NIST Interagency Report
NIST IR 8457

Calibration Stability of Data Loggers Used for Vaccine Temperature Monitoring

Michal Chojnacky
Sam W. Smith
Bethany A. Rodman
Tobias Herman

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8457

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Interagency Report
NIST IR 8457

Calibration Stability of Data Loggers Used for Vaccine Temperature Monitoring

Michal Chojnacky
Sam W. Smith
Bethany A. Rodman
Tobias Herman
Sensor Science Division
Physical Measurement Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8457

March 2023

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
Laurie E. Locascio, NIST Director and Under Secretary of Commerce for Standards and Technology

NIST IR 8457
March 2023

This publication was produced as part of the Interagency Agreement 15FED1512590 between the National Institute of Standards and Technology and the Centers for Disease Control and Prevention. The contents of this publication do not necessarily reflect the views or policies of the National Institute of Standards and Technology, the Centers for Disease Control and Prevention, or the US Government.

Publication History
Approved by the NIST Editorial Review Board on 2023-01-17

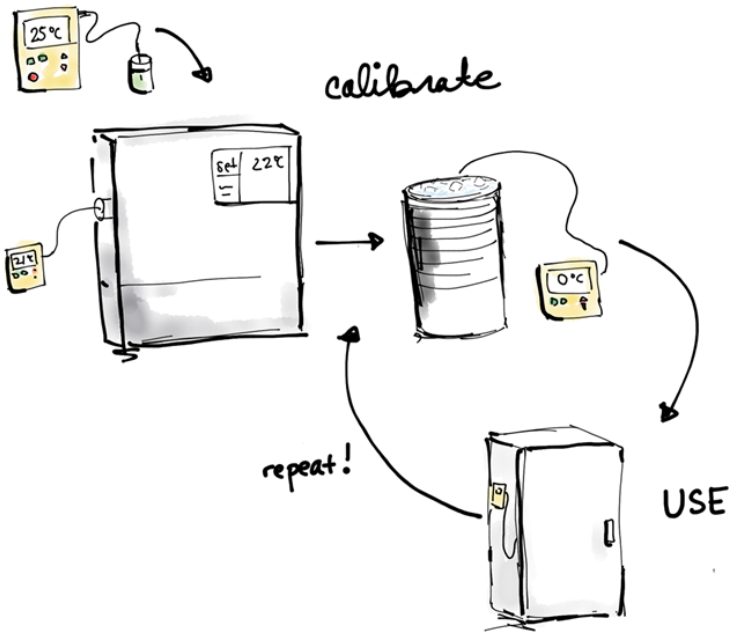
How to Cite this NIST Technical Series Publication
Chojnacky M, Smith SW, Rodman BA, Herman T (2022) Calibration Stability of Data Loggers Used for Vaccine Temperature Monitoring. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Interagency Report (IR) NIST IR 8457. https://doi.org/10.6028/NIST.IR.8457

NIST Author ORCID iDs
Michal Chojnacky: 0000-0002-7872-6413
Sam W. Smith: 0000-0003-3445-1697
Bethany A. Rodman: 0000-0002-6713-1369
Tobias Herman: 0000-0003-4296-6957

Contact Information
michalc@nist.gov

Abstract

Continuous temperature monitoring of vaccines during distribution and storage is critical to ensuring vaccination efficacy. Exposure to inappropriate storage temperatures can degrade vaccine potency, putting products and patients at risk if excursions go undetected. Digital temperature data loggers are widely used by U.S. vaccine providers to maintain a continuous record of vaccine storage temperature history. However, whether specific handling or usage conditions contribute to measurement drift in data loggers is not well known. Here we show that data logger calibration status is largely unaffected by typical usage conditions encountered at the vaccine provider level. Forty commercial digital data logger devices representing eight unique models marketed for use in vaccine temperature monitoring were evaluated in a two-year period during 2019-2021. The devices were calibrated upon receipt, subjected to trials designed to simulate typical usage conditions, and then recalibrated periodically to evaluate potential measurement drift. Simulated usage conditions included daily use in vaccine refrigerators and freezers, battery changes, local transport, cross-country shipment, and long-term storage. Out-of-tolerance calibration measurements were extremely rare, and the few failures we detected did not appear to be correlated with any of the usage trials performed during the two-year study period. These findings support extending data logger recalibration intervals when paired with concurrent manufacturer-supplied device stability data and/or intermediate data logger verification checks performed in the field.



Keywords

Calibration; Centers for Disease Control (CDC); cold chain; continuous temperature monitoring; digital data logger (DDL); ice melting point (IMP); stability; measurement drift; vaccine transport; vaccine shipment; vaccine storage; Vaccines for Children (VFC).

Table of Contents

List of Tables

List of Figures

Acknowledgments

The authors thank Christopher Meyer for his assistance with the environmental chamber.

Author Contributions

Michal Chojnacky: Conceptualization, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing; **Sam W. Smith**: Investigation, Data curation, Software, Formal analysis, Writing – original draft; **Bethany A. Rodman**: Investigation; **Tobias Herman**: Conceptualization, Writing – review & editing

1. Introduction

Vaccine temperatures can be threatened at times when staff are unable to monitor visual temperature indicators. In the absence of a continuous temperature logging system, excursions can go undetected [1]. If vaccines are exposed to instantaneously damaging temperatures, potency can be irrevocably destroyed [2, 3]. Similarly, cumulative exposure to out-of-range temperatures can degrade vaccine potency over time [4–7]. In both cases, ineffective doses might be administered to patients if the temperature excursions are not identified [8]. As a result, continuous temperature monitoring history is critical to ensuring that vaccine temperature, and consequently, vaccine efficacy, is maintained throughout the product storage lifetime [9, 10].
Since 2012, the Centers for Disease Control (CDC) has recommended the use of digital temperature data loggers to provide a continuous record of vaccine temperatures during storage and transport. In 2018, this recommendation became a requirement for Vaccines for Children (VFC) program participants [11]. The CDC's Vaccine Storage and Handling Toolkit outlines a set of minimum specifications as well as calibration and traceability requirements for data loggers used in vaccine temperature monitoring [12].

Expanded use of digital data loggers (DDLs) since the early 2010's has generated some programmatic challenges for providers and immunization managers [10, 13]. A competitive commercial market has emerged to address the need for inexpensive temperature DDLs that meet the needs of VFC providers while satisfying CDC recommendations. Even so, providers may find the process of selecting a suitable DDL burdensome, as it requires evaluating marketing claims alongside VFC requirements [14].

DDL recalibration requirements also impose additional overhead on vaccine providers. In the 2022 Vaccine Storage and Handling Toolkit, the CDC recommends that DDLs undergo recalibration "every two to three years or according to the manufacturer's suggested timeline", and suggests that DDL usage conditions like battery changes and rough handling may impact calibration status [12]. However, the effect of typical provider DDL usage on calibration status during this timeframe is not well known. In addition, recalibration can be expensive, and in some cases, may exceed the cost of the device. As a result, manufacturers may inflate recalibration intervals to satisfy customers, and users may elect to purchase new devices instead of recalibrating their old loggers. NIST guidelines have recommended ice melting-point validation testing as an accessible alternative to expensive, multi-point calibrations performed by accredited laboratories [15], but this strategy has not been widely adopted since user-performed tests can be more challenging to document and audit.

In this study, we examined the effect of typical DDL usage conditions on calibration status over a period of one to two years. The usage conditions studied included battery change, local transport, cross-country shipment, long term storage, and daily monitoring in refrigerators and freezers. Our findings provide insight into out-of-box DDL performance and reveal whether specific usage patterns contribute to temperature measurement drift. We also investigated whether ice melting-point test results are predictive of DDL calibration status over the intended temperature ranges of use for frozen and refrigerated vaccine storage, and whether ice melting-point testing alone is sufficient to detect and track temperature measurement drift in digital DDLs.

2. Method

2.1. DDL Selection

Eight different DDL models were selected for inclusion in the study. Specifications for these devices are summarized in Appendix A. The devices were chosen based on market research conducted from November

2018 – May 2019, with CDC recommendations on DDL specifications used to inform the selection process. The authors also solicited informal feedback from VFC grantees and providers regarding the use of DDLs in their states or practices. These sources identified commonly used DDL models, pointed out desirable device features, and highlighted challenges encountered in their practices. DDL market research was designed to mirror the practices of a typical healthcare purchaser, with a popular web search engine used to search for common terms like "VFC", "data logger", "DDL", and "vaccine temperature monitor."

A total of eighteen unique DDL models designed for temperature monitoring of vaccine refrigerators and/or freezers and sold by eight different manufacturers were identified in the initial market research phase. Some DDL models appeared to be rebranded and sold by multiple vendors – such duplicates were excluded from the initial pool of loggers. The identified devices ranged in price from $100 to $800, including any accessories required for data download. Problematic models, as reported by providers, were deliberately sought out for inclusion in the project, along with favorably rated models. Informal feedback indicated that most programs and providers purchase the more inexpensive devices, so devices on the upper end of the price range were eliminated. Additionally, several manufacturers offered two or more devices with identical temperature monitoring capabilities, in which the more expensive model(s) featured add-ons like paid wireless cloud accounts, remote WiFi monitoring, and cellular capabilities. In these cases, the least-expensive "base" model was selected, and the other devices were eliminated. In the end, ten unique DDL models were acquired for exploratory testing.

During the exploratory phase, the devices were tested for usability and compatibility with existing laboratory infrastructure. Two of the ten purchased DDLs included wireless data streaming, in-app data collection and/or cloud reporting features and were ultimately eliminated due to cellular and WiFi connectivity limitations in the laboratory. Since evaluating WiFi capabilities was outside the scope of the study, the exclusion of these two devices did not hinder the primary objective of assessing temperature measurement drift. The eight models used in the study ranged in price from $100 to $200. The eight unique models were each assigned a letter code (A, B, C, D, E, F, G, I), used in the device codifying scheme and throughout the results reported in this document.

2.2. DDL Setup and Standardization

The DDLs used in the study were programmed to harmonize their settings as much as possible. Several loggers required a specialized software package to adjust settings, initiate recording and download device data. All required software was provided with purchased loggers and/or available for download from vendor websites. Some devices permitted adjustment and download via push-button, display panel operation in lieu of software-based control. The time for each logger was set to local NIST Gaithersburg time, temperature units were set to Celsius, and audible alarms were disabled. Two DDL models, A and B, had a fixed recording interval of 5 minutes. The remaining devices were set to record at a rate of one reading per minute. These settings were maintained throughout the study except when noted otherwise.

Complete DDL specifications, as reported by the respective device manufacturers, are tabulated in Appendix A. As noted in Table A1, device memory capacity varied significantly, ranging from 16,000 to 24 million readings. Some devices could be programmed to stop recording new readings once the memory was full, while others automatically overwrote old readings using a rolling memory structure. One model featured a very large memory, enabling practically indefinite logging. The programmable devices were cleared, reconfigured, and restarted each time logger data was downloaded. Looping devices continued recording without requiring a restart after data download.

Data download procedures also varied widely between the different models used in this study. Several devices featured an integrated USB connection in the DDL readout, enabling download via software that opened

automatically upon USB connection, or via manual file transfer from the logger's integrated drive. Other devices required a separate USB download cable or docking station to connect to PC and initiate download. One model featured no PC connectivity, so data had to be transferred from the device to a USB thumb drive, which could then be used to transfer data to PC. Approximate download times for a typical data file are tabulated in Appendix A. In some cases, download time increased significantly along with an increase in readings stored in the device memory.

2.3. Data Outputs

Data download from each of the models generated data files with different attributes, including file type (.csv, .pdf, .txt), file storage location, and file naming conventions. Variable data structures and formatting conventions were also utilized within the files. We employed multiple strategies, detailed in Appendix B, to harmonize the DDL outputs and compare the numerous data files collected during the study.

A total of forty individual DDL devices were used in the study. Each DDL was assigned a unique name which was coded to designate the device model, the trial(s) the device was used in, and whether the device was to be used at refrigerator or freezer temperatures during the trials. This naming convention is illustrated in Figure 1, with the names and corresponding trials assigned to each of the loggers given in Table 1. Trials 0, 1 and 2 utilized the same set of DDLs, while each of the remaining trials had a dedicated set of DDLs assigned to it.

Table 1. Identifying names assigned to DDL devices used in each trial.

Trial	DDL ID							
0 - extended calibration	A0H	B0H	C0H	D0H	E0H	F0H	G0H	I0H
1 - battery change	A0H	B0H	C0H	D0H	E0H	F0H	G0H	I0H
2 - local transport	A0H	B0H	C0H	D0H	E0H	F0H	G0H	I0H
3 – shipping	A3H	-	C3H	D3H	E3H	F3H	G3H	I3H
4 - conditioned storage	A8H	-	C8H	-	E8H	F8H	G8H	I8H
4 - unconditioned storage	A9H	-	C9H	-	E9H	F9H	G9H	I9H
5 - daily use, refrigerator	A5H	-	C5H	D5H	E5H	F5H	G5H	I5H
5 - daily use, freezer	-	B5C	-	D5C	E5C	F5C	G5C	I5C

A naming convention was also developed to codify each of the 73 measurement periods performed in the study using unique test IDs. A total of six trials, listed in Table 1 and detailed in Sec. 2.5, were designed to investigate different DDL usage conditions. In some cases, trials were divided into subtrials to designate separate but related test conditions. Each trial or subtrial consisted of multiple ordered tests, which always included "initial" calibration and ice melting-point (IMP) tests, along with "final" calibration and IMP tests. Some trials also included intermediate monitoring periods and intermediate IMP tests, which were executed between the initial and final test periods. The test naming structure is reported in Fig. 1. A complete list of all the tests and corresponding ID codes are reported in Appendix C.

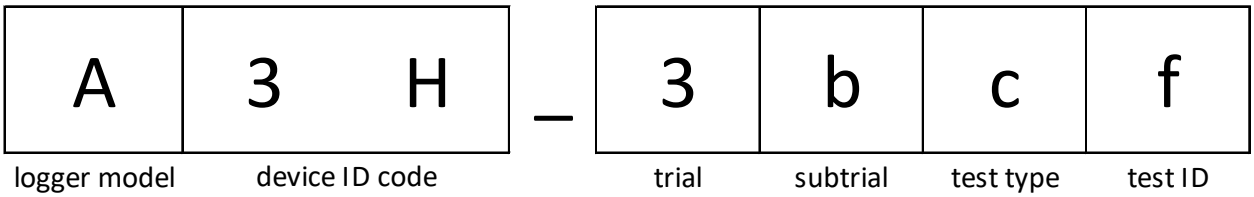


Fig. 1. Codifying scheme for individual DDL test measurements. This example shows the final calibration test (test ID "f", test type "c") in trial 3, subtrial b.

2.4. Calibration

To track possible temperature drift in response to different usage conditions, the DDLs were subjected to repeated temperature calibrations. In these calibrations, DDLs were placed in a known temperature environment, with set points chosen to mimic freezer, refrigerator, and room temperatures, along with the IMP. For each calibration temperature set point, the DDLs were set to record stabilized temperature data after a suitable equilibration period had passed. The resulting DDL temperature data were then compared to the reference temperature value to determine a temperature offset associated with each DDL. None of the DDLs evaluated in this study allowed for adjustment of the temperature output in response to a calibration process. Instead, the temperature offset was simply recorded as an indicator of the device's calibration status with respect to the stated accuracy or tolerance band.

All DDLs used in this study featured a temperature probe encased in a sealed, fluid-filled vial. All calibration measurements were performed with the probe-in-vial placed directly in a temperature-controlled environment, while the readout unit was kept in the ambient environment.

Two different calibration methodologies were used throughout the study, as described in Secs. 2.4.1 and 2.4.2.

2.4.1. Ice Melting-Point Tests

Ice melting-point (IMP) baths were prepared as follows. Ice used in the bath preparation was produced using a commercial pellet ice machine fed by a commercial reverse-osmosis (RO) filtration system connected to the laboratory tap water, which provides a > 90% reduction in metal ion contamination. NIST laboratory tap water has been shown to contribute a freezing point depression of 0.0158 °C in routine IMP measurements [16], so we estimated that the RO filtered water contributed a freezing point depression of < 0.005 °C. The RO-fed ice machine produced soft pellets (diameter ≈ 1 cm) comprised of smaller ice crystals (diameter ≈ 1 mm to 2 mm) bound together. Distilled liquid water used in ice bath preparation was obtained from the NIST Industrial Thermometer Calibration Laboratory still. The standard uncertainty ($k = 1$) of the prepared IMP is conservatively estimated to be 0.005 °C [17].

To prepare the IMP bath for measurement with DDLs, a vacuum-insulated, glass Dewar flask was first cleaned with ethanol and a lint free wipe, then rinsed with distilled water. Using a clean scoop and gloved hands, the Dewar was packed with RO pellet ice to approximately one-third full. Distilled water was added to the Dewar to ensure that all voids between the ice pellets were filled with water. If the ice appeared to float, excess water was decanted from the Dewar and/or more ice was added to ensure a homogenous mix. At this stage, two or three clean DDL probe-containing vials were embedded in the IMP bath. The vials were positioned at least 10 cm from the bottom of the flask and each vial was separated from adjacent vials and the side walls of the Dewar container by at least 5 cm. Another layer of pellet ice and water was then added to the Dewar, and additional DDL vials were positioned in the mixture, as before. This process was repeated one more time until up to six or

seven vials were embedded in the ice bath. A final layer of ice and water was used to completely fill the Dewar and cover the top layer of vials, and an insulating cover was placed over top of the Dewar. If the batch of DDLs contained more than seven loggers, additional ice baths were prepared to hold the extra vials. Once all the vials were appropriately immersed in an ice bath, the DDL displays were connected and the loggers were left to record data for at least 3 hours, ensuring plenty of time for equilibration of the fluid-filled vials with the surrounding ice bath mixture, followed by a suitable data collection period.

Calibration offsets used to evaluate and compare DDL performance were calculated as follows. First, the time/temperature data series corresponding to each DDL IMP measurement run was trimmed to remove the stabilization period as well as any rare, spurious temperature spikes resulting from a probe connection issue. The trimming process, detailed in Appendix 2, involved a combination of algorithmic processing and manual user intervention. Once the trimmed data window was logged to the processed calibration temperature database, its mean , $\bar{T}_{DDL\_IMP}$, was computed, and the difference from the mean reference temperature was calculated to obtain a calibration offset value. In the case of the IMP tests, the reference temperature value, T_{IMP}, is identically defined as the solid-to-liquid phase transition temperature of water, 273.15 K (0 °C), where the depression of the melting point temperature due to the RO filtered water is negligible. Therefore, the calibration offset (in °C) for each DDL measurement at the ice point was defined as:

$$d\bar{T}_{DDL\_IMP} = \bar{T}_{DDL\_IMP} - T_{IMP} = \bar{T}_{DDL\_IMP}$$

The standard deviation of the mean for each trimmed DDL series was used to estimate uncertainty contributions due to temperature measurement precision, $\sigma_{\bar{T}_{DDL\_IMP}}$, as described in ASTM E2593-17 [18]. DDL resolution specifications were converted to a Type B standard uncertainty by assuming a rectangular distribution [19]. The complete uncertainty budget for the IMP tests is summarized in Table 2.

Table 2. Uncertainty budget for the IMP tests. Uncertainty contributions include the preparation of the ice melting-point bath and the DDLs under test.

Item	Standard uncertainty (k = 1), °C	Description of uncertainty element
u_{IMP}	0.005	Uncertainty in the preparation of the ice melting point
δ_{DDL}	0.03	DDL resolution
$\sigma_{\bar{T}_{DDL\_IMP}}$	0.01	Precision of DDL measurements at the IMP
$u_c(k = 1)$	0.03	Combined standard uncertainty
$\boldsymbol{U(k = 2)}$	**0.06**	**Expanded uncertainty, 95% confidence level**

2.4.2. Environmental Chamber Calibrations

The environmental chamber calibrations utilized a temperature-controlled test chamber within the NIST Hybrid Humidity Generator facility. The test chamber is a stainless-steel cylindrical enclosure housed inside of a commercial environmental chamber, as described in NIST SP250-83r1 [20]. A calibrated platinum resistance thermometer (PRT) mounted in the center of the test chamber was used to record the chamber reference temperature, with a standard temperature measurement uncertainty of $u_{PRT} = 0.01$ °C. The temperature non-uniformity of the test chamber also contributed an uncertainty, conservatively estimated at $u_{chamber} = 0.03$ °C.

The environmental chamber temperature set point was controlled by front panel adjustment, and an automated data collection program was used to continuously record reference PRT temperatures approximately every 90 s. Near the end of the study, remote control capabilities were added to the measurement software, allowing chamber set point adjustment via PC.

A three-point calibration procedure was utilized throughout the study, in which the chamber was set to nominal temperature set points of −25 °C, 5 °C and 25 °C, which were selected to mimic freezer, refrigerator, and room temperatures. Reference PRT measurements were monitored during the calibration process to assess chamber temperature stability and allow for a recording interval of at least 1 h to 3 h at each temperature set point. All measurements were conducted at ambient pressure and humidity.

The DDL and reference chamber PRT calibration data series were divided into shorter series corresponding to each temperature point measured during a particular run. The temperature point series were trimmed to remove stabilization periods corresponding to the chamber set point changes, as detailed in Appendix 2. Once the trimmed data window was logged to the processed calibration temperature database, its mean, $\bar{T}_{DDL\_pt}$, was computed. The mean reference PRT temperature $\bar{T}_{PRT\_pt}$ during the trimmed timeframe was also computed. The difference between the DDL and PRT means was calculated to obtain a calibration offset value, defined as:

$$d\bar{T}_{DDL\_pt} = \bar{T}_{DDL\_pt} - \bar{T}_{PRT\_pt}$$

The standard deviation of the mean for each trimmed PRT and DDL series was used to estimate uncertainty contributions due to temperature measurement precision, $\sigma_{\bar{T}_{PRT}}$ and $\sigma_{\bar{T}_{DDL}}$, as described in ASTM E2593-17 [18]. DDL resolution specifications were converted to a Type B standard uncertainty by assuming a rectangular distribution [19]. The complete uncertainty budget for the environmental chamber calibrations is summarized in Table 3.

Table 3. Uncertainty budget for the environmental chamber DDL calibrations. Uncertainty elements include contributions from the calibrated reference PRT, the environmental chamber used to maintain a stable temperature environment, and the DDLs under test.

Item	Standard uncertainty (k = 1), °C	Description of uncertainty element
$u_{chamber}$	0.03	Temperature stability and non-uniformity of chamber
u_{PRT}	0.01	Calibrated reference PRT temperature
$\sigma_{\bar{T}_{PRT}}$	0.01	Precision of PRT measurements
δ_{DDL}	0.03	DDL resolution
$\sigma_{\bar{T}_{DDL}}$	0.02	Precision of DDL measurements
$u_c(k = 1)$	0.05	Combined standard uncertainty
$\boldsymbol{U(k = 2)}$	**0.10**	**Expanded uncertainty, 95% confidence level**

A batch of seven to thirteen DDLs was included in each chamber calibration run, according to the trial assignments in Table 1. The vials attached to each of the DDLs were arranged on the wire shelf inside the test chamber, maintaining at least 4 mm of space between adjacent vials. Any vials that tipped easily on the wire shelf were placed inside of an acrylic vial holder. Electronic leads were threaded through an exit portal on the

chamber, and connected to the DDL readouts, which were kept outside the chamber. Figure 2 shows a typical DDL installation during an environmental chamber calibration run.

Fig. 2. DDL installation during environmental calibration.

2.5. Usage Trials

Five typical usage trials were utilized to examine possible effects on DDL calibration status. An extended calibration run was also performed at the outset of the study, to examine device behavior at the limits of their stated operating ranges. Each of these trials is explained in more detail in Sec. 2.5.1 – Sec. 2.5.6.

2.5.1. Extended Calibration

One extended calibration run was conducted following the same method employed in the three-point environmental chamber calibrations, using an expanded set of temperature set points: −50 °C, −40 °C, −25 °C, 5 °C, 25 °C, 40 °C, 50 °C. As noted in Table 1, several DDLs used in this study reported a minimum operating temperature of −40 °C. The extended calibration run was designed to assess logger performance outside the intended range of use, down to the lowest temperature recommended for frozen vaccine storage of common childhood and adult vaccines (−50 °C). While some COVID-19 vaccines are maintained at temperatures near −80 °C, these vaccines had not been developed at the time of this study, so temperatures below −50 °C were not investigated.

2.5.2. Battery Replacement

A test was devised to assess the effect of battery replacement on DDL calibration. The test did not aim to evaluate the effect of draining battery life.

First, a set of eight DDLs was subjected to a "control period," during which no battery changes occurred. The loggers were first calibrated using a 3-point environmental chamber test and an IMP test, as described in the calibration section. Next, additional IMP tests were performed. Finally, a full calibration consisting of the 3-point environmental chamber test and another IMP test was performed.

Next, the same set of eight DDLs was subjected to a "change period." Once again, the loggers were calibrated in the chamber and at the IMP. Then, the batteries in each of the applicable loggers were replaced. Loggers without a replaceable battery (A0H, B0H) were left unchanged. Table 1 lists the batteries used by each of the loggers. After the battery change, the loggers were reconfigured as needed and an intermediate IMP test was performed. Finally, a full calibration consisting of the 3-point environmental chamber test and a final IMP test was performed.

2.5.3. Local Transport

A series of local transport tests was designed to simulate transport of DDLs with vaccines by providers, for delivery to off-site clinics or to facilitate intra-office transfer of vaccine stock. The CDC recommends transporting vaccines in a "qualified container and packout" consisting of an insulated cooler and passive coolant packs to maintain vaccine temperatures during transport [12]. Some VFC awardees with specialized distribution needs regularly rely on provider-level transport to deliver vaccines to satellite clinic locations. Nationwide provider-level transport also tends to increase in support of community vaccination clinics during the annual influenza season. Since 2021, provider-level vaccine transport needs have expanded significantly along with the national COVID-19 vaccine distribution efforts. However, whether the added handling associated with regular transport events could affect DDL calibration status has not been thoroughly investigated.

During the local transport trial, a total of 24 transport events were completed over a period of 31 days.

A set of eight DDLs was first calibrated using a 3-point environmental chamber test and an IMP test, as described in Sec. 2.4. The loggers were then subjected to a series of transport events and intermediate IMP tests, as shown in Appendix 3, followed by a final 3-point environmental chamber test and an IMP test.

Prior to the first transport event, a cooler was prepared for refrigerated vaccine transport. An expanded polystyrene vaccine delivery box (interior dimensions: 231 mm x 188 mm x 165 mm, 38 mm wall thickness) was lined on all interior sides with a total of six conditioned phase-change material (PCM) cold packs (5 °C phase transition temperature, flexible nylon laminated film pouch, paraffin / organic chemical mixture, 250 g net weight). The PCM cold packs were frozen solid and conditioned in a refrigerator overnight per manufacturer instructions prior to assembling the packout. The DDL vials and readouts were tightly packed in the center of the cooler and surrounded by cold packs on all sides, and the cooler lid was closed tightly. The completed cooler was maintained in a domestic refrigerator, set to 5 °C ± 3 °C, except during transport events.

Prior to each transport event, the cooler was carried outside and loaded into a vehicle. Each transport event consisted of a short local drive, typically lasting 30 min to 1 h. Upon completion of a transport event, the cooler was returned to a domestic refrigerator set to 5 °C ± 3 °C. An example of the temperature data collected by the loggers during a local transport event is shown in Fig. 3.

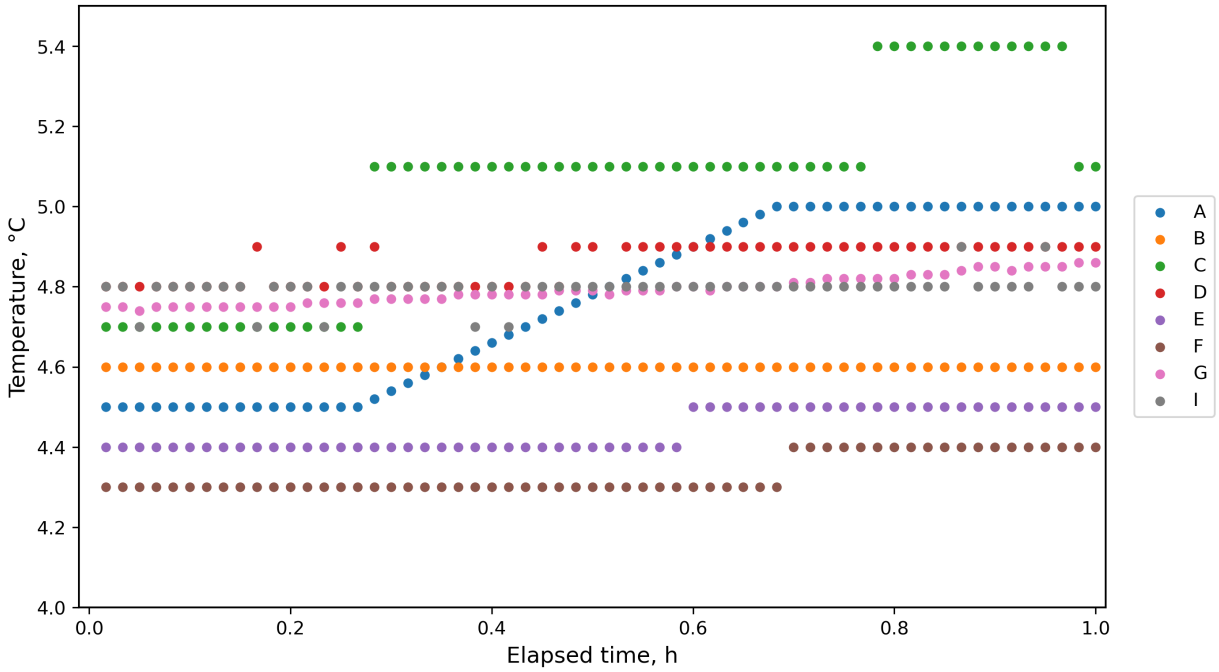


Fig. 3. DDL temperatures recorded during a 1-hour local transport event. The legend shows a different colored circle corresponding to each DDL model included in the transport trial, labeled A through I.

Eight consecutive transport events were performed utilizing the same packout. The cooler was kept closed throughout the testing period and the cold packs were maintained at their phase transition temperature of 5 °C throughout this timeframe.

After the first 8 transport events, the DDLs were removed from the cooler and measured in an IMP test. The cold packs were refrozen and reconditioned, and the packout was reassembled. A second set of 8 transport tests was then performed, followed by another ice point test. Once again, the cold packs were refrozen and reconditioned, and the packout was reassembled. After a third and final set of eight transport tests, the trial was concluded with a final 3-point environmental chamber test and an IMP test.

2.5.4. Shipping

Two cross-country shipping tests were performed to assess the effect of transport by commercial shipping carrier on DDL calibration. For each test, a set of seven DDLs were calibrated before and after a round-trip shipping event. For the first test, the DDLs were shipped from the NIST campus in Gaithersburg, MD to Boulder, CO, in January 2020. For the second test, the same set of DDLs was shipped from Clarksburg, MD to Phoenix, AZ in July 2020. Shipping details are tabulated in Table 4, along with local temperature data reported by the National Digital Forecast Database [21].

Table 4. Shipping trial details.

Shipment status	Location	Origin time	Ambient temperature
Packed	Gaithersburg, MD	15 January 2020 12:30	Indoors (22 °C ± 3 °C)
Picked up	Gaithersburg, MD	15 January 2020 15:52	12 °C
Delivered	Boulder, CO	17 January 2020 12:20	8 °C
Picked up	Boulder, CO	17 January 2020 16:29	7 °C
In transit	Commerce City, CO	18 January 2020 05:04	−7 °C
In transit	Louisville, KY	19 January 2020 00:53	0 °C
In transit	Louisville, KY	20 January 2020 14:05	−5 °C
In transit	Baltimore, MD	21 January 2020 00:13	−5 °C
Delivered	Gaithersburg, MD	21 January 2020 07:54	−4 °C
Unpacked	Gaithersburg, MD	21 January 2020 14:23	Indoors (22 °C ± 3 °C)
Packed	Clarksburg, MD	7 July 2020 15:30	Indoors (22 °C ± 3 °C)
Picked up	Clarksburg, MD	7 July 2020 16:36	30 °C
Delivered	Phoenix, AZ	8 July 2020 12:40	38 °C
Picked up	Phoenix, AZ	8 July 2020 13:40	39 °C
Delivered	Clarksburg, MD	9 July 2020 09:26	28 °C
Unpacked	Clarksburg, MD	9 July 2020 11:58	Indoors (22 °C ± 3 °C)

For each shipment, an expanded polystyrene vaccine delivery box (interior dimensions: 216 mm x 216 mm x 203 mm; 64 mm wall thickness) was lined on all interior sides with a total of six conditioned PCM cold packs (5 °C phase transition temperature, flexible nylon laminated film pouch, paraffin / organic chemical mixture, 250 g net weight). The PCM cold packs were frozen fully solid and conditioned in a refrigerator overnight per manufacturer instructions prior to assembling the packout. A layer of bubble wrap was placed over top of the cold packs to provide cushion and insulation for the DDLs, which were placed in the center of the box. Figs. 4a and 4b show the shipping packout.

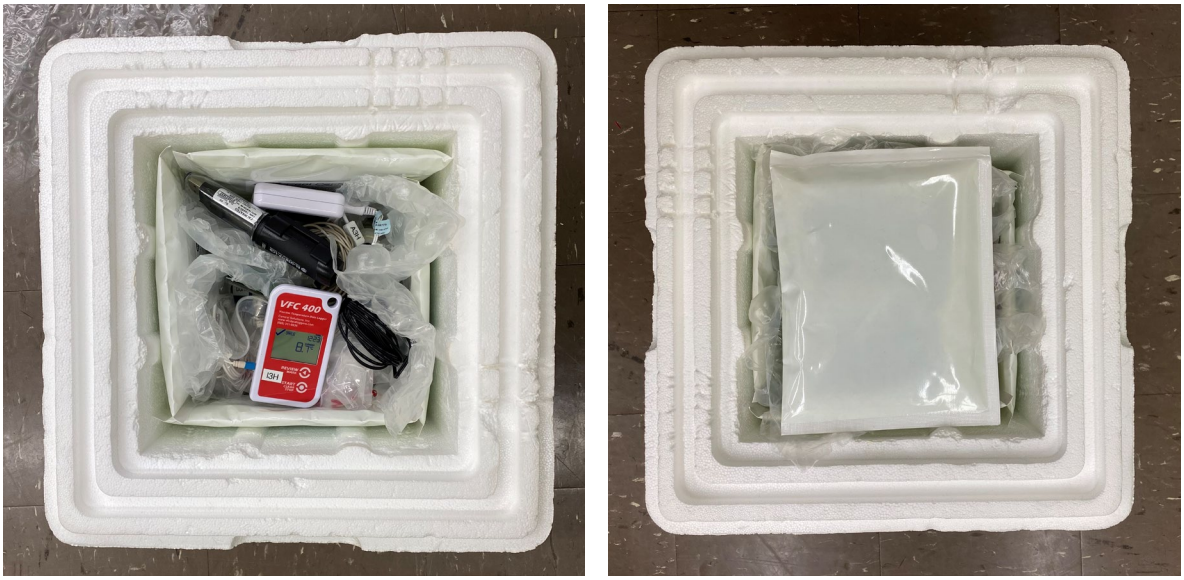


Fig. 4a, 4b. Shipping trial packout. Fig. 4a (left) shows the DDLs surrounded by bubble wrap and five PCM cold packs. Fig. 4b (right) shows the packout with the sixth cold pack placed on top, prior to closing the cooler with a lid.

Prior to the first shipment, the seal on one DDL's vial (I3H) appeared to be compromised, so this vial was sealed inside of a zip-top plastic bag to mitigate any possible leaking. For the second shipment, the leaky vial attached to I3H was exchanged for a new vial, retaining the original logger probe and readout device.

During both round-trip shipping events, the cooler was shipped from Maryland to the destination site and then returned to the origin in Maryland. The cooler was kept closed throughout the duration of the round-trip shipping event. Upon receipt of the returned cooler in Maryland, the DDLs were unpacked, the shipping data was downloaded, and the loggers were restarted prior to completion of the post-shipping calibration measurements. Figs. 5 and 6 show logger data collected during the two cross-country shipping events.

During the winter shipment to Boulder, CO (Fig. 5), local ambient temperatures remained close to 0 °C throughout the trial (Table 2). As a result, interior packout temperatures recorded by the DDLs remained close to the PCM transition temperature of 5 °C. However, during the summer shipment to Phoenix, AZ (Fig. 6), local ambient temperatures were closer to 35 °C (Table 2). As a result, the PCM cold packs most likely melted completely after about 22 h, and the interior packout temperature exceeded 25 °C during the following 10 h.

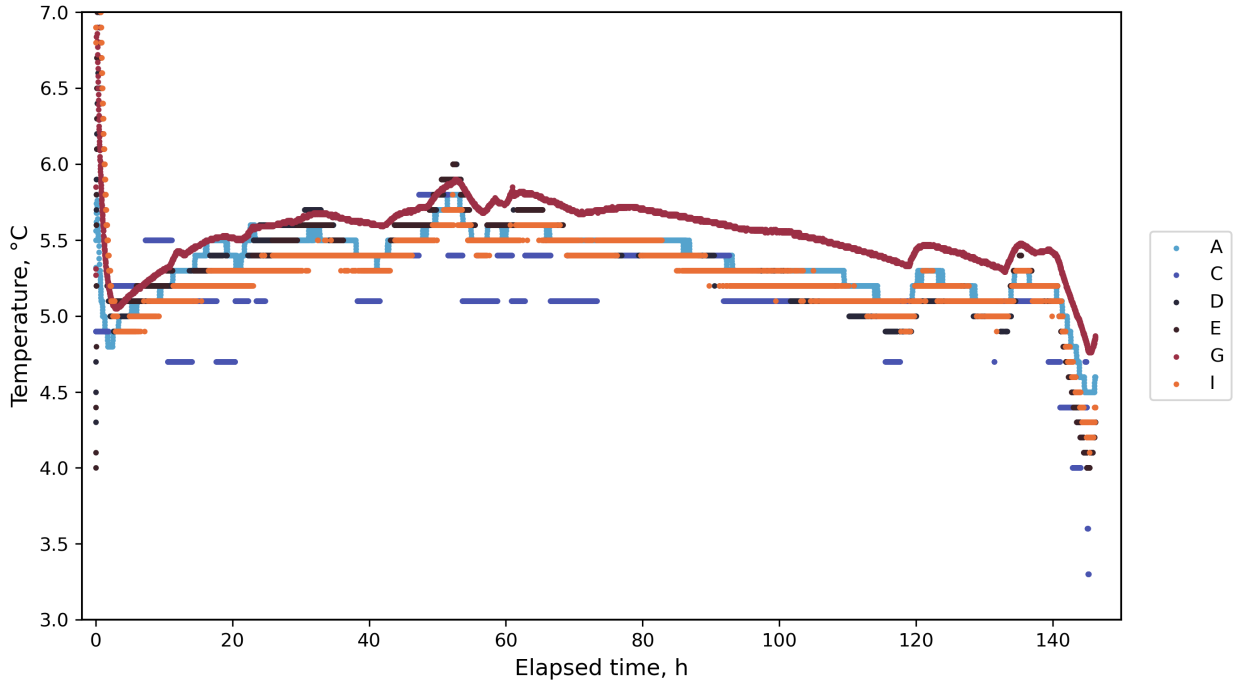


Fig. 5. DDL temperatures during the round-trip shipment from Gaithersburg, MD to Boulder, CO and back again, from January 15-21, 2020. The labels A through I represent the six DDLs used in the shipment.

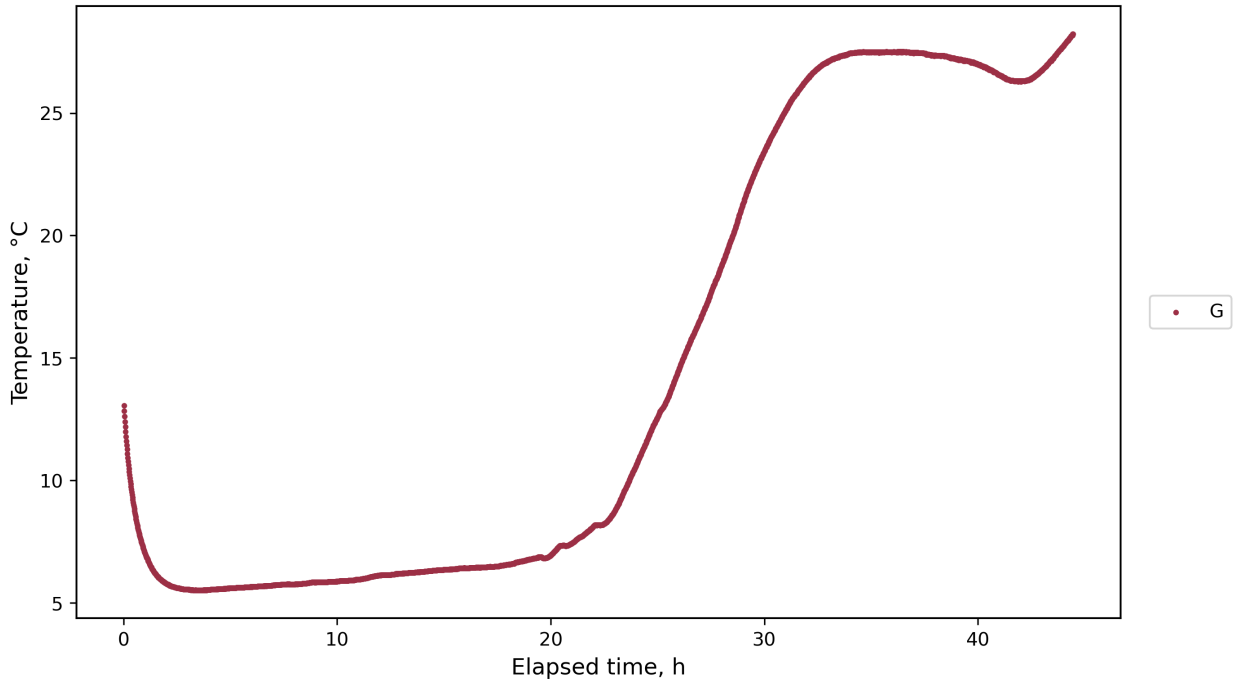


Fig. 6. Logger G temperatures during the round-trip shipment from Clarksburg, MD to Phoenix, AZ and back again, from July 7-9, 2020. The other loggers included during this shipment failed to record properly.

2.5.5. Storage

To assess the effect of long-term storage on DDL calibration, two sets of six identical DDLs were stored for a period of thirteen months. One set of loggers was placed in an unused drawer inside the NIST Vaccine Studies laboratory, to simulate storage in a temperature-controlled facility or warehouse. The second set of loggers was placed in a maintenance room which is not strictly temperature controlled, located above the mechanical level of the building in which the Vaccine Studies laboratory is housed. This set of DDLs was intended to simulate storage in an uncontrolled warehouse temperature environment. Prior to the storage period, loggers were turned off (F) or left in a standby / stopped mode (C, H, I) if possible. The remaining loggers (A, G) had no such mode and were left to run continuously.

2.5.6. Daily Use

The last trial was designed to assess the effects of long-term daily use in vaccine storage units. A set of seven DDLs intended for use at refrigerator temperatures were installed in a domestic, upright, standalone swinging door refrigerator. Although this type of unit is marketed for food storage, current CDC guidelines permit the use of standalone domestic refrigerators for vaccine storage, and these units are widely used for this application due to their low cost and availability. A second set of six DDLs intended for use at freezer temperatures was installed in an undercounter, pharmaceutical grade freezer. This unit was chosen for its ability to maintain a temperature set point of −35 °C. This is significantly colder than most domestic freezer units, and thus more likely to push the DDLs to the limits of their operating ranges. Table 2 lists the DDLs used in each type of unit. The loggers were set to log continuously in the refrigerator and freezer over a period of seven months to simulate typical daily usage, with monthly interruptions to download data, perform calibration measurements on the loggers, and restart fridge/freezer temperature logging.

DDL installations are shown in Figs. 7a and 7b. DDL vials were arranged in a stainless-steel vaccine storage tray which was placed in the center of each unit. The vials were positioned with at least 4 mm of space between them and fixed in place using Velcro, which ensured that the vials remained upright. In the domestic refrigerator, DDL cables were routed to the door seal and fixed in place using strips of Velcro to minimize bunching and disruption of the refrigerator door seal. The pharmaceutical freezer used in the study features an exit portal for electronic leads, so the DDL cables were routed through this portal, which was then re-filled with supplied insulation material. Both units also contained a moderate load of boxed vaccines, which were placed in trays, consistent with CDC guidelines [12].

Fig. 7a, 7b. DDL installation in a standalone refrigerator (left) and an undercounter freezer (right).

In the daily usage trial, all DDLs were subjected to an initial 3-point environmental calibration and an IMP test, as in the preceding trials. After calibration and verification, the DDL probes were installed in the refrigerator and freezer as described above and set to record storage unit temperatures for approximately one month. For these monitoring periods, the DDL read rate was adjusted to 1 reading every 5 minutes, ensuring that all devices would be able to capture a full month of data without exceeding memory limits. Examples of the DDL temperature measurements collected during a daily use monitoring period are shown in Figs. 8 and 9. The periodic temperature spikes observed in both figures reflect the expected defrost cycling behavior for the storage units used in this study.

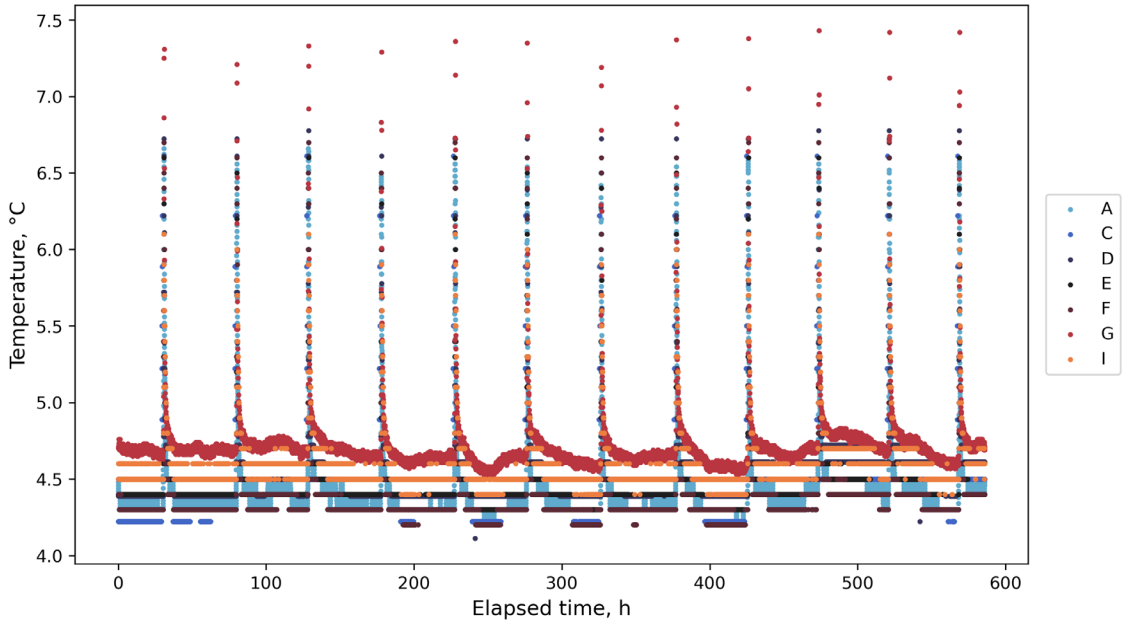


Fig. 8. Complete DDL record from the daily usage trial “5m6” in a standalone refrigerator, collected over a period of 24 days.

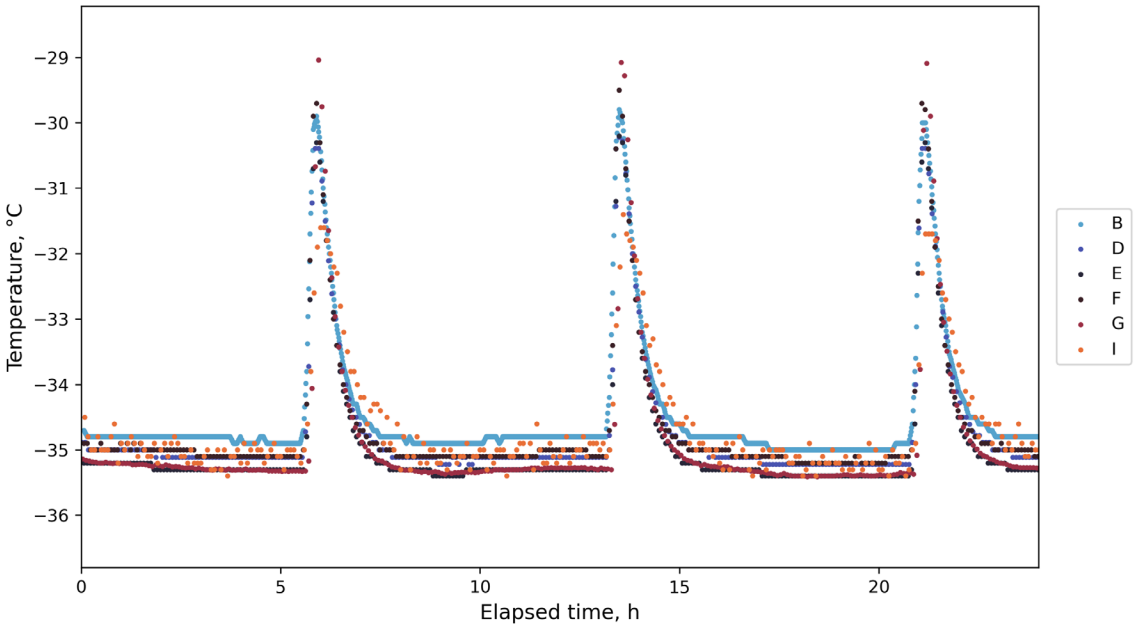


Fig. 9. DDL record from the daily usage trial “5m6” in an undercounter, pharmaceutical grade freezer. A 24-hour excerpt of data is shown.

After each monitoring period, DDL vials were removed from the storage units and subjected to an IMP test. Daily use monitoring periods followed by recalibration were repeated according to the measurement pattern shown in Appendix C, culminating in a final IMP test and environmental chamber calibration.

3. Results

The findings reported in this section are derived from the calibration offsets, $d\bar{T}$, for each calibration point measured during the five usage trials. Methods used to calculate these calibration offsets are detailed in Sec. 2.4. Complete calibration offset data corresponding to every DDL calibration performed is plotted chronologically, by trial, in Appendix D. In this section, we highlight and summarize trends identified in the calibration offset data with respect to the tested usage conditions, to determine whether specific usage patterns contribute to measurement drift. To make the data more readable, the individual DDL results shown in Appendix D have been condensed into a single average temperature offset at each calibration point, with error bars used to denote the overall range of offsets obtained from the pool loggers measured at that point.

To define DDL calibration performance at each point, we must first clarify some terminology. Per ASTM E344-20, "accuracy" is a qualitative concept describing the "closeness of agreement between the result of a temperature measurement and a true value of the temperature" [22]. The International Vocabulary of Metrology (VIM) adds that measurement accuracy is "not given a numerical value, but a measurement is said to be more accurate when it offers a smaller measurement uncertainty" [23].

The "uncertainty" of a temperature measurement is "derived from an analysis of a measurement and its result" and "characterizes the range in which the true value of temperature is estimated to lie" to within a given confidence interval [22]. Derivation of the total expanded temperature measurement uncertainties for the IMP and environmental chamber calibration tests performed in this study is detailed in Sec. 2.4.

Finally, ASTM E344-20 defines the term "tolerance," as "the permitted variation of a measured value from the correct value" [22]. Additionally, "if a measurement instrument is stated to measure correctly to within a tolerance, the instrument is classified as 'in tolerance' and it is assumed that measurements made with it will measure correctly to within this tolerance. An instrument that is not classified 'in tolerance' is classified as 'out of tolerance'" [24, 25].

In practice, usage of terms like accuracy, uncertainty, and tolerance in public health guidance and in DDL manufacturer-supplied documentation may deviate from the definitions outlined above. DDL manufacturers commonly report a "temperature accuracy" specification, while guidance for providers may describe this specification as either an "accuracy" or an "uncertainty" [12, 26].

For this work, we equate the value of the manufacturer-supplied temperature measurement accuracy specification with the ASTM E344-20-defined tolerance associated with each DDL model, which we denote τ_{DDL}. The manufacturer-supplied temperature accuracy specifications for the DDL models included in this study ranged from 0.25 °C to 0.5 °C and are reported in Table A1.

Then, to assess DDL performance, we apply a simple tolerance verification criteria, described in ASTM E2846-20 [25]. If a DDL's measured calibration offset is less than or equal to the tolerance for that model,

$$|d\bar{T}_{DDL}| \leq \tau_{DDL}$$

the calibration measurement is designated as "in-tolerance." However, if the calibration offset exceeds the tolerance,

$$|d\bar{T}_{DDL}| > \tau_{DDL}$$

the calibration measurement is designated as "out-of-tolerance." For each of the summary graphs in Secs. 3.1 through 3.6, in-tolerance points are plotted as means and ranges, whereas out-of-tolerance points are excluded from the summary statistics and are instead plotted as individual outliers.

Following ASTM E2846-20, the DDL tolerance and the total expanded measurement uncertainty of the calibration can be used to compute a Test Uncertainty Ratio (TUR),

$$TUR = \frac{\tau_{DDL}}{U(k=2)}.$$

Due to the measurement uncertainties associated with the calibration tests and the acceptance criteria outlined above, no test can verify that a device under test is within tolerance with 100% certainty. Instead, there will always be some risk of "false acceptance," where an out-of-tolerance device is accepted as in tolerance, or "false rejection," where an in-tolerance device is rejected as out of tolerance. ASTM E2846-20 describes a method for quantifying this risk, via estimates of the Probability of False Acceptance (PFA) and the Probability of False Rejection (PFR), both of which depend on the TUR. Tables 5 and 6 below summarize the TUR, PFA, and PFR for DDLs with different tolerance values, τ_{DDL}, where we have followed the calculations described in [25].

Table 5. Conformance probabilities for the IMP test tolerance verification criteria, where τ_{DDL} is the tolerance for a given DDL model, TUR is associated the Test Uncertainty Ratio, PFA is the Probability of False Acceptance, and PFR is the Probability of False Rejection.

τ_{DDL}, °C	TUR	PFA, %	PFR, %
0.25	4	2.2	2.6
0.3	5	1.8	2.1
0.5	8	1.2	1.3

Table 6. Conformance probabilities for the environmental chamber test tolerance verification criteria, where τ_{DDL} is the tolerance for a given DDL model, TUR is the associated Test Uncertainty Ratio, PFA is the Probability of False Acceptance, and PFR is the Probability of False Rejection.

τ_{DDL}, °C	TUR	PFA, %	PFR, %
0.25	2.5	3.3	4.5
0.3	3	2.9	3.4
0.5	5	1.8	2.1

As seen in Tables 5 and 6, for DDLs subjected to the same calibration process, a tighter tolerance results in a smaller TUR, and in turn, a greater risk of incorrectly classifying a DDL as in-tolerance or out-of-tolerance. Calibration laboratories often aim to achieve a $TUR > 4$, based on recommendations in ANSI/NCSL Z540.3-2006 and prior standards [27]. This ensures that the risk of an incorrect tolerance verification assessment is ≈ 2% or less. From the above, our test methodology and measurement uncertainties were sufficient for DDLs with τ_{DDL} = 0.5 °C, but could have been optimized with lower uncertainties for DDLs with tighter tolerances. From Table 4, the dominating uncertainties in the chamber calibration were the chamber temperature non-uniformity, $u_{chamber}$, and the DDL resolution, δ_{DDL}. While the DDL resolution is inherent to the devices tested in the study and thus cannot be improved, the chamber temperature non-uniformity could be reduced by restricting the placement of the DDLs under test to a reduced area within the chamber. Some additional test methodology improvements are suggested in Sec. 3.7.2.

3.1. Trial 0: Extended Calibration

The extended calibration test was intended to probe DDL performance at limits of the stated device operating ranges, since out-of-range temperature exposures could occur during routine DDL use. In particular, the lower

temperature limit for frozen vaccine storage is −50 °C, but five of the eight models included in the study reported a lower operating limit of −40 °C. Average DDL and reference PRT temperatures recorded during seven environmental chamber test periods are given in Table 37, at nominal chamber set points of −50 °C, −40 °C, −25 °C, 5 °C, 25 °C, 40 °C, and 50 °C. Average DDL temperatures that deviated from the reference temperature by more than three times the devices' stated accuracy are displayed in red text, to indicate an erroneous reading. Non-numerical outputs such as "—" are also displayed in red text, denoted by N/A. Four DDLs (A, B, D, E) generated an erroneous reading during the −50 °C measurement, and a fifth (I) generated no reading at this point. Each of these five devices was only rated for use down to −40 °C. Additionally, one DDL (C) reported no reading at the 50 °C point, which was above its maximum measurement temperature of 40 °C.

Table 7. Extended calibration test results. All reported values are temperatures in °C. The first row shows the chamber temperature as recorded by the reference PRT. Subsequent rows show the DDL temperatures recorded at each chamber setpoint. Erroneous and non-numerical readings are shown in red, italicized text. Reported reference PRT temperatures and DDL temperatures correspond to averages performed over the entire collection period at each calibration setpoint.

	Chamber temperature:	**−49.87**	**−38.38**	**−28.49**	**6.20**	**24.52**	**39.10**	**49.24**
DDL	**operating range**							
A	−40 to 60	*53.0*	−38.1	−28.5	6.0	24.0	38.7	48.8
B	−40 to 60	*53.2*	−37.9	−28.2	6.2	24.2	38.8	48.8
C	−50 to 40	−50.2	−38.8	−29.3	5.9	24.5	39.3	*N/A*
D	−40 to 40	*−40*	−37.9	−28.3	6.4	24.5	39.1	49.5
E	−40 to 125	*−40*	−38.6	−29.1	6.0	24.2	39.0	49.2
F	−50 to 70	−50.0	−38.4	−29.6	5.9	23.6	38.6	48.6
I	−40 to 40	*N/A*	−38.3	−29.3	6.1	23.9	38.8	48.8

These findings underscore the importance of selecting a device with measurement specifications appropriate for the intended range of use. DDLs used to measure temperatures outside of their stated measurement ranges are likely to generate unreliable data at those temperatures. In some cases, the device failure will be obvious, and no reading will be reported. However, in some other cases, such as seen with loggers D and E, the logger may default to a temperature reading corresponding to its allowed minimum temperature, even though the true temperature of the probe could be much lower. Loggers A and B appear to report a positive temperature rather than a negative one when their lower limit is exceeded. In either of these cases, the erroneous temperature could go undetected by a provider in the absence of a careful data review.

These findings are especially important in the current vaccine management environment, as providers and equipment manufacturers adapt to evolving needs and requirements associated with the storage of COVID-19 vaccines, where some preparations require frozen storage at temperatures as low as −80 °C.

3.2. Trial 1: Battery Replacement

As shown in Fig. 10, changing DDL batteries did not result in any appreciable temperature measurement drift or significant differences in calibration results.

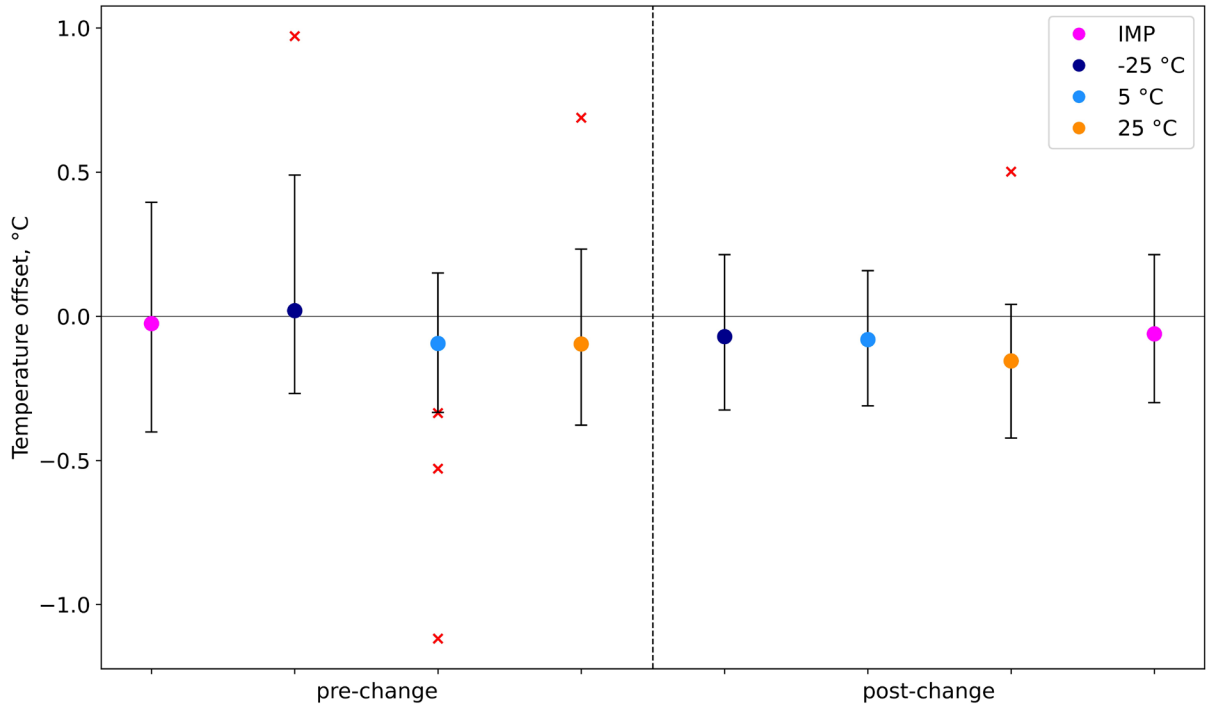


Fig. 10. DDL calibration offsets during the battery change trial, plotted in chronological order of measurement. In-tolerance calibration offset measurements are grouped together by calibration point to show overall performance trends before and after the battery change. Colored circles show the mean temperature offset at each calibration point, denoted by the nominal temperature or "IMP" for ice-melting point. The error bars show the overall range of offsets recorded by each group of DDLs. Any DDL measurements designated as out-of-tolerance ($|d\bar{T}_{DDL}| > \tau_{DDL}$) are excluded from the summary data and are instead plotted individually as red X's.

In Fig. 10, the colored circles show the mean, and the error bars show range of temperature offsets measured for each type of calibration point. Multiple calibration runs, each consisting of six to eight independent DDL measurements, have been grouped together by temperature point to compare data collected prior to the battery change, designated as "pre-change," to data collected after the battery change, designated as "post-change." These groupings are summarized in Table 8, where n represents the number of DDL points included in each summarized data point.

Table 8. DDL calibration measurement groupings used to generate Fig. 10, where *n* is the number of in-tolerance calibration measurements recorded at each calibration point. Each set of *n* calibration offset values were averaged together to produce the mean values plotted in Fig. 10.

Calibration point	n
IMP (pre)	24
−25 °C (pre)	23
5 °C (pre)	21
25 °C (pre)	23
−25 °C (post)	6
5 °C (post)	6
25 °C (post)	5
IMP (post)	18

In Fig. 10, the data appears to show more variation (e.g., larger error bars) in the calibration offset values prior to the battery change as compared to afterwards. However, this interpretation should be avoided. The larger offset range is likely due to the larger sample size included in the "pre" calibration points as compared to the "post" calibration points (see Table 4). The larger "pre" sample likely represents a better sampling of the true variation, whereas the smaller "post" sample variation is less informative due to the limited sample size.

As shown in Fig. 10, six DDL calibration measurements (out of 132 measurements total) were designated as out-of-tolerance. These points are plotted as red X's. Out-of-tolerance points indicate that the measured offset from the reference temperature exceeded the device's accuracy specification. Five of these six out-of-tolerance measurements occurred prior to the battery change, and as such most likely cannot be attributed to any particular usage condition. See Sec. 3.7 for a detailed discussion of out-of-tolerance points.

These findings are consistent with a DDL use protocol that permits battery replacement by users as needed, without necessitating a calibration or validation test after the change.

3.3. Trial 2: Local Transport

As shown in Fig. 11, repeated local transport events did not result in any appreciable temperature measurement drift or significant differences in calibration results.

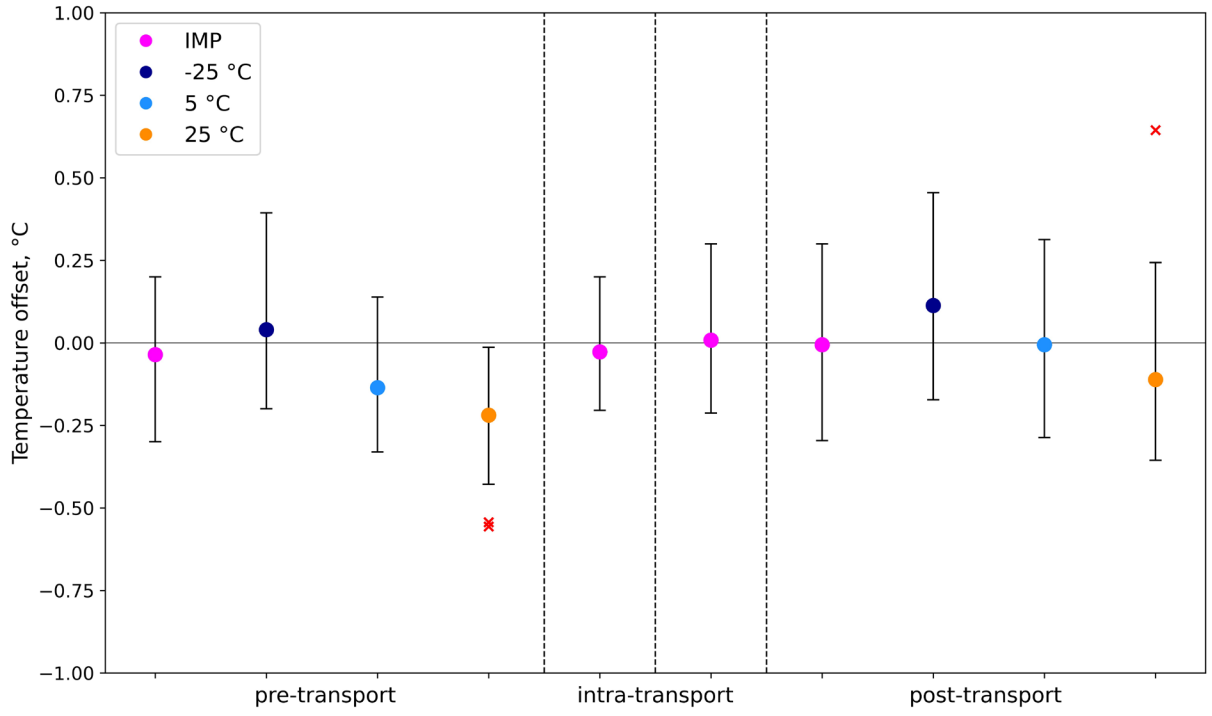


Fig. 11. Local transport trial. Colored circles show the means, and error bars show the range of temperature offsets measured for each type of calibration point. Each calibration point represents a single calibration run, summarizing the data collected from the pool of DDLs measured during that run. Eight DDLs were measured at each point. Any DDL measurements designated as out-of-tolerance ($|d\bar{T}_{DDL}| > \tau_{DDL}$) are excluded from the summary data and are instead plotted individually as red X's. The vertical dashed lines serve to indicate the chronological points at which the transport events occurred, with each line representing eight consecutive local transport events. The calibration data has been designated as "pre-transport," "intra-transport," and "post-transport," to indicate when the calibrations were performed, relative to the transport events.

As shown in the figure, three DDL calibration measurements (out of 80 measurements total) were designated as out-of-tolerance, such that the measured temperature offset from the reference temperature exceeded the device's accuracy specification. Two of these three out-of-tolerance measurements occurred prior to the transport events, and as such, cannot be attributed to the transport usage condition. See Section 3.7 for a detailed discussion of out-of-tolerance points.

The DDLs did not exhibit any significant measurement drift after completion of twenty-four transport events. These findings suggest that DDLs may be used for repeated monitoring of local vaccine transport events without negatively impacting their calibration status or requiring additional verification testing beyond the usual recommended schedule.

3.4. Trial 3: Shipping

As shown in Fig. 12, repeated cross-country shipment did not result in any appreciable temperature measurement drift or significant differences in calibration results.

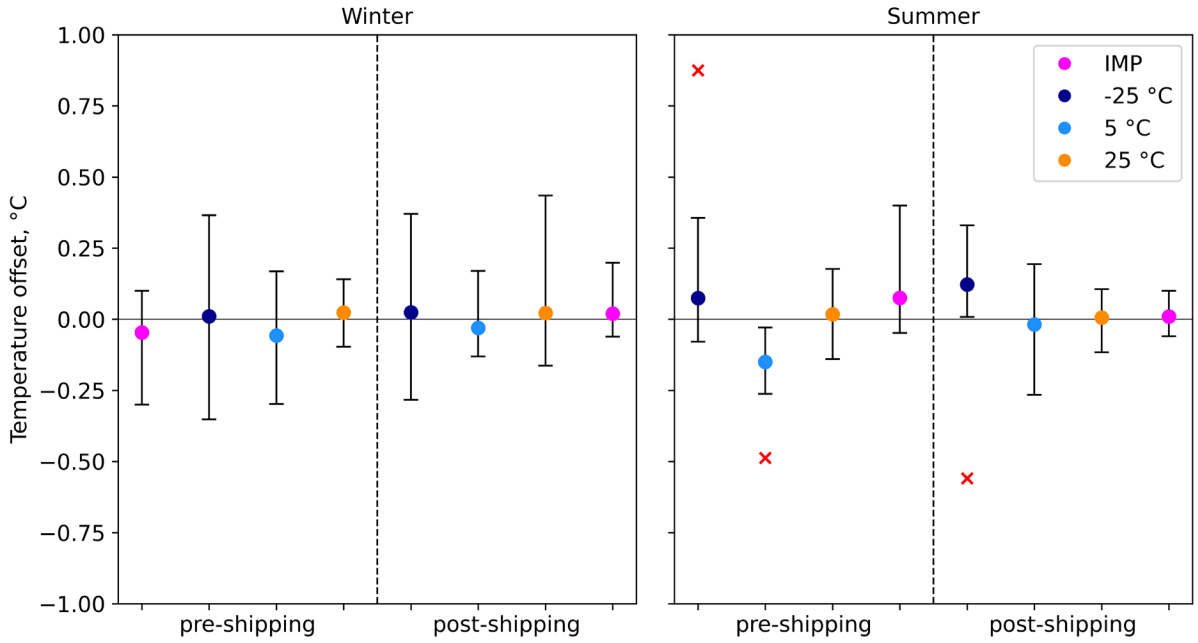


Fig. 12. Cross-country shipping trial, with the winter subtrial shown on the left, and the summer subtrial on the right. Colored circles show the means, and error bars show the range of temperature offsets measured for each type of calibration point. Each calibration point represents a single calibration run, summarizing the data collected from the pool of DDLs measured during that run. Six DDLs were measured at each point. Any DDL calibration measurements designated as out-of-tolerance ($|d\overline{T}_{DDL}| > \tau_{DDL}$) are excluded from the summary data and are instead plotted individually as red X's. The vertical dashed lines serve to indicate the chronological points where the round-trip, cross-country shipping events occurred. The calibration data has been designated as "pre-shipping," and "post-shipping," to indicate when the calibrations were performed, relative to the shipping events. The same set of DDLs was utilized for the winter and summer shipping subtrials.

As shown in the figure, three DDL calibration measurements (out of 96 measurements total) were designated as out of tolerance, such that the measured temperature offset from the reference temperature exceeded the device's accuracy specification. See Sec. 3.7 for a detailed discussion of out-of-tolerance points.

The DDLs did not exhibit any significant measurement drift after completion of two round-trip, cross-country shipping events. These findings suggest that DDLs may be used for repeated vaccine shipping events without negatively impacting their calibration status or requiring additional verification testing beyond the usual recommended schedule. However, DDLs featuring a probe encased in a fluid-filled vial should be carefully inspected before and after shipment to verify the integrity of the vial seal and its fill level, since rough handling during shipment may increase the likelihood of leaking. Leaky vials should be refilled with the manufacturer-recommended glycol solution or replaced prior to next use.

3.5. Trial 4: Long-term Storage

As shown in Fig. 13, DDL storage in either a thermally-conditioned or unconditioned environment over a period of 13 months did not result in any appreciable temperature measurement drift or significant differences in calibration results.

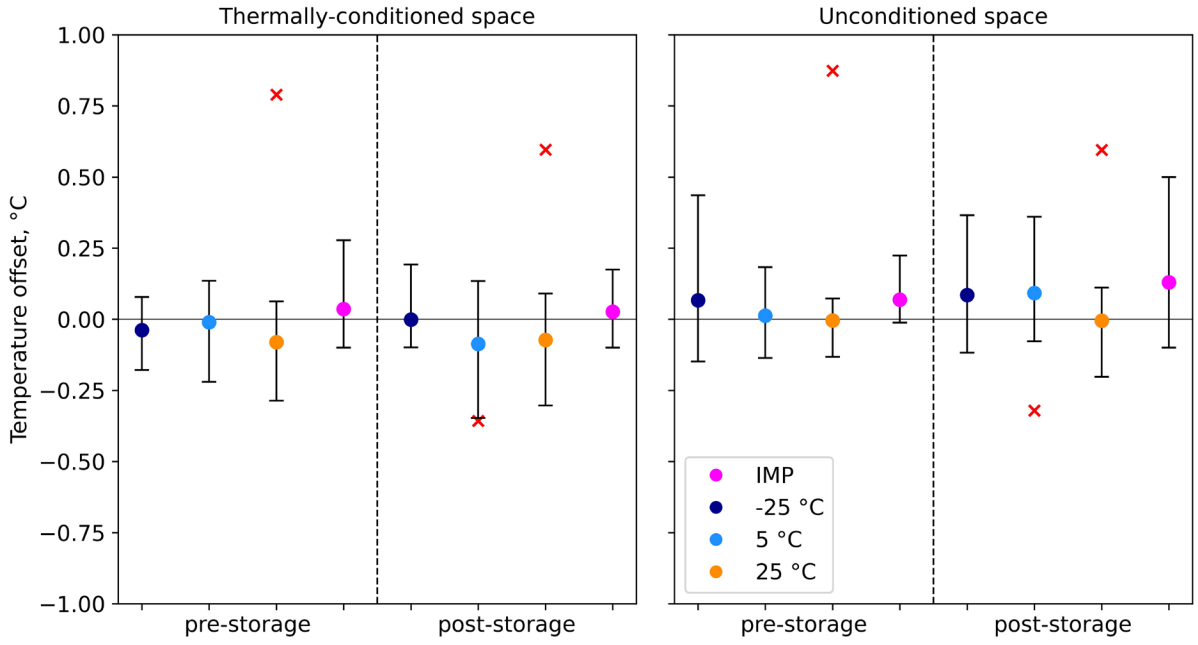


Fig. 13. Long-term storage trial. Calibration data from six loggers, stored in a thermally-conditioned space, are shown on the left, and data from five loggers, stored in an unconditioned, indoor space, are shown on the right. Colored circles show the means, and error bars show the range of temperature offsets measured for each type of calibration point. Each calibration point represents a single calibration run, summarizing the data collected from the pool of DDLs measured during that run. Any DDL calibration measurements designated as out-of-tolerance ($|d\overline{T}_{DDL}| > \tau_{DDL}$) are excluded from the summary data and are instead plotted individually as red X's. The vertical dashed lines indicate the chronological points where the 13-month storage period occurred. The calibration data has been designated as "pre-storage," and "post-storage," to indicate when the calibrations were performed, relative to the storage period.

Two identical sets of six DDL models were utilized for the two storage locations. However, one of the unconditioned space loggers malfunctioned repeatedly during post-storage testing. As a result, this device was excluded due to a lack of usable calibration data.

As shown in Fig. 13, six DDL calibration measurements (out of 88 measurements total) were designated as out-of-tolerance, such that the measured temperature offset from the reference temperature exceeded the device's accuracy specification. See Section 3.7 for a detailed discussion of out-of-tolerance points.

These findings suggest that DDLs may be stored when not in use without negatively impacting their calibration status or requiring additional verification testing beyond the usual recommended schedule.

3.6. Trial 5: Daily Use

As shown in Fig. 14, continuous logging, consistent with typical daily logger usage, in a vaccine refrigerator or freezer for a period of seven months did not result in any appreciable temperature measurement drift or significant differences in calibration results.

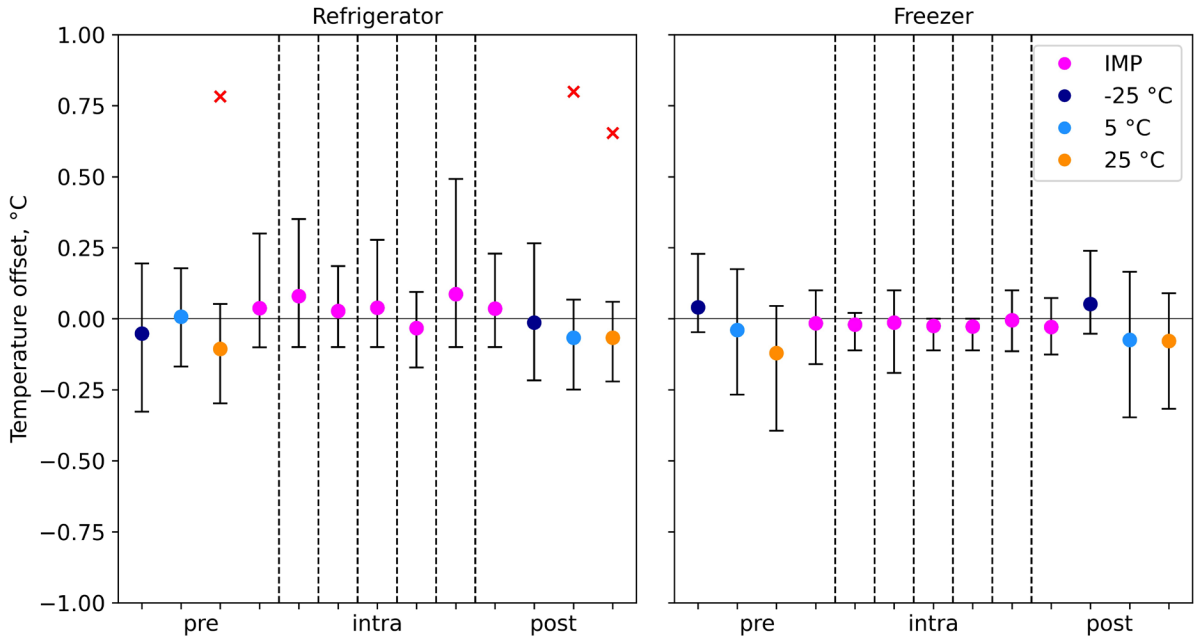


Fig. 14. Daily usage trial. Calibration data from seven loggers, used to monitor temperatures in a refrigerator, are shown on the left, and data from six loggers, used to monitor temperatures in a freezer, are shown on the right. Colored circles show the means, and error bars show the range of temperature offsets measured for each type of calibration point. Each calibration point represents a single calibration run, summarizing the data collected from the pool of DDLs measured during that run. Any DDL calibration measurements designated as out-of-tolerance ($|d\overline{T}_{DDL}| > \tau_{DDL}$) are excluded from the summary data and are instead plotted individually as red X's. The vertical dashed lines serve to indicate the chronological points where each "daily use period" occurred, consisting of at least one month of continuous operation inside a refrigerator or freezer. The calibration data has been designated as "pre," "intra," and "post," to indicate when the calibrations were performed, relative to the daily use periods.

As shown in Fig. 14, three DDL calibration measurements (out of 169 measurements total) were designated as out-of-tolerance, such that the measured temperature offset from the reference temperature exceeded the device's accuracy specification. See Sec. 3.7 for a detailed discussion of out-of-tolerance points.

These findings suggest that DDLs may be used for continuous temperature logging in vaccine refrigerators and freezers, consistent with current public health recommendations, without negatively impacting their calibration status or requiring additional verification testing beyond the usual recommended schedule.

3.7. Out-of-tolerance Points

This study included a total of 573 DDL calibration measurements at the IMP and in an environmental chamber. Ninety-six percent of these measurements were classified as in-tolerance, meaning that the calibration offset was less than or equal to the stated device accuracy or tolerance. The out-of-tolerance points that occurred were mostly limited to two specific DDL models, and these "failures" only occurred in the environmental chamber calibration tests.

Table 9. Environmental chamber calibration data, grouped by logger model. Each "logger model" set consisted of 4 to 6 individual devices of that model. During each chamber calibration run, DDLs were measured at three nominal set points (−25 °C, 5 °C, and 25 °C). These three measurements are counted as three individual calibration points. Measurements in which the calibration offset was less than or equal to the DDL tolerance ($|d\bar{T}_{DDL}| \leq \tau_{DDL}$) were designated as in-tolerance points.

Logger Model	In-tolerance Points	Total Points	Passing %
A	44	45	98
B	21	21	100
C	38	48	79
D	41	42	98
E	54	54	100
F	35	36	97
G	53	54	98
I	47	54	87

Six of the DDL models each had one or fewer failures during the entire study, for an in-tolerance rate of 97% or better during the chamber calibrations. Only two logger models, "C" and "I", experienced multiple out-of-tolerance points during the study. These models had a passing rate of 79% and 87%, respectively, across all environmental chamber calibration measurements. We will examine the data from these two models in more detail in this section.

3.7.1. Logger Model C

Figure 15 shows all calibration offsets collected using logger model "C" devices throughout the entire study.

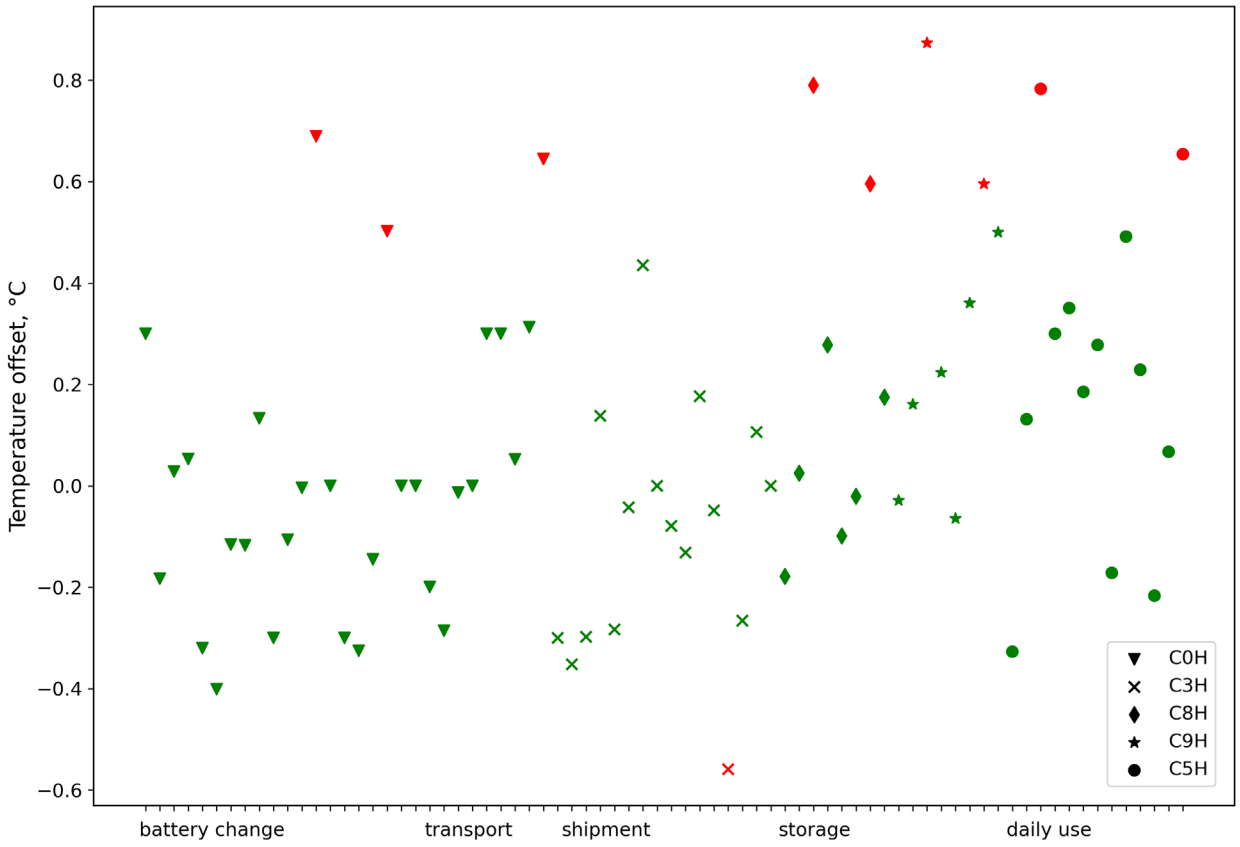


Fig. 15. Summary of all environmental chamber temperature points and IMP calibration data collected using logger model "C." The temperature offset is the difference between the DDL measurement and the reference temperature. In-tolerance points are shown in green, while out-of-tolerance points are shown in red. Individual DDL devices are distinguished by different marker symbols, as indicated in the legend. The points are shown in the chronological order in which they were measured, with the start of each corresponding trial name indicated along the x-axis.

As shown in Fig. 15, the five logger model "C" devices used in the study failed a total of ten calibration measurements. Nine of these failures occurred during a 25 °C environmental chamber calibration measurement, and in each case, the DDL recorded a temperature higher than the reference temperature. These details suggest the presence of a systematic error. The 25 °C measurements are summarized in Fig. 16.

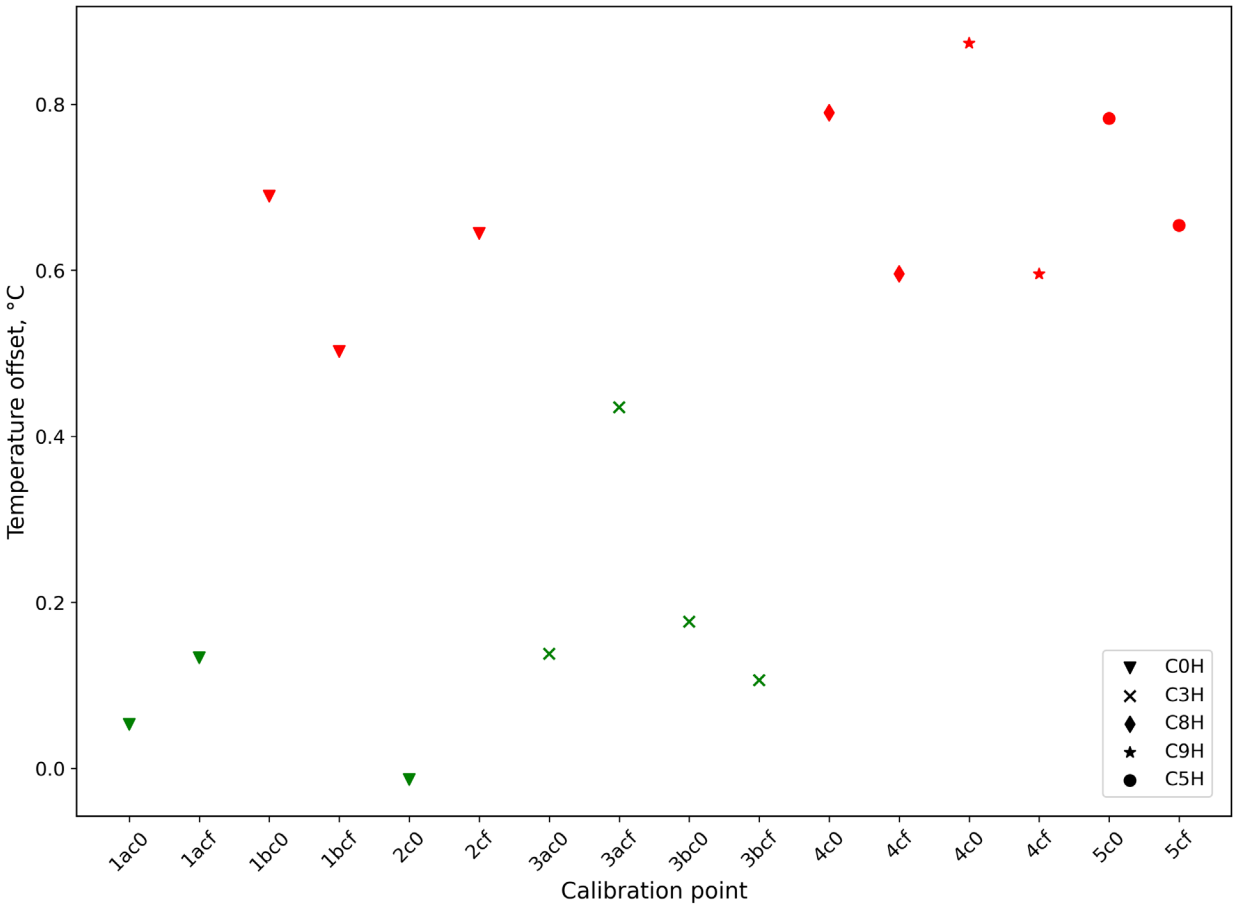


Fig. 16. Summary of logger model "C" calibration results at a nominal environmental chamber setpoint of 25 °C. The x-axis shows the corresponding test ID code (Appendix C). The temperature offset is the difference between the DDL measurement and the reference temperature. In-tolerance points are shown in green, while out-of-tolerance points are shown in red. Individual DDL devices are distinguished by different marker symbols, as indicated in the legend.

As shown in Fig. 16, logger model "C" was out of tolerance for 9 of 16 measurements at the 25 °C calibration point – a 56 % failure rate at this temperature. The failures occurred both before (measurement codes containing "0") and after (codes containing "f") most of the tested usage conditions. As such, these failures are not correlated to usage conditions and are more likely attributable to a systematic device calibration error.

From a vaccine management perspective, failures at 25 °C are not especially concerning. If a DDL in a vaccine storage unit is measuring temperatures near 25 °C, this represents a severe excursion outside allowable vaccine storage temperatures. A measurement error of approximately ± 1 °C at 25 °C is unlikely to change the standard operating procedure for vaccines exposed to an excursion of this magnitude. Conversely, temperature accuracy within ± 0.5 °C near permitted storage ranges (2 °C to 8 °C, or freezer temperatures below −15 °C) is more critical, since a small error within the target storage range could result in missed or incorrectly identified temperature excursions.

According to the materials shipped with the model C loggers, the devices were calibrated by the manufacturer at 5.6 °C, but not at 25 °C. The findings from this study suggest that DDLs should be calibrated at a temperature close to the range of use. In the case of refrigerated vaccine storage monitoring, the 5.6 °C calibration likely is sufficient. However, if the logger is intended for regular use at temperatures outside the permissible vaccine storage ranges (e.g., controlled room temperature monitoring), the calibration should also cover a range

commensurate with the intended range of use. A single point calibration is not sufficient to extrapolate performance at higher or lower temperatures.

3.7.2. Logger Model I

Figure 17 shows all calibration offsets collected using logger model "I" devices throughout the entire study.

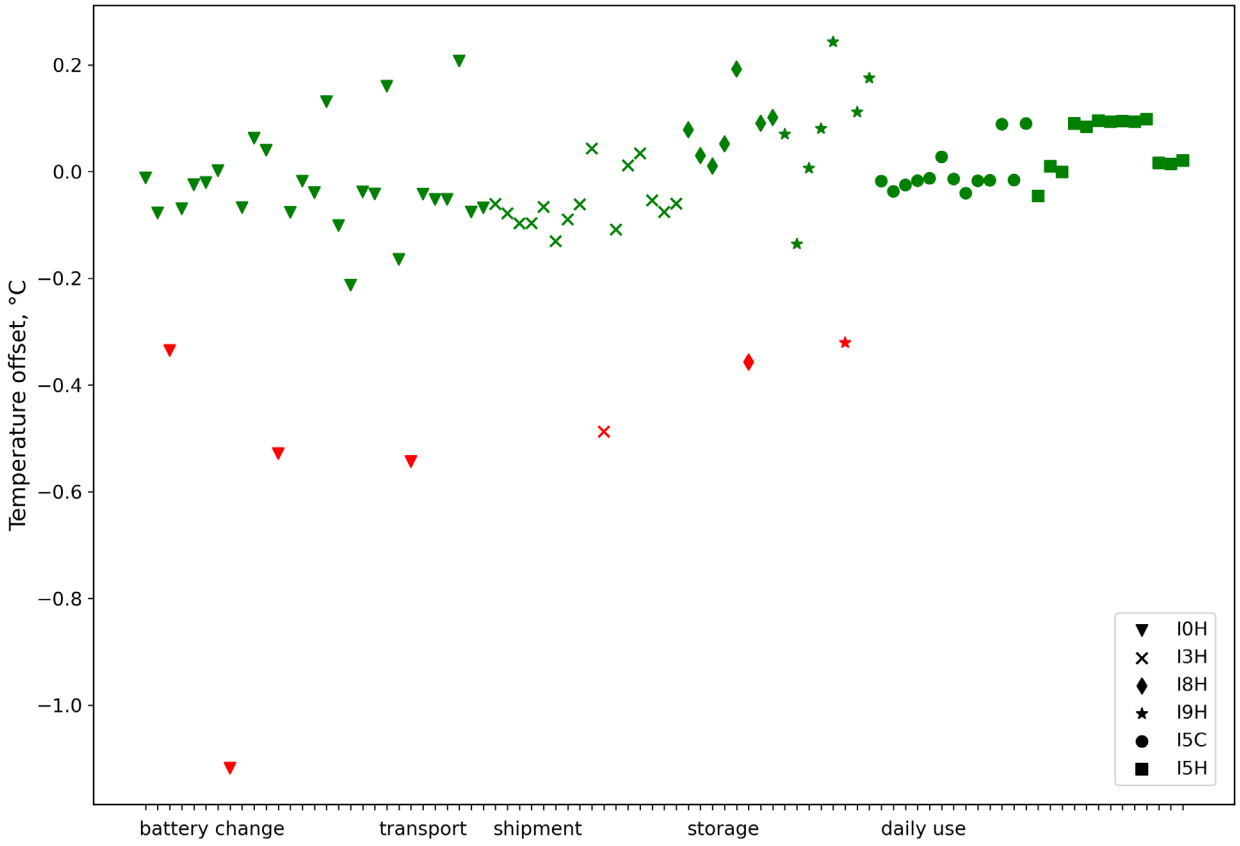


Fig. 17. Summary of all environmental chamber temperature points and IMP calibration data collected using logger model "I." The temperature offset is the difference between the DDL measurement and the reference temperature. In-tolerance points are shown in green, while out-of-tolerance points are shown in red. Individual DDL devices are distinguished by different marker symbols, as indicated in the legend. The points are shown in the chronological order in which they were measured, with the start of each corresponding trial name indicated along the x-axis.

As shown in Fig. 17, the six model "I" devices used in the study failed a total of seven calibration measurements, and in each case, the DDL recorded a temperature lower than the reference temperature. Six of these failures occurred during a 5 °C environmental chamber calibration measurement, suggesting a systematic issue. The 5 °C measurements are summarized in Fig. 18.

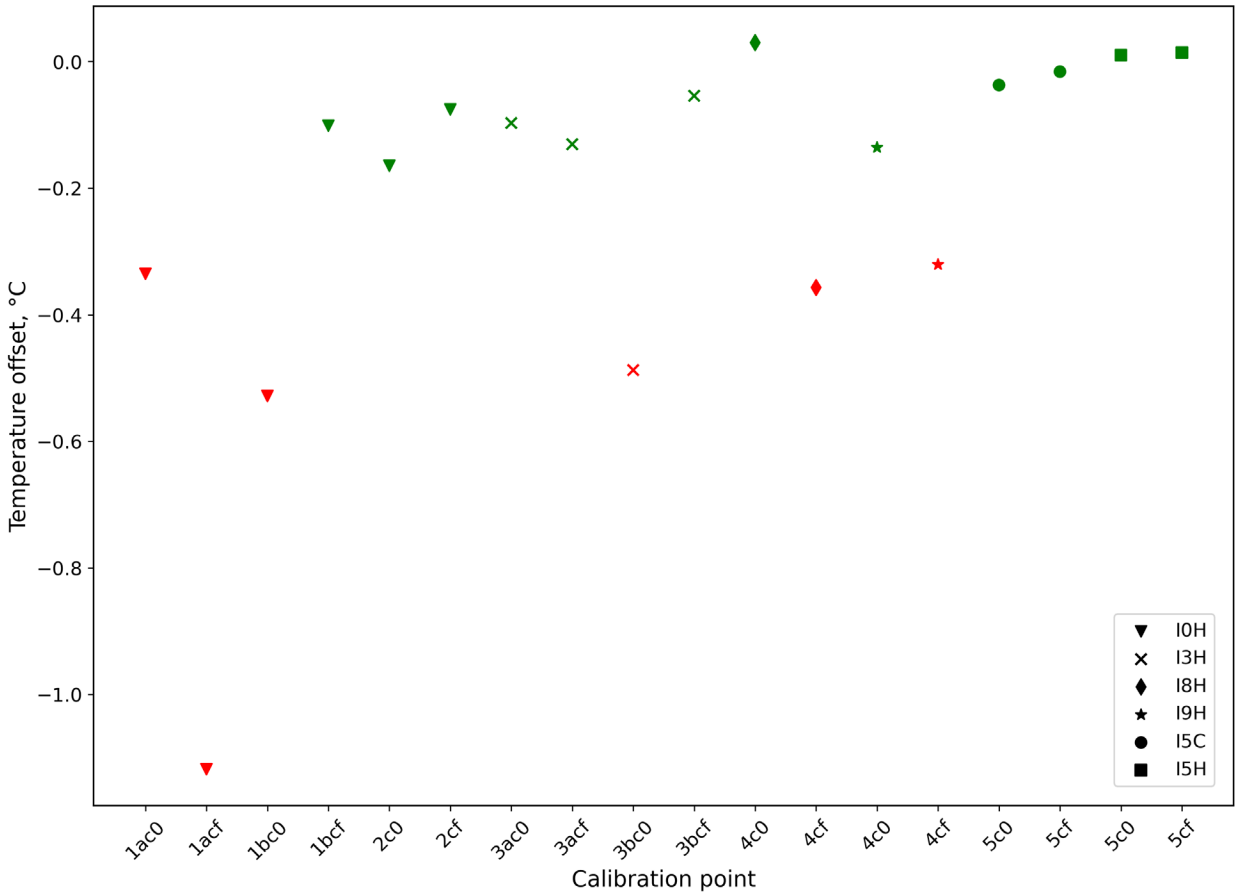


Fig. 18. Summary of logger model "I" calibration results at a nominal environmental chamber setpoint of 5 °C. The x-axis shows the corresponding test ID code (Appendix C). The temperature offset is the difference between the DDL measurement and the reference temperature. In-tolerance points are shown in green, while out-of-tolerance points are shown in red. Individual DDL devices are distinguished by different marker symbols, as indicated in the legend.

As shown in Fig. 18, the logger "I" devices failed at six of eighteen 5 °C calibration points. In each case, the DDL recorded a temperature lower than the reference temperature. These devices were supplied with an external probe immersed in a vial containing approximately 31 mL of fluid. All other devices tested in the study featured an external vial containing approximately 2 mL to 8 mL of fluid. The "I" logger vial was significantly larger than any of the other logger vials, and as a result, this device equilibrated to temperature changes at a slower rate compared to the other devices. In addition, the "I" logger had a tolerance of 0.3 °C, whereas most of the other tested models had a tolerance of 0.5 °C. For these reasons, we believe the repeated failures of this device at the 5 °C point is due to a combination of the larger vial size, the stricter tolerance band, and some procedural artifacts of our measurement methods.

During the environmental chamber calibrations, the chamber was cycled through the three calibration points (−29 °C, 5 °C, 25 °C) from the lowest to the highest temperature. The chamber requires more than 4 h to fully equilibrate and stabilize at a new temperature. To minimize the time required for each environmental chamber calibration, the runs were often compressed into a period of two workdays. As a result, the data collection time for the intermediate 5 °C point was sometimes restricted to less than 3 h, to allow for the final temperature setpoint change to occur on the same workday. By contrast, the data collection time for the −25 °C and 25 °C points often spanned upwards of 8 h, since in these cases, it was convenient to allow the chamber to equilibrate and collect data overnight, without increasing the total time for the run.

The shorter collection time for the 5 °C points means that in some cases, the larger "I" DDL vial did not fully equilibrate with the chamber during the collection period. This is evident in the logger data, where the reference PRT temperature is nominally stable, but the logger measurements are increasing. An example is shown in Fig. 19. Upon calculating the difference between the device under test and the reference temperatures, we obtain a negative temperature offset, consistent with the pattern shown in Fig. 18.

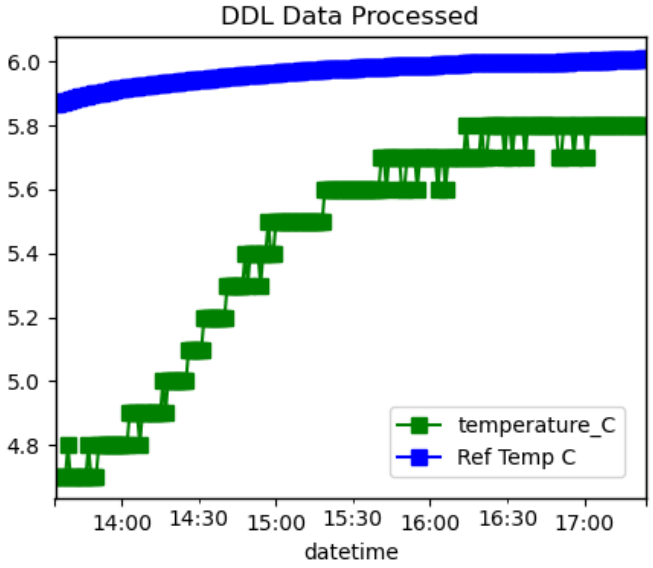


Fig. 19. Logger I during a 5 °C environmental chamber calibration point, as displayed by the data analysis graphical user interface (see Appendix B for details).

Near the end of the study, remote control capabilities were added to the environmental chamber, enabling offsite adjustment of the temperature setpoint outside regular work hours. This flexibility allowed for more uniform data collection periods. Examining the data, we see that in the final trial, during which remote control capabilities were available, logger "I" experienced zero temperature excursions. Extending the 5 °C collection period seems to have eliminated the corresponding out-of-tolerance points.

In summary, the higher incidence of out-of-tolerance points recorded by logger model "I" appears to be an artifact of the testing methodology coupled with the larger vial size and stricter tolerance band characteristic of this model, rather than a true indicator of device calibration error. A number of these apparent "failures" probably could have been eliminated by utilizing longer data collection periods at the 5 °C calibration points, consistent with the collection periods utilized at the −25 °C and 25 °C points.

From a vaccine management perspective, users may have some flexibility in determining whether a particular vial size is optimal for their vaccine temperature monitoring needs. A larger vial size will decrease temperature monitoring sensitivity to minor fluctuations, which may be advantageous in terms of eliminating "false positive" temperature excursions. False positive alarms can have significant negative consequences by increasing administrative overhead and potentially conditioning staff to ignore alarms [28]. Conversely, the reduced sensitivity associated with a larger vial also means that the logger will be slower to detect anomalous conditions, which could lead to delayed detection of temperature excursions.

NSF/ANSI 456 may provide a useful point-of-reference for future work examining the optimal setup for a vaccine temperature monitoring DDLs, including vial or weighted probe size and form factor. The standard describes a weighted temperature probe used to simulate the temperature response of a single-dose vaccine vial kept outside its original cardboard packaging, called a Vaccine Simulation Device (VSD). CDC recommendations instruct providers to keep vaccines inside their original packaging, so the VSD represents a

"worst case" vaccine storage scenario. During its development, the temperature response of the NSF 456 VSD was shown to be in good agreement with the temperature response of monitored vaccine vials [29].

4. Discussion

Typical DDL usage conditions tested within the limits of this study do not appear to be correlated with calibration measurement drift. We saw no change in the incidence of out-of-tolerance calibration measurements in response to any of the tested usage conditions (battery changes, local transport in a cooler, cross-country shipment, long-term storage, and daily use in refrigerators/freezers). We detected extremely low rates of calibration error, and rare DDL failures appeared to occur irrespective of the usage trials. Because the incidence of calibration errors does not appear to be correlated with usage conditions performed over a period of 2 years, the findings from this study support extended recalibration intervals. As of January 2023, the CDC Vaccine Storage and Handling Toolkit recommends DDL recalibration every two to three years. This timeline could be extended further if supported by manufacturer-supplied DDL stability data and/or intermediate verification checks performed in the field.

In six of eight tested DDL models, the error rate was so low that any occasional out-of-tolerance points could be explained by random statistical effects. Two tested models showed a higher rate of calibration error. One of these cases is most likely attributable to a procedural artifact of the study that was amplified for this DDL design, and not a problem inherent to the device itself. The other case appears to be a systematic calibration error confined to a particular DDL model. However, the (relatively minor) error only occurred at the 25 °C point, a temperature that is well outside the permissible ranges for vaccine storage. A small error at ambient temperatures is unlikely to have significant negative repercussions for a DDL in routine refrigerated or frozen vaccine temperature monitoring.

The study also raises some interesting questions about the role of IMP testing in DDL calibration. The IMP provides a simple mechanism to verify the performance of a DDL, without the need for specialized equipment or expertise. However, the findings from this study suggest that the IMP alone is limited in its ability to detect calibration errors at typical ambient or freezer temperatures. That is, the IMP alone does not appear sufficient to "calibrate" a DDL over a wider range of use, such as −40 °C to 30 °C. Rather, a calibration at multiple temperature points spanning the intended range of use provides the best opportunity to identify any systematic errors or anomalous behavior.

With that said, the IMP is a useful tool for detecting gross operational errors. One device was removed from the study after to failing to record data during an IMP test. Another logger appeared "stuck" measuring 1.3 °C during an IMP – an obviously erroneous result. After resetting the device and running another IMP test, the logger recorded a temperature near 0 °C, as expected. In both cases, the IMP identified a gross error, allowing for a successful correction. The IMP is perhaps most useful as a "first line of defense" strategy in situations where DDL performance is called into question, by providing a simple and inexpensive mechanism for identifying errors. Similarly, the IMP can be used as an intermediate verification method performed in the field. If a device under test records an in-tolerance result at the IMP, this result could be used to extend the time interval before a full, multi-point calibration must be repeated.

The study findings confirm that an initial, multi-point calibration covering the intended range of use provides optimal assurance of DDL performance. However, if the DDL is designated for use exclusively in monitoring refrigerated vaccine temperatures (2 °C to 8 °C), then an IMP test alone may be used in lieu of a single-point calibration near 5 °C, or a multi-point calibration spanning a larger range. Excluding the unique situation with logger "I", the study included a total of 100 DDL calibration measurements at 5 °C, and only 2 of these

measurements were identified as out-of-tolerance. This suggests that passing performance at the IMP is generally a good predictor of passing performance at 5 °C.

Conversely, if the device is intended for use in freezer monitoring or controlled room temperature monitoring, the IMP alone is not sufficient, and additional calibration point(s) covering the intended range of use are warranted. Users should examine the calibration services and certificates provided for DDLs used in vaccine temperature monitoring to ensure that the temperature range of the calibration is matched to the intended range of use. This is particularly important in the case of vaccine that is stored frozen, including COVID-19 vaccines stored at −80 °C or from −25 °C to −15 °C.

The study also highlights an open question regarding optimal weighted probe setup and/or vial size used in a vaccine temperature DDL. Clearly, vial size will affect the temperature monitoring results, with a larger vial suppressing recorded thermal fluctuations. A larger vial size may have some programmatic benefits in terms of reducing "false positive" temperature excursions, limiting "alarm fatigue," and constraining the administrative burden associated with responding to recorded temperature excursions. Conversely, a smaller vial size represents a more conservative approach, as it is more likely to detect small excursions which could have a cumulative impact on vaccine stability. This study was not designed to answer the question of an "optimal" vial size, and some variation depending on the specific needs and preferences of the user is expected. An extremely conservative, "worst case" approach designed to simulate the temperature response of an unpackaged, single-dose vaccine vial is described in NSF/ANSI 456.

5. Conclusion

We examined the effect of typical DDL usage conditions on calibration status over a period of one to two years. The usage conditions studied included battery change, local transport, cross-country shipment, long term storage, and daily monitoring in refrigerators and freezers. None of these usage conditions appeared to contribute to temperature calibration drift. These findings support the current recommendations for DDL recalibration every two to three years. This timeline could be extended further if supported by manufacturer-supplied DDL stability data and/or intermediate verification checks performed in the field.

We also investigated whether IMP test results are predictive of DDL calibration status for frozen and refrigerated vaccine monitoring applications. While passing performance at the IMP appears to be a good predictor of passing performance at refrigerated vaccine storage temperatures, IMP test performance is not a good predictor of calibration performance at freezer or room temperatures. Our findings indicate that DDLs intended for use in freezer or room temperature monitoring should be evaluated via a multi-point calibration that spans the intended range of use.

References

[1] McColloster P, Vallbona C (2011) Graphic-Output Temperature Data Loggers for Monitoring Vaccine Refrigeration: Implications for Pertussis. *Am J Public Health* 101:46–47. https://doi.org/10.2105/AJPH.2009.179853

[2] Chen D, Tyagi A, Carpenter J, Perkins S, Sylvester D, Guy M, Kristensen DD, Jones Braun L (2009) Characterization of the freeze sensitivity of a hepatitis B vaccine. *Human Vaccines* 5(1):26. https://doi.org/10.4161/hv.5.1.6494

[3] Kurzątkowski W, Kartoğlu Ü, Staniszewska M, Górska P, Krause A, Wysocki MJ (2013) Structural damages in adsorbed vaccines affected by freezing. *Biologicals* 41(2):71–76. https://doi.org/10.1016/J.BIOLOGICALS.2011.10.011

[4] Prusik T, Kumru OS, Volkin DB, Smith DE, Middaugh CR, Joshi SB (2014) Vaccine instability in the cold chain: Mechanisms, analysis and formulation strategies. *Biologicals* 42(5):237–259. https://doi.org/10.1016/j.biologicals.2014.05.007

[5] Egan W, Schofield T (2009) Basic principles of stability. *Biologicals* 37(6):379–386. https://doi.org/10.1016/j.biologicals.2009.08.012

[6] Milstien JB, Galazka AM, Kartoglu U, Zaffran M (2006) Temperature sensitivity of vaccines. *World Health Organization Dept of Immunization, Vaccines and Biologicals*:1–58. https://doi.org/WHO/IVB/06.10

[7] Boros CA, Hanlon M, Gold MS, Roberton DM (2001) Storage at −3°C for 24 h alters the immunogenicity of pertussis vaccines. *Vaccine* 19(25–26):3537–3542. https://doi.org/10.1016/S0264-410X(01)00063-9

[8] Office of Inspector General (2012) Office of the Inspector General. Department of Health and Human Services. Vaccines for Children Program: Vulnerabilities in Vaccine Management. Available at https://oig.hhs.gov/oei/reports/oei-04-10-00430.pdf

[9] Kartoğlu Ü, Nelaj E, Maire D (2010) Improving temperature monitoring in the vaccine cold chain at the periphery: An intervention study using a 30-day electronic refrigerator temperature logger (Fridge-tag®). *Vaccine* 28(24):4065–4072. https://doi.org/10.1016/j.vaccine.2010.03.076

[10] Kartoglu U, Ames H (2022) Ensuring quality and integrity of vaccines throughout the cold chain: the role of temperature monitoring. *https://doi.org/101080/1476058420222061462* 21(6):799–810. https://doi.org/10.1080/14760584.2022.2061462

[11] Wexler D (2018) TECHNICALLY SPEAKING: CDC Issues New Requirements for VFC Vaccine Storage. *Immunize.org*. Available at https://www.immunize.org/technically-speaking/20180123.asp

[12] Centers for Disease Control and Prevention, National Center for Immunization and Respiratory Diseases (2022) Vaccine Storage and Handling Toolkit. Available at https://www.cdc.gov/vaccines/hcp/admin/storage/toolkit/storage-handling-toolkit.pdf

[13] Wilson E, Zhu C, Galea S, Marko A, Victoria Urdaneta V, Straus W (2018) Turning up the heat: Effect of new vaccine for children's (VFC) program recommendations for use of temperature monitors upon incorrect product storage adverse event reporting. *Vaccine* 36(11):1516–1520. https://doi.org/10.1016/j.vaccine.2017.10.059

[14] Association of Immunization Managers (2017) Lessons from the Field: Digital Data Loggers. Available at https://www.immunizationmanagers.org/content/uploads/2021/09/AIM_LESSONS_FIELD_DDL_LYT_11.pdf

[15] Chojnacky M, Miller W, Strouse G (2012) Data Logger Thermometers for Vaccine Temperature Monitoring. *NIST Interagency/Internal Report (NISTIR) - 7899* (Gaithersburg, MD). https://doi.org/10.6028/NIST.IR.7899

[16] Mangum BW (1995) Reproducibility of the temperature of the ice point in routine measurements. *NIST Technical Note 1411* (Gaithersburg, MD). https://doi.org/10.6028/NIST.TN.1411

[17] Mercury Thermometer Alternatives: Verification Methods for Alternative Thermometers | NIST Available at https://www.nist.gov/pml/mercury-thermometer-alternatives/mercury-thermometer-alternatives-verification-methods

[18] ASTM International (2017) *ASTM E2593-17 - Standard Guide for Accuracy Verification of Industrial Platinum Resistance Thermometers.* (ASTM International, West Conshohocken, PA). https://doi.org/10.1520/E2593-17.2

[19] Taylor BN, Kuyatt CE (1994) Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results.

[20] Meyer CW, Herman T, Miller WW Calibration of Hygrometers with the Hybrid Humidity Generator. https://doi.org/10.6028/NIST.SP.250-83r1

[21] NDFD - National Digital Forecast Database - MDL - Virtual Lab Available at https://vlab.noaa.gov/web/mdl/ndfd

[22] ASTM International (2020) *ASTM E344-20 - Terminology Relating to Thermometry and Hydrometry*. (ASTM International, West Conshohocken, PA). https://doi.org/10.1520/E0344-20

[23] Joint Committee for Guides in Metrology (2012) *International vocabulary of metrology – Basic and general concepts and associated terms (VIM)*. 3rd Ed. (JCGM 200:2012 (2017 version with minor corrections and informative annotations)). Available at https://jcgm.bipm.org/vim/en/

[24] ASTM International (2019) *ASTM E2877-12(2019) - Standard Guide for Digital Contact Thermometers*. (ASTM International, West Conshohocken, PA). https://doi.org/10.1520/E2877-12R19

[25] ASTM International (2020) *ASTM E2846-20 - Standard Guide for Thermocouple Verification*. (ASTM International , West Conshohocken, PA). https://doi.org/10.1520/E2846-20

[26] California Department of Public Health (2023) Digital Data Loggers – California Vaccines for Children (VFC). *EZIZ*. Available at https://eziz.org/vaccine-storage/digital-data-loggers/

[27] NCSL International (2013) *ANSI/NCSL Z540.3-2006 (R2013) - Requirements for the Calibration of Measuring and Test Equipment*. (NCSL International, Boulder, CO). Available at https://ncsli.org/page/z5403

[28] Sendelbach S, Funk M (2013) Alarm Fatigue: A Patient Safety Concern. *AACN Advanced Critical Care* 24(4):378–386. https://doi.org/10.4037/NCI.0B013E3182A903F9

[29] NSF International (2021) *NSF/ANSI 456-2021a - Vaccine Storage* (NSF International, Ann Arbor, MI). Available at https://www.techstreet.com/standards/nsf-ansi-456-2021a?product_id=2224651

Appendix A. DDL Specifications

Table A1. DDL specifications.

Logger ID	A	B	C	D	E	F	G	I
Battery								
Battery type	CR 2032	CR 2032	AAA	CR 2032	1/2 AA	AA	AAA	CR 2032
Typical battery life, years	3.5	3.5	2	2	2	0.23	unspecified	1 year
Battery replaceable by end-user?	no	no	yes	yes	yes	yes	yes	yes
Logger display and recording								
Minimum logging rate, min:s	5:00	5:00	0:05	0:05	0:01	1:00	1:00	0:30
Log rate adjustable by end-user?	no	no	yes	yes	yes	yes	yes	yes
Device memory, data points	16,128	16,128	16,128	16,128	32,510	24,000,000	525,600	16,905
Device memory for 5 min read rate	56 days	56 days	56 days	56 days	112 days	228 years	5 years	58 days
Output filetype	PDF, TXT	PDF, TXT	CSV, PDF	CSV, PDF	CSV, JPEG, PDF, TXT	CSV	CSV	CSV, PDF, TXT
Typical data download time	1 min	1 min	25 s (657 points/s)	10 s (1642 points/s)	10 s	5 s	1 min to >10 min (55 points/s)	10 s
Recording behavior when memory full	overwrite	overwrite	programmable, stop or overwrite	programmable, stop or overwrite	programmable, stop or overwrite	logs indefinitely	overwrite	programmable, stop or overwrite
Logger body dimensions, mm	128 x 75 x 19	128 x 75 x 19	40 x 93 x 24	40 x 81 x 15	135 x 24 x 21	94 x 110 x 25	70 x 108 x 19	54 x 93 x 9
External temperature sensor								
Temperature sensor type	unspecified	unspecified	thermistor	thermistor	thermistor	thermistor	unspecified	thermistor
Probe immersion depth, mm	12.5	12.5	13	15	38	15	10	33
Probe diameter, mm	5	5	2.5	2.5	5	5	5	3.2
Alarm type	Audio/Visual	Audio/Visual	Visual	Visual	Visual	Audio/Visual	Audio/Visual	Audio/Visual
Resolution, °C	0.1	0.1	0.1	0.1	0.1	0.1	0.01	0.1
Temperature accuracy, °C	±0.5	±0.5	±0.5	±0.5	±0.5	±0.5	±0.25	±0.3
Measurement Range	−40 °C to 60 °C	−40 °C to 60 °C	−50 °C to 40 °C	−40 °C to 40 °C	−40 °C to 125 °C	−50 °C to 70 °C	−50 °C to 70 °C	−40 °C to 40 °C
Cable length, m	1.5	1.5	3.0	0.9	0.9	2.0	3.1	1.5
Buffer bottle								
Bottle material	unspecified	unspecified	PETG*	PETG*	unspecified	polyethylene	unspecified	high-density plastic
Bottle diameter, mm	26	26	25.4	25.4	22	19.5	25.4	32
Bottle height, mm	34	34	50.8	50.8	85	40	63.5	73
Bottle capacity, mL	6.2	6.2	9	8.6	5.2	2.8	4.4	35
Bottle fill volume, mL	5.5	5.5	5.5	8	5	1.8	4.2	31
Buffer fluid	"Biosafe" glycol	"Biosafe" glycol	Propylene glycol, GRAS**	Propylene glycol, GRAS**	Glycol, unspecified	"Biosafe" glycol	Nontoxic glycol, GRAS**	Glycol, unspecified
System and software requirements								
Software required?	No	No	Yes	Yes	Yes	No	No	Yes
System Requirements	PC/Mac USB port	PC/Mac USB port	Windows & USB port	Windows & USB port	Windows & USB port	PC/Mac USB port	PC/Mac USB port	Windows & USB port
Manufacturer-provided calibration								
Calibration temperature points	5 °C	−15 °C	41 °F (5.56 °C)	none specified	−15 °C, 5 °C	4 °C	−40 °C, 0.001 °C, 50 °C	−15 °C and 5 °C
Calibration uncertainty	0.12 °C	0.12 °C	0.12 °F (0.07 °C)	N/A	0.12 °C	0.05 °C	0.024 °C, 0.01 °C, 0.01 °C	0.12 °C
Logger price (2019)	$149	$149	$129	$89	$139	$144	$179	$196

Appendix B. Data Processing Methods

Data download from each of the eight data models used in the study generated data files with different attributes, including file type (.csv, .pdf, .txt), file storage location, and file naming conventions. Variable data structures and formatting conventions were also utilized within the files. This appendix details the strategies used to harmonize these DDL outputs and convert temperature time series data into usable "calibration offset" results, allowing us to compare logger calibration status across the different usage trials performed during this study.

B.1. Data Output File Harmonization

The DDLs selected for use in the study generated data files featuring different column numbers, column labels, datetime formats, and time series chronology. Different models also generated unique headers, which varied in length and reported different content. Some devices presented temperature excursion and alarm data within the header, resulting in non-standard header lengths which varied from one trial to the next. Different models also used different conventions to identify probe disconnection events. Values including −40, −50, 0, or strings such as '- -', '-.—', 'Sensor Disconnected' were recorded in the temperature column to indicate a disconnected probe.

To overcome these challenges, we developed a standard header format containing critical metadata pertaining to each device and the associated measurement file. This header included the assigned DDL name, serial number, operating range, accuracy, resolution, test logging interval, and memory capacity (readings). In addition, a standard data format was established, containing one column for a datetime string, and a second column for the temperature data series, reported in degrees Celsius. Probe disconnection events were labeled with the string "Probe Disconnected." A Python script was developed to automatically read in downloaded data files from a defined location, identify the logger model, and apply the appropriate conversion processes to extract the standard header information and time/temperature data series. The processed output was saved as a new comma separated value (.csv) data file conforming to the established formatting conventions. Finally, all available data files corresponding to the same unique DDL device were appended to a single .csv file associated with that device, named according to the DDL ID (e.g. "A0H_Appended.csv").

Each time a test was performed, the datetime associated with the start and end of that test was manually recorded in an Excel spreadsheet next to the appropriate test ID. This allowed another Python script to interrogate the "DDL_Appended.csv" file associated with each device and separate the time/temperature data into smaller chunks associated with each test ID, using the datetime ranges specified in the Excel file. These data chunks were saved as individual files using the convention "DDL_testID.csv." In addition, data was saved in an SQL database containing logger name, datetime, test ID, and temperature measurement associated with each saved data point, to enable later flexible data extraction and analysis using these parameters.

A .csv file containing the environmental chamber reference PRT data was automatically generated and saved by upon completion of each chamber run. A separate Python script was used to standardize these output files and combine them into a single .csv file which functioned as a database for all reference chamber data.

B.1.1. Calibration Offset Extraction

To evaluate DDL drift in response to the tested usage conditions, the time/temperature data series associated with each calibration test ID was translated into a calibration temperature offset value. To accomplish this, the data was first trimmed to exclude unstabilized temperature data.

Each data series corresponding to a particular test ID typically included some measurements recorded while the device sensor was equilibrating the known temperature environment. For example, during an IMP test, DDL probes were transferred from an ambient environment of approximately 22 °C to an ice bath at 0 °C. The DDL readings typically stabilized near 0 °C after a period of 30 min to 60 min – the time required for the probes immersed in glycol-filled vials to cool down and equilibrate with the ice bath. This equilibration time varied depending on the size of the vial. Similarly, in the environmental chamber tests, the chamber itself required time to cool from room temperature down to the first calibration point, nominally −25 °C. Each subsequent set point adjustment required additional time for both the chamber temperature to stabilize and the DDL probes to equilibrate with the new chamber temperature set point. These set point changes and subsequent stabilization periods were excluded from the trimmed calibration data.

We initially set out to complete the data trimming process through the exclusive use of algorithmic controls, with adjustable parameters such as rolling averages, standard deviations, and closeness to a specified nominal set point temperature used to exclude rapid temperature changes. However, taking the size of the study, variations between devices, and the potential for unpredictable errors into account, we decided that human intervention via visual inspection of graphical results to verify algorithmic filtering accuracy would significantly improve the reliability of the outputs.

To this end, a graphical user interface (GUI) was developed using the Python package Tkinter. A screenshot of the GUI is shown in Fig. A1. The GUI pulled in raw data compiled in the SQL database and plotted the measurements from each DDL and test ID in graphical form. Reference temperature data was also pulled in and overlaid with the filtered data plot.

A set of user-adjustable trim parameters adjusted the rolling average window and allowable standard deviation threshold to exclude any portions of the data series containing sudden temperature changes. After the trim parameters were set, the GUI displayed the resulting trimmed dataset in two plots: 1) overlaid with the untrimmed dataset, and 2) plotted alongside the reference temperature dataset. The user could then navigate between trials and iterate through each DDL and test ID to assess the algorithmic filtering results, adjusting trim parameters if needed to ensure that only properly stabilized data was retained in the trimmed results.

All adjustments were automatically saved to a new SQL database containing the processed data, leaving the original SQL database unaltered. Whenever the user returned to a specific test ID within the GUI, prior adjustments would be shown and replotted automatically. Preserving the original data ensured that the user could revert to the default parameters or make further adjustments at any point during the analysis process.

Once the trimming parameters were verified and adjusted as needed, the user was required to push a button to “accept” the trimmed temperature data series. In this way, the GUI required user intervention for every measurement recorded during the study. User acceptance of a data series triggered the creation of a new data object in a .csv file, which contained the trimmed

time/temperature series along with other important values. Additional filtering modifications did not affect the processed temperature point database unless the user elected to accept the new changes, at which point the old record would be overwritten. This strategy supported flexibility in processing while preventing accidental or erroneous changes to the processed data.

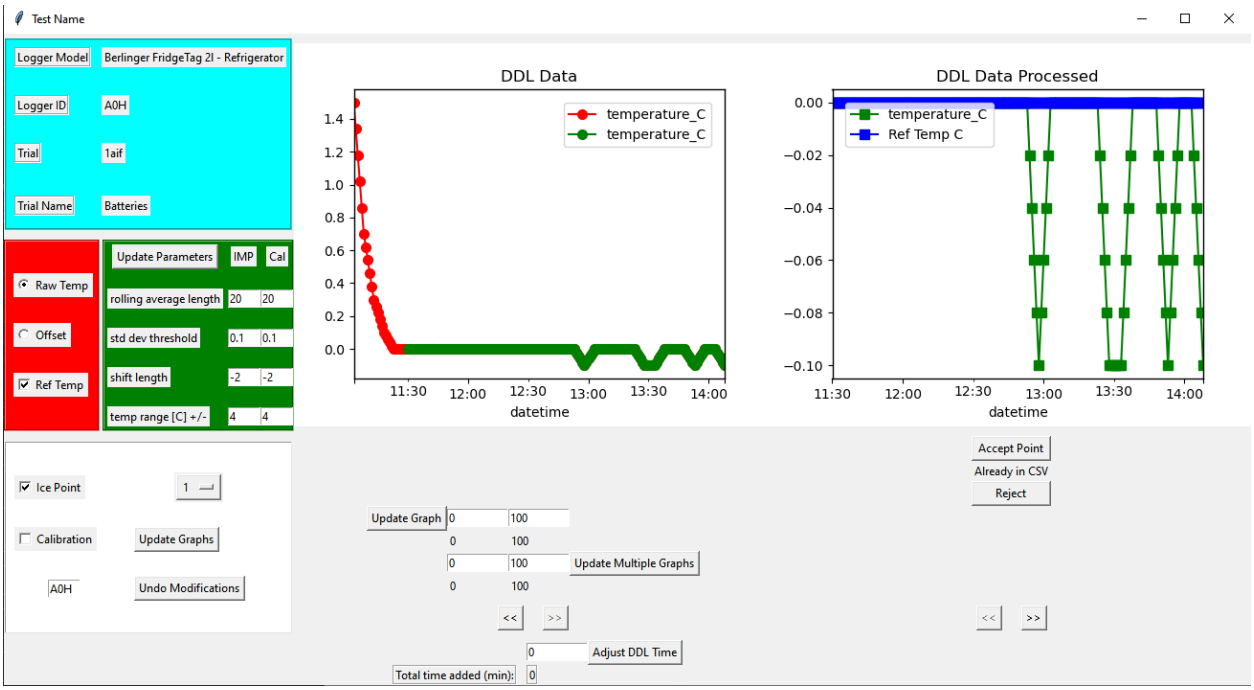


Fig. B1. Screenshot of the data trimming GUI, showing an IMP test performed with logger A0H during the battery change trial. Filtering parameters are listed in the green box. The unfiltered data, as designated by to the time window recorded for each calibration run, is displayed as red circles on the left plot. The filtered data is shown as green squares on both plots, and the reference data is shown in blue. Updating the filtering parameters changes the trimmed selection and updates the plot. Clicking the “accept point” button saves the trimmed data to the processed temperature point database.

To maintain consistency, the same filtering window was applied to all DDLs measured at a particular temperature point within a test. In exceptional cases, the filtering window was adjusted on a per-logger basis to eliminate erroneous data points. The GUI-based visual verification process exposed some previously undetected errors in the data, such as improperly adjusted DDL datetimes. Datetime errors usually arose when measurements were performed around a daylight savings time change, or occasionally due to logger malfunction. The GUI was designed to permit adjustment of time settings if needed, by shifting the DDL time (in minutes) to ensure that it lined up with the reference data time.

The filtering actions completed in the GUI saved updates to a master data summary file, called “Temperature Points.csv”. This file contained all important metadata associated with every DDL measurement series, the parameters used to filter the data series, the resulting filtered data range, and summary statistics associated with the filtered data set. These statistics included the mean DDL and reference temperatures for the filtered series, and the difference between these two values, which we’ve identified as the “calibration offset” for the series.

Appendix C. Test ID Codes

Table C1. Test ID code descriptions.

Trial	Subtrial	Test	Trial name	Subtrial name	Test description	Test id
0		ec	Extended calibration		extended calibration	0ec
0		i	Extended calibration		IMP	0i
1	a	c0	Batteries	control	initial calibration	1ac0
1	a	i0	Batteries	control	intial IMP	1ai0
1	a	i1	Batteries	control	IMP	1ai1
1	a	i2	Batteries	control	IMP	1ai2
1	a	cf	Batteries	control	final calibration	1acf
1	a	if	Batteries	control	final IMP	1aif
1	b	c0	Batteries	change	initial calibration	1bc0
1	b	i0	Batteries	change	initial IMP	1bi0
1	b	i1	Batteries	change	IMP	1bi1
1	b	cf	Batteries	change	final calibration	1bcf
1	b	if	Batteries	change	final IMP	1bif
2		i0	Transport		initial IMP	2i0
2		c0	Transport		initial calibration	2c0
2		t1_1	Transport		transport monitor	2t1_1
2		t1_2	Transport		transport monitor, append suffix if additional	2t1_2
2		t1_3	Transport		transport monitor, append suffix if additional	2t1_3
2		t1_4	Transport		transport monitor, append suffix if additional	2t1_4
2		t1_5	Transport		transport monitor, append suffix if additional	2t1_5
2		t1_6	Transport		transport monitor, append suffix if additional	2t1_6
2		t1_7	Transport		transport monitor, append suffix if additional	2t1_7
2		t1_8	Transport		transport monitor, append suffix if additional	2t1_8
2		i1	Transport		IMP (after 8 transports)	2i1
2		t2_1	Transport		transport monitor	2t2_1
2		t2_2	Transport		transport monitor	2t2_2
2		t2_3	Transport		transport monitor	2t2_3
2		t2_4	Transport		transport monitor	2t2_4
2		t2_5	Transport		transport monitor	2t2_5
2		t2_6	Transport		transport monitor	2t2_6
2		t2_7	Transport		transport monitor	2t2_7
2		t2_8	Transport		transport monitor	2t2_8
2		i2	Transport		IMP (after 8 transports)	2i2
2		t3_1	Transport		transport monitor	2t3_1
2		t3_2	Transport		transport monitor	2t3_2
2		t3_3	Transport		transport monitor	2t3_3
2		t3_4	Transport		transport monitor	2t3_4
2		t3_5	Transport		transport monitor	2t3_5
2		t3_6	Transport		transport monitor	2t3_6
2		t3_7	Transport		transport monitor	2t3_7
2		t3_8	Transport		transport monitor	2t3_8
2		cf	Transport		final calibration	2cf
2		if	Transport		final IMP	2if
3	a	c0	Shipping	summer	initial calibration	3ac0
3	a	i0	Shipping	summer	initial IMP	3ai0
3	a	s	Shipping	summer	shipping monitor	3as
3	a	cf	Shipping	summer	final calibration	3acf
3	a	if	Shipping	summer	final IMP	3aif
3	b	c0	Shipping	winter	initial calibration	3bc0
3	b	i0	Shipping	winter	initial IMP	3bi0
3	b	s	Shipping	winter	shipping monitor	3bs
3	b	cf	Shipping	winter	final calibration	3bcf
3	b	if	Shipping	winter	final IMP	3bif
4		c0	Storage		initial calibration	4c0
4		i0	Storage		initial IMP	4i0
4		cf	Storage		final calibration	4cf
4		if	Storage		final IMP	4if
5		c0	Daily use		initial calibration	5c0
5		i0	Daily use		initial IMP	5i0
5		m1	Daily use		fridge/freezer monitor	5m1
5		m1_2	Daily use		fridge/freezer monitor, append suffix if	5m1_2
5		i1	Daily use		IMP	5i1
5		m2	Daily use		monitor	5m2
5		i2	Daily use		IMP	5i2
5		m3	Daily use		monitor	5m3
5		i3	Daily use		IMP	5i3
5		m4	Daily use		monitor	5m4
5		i4	Daily use		IMP	5i4
5		m5	Daily use		monitor	5m5
5		i5	Daily use		IMP	5i5
5		m6	Daily use		monitor	5m6
5		cf	Daily use		final calibration	5cf
5		if	Daily use		final IMP	5if

Appendix D. Supplementary Figures

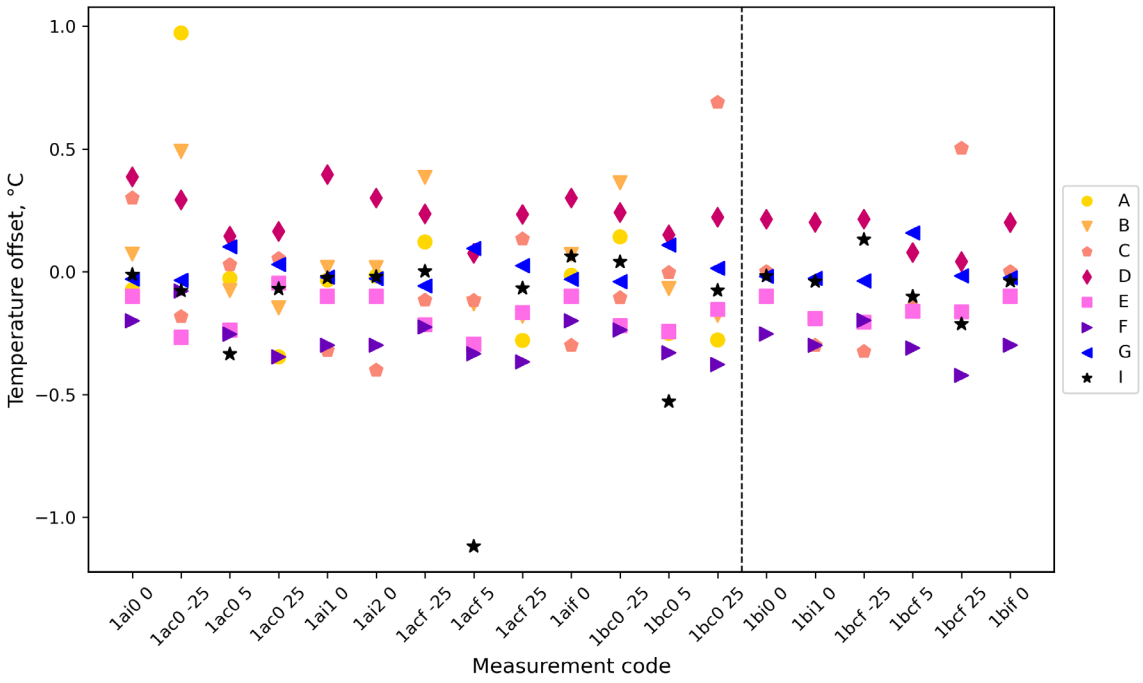


Fig. D1. Complete set of DDL calibration offsets from the battery change trial. The measurement code is the test ID code (App. C) followed by the nominal calibration point. The vertical dashed line indicates the chronological point at which the battery change was performed.

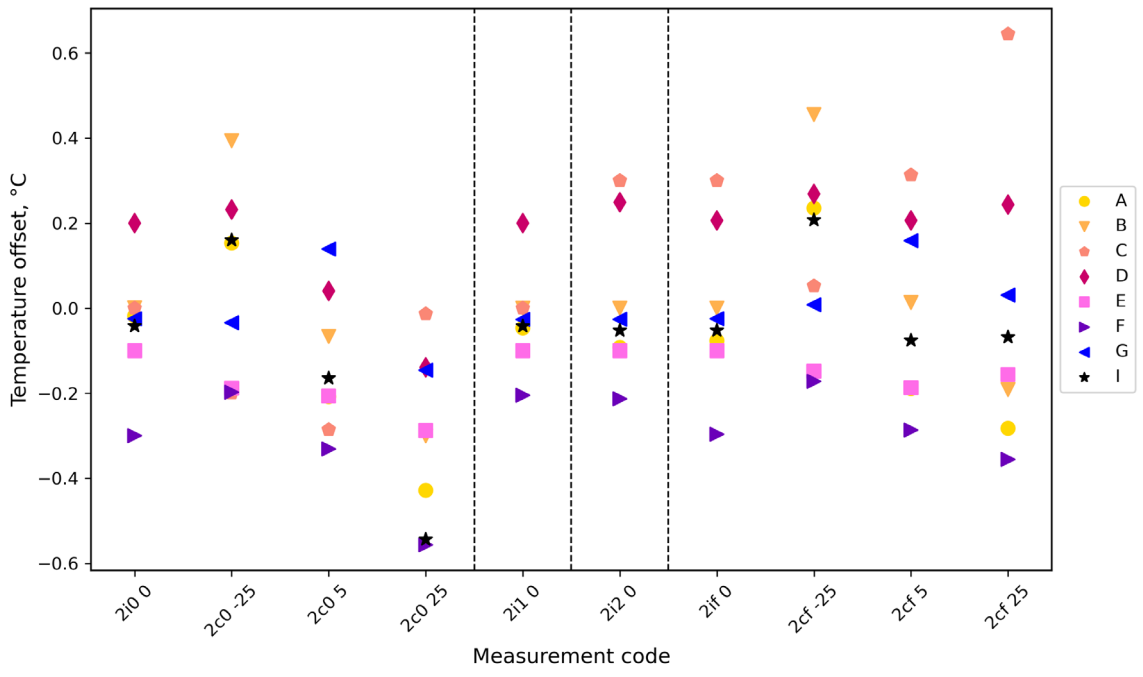


Fig. D2. Complete set of DDL calibration offsets from the local transport trial. The measurement code is the test ID code (App. C) followed by the nominal calibration point. The vertical dashed lines indicate the chronological points at which a set of 8 local transport events occurred.

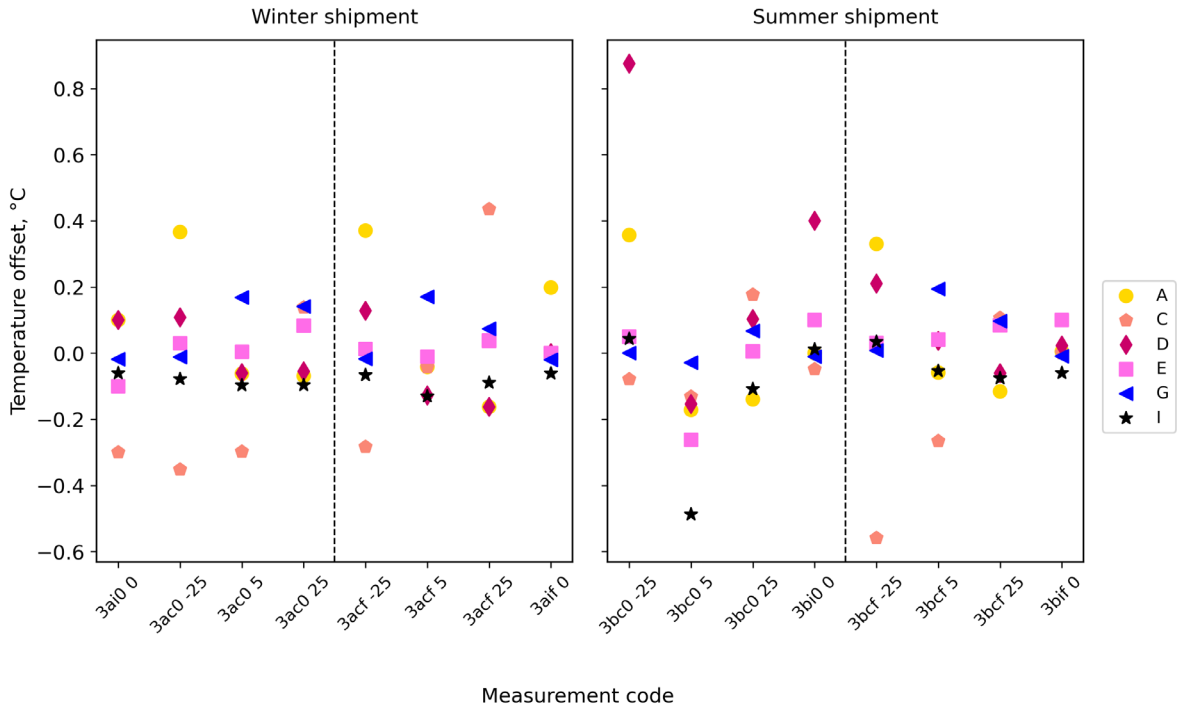


Fig. D3. Complete set of DDL calibration offsets from the shipping trials. The measurement code is the test ID code (App. C) followed by the nominal calibration point. The vertical dashed lines indicate the chronological points at which a cross-country shipping event occurred.

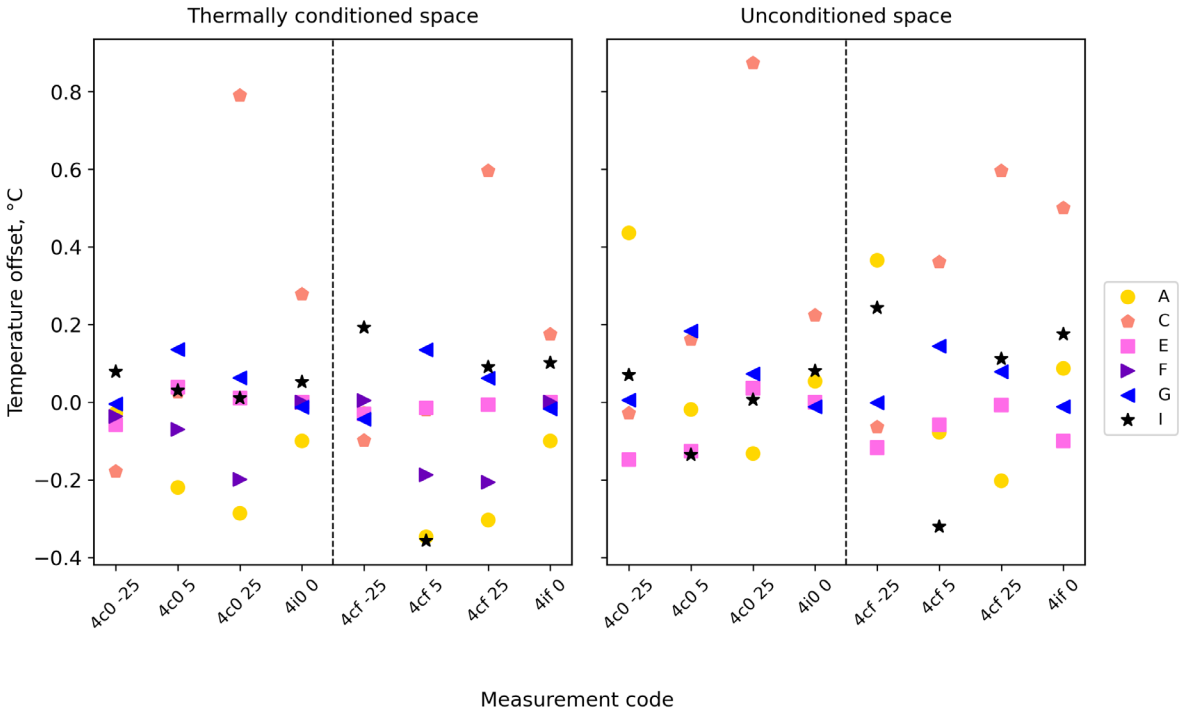


Fig. D4. Complete set of DDL calibration offsets from the long-term storage trials. The measurement code is the test ID code (App. C) followed by the nominal calibration point. The vertical dashed lines indicate the chronological points at which the DDLs were stored for 13 months.

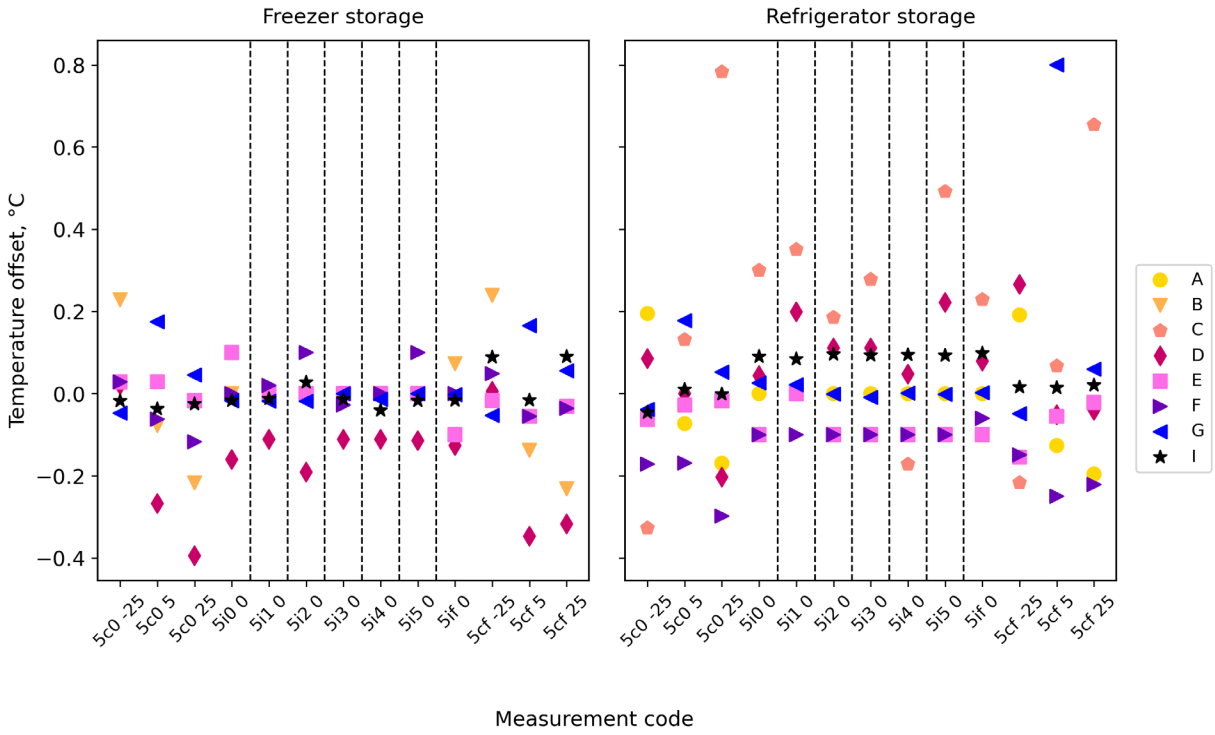


Fig. D5. Complete set of DDL calibration offsets from the daily usage trials. The measurement code is the test ID code (App. C) followed by the nominal calibration point. The vertical dashed lines indicate the chronological points at which the month-long daily use monitoring periods in a refrigerator and a freezer were performed.