the delivered intensity data before submitting their subjective judgments (in only 6% to 10% of felt responses). Motivation might be expected to favor learning, but we do not extend the result beyond this panel: ordinary residents differ in preparedness and in attention to feedback; they might over-report more, or, being less accustomed to emergencies, might judge more cautiously for fear of misjudging, and our data determine neither the size of their over-reporting bias nor whether they would calibrate. The panel provides a baseline for a prepared and motivated group, against which samples of the general population can be compared.

5.6. Relation to Aggregated Frameworks Like DYFI

Broad community aggregation, such as the USGS “Did You Feel It?” (DYFI) system in the United States, provides valuable regional overviews, whereas this study uses Japanese panel data to add a finer layer of localized detail. Specifically, within this panel we quantify individual- and structure-level variation of the kind that these aggregates average over: a person-level over-reporting bias of about +0.38 classes, concentrated at weak shaking (Table 4), a 1.4-class spread of site-year amplification, and event-to-event variation within a home (median standard deviation 0.41 classes).

The ability to track a specific panel of individuals, extending beyond traditional anonymous cross-sectional surveys, was important for assessing individual differences in sensitivity and whether the individual bias changed across multiple events. Because our questionnaire intentionally adopted shaking vocabulary from DYFI, findings based on the JMA scale can be compared with and may help enrich existing macro-scale systems internationally, linking insights from US and Japanese frameworks. The relation between felt and instrumental intensity is itself known to differ between regions [39], so such comparisons require the regional baseline to be stated.

5.7. Limitations and Robustness Check

Several structural limitations of this study warrant explicit mention. The sample size is modest and self-selected from a highly disaster-prepared population stratum; the observed seismic events were predominantly minor (JMA classes 0–3), with only 368 felt reports at municipal class 2 or above, so the calibration result is bounded by this range; and both felt reports and geographic locations rely on participant self-disclosures. Exposure to feedback is measured by cumulative felt reports or by messages delivered, and its receipt was not verified (Section 2.7). Moreover, municipal intensity functions as a coarse spatial proxy in areas without active station triggers.

The primary conclusions were unchanged when the analyses were repeated using unrounded, continuous instrumental intensities derived from device readings and municipal data from JMA bulletins (Section 3). This robustness analysis reproduced the main results with equivalent or enhanced statistical significance. Because the continuous device values lie almost exactly at the discrete class centers (with a mean within-class offset of +0.005), analytical artifacts arising from rounding are unlikely to explain the findings. The conclusions also do not depend on the events closest to the noise floor of the instruments. Repeating the analyses on the subset for which the municipal intensity was class 1 or above (600 felt reports from 104 participants; 1594 delivery responses) leaves every result in place: perceived seismic intensity still exceeds municipal JMA seismic intensity (+0.111 classes, 95% CI +0.022 to +0.199, $p = 0.015$), the probability of feeling an event remains associated with local intensity (odds ratio 2.36, 95% CI 1.84 to 3.03, $p < 0.001$), and the experience effect remains statistically unresolved (0.072 classes per standard deviation, standard error 0.040, $p = 0.068$). Restricting further to class 2 and above (368 felt reports; 671 delivery responses) gives an odds ratio of 2.16 (95% CI 1.48 to 3.14) and an experience effect of 0.062 ($p = 0.202$), while the person-level excess over municipal intensity is no longer positive (−0.091, $p = 0.057$), as expected from the intensity dependence reported in Section 4.1. We cut on the published municipal intensity rather than on the device readings, so that the subset is defined by an external reference and not by the quantity under study.

Sub-threshold censoring of public station values affects the estimates made against unrounded municipal values (the person-level excess in Section 4.1, Table 4, the household-level amplification in Section 4.2 and the nearest-station comparisons in Sections 4.1 and 4.2), which we bound using the sensitivity values given in Section 3. Finally, although our confirmatory models carry crossed random intercepts for participants and for earthquakes, and the logistic models use standard errors clustered on both participants and earthquakes, unobserved event-by-person interactions cannot be entirely ruled out.

5.8. Implications and What Remains to Be Shown

This section sets out what we take the results to imply. Except where a specific analysis is cited, these are our readings rather than findings established by the experiment.

Indoor seismic-hazard assessment: What the study provides is a distribution: the excess of indoor over municipal instrumental intensity has a median of about one-third of a class against unrounded municipal values, its site-year means against the municipal class spread over 1.4 classes, and within a single dwelling the intensity varies from event to event with a median standard deviation of 0.41 classes. Free-field hazard estimates are not intended to describe individual rooms, and our data indicate how wide the gap can be. We have not demonstrated that home-specific amplification improves any particular hazard product; that would require the amplification to be predicted from building attributes, which our sample is too small to do reliably.

The last mile of earthquake early warning: The source-based intensity predicted for each home by our alerting system correlated only moderately with the intensity

subsequently measured inside that home. The practical implication we draw is that a measurement made in the dwelling carries information that a source-based prediction does not. We have not tested whether delivering that measurement changes what residents do.

Risk communication: The excess of perceived over municipal intensity is concentrated where municipal intensity is low (Section 4.1, Table 4). Where the bulletin reported weak shaking, participants reported feeling more than the bulletin stated; at higher intensities, no difference was resolved. The gap between published and experienced intensity is therefore not uniform, and communication aimed at closing it would be most relevant at the weak end of the scale.

On continuous in-home display: Because repeated feedback did not produce detectable calibration in this panel (Sections 2.7 and 4.4), we do not rely on a display working by training perception. The argument for displaying the measured indoor intensity is the opposite one: if perception does not converge on the measurement, the measurement should be presented directly rather than inferred. This is a proposal, not a result.

Relation to earlier work in Japan: Real-time questionnaires have already been combined with building instrumentation in Japan: occupants of office buildings equipped with seismometers were asked what they felt, and ground-surface intensity was shown not to describe what they experienced [11]. That work established the feasibility of the approach. Questionnaire surveys of indoor effects in Japan have otherwise concentrated on high-rise buildings [10]. What had not been established is whether the same holds in ordinary dwellings, for the same individuals across many earthquakes, against a denominator that includes the events they did not feel, and with feedback on the measured values; these are the respects, set out in the Introduction, in which the present design differs from that work.

Transfer to other settings: Three features of this study are specific to Japan. Where the public network is sparse, the distance to the nearest station is larger, so an in-dwelling measurement gains in relative value; dense low-cost networks have been deployed for this purpose elsewhere [40,41], including in settings with little prior instrumentation [42]. Where earthquakes are less frequent, accumulating comparable numbers of events would require a longer observation period or a larger panel. And because intensity scales differ between countries, a self-assignment format would need question wording matched to the EMS-98 or MMI descriptions rather than to the JMA scale; the relation between felt and instrumental intensity has itself been shown to differ between regions [39].

What remains to be shown: This paper does not examine whether the information changed household preparedness, protective action, or trust in official information, and it does not measure risk awareness. Those questions are addressed in a companion paper using the same panel.

6. Conclusions

By directly comparing three distinct measures of intensity—municipal JMA seismic intensity, perceived seismic intensity, and indoor seismic intensity—for the same individuals and earthquake events across two study periods and 96 seismic events, this study yields three main findings: (1) indoor seismic intensity systematically exceeds official municipal JMA seismic intensity by roughly one-third of an intensity class (median offset of +0.34 against unrounded instrumental values), with site-year means spread over ~1.4 classes against the municipal class and event-to-event variation within a home (median standard deviation 0.41 classes); (2) the probability of feeling an event is associated with immediate local intensity independently of magnitude and distance, while the association with the main floor of residence was not statistically resolved; and (3) in this highly prepared volunteer panel, over mostly weak shaking, repeated personalized feedback, whose

receipt was not verified, did not produce detectable perceptual calibration over time. Perceived seismic intensity lay between the municipal and indoor values, but this position is partly implied by the definition of the JMA scale and is not counted among these findings. Together, these findings quantify how far the shaking measured inside dwellings departs from the municipal value.

From an operational perspective, automated prompts delivered within about 10 min, combined with a 72 h or next-event response threshold, yielded 96% of responses within 72 h while maintaining stable data quality. The panel reported here comprised 76 sensor hosts, with 106 installations across the two study periods; a follow-on deployment under the Tokyo Metropolitan Government program is planned at the scale of about one thousand households and will test whether the co-located design holds at a scale an order of magnitude larger. For hazard mitigation, this approach shifts intensity information from a macro-level regional proclamation toward a micro-level household measurement. What this infrastructure could offer for indoor seismic-hazard evaluation, for the last mile of earthquake early warning, and for risk communication is set out in Section 5.8 as our interpretation rather than as a result of this experiment.

The behavioral dimensions of this experiment are detailed in a companion paper (in preparation).

Author Contributions: Conceptualization, N.H. and T.F.; methodology, T.F.; software and resources, H.T. and D.S.; investigation and data curation, T.F.; formal analysis and visualization, T.F.; writing—original draft preparation, T.F.; writing—review and editing, H.T., D.S. and N.H.; supervision, project administration and funding acquisition, N.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Japan Science and Technology Agency (JST) Belmont Forum Grant Number JPMJBF2006, Japan (Belmont Forum CRA “Resilient Society through Smart-City Technology: Ultra-high-resolution Earthquake Risk Assessment (RESIST)”). The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Institutional Review Board Statement: Ethical review and approval were not required for this study under the University of Tokyo Research Ethics Review Implementation Rule, which requires review only for research covered by the national Ethical Guidelines for Medical and Biological Research Involving Human Subjects issued jointly by the Ministry of Education, Culture, Sports, Science and Technology, the Ministry of Health, Labour and Welfare, and the Ministry of Economy, Trade and Industry of Japan on 23 March 2021 (last revised in 2023) because the study does not seek knowledge that contributes to maintaining or promoting public health, to the recovery of patients from injury or illness, or to improving quality of life, and uses no specimens or genomic information of human origin; it therefore falls outside the scope of those guidelines. The study was non-interventional and consisted of voluntary questionnaire surveys on shaking perception and disaster preparedness. No medical or sensitive personal data were collected, and written informed consent, including consent to publication of the results in anonymized form, was obtained from every participant at enrollment.

Informed Consent Statement: Written informed consent was obtained from all participants at enrollment. Each participant signed (2022 study period) or submitted via web form (2023 study period) a consent document addressed to the principal investigator (Earthquake Research Institute, The University of Tokyo), confirming that they had read the study description, that they participated of their own volition, and that they consented to the publication of the results in academic form on condition that their personal information be protected.

Data Availability Statement: The anonymized panel data presented in this study are available on request from the corresponding author due to privacy and ethical restrictions. The dataset consists of individual-level questionnaire responses linked to seismic records measured inside the participants' own dwellings. Participants consented to publication of the results in anonymized form only, and because a dwelling-level intensity record corresponds to a single address, public release of the linked records could identify a household. The data are therefore not publicly available and are shared with interested researchers on request within the limits of that consent. The publicly available sources used in this study are as follows: JMA bulletins were accessed via the P2P earthquake-information application programming interface (API; p2pquake.net); municipal geocoding used the address search API of the Geospatial Information Authority of Japan (GSI). Station instrumental intensities were compiled from the seismic-intensity files of the JMA Monthly Seismological Bulletin, catalog edition (jma.go.jp) and the appendices of the JMA Monthly Report on Earthquakes and Volcanoes (jma.go.jp). The DYFI background materials cited are at usgs.gov.

Acknowledgments: This research is being carried out in cooperation with Public Interest Incorporated Association SL Disaster Volunteer Network. We thank the SL network members and the Bunkyo community participants; the QuakeSaver team for the sensor platform and data infrastructure; and Y. Ono and the project office for survey operations. We also thank M. Ishise (Yamagata University), S. Sakai (Earthquake Research Institute, The University of Tokyo), and F. Gunji for their cooperation. We thank N. Ohbo for his helpful suggestions regarding the locations of municipal seismic intensity stations. This work also contributes to the Metro-Copernicus Project, "Feel, Measure and Prepare for Shaking: Strengthening the Resilience of Tokyo and Its Citizens against a Capital-Region Earthquake", conducted under the Tokyo Metropolitan Government's Project Proposal System by University Researchers; we thank the members of its survey-design and device-and-network working groups for discussions that contributed to this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1 lists every estimate of bias or excess reported in this paper. The estimates differ because they are taken over different units of analysis (for example, responses, participants, participant-years, hosts or site-years), against different references and over different subsets; within the same unit, reference and subset they agree.

Table A1. Estimates of bias or excess reported in this paper, with the unit of analysis, the reference, the subset and the number of observations.

Quantity	Estimate	Unit	Reference	Subset	<i>n</i>	Reported In
Perceived—municipal	+0.44	Response	Municipal class	At-home felt responses	839	Section 4.1; Figure 6
Perceived—municipal	+0.38	Participant	Municipal class	At-home felt responses	107	Sections 4.1, 4.4, 5.6
Perceived—municipal	+0.36	Participant	Unrounded municipal; censored values at 0.25	At-home felt responses	107	Section 4.1
Perceived—municipal, conservative bound	+0.30	Participant	Unrounded municipal; censored values at 0.49	At-home felt responses	107	Section 4.1
Perceived—nearest station	+0.48 (2022: +0.42; 2023: +0.48)	Participant	Nearest-station instrumental	At-home felt responses	107	Section 4.1
Perceived—municipal, non-hosts/hosts	+0.33/+0.40	Participant	Municipal class	At-home felt responses	40/72	Sections 4.1, 4.6

Perceived—municipal by band: <0.5/0.5–1.0/1.0–1.5/1.5–2.0/≥2.0	+1.10/+0.55/+0.31 /+0.09/−0.07	Participant	Unrounded municipal; censored values at 0.25	At-home felt responses in each band	245/79/153/ 148/220	Table 4; Sections 4.1, 5.1; Abstract
Same bands	+1.00/+0.62/+0.33 /+0.12/−0.09	Response	Unrounded municipal; censored values at 0.25	At-home felt responses in each band	245/79/153/ 148/220	This table
Perceived—municipal, municipal class ≥1	+0.111	Participant	Municipal class	At-home felt responses	600 (104 participants)	Section 5.7
Perceived—municipal, municipal class ≥2	−0.091	Participant	Municipal class	At-home felt responses	368	Section 5.7
Perceived—municipal in 2022, continued/left	+0.324/+0.337	Participant	Municipal class	2022 at-home felt responses	42/24	Section 4.5; Figure 11c
Perceived—municipal, response rate above/below the median	+0.418/+0.291	Participant-year	Municipal class	At-home felt responses	57/57	Section 4.5
Perceived—municipal by cumulative experience: first three/thirteenth and later	+0.33/+0.66	Response	Municipal class	At-home felt responses	292/166	Section 4.4; Figure 9a
Perceived—municipal, model intercept	+0.438	Response (model-based)	Municipal class	2022, mean experience and mean municipal intensity	839	Table 6
Perceived—indoor, 2022/2023	−0.20/−0.26	Response	Indoor class	At-home felt responses of hosts	230/343	Section 4.1
Indoor—municipal, median	+0.55	Site-year	Municipal class	Site-years with at least 15 events	106	Section 4.2; Figure 7 Abstract;
Indoor—municipal, range of site-year means	−0.1 to +1.3	Site-year	Municipal class	Site-years with at least 15 events	106	Sections 4.2, 5.1, 5.2, 5.4, 5.6, 5.8, 6; Figure 7 Abstract;
Indoor—municipal, within-site SD, median	0.41	Site-year	Municipal class	Site-years with at least 15 events	106	Sections 4.2, 5.1, 5.6, 5.8, 6 Abstract;
Indoor—municipal, median	+0.34 (bounds +0.15 to +0.53)	Host	Unrounded municipal; censored values at 0.25	Hosts with at least 10 events	76	Sections 4.2, 5.1, 5.4, 5.8, 6
Indoor—nearest station, median	+0.37	Site-year	Nearest-station instrumental	Site-years with at least 15 events	104	Section 4.2

Note: All estimates are in JMA intensity classes. Perceived seismic intensity is the class assigned by the respondent; “unrounded municipal” is the municipal instrumental intensity compiled from the JMA bulletins, with values below 0.5, which are not published, set to the stated value. Participant-level estimates are averaged within each participant first and then across participants. Participants who hosted a seismometer in one study period and not in the other are counted in both the host and the non-host groups. The bands of Table 4 include the six at-home felt reports for the event without a JMA bulletin. Counts for comparisons involving indoor intensity differ slightly (568 to 573 responses) because each requires a different set of values to be available.

References

- Chen, K.H.; Bossu, R.; Liang, W.-T. Editorial: The power of citizen seismology: Science and social impacts. *Front. Earth Sci.* **2020**, *8*, 610813. <https://doi.org/10.3389/feart.2020.610813>.
- Cochran, E.S.; Lawrence, J.F.; Christensen, C.; Jakka, R.S. The Quake-Catcher Network: Citizen science expanding seismic horizons. *Seismol. Res. Lett.* **2009**, *80*, 26–30. <https://doi.org/10.1785/gssrl.80.1.26>.
- Lawrence, J.F.; Cochran, E.S.; Chung, A.; Kaiser, A.; Christensen, C.M.; Allen, R.; Baker, J.W.; Fry, B.; Heaton, T.; Kilb, D.; et al. Rapid earthquake characterization using MEMS accelerometers and volunteer hosts following the M 7.2 Darfield, New Zealand, earthquake. *Bull. Seismol. Soc. Am.* **2014**, *104*, 184–192. <https://doi.org/10.1785/0120120196>.
- Kong, Q.; Allen, R.M.; Schreier, L.; Kwon, Y.-W. MyShake: A smartphone seismic network for earthquake early warning and beyond. *Sci. Adv.* **2016**, *2*, e1501055. <https://doi.org/10.1126/sciadv.1501055>.
- Kim, A.; Uematsu, H.; Iwamoto, S. Towards constructing regional active citizen seismology network. In Proceedings of Interaction 2015, Tokyo, Japan, 5–7 March 2015; pp. 284–286. (In Japanese)
- Bossu, R.; Roussel, F.; Fallou, L.; Landès, M.; Steed, R.; Mazet-Roux, G.; Dupont, A.; Frobert, L.; Petersen, L. LastQuake: From rapid information to global seismic risk reduction. *Int. J. Disaster Risk Reduct.* **2018**, *28*, 32–42. <https://doi.org/10.1016/j.ijdr.2018.02.024>.
- Atkinson, G.M.; Wald, D.J. “Did You Feel It?” intensity data: A surprisingly good measure of earthquake ground motion. *Seismol. Res. Lett.* **2007**, *78*, 362–368. <https://doi.org/10.1785/gssrl.78.3.362>.
- Quitoriano, V.; Wald, D.J. USGS “Did You Feel It?”—Science and lessons from 20 years of citizen science-based macroseismology. *Front. Earth Sci.* **2020**, *8*, 120. <https://doi.org/10.3389/feart.2020.00120>.
- Sbarra, P.; Tosi, P.; De Rubeis, V. How observer conditions impact earthquake perception. *Seismol. Res. Lett.* **2014**, *85*, 306–313. <https://doi.org/10.1785/0220130080>.
- Nagano, M. Indoor damage at high-rise residential buildings from questionnaire survey after recent earthquakes. In Proceedings of the 17th World Conference on Earthquake Engineering (17WCEE), Sendai, Japan, 27 September–2 October 2021; Paper 2d-0117.
- Yoneda, H.; Yoshizawa, M.; Ohmoto, R.; Kuriyama, A. Results of a real-time questionnaire on the vibration sense of human during earthquakes. *Takenaka Tech. Res. Rep.* **2022**, *78*, 1–9. Available online: https://www.takenaka.co.jp/takenaka_e/services/research/pdf/no78_2022/papers/papers_01.pdf (accessed on 5 October 2026).
- Hirata, N.; Furuya, T.; Tsuruoka, H. Indoor seismic intensity measurement experiment using a plug-in compact seismometer. In Proceedings of the Seismological Society of Japan 2022 Fall Meeting, Sapporo, Japan, 24–26 October 2022; Abstract S19-01. (In Japanese)
- Hirata, N.; Schorlemmer, D.; Tsuruoka, H.; Furuya, T.; Nagao, H. Affordable smart sensor deployments in the Tokyo area: New means of communicating earthquake risk and chances for seismology. In Proceedings of the Abstracts of the Japan Geoscience Union Meeting 2023, Chiba, Japan, 21–26 May 2023; U06-P04.
- Si, H.; Midorikawa, S. New attenuation relationships for peak ground acceleration and velocity considering effects of fault type and site condition. *J. Struct. Constr. Eng. AIJ* **1999**, *64*, 63–70. https://doi.org/10.3130/aijs.64.63_2. (In Japanese)
- Wald, D.J.; Quitoriano, V.; Dengler, L.A.; Dewey, J.W. Utilization of the Internet for rapid community intensity maps. *Seismol. Res. Lett.* **1999**, *70*, 680–697. <https://doi.org/10.1785/gssrl.70.6.680>.
- Tosi, P.; Sbarra, P.; De Rubeis, V.; Ferrari, C. Macroseismic intensity assessment method for web questionnaires. *Seismol. Res. Lett.* **2015**, *86*, 985–990. <https://doi.org/10.1785/0220140229>.
- Hough, S.E. Contributed reports of widely felt earthquakes in California, United States: If they felt it, did they report it? *Front. Earth Sci.* **2021**, *9*, 770445. <https://doi.org/10.3389/feart.2021.770445>.
- Bossu, R.; Landès, M.; Roussel, F.; Steed, R. Felt reports for rapid mapping of global earthquake damage: The doughnut effect? *Seismol. Res. Lett.* **2018**, *89*, 138–144. <https://doi.org/10.1785/0220170129>.
- Japan Meteorological Agency. Method of Calculating the Instrumental Seismic Intensity. Available online: https://www.jma.go.jp/jma/kishou/kyoshin/kyoshin/kaisetsu/calc_sindo.html (accessed on 6 October 2026). (In Japanese)
- Karim, K.R.; Yamazaki, F. Correlation of JMA instrumental seismic intensity with strong motion parameters. *Earthq. Eng. Struct. Dyn.* **2002**, *31*, 1191–1212. <https://doi.org/10.1002/eqe.158>.
- Grünthal, G.; Musson, R.M.W.; Schwarz, J.; Stucchi, M. European Macroseismic Scale 1998 (EMS-98). In *Cahiers du Centre Européen de Géodynamique et de Séismologie, Volume 15*; European Center for Geodynamics and Seismology: Luxembourg, 1998. <https://doi.org/10.2312/ems-98.full.en>.

22. Musson, R.M.W.; Grünthal, G.; Stucchi, M. The comparison of macroseismic intensity scales. *J. Seismol.* **2010**, *14*, 413–428. <https://doi.org/10.1007/s10950-009-9172-0>.
23. Wong, H.L.; Trifunac, M.D. A note on an instrumental comparison of the Modified Mercalli (MMI) and the Japanese Meteorological Agency (JMA) intensity scales, based on computed peak accelerations. *Earthq. Eng. Struct. Dyn.* **1979**, *7*, 75–83. <https://doi.org/10.1002/eqe.4290070107>.
24. Wald, D.J.; Goned, T.; Hortacsu, A.; Loos, S.C. *Developing and Implementing an International Macroseismic Scale (IMS) for Earthquake Engineering, Earthquake Science, and Rapid Damage Assessment*; Open-File Report 2023-1098; U.S. Geological Survey: Reston, VA, USA, 2024. <https://doi.org/10.3133/ofr20231098>.
25. Dengler, L.A.; Dewey, J.W. An intensity survey of households affected by the Northridge, California, earthquake of 17 January 1994. *Bull. Seismol. Soc. Am.* **1998**, *88*, 441–462. <https://doi.org/10.1785/BSSA0880020441>.
26. Naito, S.; Azuma, H.; Senna, S.; Yoshizawa, M.; Nakamura, H.; Hao, K.X.; Fujiwara, H.; Hirayama, Y.; Yuki, N.; Yoshida, M. Development and testing of a mobile application for recording and analyzing seismic data. *J. Disaster Res.* **2013**, *8*, 990–1000. <https://doi.org/10.20965/jdr.2013.p0990>.
27. Anthony, R.E.; Ringler, A.T.; Wilson, D.C.; Wolin, E. Do low-cost seismographs perform well enough for your network? An overview of laboratory tests and field observations of the OSOP Raspberry Shake 4D. *Seismol. Res. Lett.* **2019**, *90*, 219–228. <https://doi.org/10.1785/0220180251>.
28. Zaharia, B.; Grecu, B.; Tolea, A.; Radulian, M. Seismic observations in Bucharest area with a Raspberry Shake citizen science network. *Appl. Sci.* **2023**, *13*, 5646. <https://doi.org/10.3390/app13095646>.
29. Japan Meteorological Agency. *Explanation Table of the JMA Seismic Intensity Scale (February 1996)*; Japan Meteorological Agency: Tokyo, Japan, 1996.
30. Sbarra, P.; Tosi, P.; De Rubeis, V.; Rovelli, A. Influence of observation floor and building height on macroseismic intensity. *Seismol. Res. Lett.* **2012**, *83*, 261–266. <https://doi.org/10.1785/gssrl.83.2.261>.
31. Cameron, A.C.; Gelbach, J.B.; Miller, D.L. Robust inference with multiway clustering. *J. Bus. Econ. Stat.* **2011**, *29*, 238–249. <https://doi.org/10.1198/jbes.2010.07136>.
32. Sbarra, P.; Tosi, P.; De Rubeis, V.; Sorrentino, D. Quantification of earthquake diagnostic effects to assess low macroseismic intensities. *Nat. Hazards* **2020**, *104*, 1957–1973. <https://doi.org/10.1007/s11069-020-04256-6>.
33. Tosi, P.; De Rubeis, V.; Sbarra, P. Earthquake perception data highlight natural frequency details of Italian buildings. *Earthq. Spectra* **2023**, *39*, 1240–1254. <https://doi.org/10.1177/87552930231164183>.
34. Du, W.; Goh, K.S.; Pan, T.-C. Methodology for estimating human perception to tremors in high-rise buildings. *J. Seismol.* **2017**, *21*, 667–682. <https://doi.org/10.1007/s10950-016-9628-y>.
35. Sbarra, P.; Fodarella, A.; Tosi, P.; De Rubeis, V.; Rovelli, A. Difference in shaking intensity between short and tall buildings: Known and new findings. *Bull. Seismol. Soc. Am.* **2015**, *105*, 1803–1809. <https://doi.org/10.1785/0120140341>.
36. Knodel, E.; Wald, D.J.; Quitoriano, V.; Loos, S. The intensity gap: A global analysis of who responds to the crowdsourced “Did You Feel It?” system. *Seismol. Res. Lett.* **2026**, *97*, 2082–2106. <https://doi.org/10.1785/0220250168>.
37. Hough, S.E. Spatial variability of “Did You Feel It?” intensity data: Insights into sampling biases in historical earthquake intensity distributions. *Bull. Seismol. Soc. Am.* **2013**, *103*, 2767–2781. <https://doi.org/10.1785/0120120285>.
38. Boatwright, J.; Phillips, E. Exploiting the demographics of “Did You Feel It?” responses to estimate the felt area of moderate earthquakes in California. *Seismol. Res. Lett.* **2017**, *88*, 335–341. <https://doi.org/10.1785/0220160041>.
39. Dangkua, D.T.; Cramer, C.H. Felt intensity versus instrumental ground motion: A difference between California and eastern North America? *Bull. Seismol. Soc. Am.* **2011**, *101*, 1847–1858. <https://doi.org/10.1785/0120100133>.
40. Wu, Y.-M.; Mittal, H.; Huang, T.-C.; Yang, B.M.; Jan, J.-C.; Chen, S.K. Performance of a low-cost earthquake early warning system (P-Alert) and shake map production during the 2018 Mw 6.4 Hualien, Taiwan, earthquake. *Seismol. Res. Lett.* **2019**, *90*, 19–29. <https://doi.org/10.1785/0220180170>.
41. Mittal, H.; Yang, B.M.; Wu, Y.-M. Progress on the earthquake early warning and shakemaps system using low-cost sensors in Taiwan. *Geosci. Lett.* **2022**, *9*, 42. <https://doi.org/10.1186/s40562-022-00251-w>.
42. Calais, E.; Boisson, D.; Symithe, S.; Pr  petit, C.; Pierre, B.; Ulyse, S.; Hurbon, L.; Gilles, A.; Th  odat, J.-M.; Monfret, T.; et al. A socio-seismology experiment in Haiti. *Front. Earth Sci.* **2020**, *8*, 542654. <https://doi.org/10.3389/feart.2020.542654>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.