

Article

Heat Metrics and Heat-Release Pathways Change Thermal Runaway Rankings of Commercial Cylindrical Lithium-Ion Cells

Yunfei Lou ^{1,2} and Ningning Wei ^{2,*} ¹ Intelligent Manufacturing School, Jilin Tiedao University, Jilin 132299, China² School of Energy and Mechanical Engineering, Dezhou University, Dezhou 253023, China

* Correspondence: weiningning@dzu.edu.cn

Abstract

Thermal runaway (TR) severity can be ranked using absolute or energy-normalised heat and by the pathway through which heat leaves the cell, but whether these definitions identify the same high-hazard cells remains unclear. We analysed 98 non-ISC heater-triggered tests from 13 commercial or near-commercial cylindrical cell models in the Battery Failure Databank. Cathode chemistry was not analysed because it was incompletely documented. Total, body, and ejecta heat were evaluated using allometric scaling, S-FTRC-generation-adjusted fixed-effects models, hierarchical bootstrap ranking, and isometric log-ratio analysis. Scaling exponents for total, ejecta, and body heat were 0.899, 1.003, and 0.879, respectively, with all 95% bootstrap intervals including $b = 1$. Despite this approximate proportional scaling, generation-adjusted absolute-total and total-kJ/Wh rankings disagreed for 46.2% of cell-model pairs. Body- and ejecta-heat rankings diverged even more strongly, with 62.8% point discordance, a bootstrap median of 61.5% (95% interval 53.8–67.9%), and 28 of 78 pairs showing reversal probabilities $\geq 95\%$. These findings show that absolute heat, energy-normalised heat, body heat, and ejecta heat address different engineering questions and should be reported and interpreted separately rather than combined into a universal cell-safety ranking.

Keywords: thermal runaway; lithium-ion battery; fractional calorimetry; heat-release pathway; hazard ranking

1. Introduction

Lithium-ion batteries (LIBs) have enabled the rapid expansion of electric vehicles, portable electronics, aerospace systems, and stationary energy storage [1–3]. Their increasing energy density, however, places greater demands on the prevention and mitigation of thermal runaway (TR) [4–6]. Once initiated by an internal defect, external heating, mechanical penetration, overcharge, or another abuse condition, TR can produce rapid self-heating, venting, material ejection, gas release, and fire [7–9]. In a multicell system, the heat released by the first failed cell may initiate TR in neighbouring cells, making quantitative characterisation of cell-level heat release central to propagation-resistant battery design.

TR heat is often summarised using a single total-energy value [10–12]. Such a representation is convenient but does not describe how the energy leaves the cell. For a cylindrical cell, part of the heat is transferred through the casing, while another part is carried by hot gases, particles, electrolyte, and electrode materials ejected through the positive or negative end [13–15]. These pathways impose different thermal boundary conditions on



Academic Editor: Yong-Joon Park

Received: 11 August 2026

Revised: 4 September 2026

Accepted: 8 September 2026

Published: 9 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

surrounding components. Heat transferred through the cell body is more directly associated with conductive loading of adjacent cells, supports, thermal barriers, and cooling structures [16–18]. Ejecta heat is instead associated with directional jets, hot particles, flame impingement, and localised heating away from the cell surface. Fractional TR calorimetry was developed to separate these contributions and quantify total heat, cell-body heat, and positive- and negative-end ejecta heat [19]. Subsequent studies showed that cell design, failure mechanism, abuse method, internal propagation dynamics, and the timing and amount of mass ejection can all influence the measured thermal response [19–21].

The open-access Battery Failure Databank substantially expanded the evidence available for studying these effects [21–23]. It contains hundreds of fractional TR calorimetry experiments spanning multiple commercial cell models, cell formats, and trigger mechanisms, together with post-test mass measurements and, for many tests, high-speed synchrotron radiography [21]. Analysis of the databank has demonstrated considerable variation between nominally identical cells tested under similar conditions. It has also shown that unusually severe responses can reflect physical failure processes, including vent obstruction and the time-resolved interaction between internal TR propagation and material ejection, rather than simple measurement error [20,21]. The existence of such repeated-test variability means that hazard comparisons should be expressed probabilistically rather than through a single deterministic rank.

Recent work has used Battery Failure Databank measurements to predict the distributions of total and fractional heat output from ejected mass and cell metadata [23]. These studies confirmed that mean heat release, variance, and the relationships between mass and heat are strongly dependent on cell model. They also showed that body, positive-end, and negative-end heat outputs have different levels of predictability and variability [24]. Parallel calorimeter development has continued to emphasise the importance of separating ejected and non-ejected heat, while newer experimental methods seek to resolve not only total energy, but also its temporal evolution [25].

Although heat-path separation is increasingly recognised, the choice of scale used to report TR heat introduces a second source of ambiguity [26–28]. Absolute heat release, expressed as kJ per failed cell, characterises the magnitude of one cell-level failure event. In contrast, heat normalised by nominal capacity or energy, commonly expressed as $\text{kJ}\cdot\text{Ah}^{-1}$ or $\text{kJ}\cdot\text{Wh}^{-1}$, characterises heat release relative to the electrical size of the cell [5,18,29]. Capacity normalisation has been widely used in fractional calorimetry studies, while energy normalisation has been used in cross-study meta-analyses of heat release, heat-release rate, and gas emissions [21,24,30]. These transformations facilitate comparison among cells with different capacities or formats, but they do not necessarily answer the same engineering question as absolute heat.

Dividing heat release by capacity or nominal energy also introduces an implicit scaling assumption. If total heat is proportional to nominal cell energy, $\text{kJ}\cdot\text{Wh}^{-1}$ provides a reasonable linear adjustment for electrical size. If the relation is sublinear or superlinear, simple division may systematically alter comparisons between small and large cells. Even when proportional scaling is approximately valid at the population level, ranking reversals may still occur, because absolute and normalised quantities represent different system boundaries. A cell producing a relatively large amount of heat per failed unit may produce a lower heat intensity per unit nominal energy, or vice versa. Previous studies have reported normalised heat distributions, correlations with mass ejection, and differences associated with cell geometry and abuse type [21,24,30]. However, the extent to which alternative metric definitions reverse pairwise hazard rankings, and whether those reversals exceed repeated-test uncertainty, has not been systematically quantified.

A similar unresolved issue applies to heat pathways. Total, body, and ejecta heat are physically related, but they are not interchangeable engineering outcomes [30–32]. A cell with high body heat may not have high ejecta heat, and a cell dominated by positive-end ejection may require different mitigation from one exhibiting bilateral or negative-end ejection. Conventional analyses based on individual percentages also face a statistical constraint: body, positive-end, and negative-end heat fractions sum to unity and therefore form compositional data. Treating these fractions as independent variables can produce misleading associations. Log-ratio coordinates provide a more appropriate framework for separating body-versus-ejecta balance from positive-versus-negative directional balance.

Experimental comparability must also be considered. TR measurements depend not only on the cell, but also on the trigger protocol, confinement, oxygen availability, calorimeter geometry, heat-loss correction, and measurement configuration. Recent comparative work has shown that experimental setup can influence measured mass loss, gas composition, maximum temperature, and energy-related outcomes, even when the same cell and nominal heating procedure are used [33–35]. Consequently, analyses combining experiments collected over multiple test campaigns should adjust for apparatus generation and avoid interpreting device effects causally when cell models are not tested in a fully crossed design.

The present study reanalysed repeated fractional calorimetry tests from commercial and near-commercial lithium-ion cells in the Battery Failure Databank. Because cathode chemistry was not consistently documented for all included cell models, chemistry-stratified comparisons, such as LFP versus NMC/NCA cells, were not performed; the analysis instead focused on the directly reported cell-level thermal quantities. The objectives were to (1) test whether total, body, and ejecta heat scaled proportionally with nominal cell energy; (2) quantify changes in hazard ranking produced by absolute versus energy-normalised metrics; (3) determine whether body-heat and ejecta-heat rankings identified the same high-hazard cells; (4) separate ranking changes attributable to scale definition from those attributable to heat pathway; and (5) estimate the probability that individual pairwise ranking reversals persisted after accounting for repeated-test variability and S-FTRC generation. Heat-path composition was further examined using isometric log-ratio coordinates, while trigger-method, leave-one-out, cell-format, and data-quality analyses were used to evaluate robustness.

All quantities analysed in this study originated from single-cell experiments. Energy-normalised values were therefore interpreted as cell-level heat release per unit nominal cell energy and not as measured module- or pack-level thermal performance. The central hypothesis was that no single TR heat metric provides a universal cell ranking because absolute event magnitude, nominal-energy-normalised intensity, conductive body heat, and ejecta-mediated heat represent distinct engineering questions. The complete analytical framework is summarised in Figure 1.

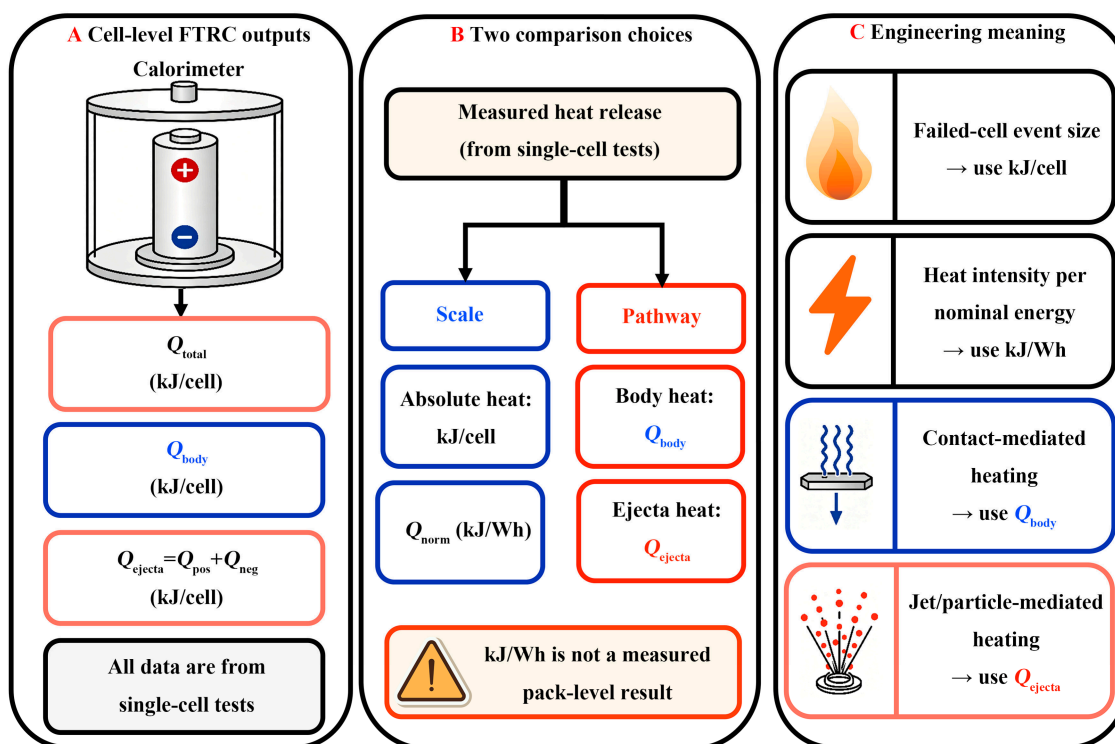


Figure 1. Simplified framework for metric-dependent TR hazard assessment.

2. Materials and Methods

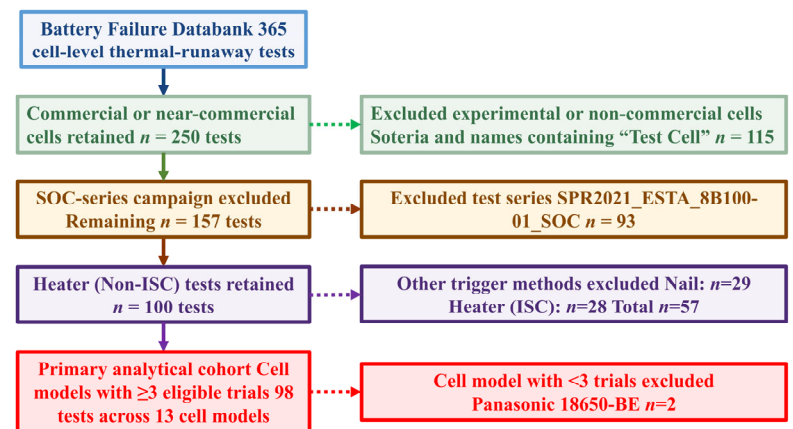
2.1. Data Source and Analytical Cohort

Data were obtained from Revision 2 of the open-access Battery Failure Databank, which contains 365 single-cell fractional TR calorimetry experiments and 43 recorded variables. The database includes cell identification, form factor, nominal capacity, nominal voltage, nominal energy, initial mass, trigger mechanism, S-FTRC generation, corrected total heat release, cell-body heat, positive-end ejecta heat, negative-end ejecta heat, and post-test mass measurements. All observations represented cell-level experiments; no module- or pack-level TR tests were included. Because cathode chemistry was not consistently reported for all included cell models in the source databank, chemistry-specific subgroup analyses, including an LFP-versus-layered-oxide comparison, were not performed.

The analytical cohort was constructed before outcome comparisons. Soteria experimental cells and records explicitly identified as test cells were excluded because they did not represent conventional commercial or near-commercial cell designs. A 93-test series designed to examine variable state-of-charge or overcharge conditions was also excluded because its pre-test voltage range was not comparable with the standard fully charged experiments. The primary analysis was restricted to non-ISC heater-triggered tests to minimise heterogeneity arising from the initiation method. Cell models represented by fewer than three repeated tests were excluded from model-level ranking analyses.

The final primary cohort comprised 98 experiments from 13 commercial or near-commercial cell models, including ten 18650 models, two 21700 models, and one D-format model. The number of repeated tests per model ranged from 3 to 15. Corrected total heat, positive-end ejecta heat, negative-end ejecta heat, nominal energy, and initial cell mass were complete for all primary-cohort records. One record contained an inconsistency between the cell name and the recorded form factor, and one record showed an 8.3% difference between corrected total heat and the sum of its reported heat components. Both records were retained in the primary analysis and removed in a prespecified data-quality sensitivity analysis. The cohort-selection process is summarised in Figure 2, and the analytical sample

is described in Table 1. As summarised in Table 1, all observations were derived from single-cell tests. The primary cohort excluded Soteria cells, cells explicitly designated as test cells, experiments conducted under variable SOC conditions, non-primary triggering conditions, and cell models represented by fewer than three replicate tests.



All included observations were single-cell tests; no module-or pack-level experiments

Figure 2. Analytical cohort selection flow diagram.

Table 1. Analytical cohort and data structure.

Characteristic	Value
Original database records	365
Primary analytical cohort	98 trials
Cell models	13
Cell formats	10 × 18,650; 2 × 21,700; 1 × D-format
Repeated tests per model	3–15
Primary trigger	Heater (non-ISC)
S-FTRC generations represented	4
Primary thermal outcomes	Total, body, positive ejecta, negative ejecta heat

Cell-model-specific sample sizes, formats, nominal energies, corrected total heat release, and post-test mass loss are summarised in Table 2. For each test, mass loss was calculated as $100 \times (m_{\text{initial}} - m_{\text{post}})/m_{\text{initial}}$, where m_{post} denotes the reported post-test cell-body mass.

Table 2. Characteristics and repeated-test thermal outcomes of the 13 cell models in the primary cohort.

Code	Cell Model	Format	n	Nominal Energy (Wh)	Corrected Total Heat, Median [IQR] (kJ)	Mass Loss, Median [IQR] (%)
C1	LG 18650-MJ1	18,650	15	12.70	88.68 [83.60–92.03]	75.96 [69.04–78.94]
C2	LG 21700-M50 (BV)	21,700	8	18.20	85.36 [78.86–92.72]	61.19 [50.31–67.08]
C3	KULR 21700-K500	21,700	8	18.00	82.60 [79.24–84.88]	49.91 [40.67–54.36]
C4	LG 18650-M36	18,650	3	12.49	75.60 [73.74–75.73]	73.05 [69.47–76.32]
C5	Sony 18650-VC7	18,650	12	12.70	75.96 [71.78–77.13]	62.09 [57.56–71.90]
C6	Saft D-Cell-VES16	D-Cell	9	16.20	68.81 [66.45–73.65]	34.27 [33.73–34.53]
C7	Samsung 18650-30Q	18,650	5	10.80	67.26 [65.96–72.23]	81.46 [75.21–82.76]
C8	MOLiCEL 18650-M35A	18,650	6	12.60	66.48 [65.11–67.82]	82.64 [82.18–82.89]
C9	LG 18650-HG2	18,650	4	10.80	61.51 [59.63–66.05]	76.87 [69.03–83.20]
C10	Sony 18650-VTC6	18,650	4	11.27	61.70 [59.69–64.31]	69.61 [64.08–77.73]
C11	KULR 18650-K330	18,650	10	11.88	61.75 [59.12–63.32]	54.91 [50.74–56.68]
C12	Samsung 18650-26J	18,650	10	9.44	41.34 [37.12–43.79]	53.61 [36.89–74.89]
C13	MOLiCEL 18650-J	18,650	4	8.77	35.97 [35.13–36.91]	26.15 [25.73–26.67]

2.2. Thermal Metrics and Ranking Framework

Corrected total heat release was denoted by Q_{total} . The heat transferred through the cell body was denoted by Q_{body} , and total ejecta heat was defined using Equation (1).

$$Q_{\text{ejecta}} = Q_{\text{pos}} + Q_{\text{neg}} \quad (1)$$

where Q_{pos} and Q_{neg} represent positive- and negative-end ejecta heat, respectively.

Absolute thermal outcomes were expressed in kJ per failed cell. Energy-normalised outcomes were calculated using Equation (2), and nominal cell energy was checked using Equation (3).

$$q_k = \frac{Q_k}{E_{\text{nominal}}} \quad (2)$$

$$E_{\text{nominal}} = C_{\text{Ah}} V_{\text{nominal}} \quad (3)$$

where Q_k denotes the absolute heat release (kJ) for heat metric k , E_{nominal} denotes the nominal electrical energy of the cell (Wh), and q_k denotes the corresponding energy-normalised heat release ($\text{kJ} \cdot \text{Wh}^{-1}$). Here, k identifies the heat metric considered, including total, body, and ejecta heat. Where C_{Ah} denotes the nominal cell capacity (Ah), V_{nominal} denotes the nominal cell voltage (V), and E_{nominal} denotes the nominal electrical energy (Wh) used for normalisation of the heat-release metrics. The nominal-energy value provided in the database was used for the primary calculations, while the product of nominal capacity and nominal voltage was used as a consistency check. The resulting $\text{kJ} \cdot \text{Wh}^{-1}$ values represent cell-level heat release per unit nominal cell energy and were not interpreted as measured module- or pack-level heat release.

To examine whether linear normalisation by nominal energy was supported by the data, the allometric model in Equation (4) was fitted separately for total, body, and ejecta heat:

$$Q = a E_{\text{nominal}}^b \quad (4)$$

where a is the fitted scaling coefficient, and b is the scaling exponent describing how heat release changes with nominal cell energy.

A value of $b = 1$ indicates proportional scaling, whereas $b < 1$ and $b > 1$ indicate sublinear and superlinear scaling, respectively. Thus, $b = 1$ provides the reference condition under which normalisation by nominal energy (kJ/Wh) is consistent with proportional scaling of heat release with cell energy. After logarithmic transformation, Equation (4) was fitted in the linear form shown in Equation (5).

$$\ln(Q) = \alpha + b \ln(E_{\text{nominal}}) + \varepsilon \quad (5)$$

where $\alpha = \ln(a)$ is the intercept on the log scale, b is the scaling exponent, and ε is the residual error term representing variation in log-transformed heat release not explained by nominal cell energy. Confidence intervals were obtained using 5000 hierarchical bootstrap replications. The 13 cell models were treated as the independent top-level clusters; in each replicate, 13 cell-model clusters were sampled with replacement, followed by resampling of repeated trials within each selected model.

Cell models were ranked from highest to lowest hazard based on six definitions: absolute total heat, total heat per Wh, absolute body heat, body heat per Wh, absolute ejecta heat, and ejecta heat per Wh. The two primary ranking contrasts were absolute total heat versus total heat per Wh, representing the scale effect, and absolute body heat versus absolute ejecta heat, representing the pathway effect. Other metric contrasts were treated as secondary analyses. Kendall's τ -a coefficient was used to quantify the consistency of rankings between each pair of hazard definitions. Because no tied ranks were present in the

rankings of the complete set of cell models, tau-a and tau-b were numerically equivalent in the present analysis. Pairwise ranking discordance was calculated using Equation (6). In the absence of tied ranks, this metric is exactly equal to the proportion of cell-model pairs whose relative order is reversed under the two hazard definitions.

$$D = \frac{1 - \tau}{2} \quad (6)$$

where D is the pairwise rank-discordance proportion and τ is Kendall's rank-correlation coefficient. For the complete rankings considered here, which contained no tied ranks, D corresponds to the proportion of cell-model pairs ordered in opposite directions under the two hazard definitions. Thus, $D = 0$ indicates identical ordering, $D = 0.5$ indicates no overall rank concordance, and $D = 1$ indicates complete reversal of the ranking.

To account for potential systematic differences among S-FTRC generations, the trial-level fixed-effects model in Equation (7) was fitted on the logarithmic scale:

$$\ln(Y_{ijk}) = \alpha + \beta_i^{\text{cell}} + \gamma_j^{\text{S-FTRC}} + \varepsilon_{ijk} \quad (7)$$

where Y_{ijk} is the thermal outcome measured in trial k for cell model i tested using S-FTRC generation j ; α is the model intercept; β_i^{cell} is the fixed effect associated with cell model i ; $\gamma_j^{\text{S-FTRC}}$ is the fixed effect associated with S-FTRC generation j ; and ε_{ijk} is the residual error term on the logarithmic scale. Cell-specific marginal estimates were then standardised to the overall observed distribution of S-FTRC generations so that comparisons among cell models were not driven by differences in their S-FTRC-generation composition. Estimates were retransformed to the original heat scale using Duan's smearing estimator to account for bias introduced by exponentiating predictions from the log-transformed model. The model assumed an additive S-FTRC-generation effect without a cell-model-by-generation interaction because the available data did not provide a fully crossed design for estimating such interactions reliably. The resulting design matrix was full rank (rank = 16 of 16), with 82 residual degrees of freedom. Thus, all cell-model adjusted estimates were estimable under the additive cell-model and S-FTRC-generation fixed-effects model. For the generation-adjusted analyses, bootstrap uncertainty was conditional on the 13 observed cell models. A total of 5000 bootstrap replications were performed by retaining the observed cell-model identities and resampling repeated trials with replacement within each observed cell-model-by-S-FTRC-generation stratum, thereby preserving the generation composition and repeated-test structure of the dataset.

For each pair of cell models and each pair of hazard metrics, the bootstrap probability of a rank reversal was defined as Equation (8).

$$P_{\text{rev},ij}^{(Y,Z)} = P_b \left[\text{sign}(\hat{Y}_i^{(b)} - \hat{Y}_j^{(b)}) \neq \text{sign}(\hat{Z}_i^{(b)} - \hat{Z}_j^{(b)}) \right] \quad (8)$$

where i and j denote two cell models, Y and Z denote the two hazard metrics being compared, b denotes a bootstrap replicate, and $\hat{Y}_i^{(b)}$ and $\hat{Z}_i^{(b)}$ are the corresponding bootstrap estimates for cell model i . Thus, $P_{\text{rev},ij}^{(Y,Z)}$ represents the proportion of bootstrap replicates in which the relative ordering of cell models i and j is opposite under metrics Y and Z . A value approaching 1 indicates high resampling stability of an opposite ordering under the two metrics, whereas a value approaching 0 indicates stable agreement in their relative ordering. Reversal probabilities of at least 95% were classified as robust reversals, probabilities from 80% to below 95% as likely reversals, probabilities above 20% but below 80% as uncertain ordering, and probabilities of 20% or less as stable agreement. These categories were used as descriptive measures of resampling stability rather than as formal hypothesis-testing thresholds. Accordingly, no multiplicity correction was applied

across the 78 cell-model pairs, because the pairwise reversal probabilities were not treated as p values or as a family of null-hypothesis tests; counts of robust or likely reversals should therefore be interpreted as descriptive measures of ranking stability rather than family-wise-error-controlled discoveries.

2.3. Heat-Path Composition and Robustness Analyses

Body and ejecta heat fractions were calculated using Equation (9) and Equation (10), respectively.

$$f_{\text{body}} = \frac{Q_{\text{body}}}{Q_{\text{total}}} \quad (9)$$

$$f_{\text{ejecta}} = \frac{Q_{\text{ejecta}}}{Q_{\text{total}}} \quad (10)$$

where Q_{body} is the heat released through the cell body, and Q_{total} is the corrected total heat release. Thus, the heat-release pathway was represented by the three components Q_{body} , Q_{pos} , and Q_{neg} . Because these three components jointly describe the partitioning of total heat release and are therefore compositional rather than independent quantities, their joint structure was analysed using isometric log-ratio (ILR) coordinates. The ILR basis was defined using two sequential binary partitions. The first contrasted body heat with the combined ejecta pathway, $\{Q_{\text{body}}\} \mid \{Q_{\text{pos}}, Q_{\text{neg}}\}$, whereas the second contrasted the two ejecta directions, $\{Q_{\text{pos}}\} \mid \{Q_{\text{neg}}\}$.

The body-versus-ejecta pathway coordinate (z_{path}) was defined using Equation (11).

$$z_{\text{path}} = \sqrt{\frac{2}{3}} \ln \left[\frac{Q_{\text{body}}}{Q_{\text{pos}} Q_{\text{neg}}} \right] \quad (11)$$

Higher values of Equation (11) indicate a more body-dominated thermal pathway, whereas lower values indicate a more ejecta-dominated pathway. Directional asymmetry ($z_{\text{direction}}$) was quantified using Equation (12).

$$z_{\text{direction}} = \frac{1}{2} \ln \left(\frac{Q_{\text{pos}}}{Q_{\text{neg}}} \right) \quad (12)$$

where positive values indicate positive-end dominance, and negative values indicate negative-end dominance. All three heat components were strictly positive across all 98 primary-cohort tests. The minimum observed values were 2.95 kJ for body heat, 1.59 kJ for positive-end ejecta heat, and 0.35 kJ for negative-end ejecta heat; therefore, neither zero replacement nor pseudocount addition was required before ILR transformation. Generation-adjusted cell estimates and confidence intervals for both coordinates were obtained using the same stratified bootstrap procedure.

The contributions of scale definition and heat pathway were evaluated through the sequential transformation in Equation (13).

$$Q_{\text{total}} \rightarrow Q_{\text{ejecta}} \rightarrow Q_{\text{ejecta}}/E_{\text{nominal}} \quad (13)$$

Each cell-model pair was classified according to whether its relative ordering changed when the ranking basis was altered. Four mutually exclusive categories were defined: no reversal, a pathway-only reversal, a scale-only reversal, or reversals at both steps. This classification was used to distinguish ranking changes arising from normalisation by nominal energy from those arising from separation of body and ejecta heat and to identify cell pairs for which both effects occurred simultaneously.

Sensitivity analyses were performed to evaluate whether the main ranking-discordance results depended on specific features of the analytical cohort. These analyses included restriction to 18,650 cells, exclusion of the two data-quality-flagged records, restriction to cell models represented by at least five repeated tests, and stratification by S-FTRC generation. Additional analyses restricted comparisons to 18,650 cells within individual generation strata to reduce potential confounding by cell format. Leave-one-cell, leave-one-format, and leave-one-generation analyses were also conducted to assess whether the observed discordance was disproportionately influenced by any single cell model, cylindrical format, or S-FTRC generation. The resulting discordance estimates were compared with the corresponding full-cohort values as measures of robustness rather than as independent hypothesis tests.

Trigger-method validation was conducted separately for cell models with at least three repeated experiments using both non-ISC heater triggering and at least one alternative trigger method. For each eligible cell model, mean heat values using the alternative trigger were expressed as ratios relative to the corresponding non-ISC heater mean, allowing the direction and magnitude of trigger-related changes to be compared across models. Confidence intervals for these ratios were obtained by within-group bootstrap resampling so that uncertainty arising from repeated-test variability was retained. Cross-trigger rank correlations were also calculated to examine whether the relative ordering of cell models was preserved across trigger methods. Because only three to four cell models were available for direct cross-trigger comparison, these rank-correlation results were treated as exploratory and were not used as primary evidence for trigger-independent ranking stability.

3. Results and Discussion

3.1. Cohort Characteristics, Thermal Scaling, and Metric-Scale Effects

The primary cohort comprised 98 thermal-runaway experiments from 13 commercial or near-commercial cylindrical cell models, with 3–15 repeated tests available for each model (Tables 1 and 2). Cell-model-specific nominal energy, corrected total heat, and post-test mass loss are summarised in Table 2. This repeated-test structure allowed the analysis to distinguish systematic differences among cell models from the experimental variability observed within the same model. As shown in Figure 3, substantial within-model dispersion was present in total, body, and ejecta heat, and the magnitude of this dispersion differed among cell models and heat-release pathways. In several cases, repeated tests from the same model covered a comparatively broad range of heat outputs, indicating that the thermal-runaway response of a given cell model cannot be fully characterised by a single test result. Importantly, variability was not restricted to corrected total heat; it was also evident in the partitioning of heat between the cell body and ejecta, which is directly relevant to the pathway-dependent comparisons developed later in this study. These observations motivated the use of repeated-test distributions rather than single point estimates in the subsequent ranking, bootstrap, and sensitivity analyses. Accordingly, cell-model rankings were evaluated together with their resampling uncertainty so that apparent differences among models could be distinguished from ordering changes arising from repeated-test variability. Post-test mass loss also varied substantially among cell models (Table 2). At the trial level, mass loss was positively associated with ejecta heat (Spearman's $\rho = 0.601$, $p < 0.001$) and negatively associated with body heat ($\rho = -0.766$, $p < 0.001$), whereas its association with total heat was weaker ($\rho = 0.221$, $p = 0.029$). These associations were treated as exploratory because mass loss is physically coupled to the thermal-runaway and material-ejection processes rather than representing an independent experimental factor. The between-model distributions in Table 2 further show that differences in absolute heat release cannot be attributed solely to nominal electrical energy. Cell models with

similar nominal energies exhibited appreciable differences in both corrected total heat and post-test mass loss, indicating that cell-specific failure and ejection behaviour contributes to the observed thermal variability. The exploratory correlations further suggest that mass loss was more closely associated with the partitioning of heat toward the ejecta pathway than with total heat alone. Accordingly, reversals between absolute and energy-normalised rankings should not be interpreted as a simple consequence of cell size; they reflect the combined influence of nominal energy, total heat release, and cell-specific heat-release pathways.

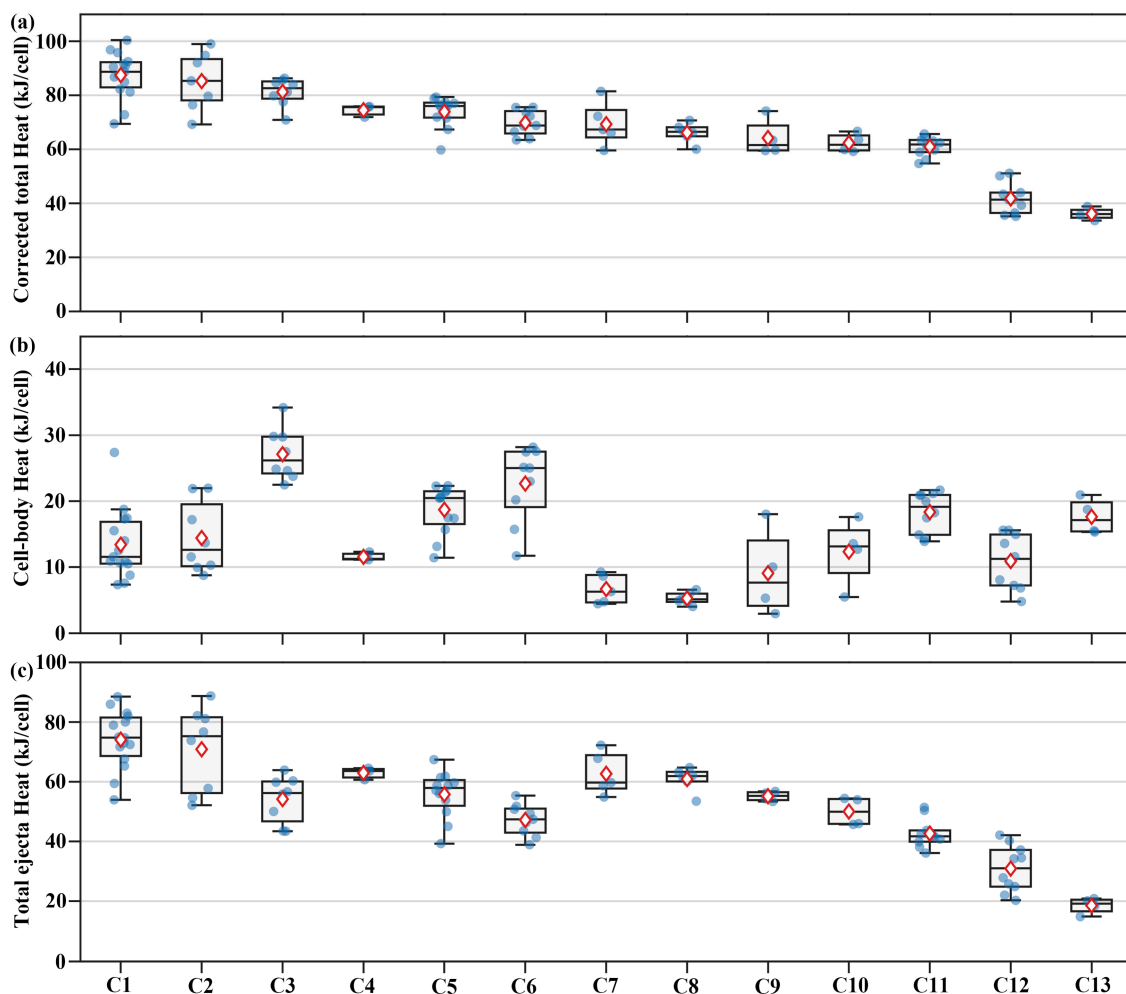


Figure 3. Trial-level distributions of corrected total heat, cell-body heat, and total ejecta heat across the 13 cell models in the primary cohort. (a) Corrected total heat. (b) Cell-body heat. (c) Total ejecta heat.

Figure 3 presents the trial-level heat-release distributions rather than only cell-model averages, thereby retaining the substantial within-model variability observed across repeated tests. This repeated-test variability was propagated through the subsequent hierarchical bootstrap and sensitivity analyses rather than being discarded by reducing each cell model to a single deterministic heat-release value. Total heat release showed an overall increasing trend with nominal cell energy. This relationship was quantified using the allometric model ($Q = aE^b$), with 95% confidence intervals estimated from 5000 hierarchical bootstrap replications. In the all-format analysis, the estimated allometric exponent was 0.899 for total heat, with a 95% hierarchical-bootstrap interval of 0.320–1.576. The corresponding exponents were 0.879 for body heat and 1.003 for ejecta heat, as summarised in Table 3 and Figure 4. Because the 95% intervals for all three thermal outcomes encompassed $b = 1$, the data were consistent with an approximately proportional relationship between

nominal cell energy and heat release. However, the broad uncertainty intervals, particularly for body and ejecta heat, indicate that the scaling exponents remain imprecisely estimated, largely reflecting the limited number of independent cell models available for analysis.

Table 3. Allometric scaling of heat release with nominal cell energy.

Outcome	Exponent b	95% Bootstrap Interval	Interpretation
Total heat	0.899	0.320–1.576	Interval includes $b = 1$
Ejecta heat	1.003	−0.118–2.235	Interval includes $b = 1$
Body heat	0.879	−0.321–2.192	Interval includes $b = 1$

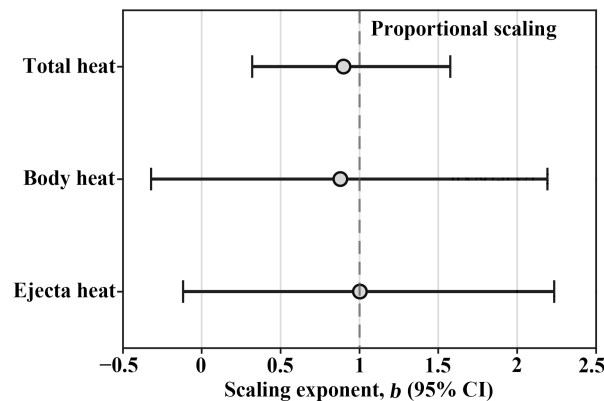


Figure 4. Scaling of TR heat release with nominal cell energy.

This finding is important for interpreting normalised heat metrics. It does not imply that $\text{kJ}\cdot\text{Wh}^{-1}$ is invalid or misleading. Rather, $\text{kJ}\cdot\text{cell}^{-1}$ and $\text{kJ}\cdot\text{Wh}^{-1}$ describe different engineering quantities. Absolute heat characterises the integrated magnitude of a single failed-cell event, whereas energy-normalised heat characterises TR heat relative to the nominal electrical energy represented by the cell. Consequently, approximate proportional scaling of heat with nominal energy does not require the two metrics to produce identical hazard rankings.

After adjustment for S-FTRC generation, the ranking based on absolute total heat showed only weak agreement with the ranking based on energy-normalised total heat. Kendall's τ was 0.077, indicating that the relative ordering of cell models changed substantially when the ranking basis was shifted from absolute heat release to heat release per unit nominal electrical energy. This corresponded to a point-estimate pairwise discordance of 46.2%, meaning that nearly half of the 78 possible cell-model pairs changed their relative ordering between the two metrics. The effect was not limited to a small number of unstable comparisons. Across 5000 bootstrap replications, the median discordance remained high at 43.6%, with a 95% bootstrap interval of 35.9–51.3%, showing that substantial ranking disagreement persisted after accounting for repeated-test variability. Moreover, 20 of the 78 cell-model pairs had reversal probabilities of at least 95%, indicating that their change in relative ordering was reproduced consistently across resampled datasets rather than arising from a single point estimate. These results demonstrate that normalisation by nominal electrical energy can materially alter the apparent comparative severity of thermal runaway, even when the same underlying heat-release measurements are used. In practical terms, a cell model ranked relatively high by absolute total heat may not retain the same position when heat release is expressed on a per-Wh basis, because the two metrics emphasise different aspects of thermal-runaway severity. Because the primary inferential comparison was based on S-FTRC-generation-adjusted estimates, Figure 5 is presented as a descriptive

visualisation of the corresponding unadjusted rank shifts and bootstrap uncertainty rather than as the basis for the adjusted statistical inference.

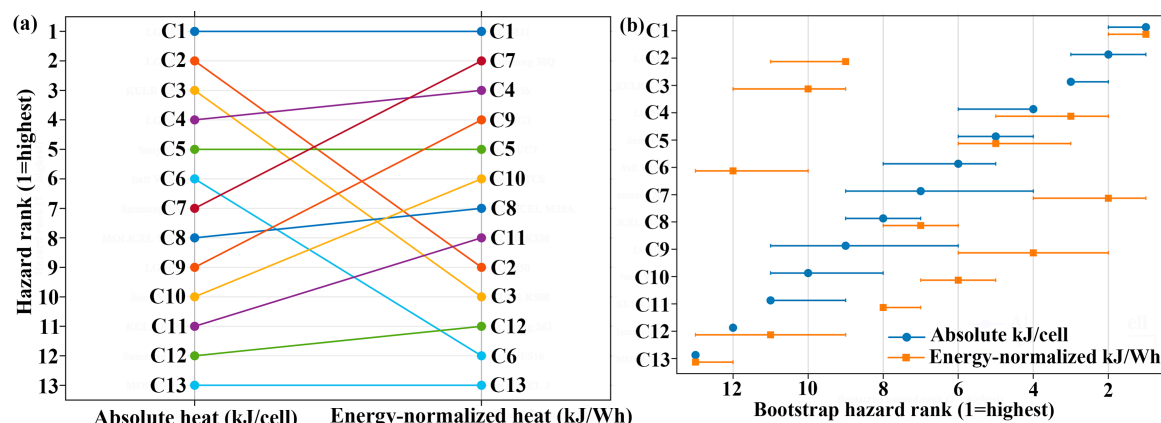


Figure 5. Scale-dependent changes in total-heat ranking. (a) Unadjusted absolute-total versus energy-normalised ($\text{kJ}\cdot\text{Wh}^{-1}$) ranking slope plot. (b) Bootstrap uncertainty intervals for the corresponding unadjusted rankings.

The distinction is particularly important for cross-format comparisons. Previous Battery Failure Databank analyses and related statistical studies have used capacity- or energy-normalised heat metrics to facilitate comparison among cells differing in size, geometry, and nominal energy [21,24], while cross-study meta-analysis has similarly employed normalised heat-release quantities to compare heterogeneous cell populations [30]. These approaches are appropriate when the objective is to express thermal output relative to electrical size. The present results, however, show that normalisation is not rank-neutral: a cell that releases more absolute heat per failed unit may rank lower after division by nominal energy. Importantly, this occurred even though the allometric exponents for total, body, and ejecta heat were statistically compatible with proportional scaling. Thus, approximate proportionality at the cohort level does not imply equivalence of absolute and normalised hazard rankings at the individual cell-model level. Robustness of the scale-dependent result is evaluated together with the pathway sensitivity analyses in Section 3.3.

3.2. Pathway-Dependent Rankings and Heat-Path Composition

The strongest pathway-dependent divergence was observed between the body-heat and ejecta-heat rankings. After adjustment for S-FTRC generation, the two rankings showed a negative Kendall's τ of -0.256 , indicating that cell models ranked highly according to heat released through the cell body did not necessarily rank highly according to ejecta heat, and in many pairwise comparisons the relative ordering was reversed. This corresponded to a point-estimate discordance of 62.8%, meaning that nearly two-thirds of the 78 possible cell-model pairs changed their relative ordering when the ranking basis was switched from body heat to ejecta heat. The disagreement remained pronounced after accounting for repeated-test variability. Across 5000 bootstrap replications, the median discordance was 61.5%, with a 95% bootstrap interval of 53.8–67.9%, demonstrating that the pathway-dependent ranking conflict was not driven by a small number of unstable observations. In addition, 28 of the 78 pairwise comparisons were classified as robust reversals, with bootstrap reversal probabilities of at least 95%, while a further 10 comparisons were classified as likely reversals. Thus, a substantial fraction of the observed rank changes were reproduced consistently across resampled datasets. These results show that body heat and ejecta heat capture distinct aspects of thermal-runaway severity and should not be treated as interchangeable ranking metrics. The S-FTRC-generation-adjusted rankings are shown

in Figure 6a, while Figure 6b summarises the corresponding pairwise bootstrap reversal probabilities. Detailed cell-model ranking comparisons are provided in Table 4.

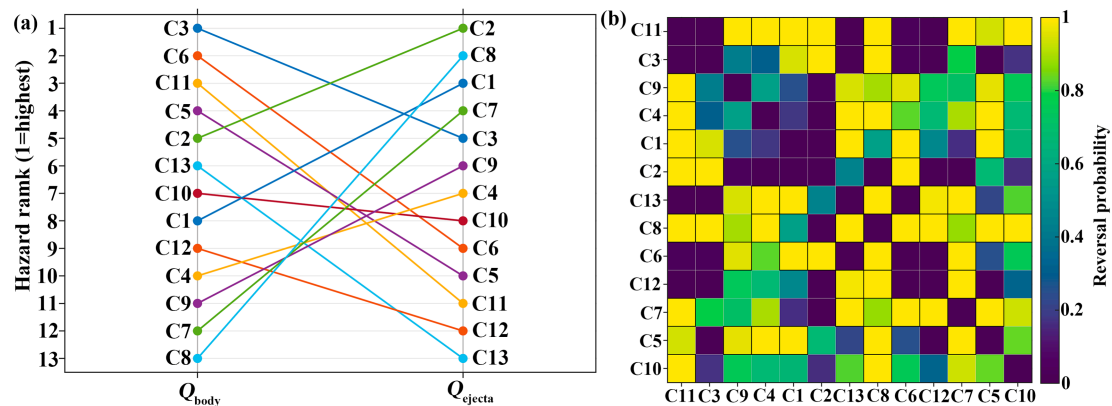


Figure 6. Pathway-dependent divergence in TR hazard rankings. (a) S-FTRC-generation-adjusted body-heat and ejecta-heat hazard rankings. (b) Pairwise bootstrap probability that the body-heat and ejecta-heat rankings of a cell-model pair reverse after S-FTRC-generation adjustment.

Table 4. S-FTRC-generation-adjusted ranking comparisons.

Comparison	Point τ	Point Discordance	Bootstrap Median Discordance	95% Interval	Robust Reversals
Absolute total vs. total kJ/Wh	0.077	46.2%	43.6%	35.9–51.3%	20/78
Total vs. absolute ejecta	0.590	20.5%	20.5%	14.1–26.9%	4/78
Body vs. absolute ejecta	−0.256	62.8%	61.5%	53.8–67.9%	28/78
Absolute total vs. ejecta kJ/Wh	0.051	47.4%	43.6%	35.9–51.3%	19/78

A smaller but still meaningful divergence was observed when total heat was replaced by absolute ejecta heat. Total-heat and ejecta-heat rankings had Kendall's $\tau = 0.590$, equivalent to a point-estimate discordance of 20.5%. The bootstrap median was also 20.5%, with a 95% interval of 14.1–26.9%. Four cell-model pairs showed robust reversals and two additional pairs showed likely reversals.

Generation-adjusted estimates illustrate the practical size of these changes. KULR K500 had the highest adjusted body heat but ranked fifth for ejecta heat, while Saft VES16 ranked second for body heat and ninth for ejecta heat. In contrast, LG M50 ranked fifth for body heat but first for ejecta heat, and MOLiCEL M35A ranked last for body heat but second for ejecta heat. The rank shifts were not confined to adjacent positions; for many cell-model pairs, the reversal in relative ordering persisted across bootstrap resamples.

This divergence is consistent with, but extends, previous fractional-calorimetry studies. Walker et al. [19] demonstrated that heat transferred through the cell casing and heat carried by ejecta constitute experimentally separable components of thermal-runaway energy release. Sharp et al. [20] subsequently showed that heat output and mass ejection vary with cell geometry and abuse type, and that greater mass ejection can accompany greater heat release even for the same cell model under comparable conditions. This is qualitatively consistent with the present exploratory finding that post-test mass loss was more strongly associated with ejecta heat than with total heat. More recently, Willstrand et al. [25] reported that the fraction of heat carried by ejecta was influenced by both state of charge and cell mass loss. The present analysis extends these observations from heat partitioning itself to hazard ranking, showing that body-heat and ejecta-heat metrics can identify different cell models as the higher-hazard option.

The engineering implications are pathway-specific. Body heat is more directly relevant to conductive thermal loading of contacting structures, cell holders, cooling elements, and radial thermal barriers, whereas ejecta heat is more directly relevant to axial spacing, vent channels, shields, particle interception, and components located along the ejecta trajectory. Positive- and negative-end ejecta are likewise not equivalent when a cylindrical cell is mounted asymmetrically. The actual module-level consequence will additionally depend on geometry, temporal heat release, confinement, and neighbouring-cell susceptibility. A universal weighted severity score would therefore require architecture-specific weights derived from a defined module geometry and heat-transfer model rather than from the cell data alone.

The heat-path composition provided a physical basis for understanding the observed divergence between body-heat and ejecta-heat rankings. Trial-level heat-path fingerprints revealed substantial differences among cell models in the proportions of heat released through the cell body and through positive- and negative-end ejecta (Figure 7a). For some cell models, a larger fraction of the measured heat remained associated with the cell body, whereas others released a greater proportion through ejecta. The relative contribution of the positive and negative ends also varied among models, indicating that thermal-runaway heat release was not only different in magnitude, but also different in its spatial pathway. These differences are important because body-mediated heat and ejecta-mediated heat represent distinct routes by which a failed cell can transfer thermal energy to its surroundings. The fractions shown in Figure 7a are compositional rather than statistically independent quantities, because the three components collectively describe the distribution of released heat, and an increase in the relative contribution of one component necessarily changes the relative shares of the others. For this reason, direct statistical analysis of the individual fractions could be affected by the constant-sum constraint inherent to compositional data. Figure 7a was therefore used primarily as a descriptive representation of the component-to-total heat fractions, whereas the formal compositional analysis was performed using isometric log-ratio (ILR) coordinates calculated directly from the three positive component heat values. Accordingly, the ILR analysis did not rely on normalisation by corrected total heat and instead characterised the relative balance among body heat, positive-end ejecta heat, and negative-end ejecta heat.

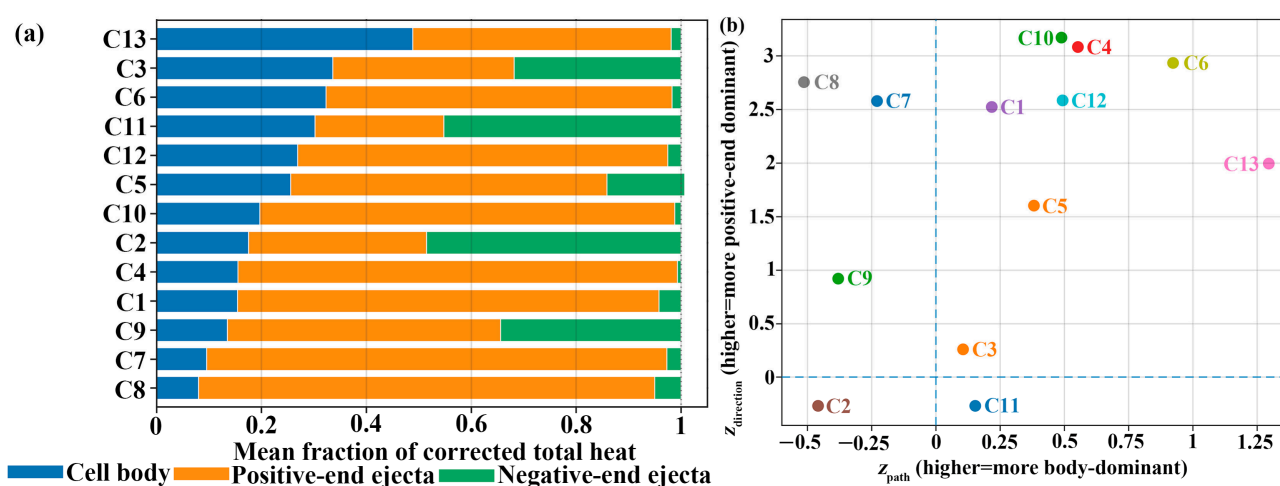


Figure 7. Heat-path composition across cell models. (a) Mean fractions of corrected total heat released through the cell body and through positive- and negative-end ejecta. (b) S-FTRC-generation-adjusted isometric log-ratio map of body-versus-ejecta composition and ejecta direction.

After adjustment for S-FTRC generation, substantial cell-specific variation remained in the body-versus-ejecta composition. The model containing S-FTRC generation alone explained 15.0% of the variation in z_{path} , whereas adding cell identity increased the model R^2 to 60.2%. The corresponding adjusted R^2 increased from 0.123 to 0.529, and the improvement in model fit was significant [$F(12,82) = 7.764, p < 0.001$]. This large increase indicates that differences among cell models accounted for a considerable proportion of the observed variation in heat-release pathway beyond that associated with S-FTRC generation. In the z_{path} coordinate, more positive values indicate a relatively more body-dominated heat-release composition, whereas more negative values indicate a relatively greater contribution from ejecta. MOLIcel 18650-J showed the strongest body-dominated composition, with an adjusted z_{path} of 1.295 and a 95% interval of 1.043–1.552, followed by Saft VES16 at $z_{\text{path}} = 0.923$ (0.744–1.124). In contrast, MOLIcel M35A showed a distinctly ejecta-dominated composition, with $z_{\text{path}} = -0.514$ (−0.819 to −0.278), followed by LG M50 at −0.459 (−0.677 to −0.205) and Samsung 30Q at −0.229 (−0.408 to −0.046). These contrasts show that cells with different thermal-runaway behaviours can partition released heat very differently between the casing and ejecta pathways even after accounting for S-FTRC generation. The generation-adjusted ILR coordinates are summarised in Figure 7b and provide a complementary compositional interpretation of the pathway-dependent ranking differences observed in Figure 6.

For the directional coordinate, the S-FTRC-generation-only model explained 24.9% of the variation, whereas inclusion of cell identity increased the model R^2 to 64.0%. The corresponding adjusted R^2 increased from 0.225 to 0.574, and the improvement in model fit was significant [$F(12,82) = 7.420, p < 0.001$]. This increase indicates that cell-specific differences accounted for a substantial proportion of the observed variation in ejecta direction beyond that associated with S-FTRC generation. Most cell models showed positive-end-dominant ejecta heat, while KULR K330 and LG M50 were closer to a balanced distribution or showed a slight negative-end bias, further demonstrating that the spatial direction of heat release differed among cell models.

When absolute total heat was compared directly with energy-normalised ejecta heat, the point-estimate discordance was 47.4%. Across 5000 bootstrap replications, the median discordance was 43.6%, with a 95% interval of 35.9–51.3%, and 19 cell-model pairs showed robust reversals. Thus, changing both the scale of the metric and the heat-release pathway produced substantial and reproducible changes in relative cell ordering. The sequential decomposition therefore indicates that scale normalisation and pathway selection represent distinguishable sources of ranking variation rather than redundant transformations of the same underlying thermal outcome.

3.3. Robustness, Trigger Sensitivity, Limitations, and Engineering Implications

Adjustment for S-FTRC generation did not eliminate the scale-dependent conflict between absolute and energy-normalised hazard rankings, indicating that the observed ranking reversals could not be explained simply by differences in the distribution of experiments across S-FTRC generations. This result, however, should not be interpreted as evidence that S-FTRC generation itself caused, or did not cause, the reversals. The experimental design was observational rather than fully crossed: not every cell model was tested under every S-FTRC generation, and the composition of cell models therefore differed among generation subsets. Consequently, cell identity, format, test availability, and apparatus generation were not completely separable. The adjusted analysis should instead be interpreted more narrowly as showing that accounting for the observed generation composition was insufficient to remove the metric-dependent ranking disagreement. In other words, the conflict between absolute and energy-normalised rankings persisted after

statistical adjustment for generation, supporting the interpretation that metric definition itself remained an important source of ranking variation within the available dataset. Figure 8a summarises the scale-dependent ranking sensitivity across alternative subsets defined by cell format, data-quality restrictions, and S-FTRC generation. The magnitude of absolute-versus-energy-normalised discordance was strongly dependent on subset composition: it was 32.1% in the full primary cohort but decreased to 11.1% when the analysis was restricted to the ten 18,650 cell models. In contrast, exclusion of the two data-quality-flagged trials produced little change, with discordance increasing slightly to 33.3%. These results indicate that the scale-dependent ranking effect was reproducible in the full cohort, but that its magnitude was sensitive to the cell formats and cell models included in the comparison. These subgroup estimates were treated descriptively because the available cell-model sets were not identical across S-FTRC-generation subsets; direct numerical comparison among them would therefore confound generation with differences in the cells represented in each subset and should not be interpreted as a causal estimate of apparatus-generation effects. For transparency, the primary cohort contained 98 trials from all 13 cell models (C1–C13), whereas the 18,650-only subset contained 72 trials from 10 models (C1, C4, C5, and C7–C13). Exclusion of the two data-quality-flagged records retained 96 trials from all 13 models, and the corresponding 18,650-only quality-restricted subset contained 71 trials from the same 10 models. The generation-specific subsets comprised 33 trials from six models for V10.0 (C3, C6, C8, C10, C11, and C12), 35 trials from five models for V7.0/8.0 (C1, C4, C5, C7, and C13), and 19 trials from five models for V9.0 (C2, C3, C7, C9, and C11); the pooled V9.x subset contained 30 trials from seven models (C2, C3, C5, C6, C7, C9, and C11).

A complementary leave-one-out analysis was used to assess whether the body-versus-ejecta ranking conflict was driven by any single cell model, cell format, S-FTRC generation, or analysis scenario. The S-FTRC-generation-adjusted discordance remained substantial across all identifiable omission scenarios, ranging from 56.1% to 77.8% (Figure 8b). When the analysis was restricted to 18,650 cells, the discordance increased to 77.8%, indicating that the pathway-dependent ranking disagreement persisted even after excluding the 21,700 models and the D-format cell. These results suggest that the observed body-versus-ejecta divergence was not dominated by a single influential subset and remained evident within a common cylindrical format.

Exclusion of the two data-quality-flagged records did not materially alter the unadjusted ranking results. The pairwise discordance for total heat increased only slightly, from 32.1% to 33.3%, whereas the discordance for ejecta heat remained unchanged at 20.5%. These small changes indicate that the observed ranking disagreements were not driven by the two flagged observations. Removal of the identified data inconsistencies therefore did not materially change either scale-dependent or pathway-dependent ranking discordance, although the separate 18,650-only analysis showed that the magnitude of the scale-dependent effect was more sensitive to cohort composition.

Trigger-method sensitivity analysis indicated that the effect of triggering mode on absolute heat release varied across cell models. Relative to non-ISC heater triggering, nail triggering increased mean total heat by 28.4% for Saft VES16, 18.0% for KULR K500, and 15.3% for LG M50. After Holm correction for the seven primary total-heat trigger comparisons, all three nail-trigger effects remained significant (adjusted $p = 0.003$, 0.003 , and 0.004 , respectively). Heater-ISC triggering increased mean total heat by 8.7% for Saft VES16 and 6.6% for KULR K500; the former remained significant after multiplicity correction (adjusted $p = 0.004$), whereas the latter did not (adjusted $p = 0.082$). KULR K330 showed no significant change under either alternative trigger. Despite these differences in heat magnitude, the ordering of the three to four directly comparable cell models by total

and ejecta heat remained unchanged across trigger methods, yielding exploratory Kendall coefficients of 1.0. These comparisons are summarised in Figure 9 and Table 5. In Figure 9, points indicate mean heat ratios and horizontal error bars represent 95% bootstrap intervals; a ratio of 1 indicates no change relative to non-ISC heater triggering. Ratios greater than 1 indicate higher mean total heat under the alternative trigger relative to non-ISC heater triggering. Cross-trigger ranking analyses were considered exploratory because only three to four cell models overlapped across trigger methods; therefore, the observed ranking stability should not be interpreted as evidence of generally trigger-invariant rankings.

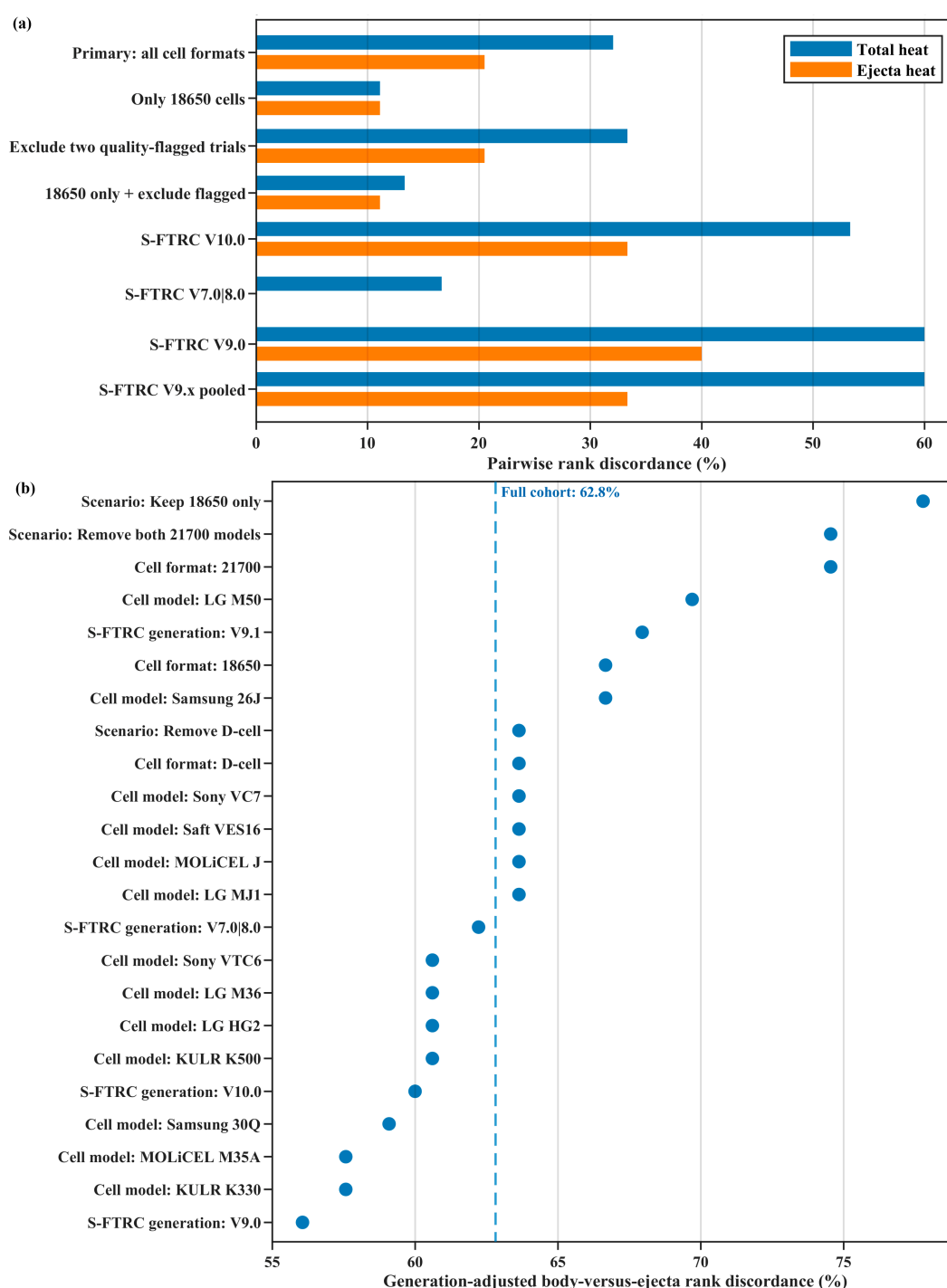


Figure 8. Robustness of ranking discordance. (a) Absolute-versus-normalised discordance across sensitivity subsets. (b) Leave-one sensitivity of adjusted body-versus-ejecta discordance; the dashed line shows the full-cohort estimate.

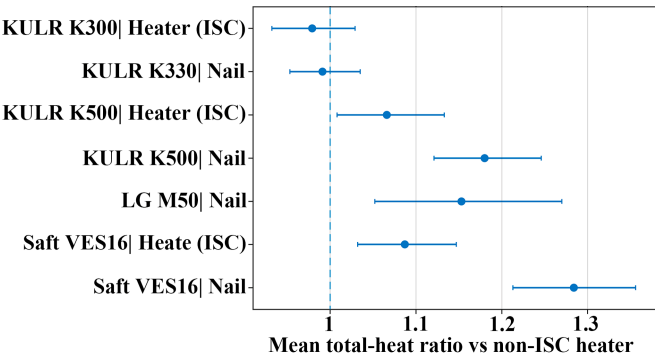


Figure 9. Trigger-method sensitivity of mean total heat relative to non-ISC heater triggering.

Table 5. Trigger-method effects on mean total heat relative to non-ISC heater triggering.

Cell Model	Alternative Trigger	<i>n</i> (Alternative)	<i>n</i> (Non-ISC Heater)	Mean Ratio	95% Bootstrap CI	Raw <i>p</i>	Holm-Adjusted <i>p</i>
KULR 18650-K330	Heater (ISC)	9	10	0.979	0.932–1.029	0.400	0.800
KULR 18650-K330	Nail	3	10	0.991	0.953–1.032	0.655	0.800
KULR 21700-K500	Heater (ISC)	10	8	1.066	1.008–1.132	0.027	0.082
KULR 21700-K500	Nail	12	8	1.180	1.122–1.247	<0.001	0.003
LG 21700-M50 (BV)	Nail	5	8	1.153	1.057–1.269	<0.001	0.004
Saft D-Cell-VES16	Heater (ISC)	9	9	1.087	1.032–1.147	<0.001	0.004
Saft D-Cell-VES16	Nail	9	9	1.284	1.213–1.357	<0.001	0.003

NOTE: Ratios were calculated relative to the corresponding non-ISC heater-triggered tests for the same cell model. Raw bootstrap *p* values were based on 5000 resamples. Holm-adjusted *p* values control the family-wise error rate across the seven primary total-heat trigger comparisons.

To translate the metric-dependent ranking differences into practical engineering use, Table 6 summarises the design question addressed by each primary heat metric and its most relevant application context.

Table 6. Decision-oriented interpretation of thermal-runaway heat metrics for cell-level safety assessment.

Heat Metric	Primary Engineering Question	Representative Application	Key Interpretation
Absolute total heat (kJ per failed cell)	How much heat is released by one failed cell?	Assessing the total cell-level thermal load generated by a failure event	Represents absolute event magnitude; larger cells may rank highly simply because more total energy is released
Energy-normalised total heat (kJ·Wh ^{−1})	How much heat is released relative to nominal electrical size?	Comparing thermal output among cells with different nominal energies or formats	Useful for size-normalised comparison, but should not be interpreted as equivalent to absolute failed-cell severity
Body heat (kJ)	How much heat is transferred through the cell casing?	Evaluating thermal loading of contacting structures, cell holders, cooling elements, and radial thermal barriers	Most directly represents casing-mediated thermal loading; does not capture directional ejecta heating
Ejecta heat (kJ)	How much heat is carried away by vented gases, particles, electrolyte, and other ejected material?	Designing axial spacing, vent channels, shields, particle interception, and protection of components along the ejecta trajectory	Represents directional vent- and particle-mediated loading; its ranking can differ substantially from body-heat ranking

Several limitations should be recognised. First, this was a secondary analysis of an experimental databank rather than a prospectively balanced comparison. Cell models were tested in different campaigns and were not fully crossed with S-FTRC generation, trigger mechanism, vent configuration, or test date. The fixed-effects adjustment standardises observed outcomes for apparatus generation but cannot identify a causal device-generation effect.

Second, the number of independent cell-model clusters was modest ($n = 13$), even though 98 repeated tests were included. Repeated experiments improve estimation of within-model variability but do not replace diversity at the cell-model level; consequently, population-level generalisation beyond the commercial cell models represented in the present dataset is limited. Replication was also unequal across cell models ($n = 3\text{--}15$); for example, C9 (LG HG2) was represented by four primary-cohort tests. Because this study was a secondary analysis of a fixed public databank, additional experiments could not be generated for sparsely represented cell models. We therefore evaluated robustness using the prespecified restriction to models with at least five repeated tests and leave-one-cell analyses, although estimates for sparsely represented models should still be interpreted with greater caution. This limitation is reflected in the wide uncertainty intervals for the allometric exponents, so the results support, but do not prove, proportional scaling of heat with nominal energy.

Third, nominal cell energy was used as an electrical-size descriptor. Nominal Wh is readily available and directly comparable across commercial cells, but it is not identical to actual discharge energy immediately before TR. Charge protocol, state of health, temperature, and usable capacity may alter the actual stored electrochemical energy. Future studies should therefore report both nominal energy and measured pre-test discharge energy where available.

Fourth, the analysed outcomes were integrated heat quantities. They do not directly describe peak heat-release rate, instantaneous heat flux, ejecta velocity, particle temperature, gas composition, flame radiation, or the timing of body and ejecta energy release. Recent calorimeter developments indicate that temporal resolution and external combustion can materially change interpretation of the thermal hazard [25]. A lower integrated ejecta energy could still correspond to a short, high-intensity jet capable of initiating a neighbouring cell, while a high body-heat total could be released over a longer duration.

Fifth, the analysis was restricted to single cylindrical cells. Energy-normalised quantities should not be interpreted as measured module- or pack-level outcomes. Module propagation depends on the interaction between the source-cell response and system geometry, thermal barriers, cooling, electrical connections, confinement, vent paths, and the susceptibility of neighbouring cells. Comparisons among calorimeter configurations further indicate that environmental and apparatus conditions can affect measured TR characteristics.

These limitations define several useful research directions. The most direct next step is to combine probability-based cell rankings with a module heat-transfer model in which body heat is applied as a distributed casing boundary condition and positive- and negative-end ejecta heat are represented as orientation-dependent localised sources. High-speed radiography available for many Battery Failure Databank tests could also be used to examine whether vent blockage, jelly-roll ejection, bottom rupture, or delayed internal propagation explains cell pairs with robust pathway reversals. Future calorimetry should additionally report time-resolved body and ejecta heat, directional heat flux, ejecta mass and velocity, and gas composition within repeated-test designs.

From a reporting perspective, studies comparing commercial cells should report absolute total heat, body heat, positive- and negative-end ejecta heat, nominal or measured electrical energy, and repeated-test distributions. Normalised metrics should complement

rather than replace absolute quantities, and the intended engineering interpretation of each ranking should be stated explicitly.

4. Conclusions

This study demonstrates that the TR hazard ranking of commercial cylindrical lithium-ion cells is not invariant, but changes according to both the scale used to express heat release and the pathway through which heat leaves the cell. Although the allometric analyses did not reject proportional scaling of total, body, or ejecta heat with nominal cell energy, replacing absolute total heat with $\text{kJ}\cdot\text{Wh}^{-1}$ changed the relative ordering of 46.2% of cell-model pairs after S-FTRC-generation adjustment. The pathway effect was even larger: body- and ejecta-heat rankings disagreed for 62.8% of pairs, and 28 of 78 pairwise reversals were reproduced in at least 95% of bootstrap samples. The compositional analysis further showed marked cell-specific differences in the balance between body heat and ejecta heat and in the directionality of ejecta.

These results indicate that TR severity should be evaluated using metrics matched to the engineering question rather than condensed into a universal cell-safety ranking. Absolute heat is relevant to the energy released by a failed cell, energy-normalised heat to heat output relative to nominal electrical size, body heat to casing-mediated thermal loading, and ejecta heat to directional vent- and particle-mediated loading. For cell-level safety screening, these quantities should therefore be reported together and interpreted separately. Extending this framework to fully crossed experimental designs, measured pre-test stored energy, time-resolved heat flux, and module-level geometries would help translate cell-level ranking differences into propagation-relevant design criteria.

Author Contributions: Conceptualisation, Y.L.; methodology, N.W. and Y.L.; data collection, Y.L. and N.W.; data curation, Y.L.; formal analysis, N.W. and Y.L.; validation, Y.L.; visualisation, N.W.; writing—original draft preparation, Y.L.; writing—review and editing, Y.L. and N.W.; supervision, N.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data analysed in this study are publicly available in the Battery Failure Databank, available at <https://www.nlr.gov/transportation/battery-failure> (accessed on 11 August 2026). The present study is a secondary analysis of the publicly available February 2024 revision of the databank; no new experimental data were generated.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT (GPT-5.6 Sol) to improve the English language, including grammar, clarity, concision, and overall readability. After using this tool, all authors carefully reviewed and edited the content as necessary and take full responsibility for the content of the publication.

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

References

1. Xu, W.; Zhou, K.; Wang, H.; Lu, L.; Gao, B.; Wang, Y.; Li, Y. Experimental and Modeling Study of Arc Fault Induced Thermal Runaway in Prismatic Lithium-Ion Batteries. *Batteries* **2024**, *10*, 269. [CrossRef]
2. Li, Z.; Cheng, Z.; Yu, Y.; Wang, J.; Wang, L.; Mei, W.; Wang, Q. Thermal Runaway Comparison and Assessment between Sodium-Ion and Lithium-Ion Batteries. *Process Saf. Environ. Prot.* **2025**, *193*, 842–855. [CrossRef]
3. Yang, J.; Pan, T.; Li, Y.; Wang, J.; Wu, Z.; Duan, Q.; Jin, K.; Wang, Q.; Sun, J. Thermal Runaway in Large-Capacity Sodium-Ion Batteries: Safety Performance Evaluation under Thermal, Electrical, and Mechanical Abuse Conditions. *J. Energy Storage* **2025**, *130*, 117405. [CrossRef]
4. Qi, C.; Wang, H.; Li, M.; Li, C.; Li, Y.; Wei, N.; Zhang, H. Study on the Multi-Stage Evolution of Thermal Runaway and the Flammability Threshold of Gas Generation in Lithium Iron Phosphate Batteries Based on SOC Gradient. *Micromachines* **2025**, *16*, 544. [CrossRef] [PubMed]

5. Li, H.; Shen, H.; Li, M.; Gao, Q.; Xu, C.; Feng, X. A Simplified Mathematical Modeling Approach for Thermal Runaway of $\text{Li}(\text{Ni}_{0.8}\text{Co}_{0.1}\text{Mn}_{0.1})\text{O}_2$ Pouch Cells Based on Thermal Runaway Experimental Data. *J. Energy Storage* **2024**, *90*, 111829. [\[CrossRef\]](#)
6. Wei, N.; Zhang, F.; Zhang, W.; Li, X. Comparative Study on the Thermal Runaway Characteristics of $\text{Li}(\text{Ni}_x\text{Co}_y\text{Mn}_z)\text{O}_2$ Batteries. *Heliyon* **2024**, *10*, e31203. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Ren, X.; Kong, D.; Ping, P.; Zhao, X.; Gao, W.; Zhang, Y.; Yang, C.; Feng, Z.; Wang, D. Experimental Study on the Evolution Patterns of Thermal Runaway and Safety Boundaries of Commercial Lithium-Ion Batteries under Overheating Failure. *Mater. Today Energy* **2025**, *52*, 101937. [\[CrossRef\]](#)
8. Feng, X.; Lu, L.; Ouyang, M.; Li, J.; He, X. A 3D Thermal Runaway Propagation Model for a Large Format Lithium Ion Battery Module. *Energy* **2016**, *115*, 194–208. [\[CrossRef\]](#)
9. Xu, W.; Li, Y.; Wang, H.; Lu, L.; Ouyang, M. Study of the Arc Generation Mechanism Induced by Liquid Electrolyte in Battery Energy Storage Systems. *eTransportation* **2026**, *27*, 100533. [\[CrossRef\]](#)
10. Liu, S.; Lu, S.; Huang, J.; Zhao, J.; Zhang, H. A Review on Thermal Runaway Warning Technologies for Aviation Lithium-Ion Batteries: Progress, Challenges, and Prospects. *Renew. Sustain. Energy Rev.* **2026**, *239*, 117149. [\[CrossRef\]](#)
11. Wang, Q.; Wang, H.; Xu, C.; Jin, C.; Wang, S.; Xu, L.; Ouyang, J.; Feng, X. Multidimensional Fire Propagation of Lithium-Ion Phosphate Batteries for Energy Storage. *eTransportation* **2024**, *20*, 100328. [\[CrossRef\]](#)
12. Zhao, L.; Hou, J.; Feng, X.; Xu, J.; Xu, C.; Wang, H.; Liu, H.; Hou, B.; Rui, X.; Gu, Y.; et al. The Trade-off Characteristic between Battery Thermal Runaway and Combustion. *Energy Storage Mater.* **2024**, *69*, 103380. [\[CrossRef\]](#)
13. Vujnović, N.; Hohenberg, K.; Höschel, P.; Moser, J.; Ellersdorfer, C. A Comprehensive Review of Thermal Runaway and Thermal Runaway Propagation Modelling Approaches in Lithium-Ion Batteries. *J. Energy Storage* **2025**, *139*, 118685. [\[CrossRef\]](#)
14. Luo, Y.; Li, P.; Wu, Z. Rational Design of Polymer Electrolytes for Thermal Runaway Suppression: A Review. *Sci. China Chem.* **2026**. [\[CrossRef\]](#)
15. Dai, Y.; Panahi, A. Thermal Runaway Process in Lithium-Ion Batteries: A Review. *Next Energy* **2025**, *6*, 100186. [\[CrossRef\]](#)
16. Wei, N.; Li, M. Experimental Investigation of Thermal Runaway in 40Ah Prismatic Lithium Batteries at Different SOC. *An. Acad. Bras. Ciênc.* **2024**, *96*, e20230648. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Wei, N.; Huo, L. Effect of Initial Confined-Space Oxygen Concentration on Vent-Gas Combustion During Thermal Runaway of NCM811 Lithium-Ion Cells. *World Electr. Veh. J.* **2026**, *17*, 212. [\[CrossRef\]](#)
18. Wang, G.; Ping, P.; Kong, D.; Peng, R.; He, X.; Zhang, Y.; Dai, X.; Wen, J. Advances and Challenges in Thermal Runaway Modeling of Lithium-Ion Batteries. *Innovation* **2024**, *5*, 100624. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Walker, W.Q.; Darst, J.J.; Finegan, D.P.; Bayles, G.A.; Johnson, K.L.; Darcy, E.C.; Rickman, S.L. Decoupling of Heat Generated from Ejected and Non-Ejected Contents of 18650-Format Lithium-Ion Cells Using Statistical Methods. *J. Power Sources* **2019**, *415*, 207–218. [\[CrossRef\]](#)
20. Sharp, M.; Darst, J.J.; Hughes, P.; Billman, J.; Pham, M.; Petrushenko, D.; Heenan, T.M.M.; Jarvis, R.; Owen, R.; Patel, D.; et al. Thermal Runaway of Li-Ion Cells: How Internal Dynamics, Mass Ejection, and Heat Vary with Cell Geometry and Abuse Type. *J. Electrochem. Soc.* **2022**, *169*, 020526. [\[CrossRef\]](#)
21. Finegan, D.P.; Billman, J.; Darst, J.; Hughes, P.; Trillo, J.; Sharp, M.; Benson, A.; Pham, M.; Kesuma, I.; Buckwell, M.; et al. The Battery Failure Databank: Insights from an Open-Access Database of Thermal Runaway Behaviors of Li-Ion Cells and a Resource for Benchmarking Risks. *J. Power Sources* **2024**, *597*, 234106. [\[CrossRef\]](#)
22. Lin, L.S.; Li, J.L.; Fishman, I.M.; Torres-Castro, L.; Preger, Y.; De Angelis, V.; Lamb, J.; Zhu, X.Q.; Allu, S.; Wang, H. Mechanically Induced Thermal Runaway Severity Analysis for Li-Ion Batteries. *J. Energy Storage* **2023**, *61*, 106798. [\[CrossRef\]](#)
23. Elsner, F.; Pischinger, S. Thermal Runaway in Batteries: A Database-Driven Literature Review and Exploratory Statistical Analysis. *Batteries* **2026**, *12*, 240. [\[CrossRef\]](#)
24. Masalkovaitė, K.; Gasper, P.; Finegan, D.P. Predicting the Heat Release Variability of Li-Ion Cells under Thermal Runaway with Few or No Calorimetry Data. *Nat. Commun.* **2024**, *15*, 8399. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Willstrand, O.; Pushp, M.; Andersson, P.; Brandell, D. A Calorimeter for Analyzing Ejected and Non-Ejected Heat during Li-Ion Battery Thermal Runaway. *iScience* **2025**, *28*, 112941. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Gillich, E.I.; Steinhart, M.; Fedoryshyna, Y.; Jossen, A. Mechanical Measurement Approach to Characterize Venting Behavior during Thermal Runaway of 18650 Format Lithium-Ion Batteries. *Batteries* **2024**, *10*, 142. [\[CrossRef\]](#)
27. Wang, Z.; Yin, B.; Zhao, Q.; Wang, J.; An, W.; Zhai, H. External Heating-Induced Thermal Runaway and Gas Venting Characteristics of Polymer Lithium-Ion Cells with $\text{LiNi}_x\text{Co}_y\text{Mn}_z\text{O}_2$ Cathode. *Process Saf. Environ. Prot.* **2023**, *174*, 745–755. [\[CrossRef\]](#)
28. Wang, Y.; Gao, Q.; Wang, G.; Lu, P.; Zhao, M.; Bao, W. A Review on Research Status and Key Technologies of Battery Thermal Management and Its Enhanced Safety. *Int. J. Energy Res.* **2018**, *42*, 4008–4033. [\[CrossRef\]](#)
29. Azuaje-Berbeci, B.J.; Ertan, H.B. A Model for the Prediction of Thermal Runaway in Lithium-Ion Batteries. *J. Energy Storage* **2024**, *90*, 111831. [\[CrossRef\]](#)
30. Rappsilber, T.; Yusfi, N.; Krüger, S.; Hahn, S.-K.; Fellingner, T.-P.; Krug Von Nidda, J.; Tschirschwitz, R. Meta-Analysis of Heat Release and Smoke Gas Emission during Thermal Runaway of Lithium-Ion Batteries. *J. Energy Storage* **2023**, *60*, 106579. [\[CrossRef\]](#)

31. Li, W.; Xue, Y.; Feng, X.; Liu, J.; Zhang, F.; Rao, S.; Zhang, T.; Gao, Z.; Du, Z.; Ni, C.; et al. Enhancing Understanding of Particle Emissions from Lithium-Ion Traction Batteries during Thermal Runaway: An Overview and Challenges. *eTransportation* **2024**, *22*, 100354. [[CrossRef](#)]
32. Huang, W.; Feng, X.; Pan, Y.; Jin, C.; Sun, J.; Yao, J.; Wang, H.; Xu, C.; Jiang, F.; Ouyang, M. Early Warning of Battery Failure Based on Venting Signal. *J. Energy Storage* **2023**, *59*, 106536. [[CrossRef](#)]
33. Lian, T.; Mikalsen, R.F.; Valset, K.; Meraner, C.; Li, T.; Snersrud, D.O.; Stølen, R.; Vie, P.J.S. Influence of Abuse Methods on Thermal Runaway in Lithium-Ion Cells: Measured Heats from Battery, Jet Flame, and Oxygen Depletion Calorimetry. *J. Electrochem. Soc.* **2025**, *172*, 080517. [[CrossRef](#)]
34. Nouhra, C.; Bengaouer, A.; De Persis, S.; Chaumeix, N.; Cognard, J.; Elezaar, K.B.; Maire, P.; Marteau, D.; Dubourg, S. Elucidation of the Impact of Experimental Setup on Thermal Runaway Characterization and Risk Assessment of Li-Ion Batteries. *J. Therm. Anal. Calorim.* **2026**, *151*, 5633–5651. [[CrossRef](#)]
35. Mallarapu, A.; Sunderlin, N.; Boovaragavan, V.; Tamashiro, M.; Peabody, C.; Pelloux-gervais, T.; Li, X.X.; Sizikov, G. Effects of Trigger Method on Fire Propagation during the Thermal Runaway Process in Li-Ion Batteries. *J. Electrochem. Soc.* **2024**, *171*, 040514. [[CrossRef](#)]

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