Statistical Uncertainty in Bench-scale Flammability Tests

Citation for published version:
Morrisset, D, Thorncroft, G, Hadden, RM, Law, A & Emberley, R 2021, 'Statistical Uncertainty in Bench-scale Flammability Tests ', *Fire Safety Journal*, vol. 122, 103335.
https://doi.org/10.1016/j.firesaf.2021.103335

Digital Object Identifier (DOI):
10.1016/j.firesaf.2021.103335

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Fire Safety Journal

Statistical Uncertainty in Bench-scale Flammability Tests

David Morrisset[1,2*], Glen Thorncroft[1], Rory Hadden[2], Angus Law[2], Richard Emberley[1]

[1] California Polytechnic State University – San Luis Obispo, Mechanical Engineering Department and Fire Protection Engineering Program
[2] The University of Edinburgh, School of Engineering

Abstract

Standardized testing procedures and experimental studies in fire safety science specify repeat trials in an effort to minimize the impact of statistical variation on the presented results. Bench-scale flammability testing procedures typically require a constant number of trials – generally a total of three – across all experimental conditions. A series of experiments was conducted to quantify the influence of trial quantity on the statistical variation in Cone Calorimeter results; sets of 100 identical trials were conducted using black PMMA for three different heat flux exposures (20, 40, and 60 kW/m^2). Gaussian statistics were applied to the time to ignition data and time-resolved mass loss rate data to investigate and illustrate the influence of trial quantity on predicted statistical uncertainty. Statistical variation observed for time to ignition data was found to decrease with increasing heat flux while the variation observed in the MLR data increased with increasing heat flux. The statistical uncertainty was found to decrease as a function of $1/\sqrt{n}$ for *n* trials. The presented statistical analysis requires population characteristics to be known, or assumed, *a priori*; the results, however, demonstrate the importance of considering statistical uncertainty in addition to the variation within data collected under different exposure conditions.

Key Words: Statistical uncertainty, Cone Calorimeter, bench-scale testing, time to ignition, mass loss rate

1. Introduction

Bench-scale flammability testing is widely used for material classification in fire safety science. Flammability assessment includes the ignition of materials under radiant heat exposures and the measurement of burning behavior (e.g., mass loss rate, heat release rate) under various conditions. One of the most commonly used experimental apparatuses for bench-scale flammability is the Cone Calorimeter, which is used for both standardized testing and in novel experimental work. Results from testing procedures or experiments in the Cone Calorimeter are likely to show variation from trial to trial; as such, multiple trials are generally conducted, and the results are then averaged across all trials. While results from bench-scale flammability assessment are readily found in the literature (e.g., [1]–[4]), relatively little discussion has been dedicated to the influence of trial quantity on the statistical confidence of experimental results.

Procedures for bench-scale ignition testing are outlined in various testing standards (e.g., [5]–[14]). Each standard stipulates a specific number of trials to be conducted (*n*); the number of trials defined in selected

* Corresponding Author: d.morrisset@ed.ac.uk; John Muir Building, Edinburgh EH9 3DW, United Kingdom

standards are shown in Table 1. This shows that most standardized tests require a total of three trials to be conducted for a given material and experimental condition. Past work has investigated the reproducibility of ignition data in round robin studies [15]–[17] – all of which were conducted using three trials. To the knowledge of the authors, there has been no quantification of the statistical uncertainty associated with a low number of trials or the marginal confidence that is gained with increasing the number of trials conducted in flammability studies.

Table 1. Required number of trials for different ASTM ignitability tests.

Standardized Test	Minimum Number of Trials
Cone Calorimeter **ASTM E1354/ISO 5660**	3 Samples
Mass Loss Cone **ASTM E2102/ISO 17554**	3 Samples
Lateral Ignition and Flame Spread (LIFT) **ASTM E1321/ISO 5658**	3 Samples (Flame Spread) 6 Samples (Ignition)
Fire Propagation Apparatus (FPA) **ASTM E2058**	3 Samples
OSU Apparatus **ASTM E906**	3 Samples
Intermediate Scale Calorimeter (ICAL) **ASTM E1623/ISO 14696**	3 Samples
Critical Radiant Flux of Floor-covering Systems using a Radiant Heat Energy Source **ASTM E648/ISO 9269**	3 Samples
Surface Flammability of Materials Using a Radiant Heat Energy Source **ASTM E162**	4 Samples
Vertical Tube Furnace **ASTM E136/ISO 1182**	4 Samples

Reported results from Cone Calorimeter testing include time to ignition and time resolved mass loss rate (MLR), heat release rate (HRR), effective heat of combustion (HOC), and soot yield. Results such as the HRR and HOC require multiple measurements including gas analysis of the products of combustion and therefore the uncertainty in such results is driven by multiple measurements. Due to the calculation of these parameters requiring inputs from multiple measurements, each with its own uncertainty, this present study investigates the statistical uncertainty of single parameter measurements (e.g., time to ignition and mass loss rate). Quantifying the statistical uncertainty as a function of trials conducted can provide researchers with the ability to design experiments and testing procedures around a desired threshold of uncertainty, as opposed to a single fixed trial quantity across all conditions.

This study outlines a series of experiments that were conducted with high trial quantities (100 trials per experimental condition) and then presents the statistical analysis required to explore the influence of sample number on statistical variation. The statistical analysis presented in this work is not limited to bench-scale flammability assessment studies and can be applied to many other experimental fields within and outside fire safety science.

2. Previous Studies

While previous studies have not explicitly investigated the reduction in statistical uncertainty with increasing trial quantity, studies have been conducted for apparatuses such as the Cone Calorimeter to investigate repeatability and reproducibility. Repeatability (r) refers to the variation in results taken by the same instrument, while reproducibility (R) refers to variation in results compared between different instruments (often the same device, e.g., the Cone Calorimeter at different laboratories) [16].

The ASTM standard for the Cone Calorimeter, ASTM E1354, presents interlaboratory studies on the repeatability and reproducibility of results from the Cone Calorimeter. The results were obtained from six different laboratories and included results for five different materials. It is noted that "for most of these materials, three replicates each were tested in two orientations (horizontal and vertical) and at two irradiance levels (25 and 50 kW/m^2)". No further information is provided to comment on whether three trials were in fact conducted for each material and heat flux, as the term "most" is not precise. A similar round-robin study was coordinated by Urbas and was conducted using the Cone Calorimeter with 16 materials across four laboratories [15]. Trials were conducted at a single heat flux, 75 kW/m^2, with a total of three trials for each material. The study compared the repeatability and reproducibility across the participating labs and compared the results to three previously reported round robin studies (including the study discussed in the annex of ASTM E1354). The resulting repeatability and reproducibility results for time to ignition data were presented as a linear function of the average time to ignition; the linear models are illustrated in Figure 1. Additional round robin studies can be found in literature for bench-scale flammability assessment [16], [17], but were again only conducted in triplicate and offer no insight into the influence of trial quantity.

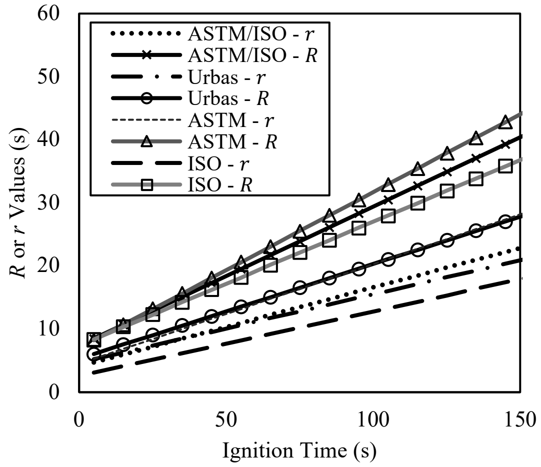

Figure 1. The round robin studies provided in Urbas' work showing the repeatability (*r*) and reproducibility (*R*) of time to ignition data from the Cone Calorimeter [15]. See referenced work for further information on individual round robin studies.

Repeatability and reproducibility are useful metrics when commenting on the capacity of an apparatus to produce consistent results in terms of material classification. These quantities do not, however, allow inference regarding the degree to which the results from a defined number of trials is representative of material behavior and whether or not the trial quantity truly captures the range of expected results. Results from the round robin studies do not differentiate between the various materials used in the studies, which ranged from gypsum board to PVC. The studies were also conducted over a limited range of external heat fluxes. Some of which were limited to a single heat flux as high as 75 kW/m^2 - a relatively high level of irradiation compared to typical exposures in the Cone Calorimeter. Therefore, the results of previous round robin studies do not isolate the influence of material variation or the influence of external radiation (i.e., experimental conditions) on repeatability. These results do not indicate how the statistical variation would change as a function of the trials conducted and are only applicable to studies conducted in triplicate.

The aim of this current study is to provide information on statistical variation outside the context of repeatability and reproducibility and instead quantify the statistical scatter in flammability testing results for a given experimental condition as a function of the number of trials conducted. This study instead focuses on a single material exposed to various heat fluxes to isolate the influence of trial quantity on statistical variation. The results presented may vary for different materials and experimental conditions, but the analysis can be applied to any experimental procedure.

3. Application of Gaussian Statistics

A Gaussian distribution, or a normal distribution, can be used to predict the probability of a given result based on the distribution and mean of the data. The degree of scatter from the mean within a set of experimental results can be quantified through either the standard deviation or a specified confidence interval (e.g., a 95% confidence interval suggests the specified range captures 95% of the expected results). In the context of statistics, a set of *n* trials is generally referred to as a sample; however, in the context of fire safety science, this term can be confused with an individual specimen tested (i.e., trial). To avoid this confusion, this work will not use the term “sample” to refer to a set number of trials; instead “dataset” is used throughout to refer to the collective results for a given experimental condition (e.g., all time to ignition data for a given heat flux).

If a normal distribution is assumed, the population mean of the experimental data is generally predicted given a set of *n* trials using Equation 1, where $\boldsymbol{\sigma}$ is the standard deviation of the population and $\boldsymbol{z}$ is the *z*-statistic which is a function of the desired confidence interval (e.g., 1.96 for a 95% confidence interval).

$$\mu = \bar{x} \pm \frac{z\,\sigma}{\sqrt{n}} \qquad \text{Equation 1}$$

The statistical uncertainty associated with predicting the population mean from the dataset is captured in the $z\sigma \,/\, \sqrt{n}$ term. This equation suggests that the statistical uncertainty behaves with a $1 \,/\, \sqrt{n}$ relationship, decreasing as the number of trials increases; this behavior is demonstrated in Figure 2. For a given standard deviation and confidence interval, the statistical uncertainty can be plotted as a function of the number of trials. This model requires the population standard deviation to be known *a priori*; any finite sample will only result in an estimate of the true population mean. Regardless of the value of sigma, Figure 2 shows that increasing *n* reduces the statistical uncertainty of predicting the population mean. Equation 1 can be modified to incorporate the *t*-statistic and the standard deviation of a finite set of trials. While the use of the *t*-statistic eliminates the need for the population standard deviation to be known *a priori*, the standard deviation of a finite set of trials varies between any set of n trials (particularly for low values of n) which introduces further complexity in the analysis. The following analysis focuses on the z-statistic to describe the variation in statistical uncertainty as a function of n, thus allowing for a robust and simplified analysis. Further elaboration on the use of the t-statistic can be found elsewhere [19].

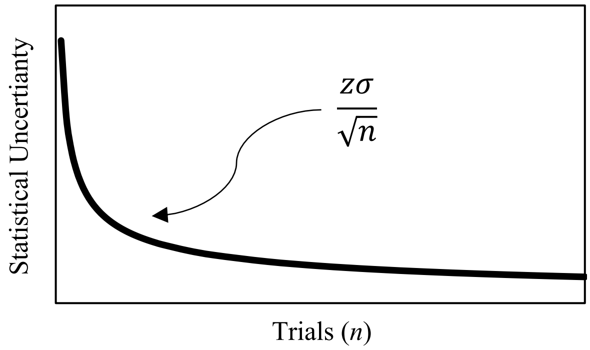

Figure 2. Statistical uncertainty plotted as a function of n using Equation 1.

Figure 2 demonstrates that the decrease in statistical uncertainty becomes more gradual as n increases due to the $1 / \sqrt{n}$ relationship. This behavior also suggests that additional trials provide diminishing marginal benefit to reducing the statistical uncertainty in the results. The rate at which the marginal benefit diminishes with each trial is dependent on the confidence interval and the standard deviation. The gain in confidence from any additional trial n will be referred to as the marginal gain in confidence, or simply marginal gain, and can be calculated by taking the difference of the statistical uncertainty of n trials and for $n - 1$ trials. The marginal gain can be used as a metric for balancing the gain in statistical certainty from additional trials with the practical limit of time and resources that must be dedicated to additional trials.

4. Methodology

A methodology was developed to investigate the impact of trial quantity on the statistical confidence of results from bench scale flammability testing. The Cone Calorimeter was selected as the testing apparatus for this study due to its widespread use within the fire safety science community. Of the results taken from Cone Calorimeter testing, the time to ignition and mass loss rate were chosen for this study since both are based on a single measurement (i.e., no calculated values). Experiments were conducted with 6 mm thick black PMMA with a heated area of 100 x 100 mm. All samples were wrapped in a single layer of aluminum foil and placed on an aluminum block (100 x 100 x 20 mm) within the ASTM/ISO standard sample holders. Three heat flux exposures were used for this study – 20 kW/m^2, 40 kW/m^2 and 60 kW/m^2. A total of 100 trials were conducted at each heat flux with approximately 15-25 trials conducted a day to mitigate variation from set up or calibration. In undertaking this approach, it is assumed that the results can be characterized using Gaussian statistics. No finite sample size can truly represent population behavior; however, in this particular application a dataset of 100 is a reasonable size to test Gaussian behavior.

4.1. Experimental Procedure

Samples were ignited with a spark ignitor in accordance with the ASTM/ISO standard procedure. The top half of the retainer frame was not used in these experiments and the samples were set flush to the top of the sample holder base. The heating element of the Cone Calorimeter was calibrated using a Schmidt-Boelter heat flux gauge (uncertainty of 2.5%) and was calibrated to within 1% of the desired heat flux. Heat flux readings were also taken at various times between trials to ensure the heat flux calibration remained within the desired 1% range. The time to ignition was recorded for each trial; ignition times were determined through visual observation during each trial and were reported within the nearest second. Mass measurements were recorded at 1 Hz using a Mettler Toledo WM3002-L22 load cell (accuracy of 0.01 g). All MLR data were smoothed with a 10-point moving average.

5. Experimental Results

Ignition time results for each of the 100 trials conducted at 20, 40, and 60 kW/m^2 are presented in Figure 3. Scatter in the data around the mean can be described through the standard deviation of the data ($\boldsymbol{\sigma}$) or expressed as a confidence interval. The experimental results in this study are presented as a 95% confidence interval (i.e., variation from the mean value in which 95% of the data lie). As expected, the ignition results show a high time to ignition at low heat fluxes accompanied with higher scatter in the data when compared to higher heat fluxes. As the external heat fluxes increase, both the time to ignition and scatter in the data decrease. The 40 and 60 kW/m^2 cases both show discrete bands of ignition times; this illustrates that the degree of scatter in the experimental data is similar to the resolution of the recorded data, which was rounded to the nearest second. Figure 3 also shows histograms for each heat flux, indicating a generally unimodal, Gaussian distribution (see the following section for Gaussian goodness-of-fit test). The “assumed population mean” ignition time was determined by excluding recorded values two standard deviations from the mean; the values of which are presented in Table 2.

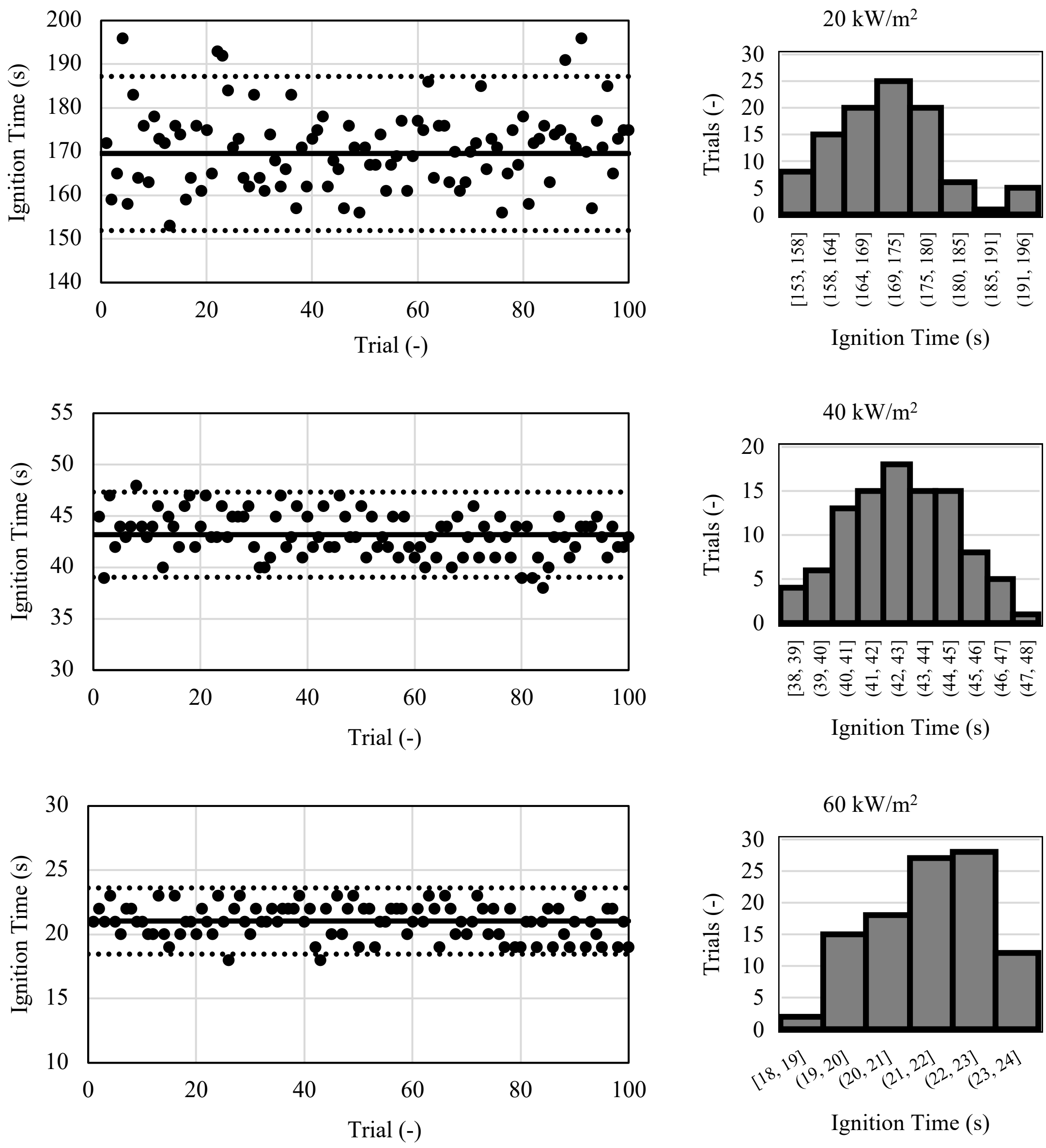

Figure 3. Resulting ignition times for black PMMA exposed to 20, 40, and 60 kW/m^2. Dashed lines indicate a 95% confidence interval and the solid line indicates the assumed population mean. Histograms are provided to demonstrate the approximately Gaussian behavior for each set of data.

Table 2. The means for the collected time to ignition (t_{ig}) data at each heat flux tested. Values corresponding to two standard deviations from the mean and the bounds in which 95% of the data fall are also presented.

Heat Flux [kW]	t_{ig} [s]	2σ [s]	95% [s]
20	169.56	18.00	17.64
40	43.18	4.23	4.14
60	21.04	2.63	2.58

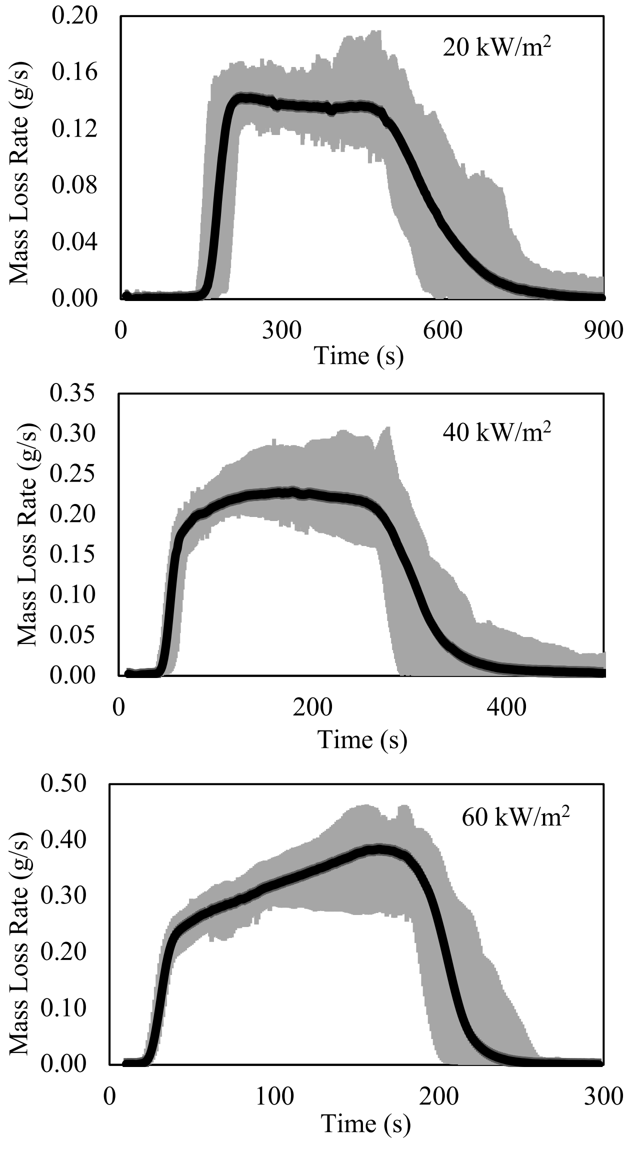

Figure 4. The resulting mass loss rate curves for each experimental heat flux. Data were smoothed with a 10-point moving average. The black line indicates the average MLR for all 100 trials at each heat flux and the shaded region indicates the range of all MLR results.

The MLR results for each heat flux are illustrated in Figure 4. Previous studies have indicated that non-charring materials such as PMMA achieve a relatively steady burning rate after ignition, and therefore a plateau in the MLR data is anticipated for PMMA [1]. The results corresponding to 20 kW/m^2 and 40 kW/m^2 indicate steady burning behavior immediately after the growth following ignition; however, the MLR data recorded at 60 kW/m^2 shows a steady increase in MLR until the burning rate begins to decay towards the end of the experiment. A steady increase instead of a MLR plateau has also been observed in previous studies [18]. This increasing behavior seen in the 60 kW/m^2 case and the increased statistical variation in all cases toward the end of burning are indicative of thermal penetration of the PMMA sample and influence of the back-face boundary condition. The resulting MLR curves are expected for PMMA based on literature; the interest of this study instead lies in the following statistical analysis.

6. Statistical Analysis

The following analysis is based upon the Gaussian statistics presented in Section 3. A Chi-square goodness-of-fit test (significance criteria of 1%) was conducted on each data set to determine if data could be assumed to be normally distributed. The time to ignition data for both the 20 and 40 kW/m^2 passed the test as normal distributions while the 60 kW/m^2 data did not. The unimodal distribution seen in the histogram of Figure 3 suggests that if the time to ignition data had been recorded to a higher resolution than one second and to a higher degree of accuracy than visual observation, the data set would likely have passed the Chi-square test. Thus, the failure of the test is likely artifact to the resolution and accuracy of the measurement approaching the resolution of the statistical variation for that specific experimental condition. Therefore, each time to ignition data set was analyzed as a Gaussian distribution.

The Chi-squared goodness-of-fit test was then applied to the time resolved MLR data to determine if the distribution in results at each time step can be treated as Gaussian. The results indicated that the distribution of the MLR for each heat flux was found to be Gaussian for approximately 50% of the time steps for 20 and 60 kW/m^2, and the data at 40 kW/m^2 reporting Gaussian for approximately 25% of the time steps. In the absence of data that is entirely Gaussian, statistical analysis was employed to illustrate general trends in the data, such as the $1 / \sqrt{n}$ decrease in statistical uncertainty with increasing trial quantity; therefore, the analysis of the whole time series is an approximation for illustrative purposes. More specific analysis was provided at specified points in time where the data is known to be normally distributed. The MLR results are limited to the smoothing technique used in this analysis; exact results may vary depending on the noise of the raw MLR data and the smoothing techniques chosen, but the general trends and analysis developed here apply outside the context of this study alone. Further work can be done to develop a more rigorous approach to the statistical analysis of time resolved MLR data.

Throughout the following analysis, "statistical uncertainty" will be quantified through a 95% confidence interval in the distribution of predicted values from the population mean (e.g., the $z\sigma \,/\, \sqrt{n}$ term seen in Equation 1 assuming 95% confidence). While this study focuses on statistical uncertainty, other forms of uncertainty (e.g., measurement error and calibration error) must also be considered if experimental results are to be used in further analysis or calculation. Such consideration ensures that an appropriate degree of uncertainty is propagated to the calculated results. Other forms of error, calibration error in particular, can influence the perceived statistical variation.

6.1. Ignition Time Results

Equation 1 was used to determine the 95% confidence interval for time to ignition as a function of trials based on the mean and standard deviations presented in Table 2. These results are presented in Table 3 and illustrated in Figure 5. The statistical variation is highest at low numbers of trials and decreases with the $1 \,/\, \sqrt{n}$ relationship predicted by Equation 1. Higher heat fluxes result in lower degrees of statistical scatter for any value of *n* trials, however, the variation between heat fluxes is non-linear showing a much higher difference in behavior when comparing the 20 and 40 kW/m^2 cases to the results from 40 and 60 kW/m^2. Such a nonlinear behavior is to be expected due to the rapid increase in the time to ignition as the incident heat flux approaches the critical heat flux for ignition. Even when considering 100 trials, the statistical uncertainty for the 20 kW/m^2 results remains approximately 1.75 seconds and would require over 300 trials to reduce the statistical variation from the true mean below one second (i.e., using Equation 1 assuming *n*=300).

Table 3. The 95% confidence interval and marginal gain in uncertainty for the time to ignition data conducted at 20, 40, and 60 kW/m^2 determined using Equation 1 ($z = 1.96$ for 95% confidence).

Trials (n)	20 kW/m^2 (σ=9.00 s)		40 kW/m^2 (σ=2.12 s)		60 kW/m^2 (σ=1.31 s)	
	Uncertainty	Marginal Gain	Uncertainty	Marginal Gain	Uncertainty	Marginal Gain
[-]	[s]	[s]	[s]	[s]	[s]	[s]
1	17.64	-	4.15	-	2.58	-
2	12.48	5.17	2.93	1.21	1.82	0.75
3	10.19	2.29	2.39	0.54	1.49	0.33
4	8.82	1.36	2.07	0.32	1.29	0.20
5	7.89	0.93	1.85	0.22	1.15	0.14
10	5.58	0.30	1.31	0.07	0.81	0.04
15	4.56	0.16	1.07	0.04	0.67	0.02
25	3.53	0.08	0.83	0.02	0.52	0.01

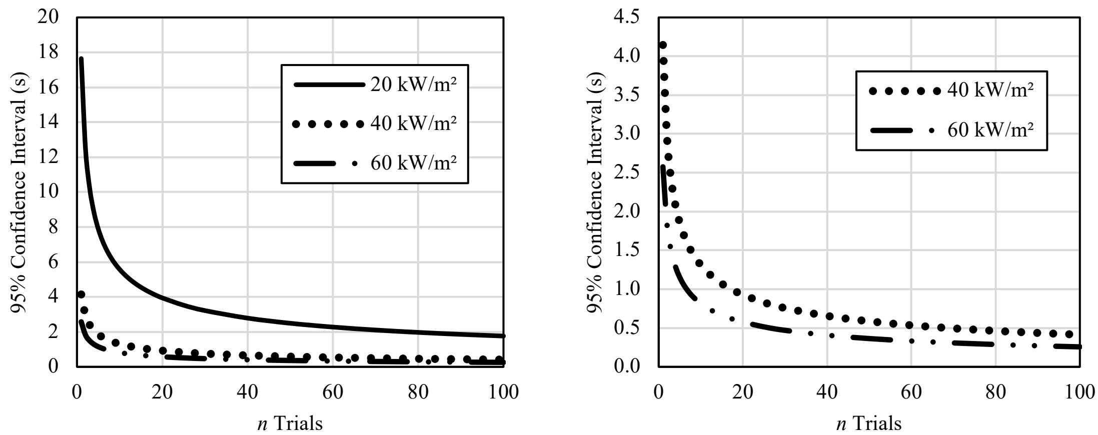

Figure 5. The uncertainty for each external heat flux tested plotted using Equation 1 (assuming 95% confidence). Additional plot showing only the 40 and 60 kW/m^2 cases is provided to show these results with higher resolution.

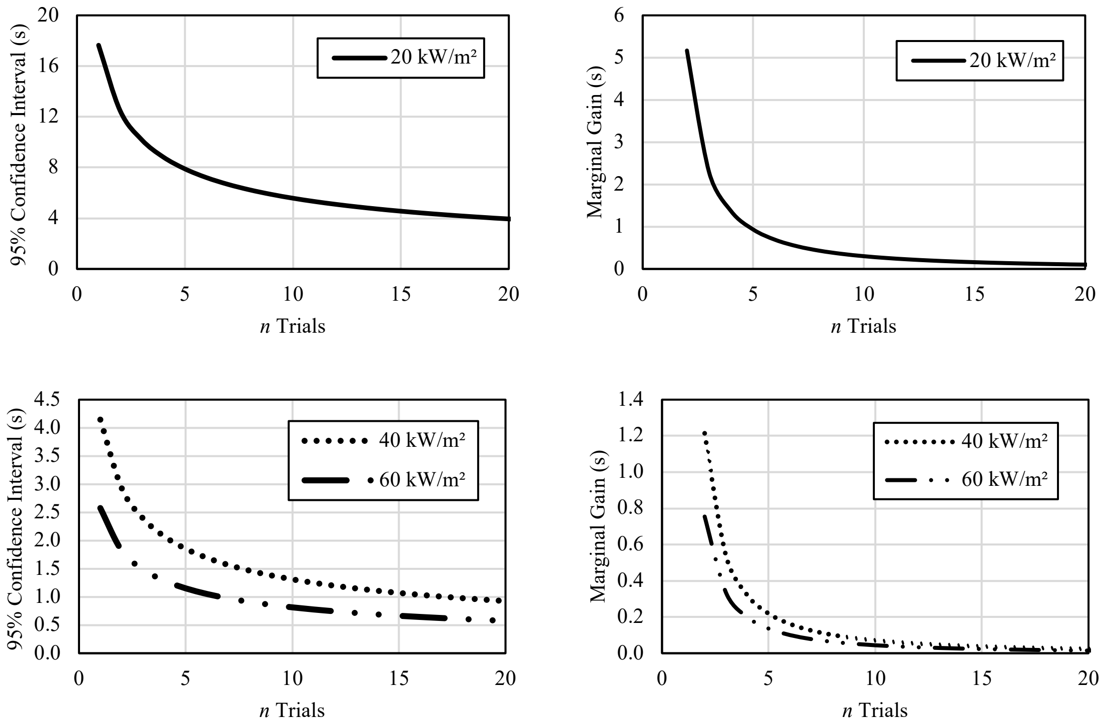

Figure 6. The statistical uncertainty and marginal gain as a function of n for each heat flux used (assuming 95% confidence). The statistical uncertainty plots are identical to those of Figure 5, showing only n<20.

Figure 6 illustrates both the statistical variation and the marginal gain as a function of trials conducted. The results conducted at 40 and 60 kW/m^2 suggest that the statistical variation in the ignition time can be reduced well below one second by conducting 20 trials. The results for 20 kW/m^2, however, continue to show uncertainty of approximately four seconds even when considering 20 trials being conducted. Results at all heat fluxes, however, suggest that the marginal gain in confidence is significantly reduced beyond 10-15 trials.

The results suggest two different approaches to guide the choice of an appropriate value for n – the statistical uncertainty (quantified through a 95% confidence interval) and the marginal gain. While the uncertainty in the ignition time reduces appreciably when considering 100 trials, such a large trial set is not necessarily practical. The reduction in uncertainty with each additional trial diminishes as the value of n increases – calculating the marginal gain can provide a criterion for quantifying at what point additional trials do not result in a commensurate improvement in uncertainty. If the criterion for trial quantity is set by uncertainty alone, then more trials would be expended with diminishing marginal benefits in statistical certainty. Therefore, the marginal gain can serve to optimize the reduction in statistical uncertainty with consideration to practical limitations such as time and resources. The statistical variation in the 20 kW/m^2 results suggest that 20 trials still result in an uncertainty of approximately four seconds. However, additional trials beyond 20 are likely to add less than 0.1 seconds of marginal gain. Additional trials beyond this threshold of 20 trials may or may not be warranted depending on the goals of the researcher.

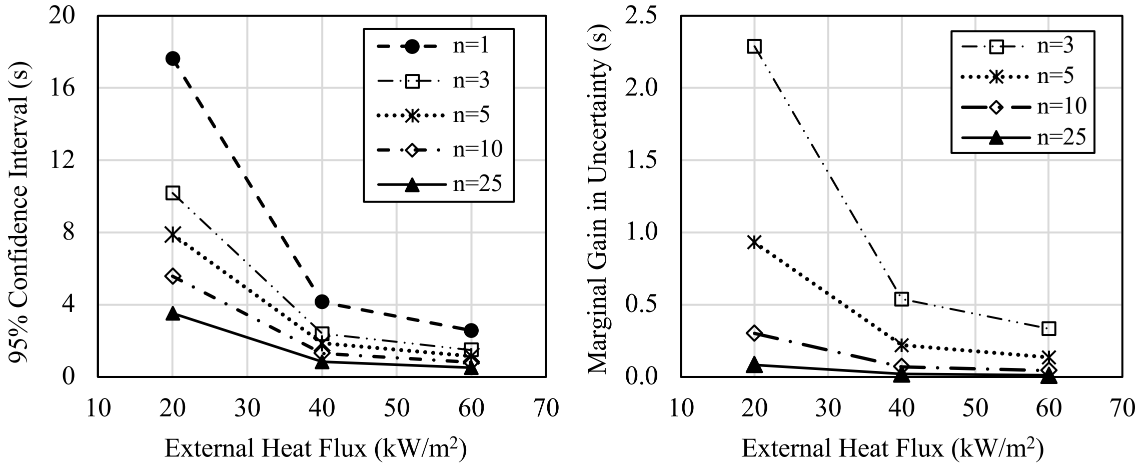

Figure 7. The uncertainty and marginal gain as a function of heat flux.

The statistical variation was found to be significantly lower when considering the 40 and 60 kW/m^2 cases when compared to 20 kW/m^2 (approximately 75% and 85% lower, respectively, for all values of n). Figure 7 illustrates the nonlinear behavior between the predicted statistical uncertainty and heat flux used. The nonlinear behavior of the uncertainty matches the general behavior of the time to ignition results expected for a material exposed to various heat fluxes (i.e., a plot of ignition time vs. external heat flux). For any value of n, the increase in uncertainty from 60 to 40 kW/m^2 was found to be as low as one or two seconds; the increase in uncertainty from 40 to 20 kW/m^2 was found to be as high as 13.5 seconds (for $n = 1$). Further experimentation would be required to definitively quantify the behavior of heat fluxes between 20 kW/m^2 and the critical heat flux for ignition. If the user desired to reduce the statistical variation of the results to be within two seconds, experiments conducted at 20, 40, and 60 kW/m^2 would require 80, 5, and 3 trials, respectively. Thus, reducing the statistical uncertainty to only twice the resolution of the measurement still requires more trials than the 1-3 trials prescribed in current flammability standards for each of the experimental conditions (apart from 60 kW/m^2 which corresponds to 3 trials).

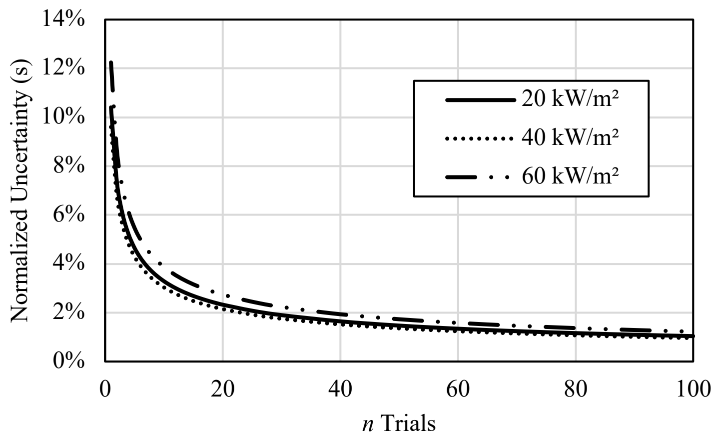

Figure 8. The normalized uncertainty for each heat flux as a function of *n*.

Further insight into the results are gained by presenting the normalized uncertainty (i.e., the statistical uncertainty for any given value of *n* normalized by the mean ignition time for a given heat flux) as seen in Figure 8. While the statistical variation for any value of *n* was found to vary between heat fluxes, the normalized uncertainty remains constant for all values of *n* across all heat fluxes. This suggests that for any heat flux, the uncertainty of predicting the population mean can be approximated as 10-12% of the recorded mean ignition time when considering a single trial. Assuming the normalized uncertainty can be applied across all heat fluxes, then the analysis presented previously can be potentially extrapolated to lend insight into heat fluxes between those that were conducted experimentally in this study; further information on this concept is presented by Morrisset [19].

Figure 7 demonstrates the importance of trial quantity at particularly low heat fluxes – in this case 20 kW/m^2 and below. If the observed trends for any value of n continue beyond 20 kW/m^2, then the statistical uncertainty for heat fluxes approaching the critical heat flux could be substantial. This behavior has implications to experiments or standardized tests that operate at low heat fluxes. For example, the ASTM/ISO standardized procedure for the Cone Calorimeter suggests the process of bracketing to experimentally determine the critical heat flux which inherently uses low heat flux exposures close to the critical heat flux. When conducting experiments near the critical heat flux, experimental ignition times can be on the order of 600-1000 seconds. The statistical uncertainty in determining such long ignition times could result in errors of 60-120 seconds assuming the normalized uncertainty remains constant at heat fluxes lower than 20 kW/m^2. Experimental validation of these results at low heat fluxes is required for further insight; however, these results do suggest that care should be taken at low heat fluxes and a higher trial quantity may be required to mitigate error.

6.2. Mass Loss Rate Results

While the analysis presented in the previous section on the time to ignition focuses on a single data point for each trial, a similar statistical approach was used to investigate the time resolved mass loss rate data. Figure 4 illustrates the range of recorded MLR time-histories for each heat flux and suggests that the variation between trials is not constant with time. The statistical analysis used previously was applied to the spread of MLR trials at each time step to quantify the temporal variation in MLR data. The MLR data for each trial was aligned at the time to ignition for the following analysis. At each time step following ignition, the spread of MLR data was described through a 95% confidence interval calculated using Equation 1. In doing so, the distribution of MLR data between the 100 trials at any given time step is assumed to be Gaussian based on the Chi-square test discussed previously.

Figure 9 demonstrates how the statistical variation changes as a function of time for 20, 40, and 60 kW/m^2. The statistical variation increases in time for each of the heat fluxes used in the experiments, showing particularly low variation for the growth phase immediately after ignition. The statistical variation then increases before the MLR decays from the peak value and the magnitude of the statistical variation increases with increasing heat flux. Equation 1 was used to predict how the time resolved statistical uncertainty would likely vary with increasing trial quantity; increasing trial quantity reduces the statistical uncertainty as a function of $1 / \sqrt{n}$, as seen in Figure 9 by the predicted error bars for n=1, n=3, and n=10.

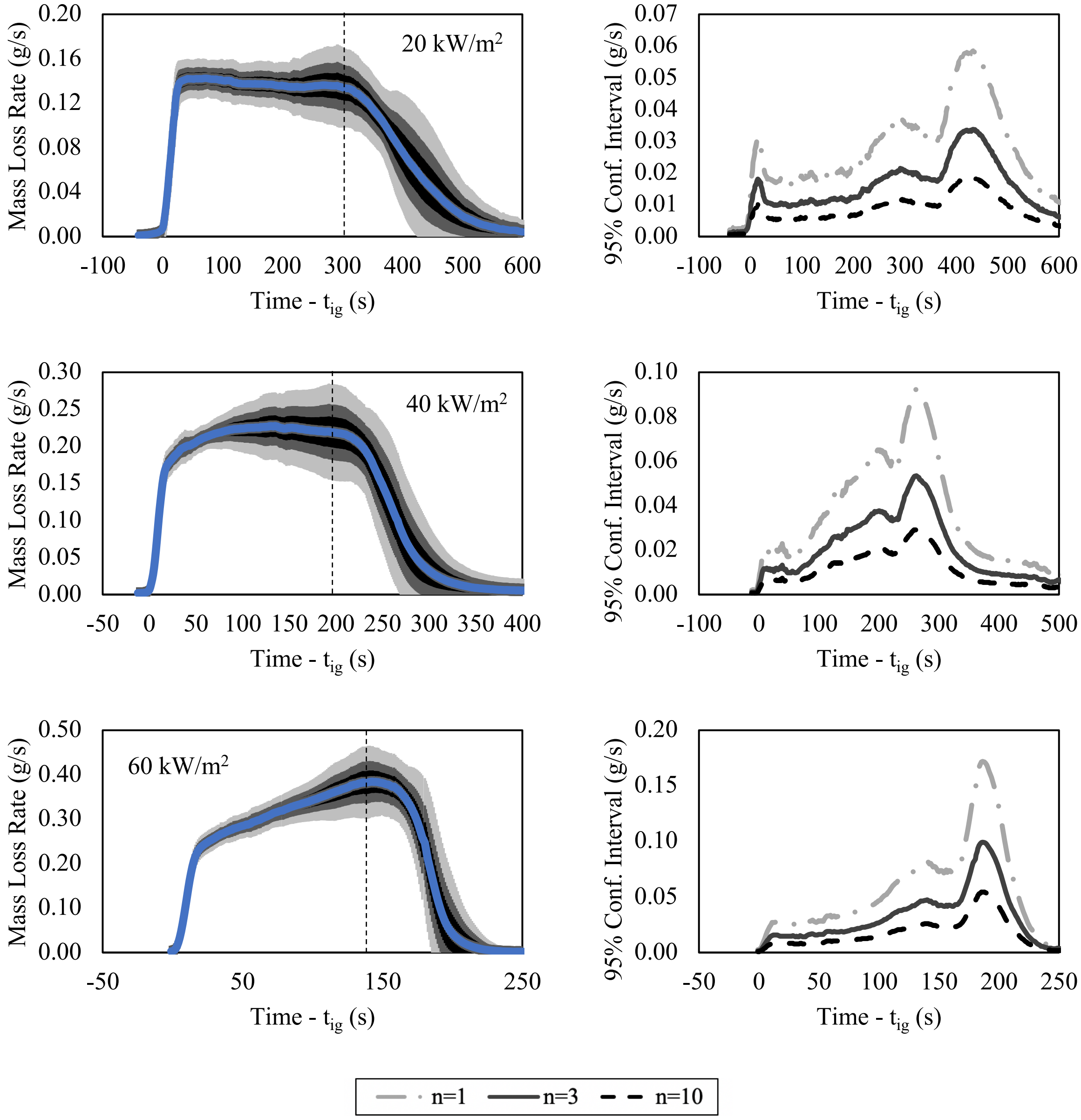

Figure 9. The time resolved statistical variation in recorded MLR for 20 kW/m^2, 40 kW/m^2, and 60 kW/m^2. The shaded regions indicate a 95% confidence interval for *n*=1, *n*=3, and *n*=10. The vertical dashed lines indicate the average point of 75% mass loss.

As seen in the statistical analysis for the time to ignition data, the marginal increase in statistical confidence decreases as *n* increases. Figure 9 also illustrates the variation in statistical uncertainty over the duration of the experiments. For each heat flux used in this study, the statistical uncertainty increased in time for any value of *n*. The statistical uncertainty increased just prior to the beginning of the decay phase while the highest degree of statistical uncertainty in all heat flux conditions was observed during the decay phase towards the end of each trial. Increasing values of *n* reduce the statistical uncertainty at all times and high values of *n* (e.g., *n*=10) significantly reduce the peak uncertainty observed in the decay phase.

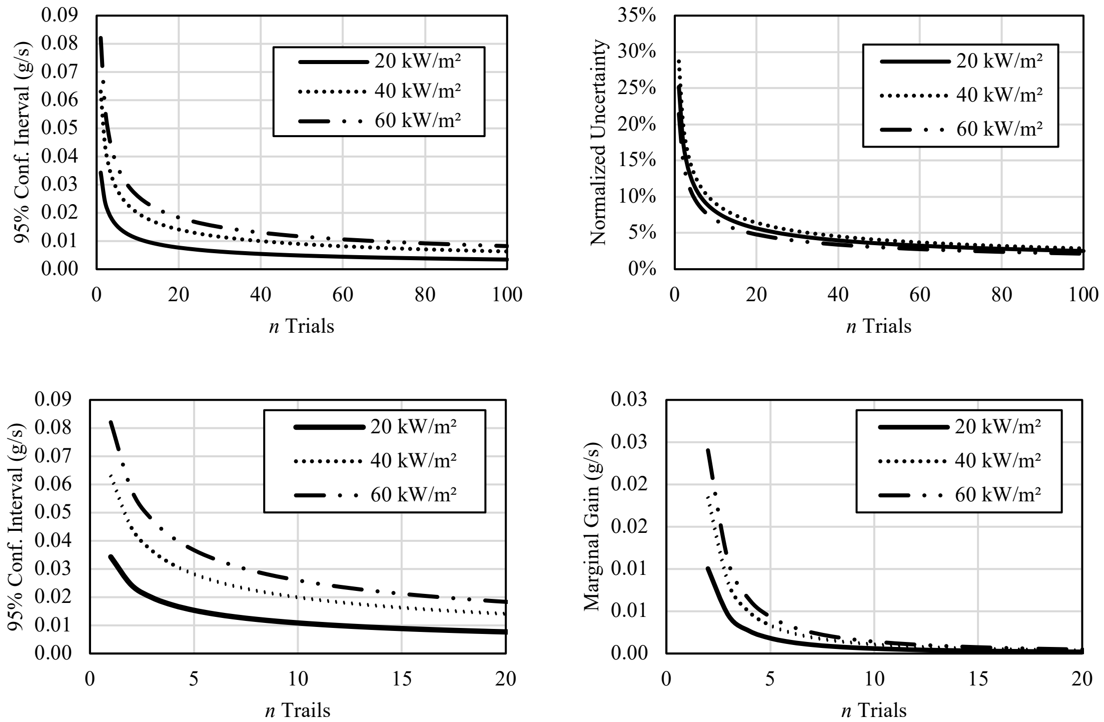

Figure 10. The statistical uncertainty, marginal gain, and normalized uncertainty in the MLR data at 75% mass loss as a function of trials conducted (assuming 95% confidence).

The precision to which the time resolved uncertainty can be predicted for various values of *n* is limited due to the time resolved MLR data not following a Gaussian distribution at all time steps. For the purposes of this study, the statistical variation at only a single point in time was used for further analysis – this allowed the same method analysis that was applied to the time to ignition data to be used for the single point in time of the mass loss data. The point chosen for this analysis corresponds to the point at which the average sample had lost 75% of the initial mass (indicated by the dashed lines in Figure 9). This point is of interest as it occurs during the steady burning regime, towards the point at which the sample transitions to the decay phase as seen in Figure 9 (occurring at approximately 310, 200, and 140 seconds for 20, 40, and 60 kW/m^2, respectively).

Figure 10 illustrates the statistical variation at 75% mass loss for each heat flux exposure. The highest statistical variation was observed at 60 kW/m^2 and decreased with decreasing heat flux. When normalized by the corresponding average MLR value for each heat flux, the normalized uncertainty ranges from 21-29% for n=1, then decreases to 12-17% for n=3, and further to 7-9% for n=10. As previously observed in the time to ignition data, the marginal gain in confidence for the statistical uncertainty in MLR, seen in Figure 10, is significantly reduced beyond 10-15 trials.

7. Implications to Experimental Design

The analysis in this study has demonstrated how statistical uncertainty in bench-scale flammability testing varies with increasing values of n; the following section outlines possible techniques to choose a relevant trial quantity for a given experiment based on these results. At this time, however, the techniques developed in this study require an approximation of population characteristics for the data collected.

The simplest method in selecting n is to follow the current standard approach and choose a single trial size to be used for all experiential conditions and materials (e.g., conduct trials in triplicate). There are instances in which this approach may be desired; for example, this study specified the same number of trials across all heat fluxes as an exploratory study into statistical variation. However, the results of this paper demonstrate that the use of a single trial size across all experimental conditions would yield data of varying statistical certainty for each of the different conditions. Thus, when considering time to ignition, a constant value of n may result in an insufficient amount of certainty at lower heat fluxes or an unnecessarily high degree of certainty at higher heat fluxes. In either case, the balance of statistical certainty and finite time and resources is not optimized.

An improved approach to choosing a trial quantity is based on matching an acceptable range of statistical uncertainty to the predicted degree of statistical scatter. This approach then places the objective of obtaining a uniform degree of statistical variation at the center of choosing a trial size. At this time, the operator must first assume that the population of possible results is a normal distribution and must be able to predict the approximate statistical scatter in the ignition results (i.e., population standard deviation). Equation 1 can then be used to estimate a number of trials based on the desired degree of uncertainty. ASTM E122 (Standard Practice for Calculating Sample Size to Estimate, with Specified Precision, the

Average for a Characteristic of a Lot or Process) more explicitly outlines this process of specifying the required number of trials based upon a specified range of acceptable uncertainty (*E*); Equation 2 can be used to approximate the required trials to achieve a specified uncertainty as a function of the predicted standard deviation and the operators confidence interval. The value of z can be modified based on the assumed degree of certainty.

$$n = (z\sigma/E)^2 \qquad \text{Equation 2}$$

The choice of trial quantity can be optimized further by considering both the statistical uncertainty and the marginal gain, as discussed previously. This consideration would allow the operator to minimize uncertainty while optimizing time and resources. The marginal gain does not require any further information than quantifying the statistical uncertainty itself and therefore using the two criteria in combination is the recommended approach to choosing a trial quantity.

Knowing the population standard deviation *a priori* is a major limitation to this approach, as the analysis is sensitive to small variations in the population standard deviation. Even when conducting 100 trials, as was in this study, the predicted population standard deviation is still only within approximately 16% of the true value (using the chi-squared distribution and assuming 95% confidence). Therefore, smaller sets of trials will result in even higher uncertainty in predicting the population standard deviation. Uncertainties presented in published work must be considered in the context of the given experiment and the number of trials conducted. Further work is required to establish a methodology for selecting trial quantity in experiment design if little is known about the population behavior.

8. Conclusions

The analysis presented in this work demonstrates the influence of trial quantity on the statistical uncertainty of bench-scale flammability testing and presents guidance through which trial quantity can be selected in order to satisfy a target threshold of uncertainty. Both the statistical uncertainty and marginal gain at any given value of n can be used together to choose a number of trials that minimize uncertainty while still optimizing the time and resources dedicated to the experiment or testing procedure. The results presented here can guide researchers to understand the statistical scatter in experimental data for a given number of trials for various experimental conditions. Experiments were conducted with relatively thin samples (6 mm) and were not conducted with insulation as specified in standardized testing procedures, both of which may have influenced the degree of observed statistical scatter. Thus, while the statistical analysis outlined in this work is valid in other applications involving known population characteristics, the distribution of data presented in this work is only limited to the experimental conditions of this study. While the general trends can be used for insight into other experimental conditions, further experimentation is required to determine if the same degree of uncertainty can be expected for any value of n trials when using other materials (e.g., charring) than those used in these experiments. Adapting this analysis to other experiential conditions would, however, require an approximation of the predicted population characteristics (e.g., standard deviation and mean). Further work is required to develop similar analysis for finite samples.

The criteria for determining an appropriate number of trials depends on the experimental conditions (e.g., apparatus used, material considered, heat fluxes to be used), the acceptable range of uncertainty, and the practical limit of time and resources that can be dedicated to the experimental series. Therefore, it is difficult to establish a single benchmark criterion for a "statistically significant amount of trials". The number of trials conducted should be based on the goals of the researcher as opposed to precedent. The calculated uncertainty for a value of n can be used in combination with the marginal gain, but no single combination of criteria can capture the best practice for all applications. The selection of n can be guided by this analysis but is ultimately dependent on the acceptable degree of uncertainty required by the researcher. The reported uncertainty on experimental data must explicitly state the quantity of trials conducted to give context with respect to the possible statistical variation not accounted for.

Similar statistical analysis may be beneficial in other aspects of fire safety science. Often theory, codes and standards, and other procedures are developed from a limited number of experiments. The ability to quantify statistical uncertainty is possible for all scales of flammability assessment. This does not suggest that experiments and testing at all scales must be conducted in trial sets of 100, as this is an impractical use of time and resources. However, a better understanding of the uncertainty and limitations of experimental procedures in the field of fire safety may allow researchers to more readily design their experimental programs and interpret the results accurately.

References

[1] D. Drysdale, *An Introduction to Fire Dynamics: Third Edition*. 2011.

[2] J. Quintiere, *Fundamentals of Fire Phenomena*. Wiley, 2006.

[3] V. Babrauskas, *Ignition handbook*, vol. 318. Fire science publishers: Issaquah, WA, 2003.

[4] M. J. Hurley *et al.*, *SFPE handbook of fire protection engineering*. Springer, 2015.

[5] ISO 5560-1:2015+A1:2019, “Reaction-to-fire tests — Heat release, smoke production and mass loss rate,” 2019.

[6] ASTM E1321-13, “Standard Test Method for Determining Material Ignition and Flame Spread Properties,” 2013.

[7] ASTM E2058-13a, “Standard Test Method for Measurement of Material Flammability Using a Fire Propagation Apparatus (FPA),” 2013.

[8] ASTM E1354-17, “Standard Test Method for Heat and Visible Smoke Release Rates for Materials Products Using an Oxygen Consumption Calorimeter,” 2017.

[9] ASTM E2102-17, “Standard Test Methods for Measurement of Mass Loss and Ignitability for Screening Purposes Using a Conical Radiant Heater,” 2017.

[10] ASTM E906-17, “Standard Test Method for Heat and Visable Smoke Release Rates for Materials and Products Using a Thermopile Method,” 2017.

[11] ASTM E1623-16, “Standard Test Method for Determination of Fire and Thermal Parameters of Materials, Products, and Systems Using an Intermediate Scale Calorimeter (ICAL),” 2016.

[12] ASTM E648-19ae1, “Standard Test Method for Critical Radiant Flux of Floor-Covering Systems Using a Radiant Heat Energy Source,” 2019.

[13] ASTM E162-16, “Standard Test Method for Surface Flammability of Materials Using a Radiant Heat Energy Source,” 2016.

[14] ASTM E136-19a, “Standard Test Method for Behavior of Materials in a Vertical Tube Furnace at 750°C,” 2019.

[15] J. Urbas, “BDMC interlaboratory cone calorimeter test programme,” *Fire Mater.*, vol. 26, no. 1, pp. 29–35, 2002.

[16] J. Urbas, “Precision of the Cone Calorimeter and ICAL Test Methods,” in *Uncertainty in Fire Standards and What to Do About It*, ASTM International, 2012.

[17] J. Urbas, “Effects of retainer frame, irradiance level and specimen thickness on cone calorimeter test results,” *Fire Mater. An Int. J.*, vol. 29, no. 1, pp. 1–13, 2005.

[18] J. Luche, T. Rogaume, F. Richard, and E. Guillaume, “Characterization of thermal properties and analysis of combustion behavior of PMMA in a cone calorimeter,” *Fire Saf. J.*, vol. 46, no. 7, pp. 451–461, 2011.

[19] D. Morrisset, “The Effect of Orientation on the Ignition of Solids (M.S. Thesis),” 2020. https://doi.org/10.15368/theses.2020.95