scientific reports

OPEN

Evaluation method for asymmetric uncertainty of quantitative polymerase chain reaction measurements of deoxyribonucleic acids with low copy number

Unoh Ki✉, Takeru Suzuki, Satoshi Nakazawa, Yuuki Yonekawa, Kazuki Watanabe, Michie Hashimoto, Shigeo Hatada & Hirotaka Unno

Recently, in food safety and various other fields, qualitative and quantitative gene analysis using real-time polymerase chain reaction (PCR) method has become increasingly popular. The limit of detection (LOD) and quantifiable range for these measurements depends on the range and precision of DNA calibrators' concentrations. Low-copy-number nucleic acid reference materials with low uncertainty produced by an inkjet system have been developed to allow for precise measurements in a low-copy-number region. However, when using a calibrator with a low copy number near one, the copy number distribution is asymmetric. Consequently, the confidence intervals of estimated copy numbers can include negative values when conventional methods of uncertainty estimation are used. A negative confidence interval is irrelevant in the context of copy number, which is always positive value or zero. Here, we propose a method to evaluate the uncertainty of real-time PCR measurements with representative values and an asymmetric 95% confidence interval. Moreover, we use the proposed method for the actual calculation of uncertainty of real-time PCR measurement results for low-copy-number DNA samples and demonstrate that the proposed method can evaluate the precision of real-time PCR measurements more appropriately in a low-copy-number region.

The PCR method developed by Mullis in 1983[1], a technique to amplify small amounts of DNA for analysis, has been broadly applied in both science and engineering as well as in agriculture and medicine. In particular, the spread of the quantitative PCR method has greatly contributed to the development of various fields that require the quantitative analysis of trace amounts of nucleic acids, including the field of food inspection[2–6], such as genetically modified foods and microbiological analysis; the field of laboratory tests[7], such as cancer and infection; and the field of forensic medicine[8].

Because PCR has become increasingly popular worldwide, the appropriate quality control of PCR-based analytical methods has become necessary, and international standardization activities have been particularly active in recent years. ISO20395 states the importance of evaluating the limit of detection (LOD) for qualitative analysis of the target nucleic acid sequence or the limit of quantification (LOQ) for quantitative analysis as well as understanding the performance of the PCR system, including reagents and instruments, before analysis[9]. This can be interpreted to indicate that the performance of the reaction system should be assessed using calibrators (reference nucleic acid materials) covering a full range of concentrations to be quantified to obtain more reliable analysis results from quantitative real-time PCR[10,11] analysis with extremely small amounts of nucleic acids[12]. To meet such demand, we developed a new cell-based DNA reference material (hereinafter referred to as the new reference material) containing the target DNA sequence as low as one copy[13]. Note that "copy" is a unit generally used to express one DNA molecule containing a certain sequence.

The uncertainty of real-time PCR measurements is primarily composed of two components, calibrator uncertainty and measurement process uncertainty; both these components are assumed to follow a normal distribution[14,15]. Therefore, measurement results are represented by the mean DNA concentration of each sample and symmetrical measurement uncertainty. However, in an extremely low concentration range, quantitative

analysis has a problem that a confidence limit of DNA concentration can be a negative value. Therefore, methods to impart more probable uncertainty information are required. One solution is to measure the same sample enough number of times to reduce measurement uncertainty. However, in reality, it is desirable to have a method to evaluate the uncertainty information efficiently and reproducibly with at most 4–5 repeat measurements. This study reports that the inclusion of asymmetry in uncertainty can provide more reliable uncertainty information in an extremely low concentration range with one copy.

Results

In this study, we compared the new method of estimating asymmetric uncertainty with the conventional method of uncertainty estimation based on normal distribution. After real-time PCR quantification experiments, the uncertainty of quantification results was estimated with the two methods. To demonstrate the versatility of the new method, we created calibration curves with two different calibrators: one prepared by the conventional dilution method and another prepared with the new reference material.

Sources of the uncertainty of real-time PCR.
As mentioned above, the uncertainty of real-time PCR measurements has the following two major components: (1) uncertainty of calibrator copy number and (2) uncertainty from the measurement process. Generally, the calibrator uncertainty is combined with the uncertainty of the original solution's concentration used to prepare serial dilutions and the uncertainty of the measuring instrument used for dilution. After one step of dilution, the measuring instrument's uncertainty is from a total of two measurements, i.e., one from the original solution and another from the buffer used for dilution. It is desirable to select an appropriate range and avoid using an excessively small volume for measurement to reduce the uncertainty of the measuring instrument[12,14]. Furthermore, when low-copy-number samples are used, the variability in original solution sampling becomes significant because of an effect of the Poisson distribution[16] and is desirable to consider as the third source of uncertainty[12,17]. However, the new reference material was prepared using inkjet to dispense a cell suspension into a container to yield a specified copy number. We considered the specified number of copies to be dispensed as a measured value and calculated the uncertainty of measurements from the production variability and the number of samples used. In addition to the variability of the copy number of intracellular target DNA and the variability because of contamination, which have been evaluated in the previous study[13], cell aggregation was identified to be another source of the production variability. To increase the production efficiency, two avalanche photodiodes (APD410A/M, Thorlabs; hereinafter referred to as APD) placed oppositely were used to detect any cells in flying droplets (Figure S1 in Supplementary Information File 1). APDs convert detected fluorescent energies of cells into voltage values. When output voltage value exceeded the threshold (set to 0.2 V), it was considered that a cell was detected. Count errors occur depending on the detection rate (probability of detecting droplets containing a particle) and false detection rate (probability of detecting particles that are not cells) of APD as well as the possibility that one droplet contains two or more cells (dependent on the cell suspension concentration). Therefore, the variation because of APD count errors was calculated. The term DNA below always refers to DNA molecules containing the target sequence.

The conventional method to calculate the uncertainty of concentrations in a serial dilution.
We calculated the symmetric uncertainty of the concentrations of serially diluted, certified DNA reference material as described in the Guide to the Expression of Uncertainty in Measurement[18] (Sect. Determining combined standard uncertainty), briefly introduced in Supplementary Information File 2. Table 1 shows all variables and parameters used in this paper.

The concentration C_{after} (unit: μL^{-1}) after each dilution step is described using the following equation:

$$C_{after} = \frac{P \cdot C_{before} + E}{P + Q} \tag{1}$$

In this equation, C_{before} is the DNA concentration of the original solution at each step (unit: μL^{-1}); P and Q are volumes of the original solution and the buffer (unit: μL), respectively; and E is the error of DNA copy number. The probability mass function of E was defined by $P_E(e) = \Pr(E = e) = \Pr(N_P = e + \lambda)$, where e is a specific value of E, and N_P denotes the DNA copy number in the P volume of original solution. N_P is generally assumed to follow the Poisson distribution with an expectation $\lambda = p \cdot c_{before}$, where p and c_{before} are estimates of the expectations of P and C_{before}, respectively. Therefore, the expectation of E is zero. Note that because dilution was performed serially, C_{before} at each step is equivalent to C_{after} at the previous step. Equation (1) corresponds to f in equation (S1) in Supplementary Information File 2.

According to the combined uncertainty formula using the sensitivity coefficient of the Guide to the Expression of Uncertainty in Measurement[18], based on equation (S1), the following equation was used to calculate the uncertainty $u(c_{after})$ of a concentration after dilution:

$$
\begin{aligned}
u(c_{after}) &= \sqrt{\left[c_1 \cdot u(c_{before})\right]^2 + \left[c_2 \cdot u(p)\right]^2 + \left[c_3 \cdot u(q)\right]^2 + \left[c_4 \cdot u(e)\right]^2} \\
&= \sqrt{\left[\frac{p}{p+q} \cdot u(c_{after})\right]^2 + \left[\frac{q}{(p+q)^2} c_{before} \cdot u(p)\right]^2 + \left[-\frac{p}{(p+q)^2} c_{before} \cdot u(q)\right]^2 + \left[\frac{1}{p+q} \cdot u(e)\right]^2}
\end{aligned} \tag{2}
$$

In this equation, c_{after} and q are estimates of expectations of C_{after} and Q; $u(c_{before})$, $u(p)$, $u(q)$, and $u(e)$ are the uncertainties of their respective variables; and c_1, c_2, c_3, and c_4 are sensitivity coefficients of C_{before}, P, Q, and E, respectively. Because p and q can only be measured once, the only component of $u(p)$ and $u(q)$ is the uncertainty of the pipette. Therefore $u(p)$ and $u(q)$ are obtained directly from the pipette calibration certificate.

Variable	Definition	Equation no.
C_{after} (unit: μL^{-1})	DNA concentration of diluted solution after each dilution step	(1)
C_{before} (unit: μL^{-1})	DNA concentration of the original solution at each dilution step	(1)
P (unit: μL)	Volume of the original solution at each dilution step	(1)
Q (unit: μL)	Volume of the buffer added at each dilution step	(1)
E	Error of DNA copy number in the P volume of the original solution due to the Poisson distribution at each dilution step	(1)
N_P	DNA copy number in the P volume of the original solution at each dilution step	
e	Specific value of E	(2)
c_{after} (unit: μL^{-1})	Estimate of the expectation of C_{after}	(2)
c_{before} (unit: μL^{-1})	Estimate of the expectation of C_{before}	(2)
p (unit: μL)	Estimate of the expectation of P	(2)
q (unit: μL)	Estimate of the expectation of Q	(2)
$u(c_{after})$	Symmetric uncertainty of c_{after}	(2)
$u(c_{before})$	Symmetric uncertainty of c_{before}	(2)
$u(p)$	Symmetric uncertainty of p	(2)
$u(q)$	Symmetric uncertainty of q	(2)
$u(e)$	Symmetric uncertainty of e	(2)
c_1	Sensitivity coefficient of C_{before}	(2), (5) and (6)
c_2	Sensitivity coefficient of P	(2), (5) and (6)
c_3	Sensitivity coefficient of Q	(2), (5) and (6)
c_4	Sensitivity coefficient of E	(2), (5) and (6)
N_{dil}	DNA copy number of a calibrator prepared with dilution series in a well	
$\overline{N_{dil}}$	Mean of N_{dil} in each well	(3)
$\overline{n_{dil}}$	Estimate of the expectation of $\overline{N_{dil}}$	
C_{final} (unit: μL^{-1})	DNA concentration of a diluted solution used to prepare a calibrator	(3)
P_{final} (unit: μL)	Volume of a diluted solution dispensed in each well to prepare a calibrator	(3)
E_{final}	Error of DNA copy number of a calibrator prepared with dilution series due to the Poisson distribution	
e_{final}	Specific value of E_{final}	
$\overline{E_{final}}$	Mean of E_{final} in each well	(3)
c_{final} (unit: μL^{-1})	Estimate of the expectation of C_{final}	(4)
p_{final} (unit: μL)	Estimate of the expectation of P_{final}	(4)
$\overline{e_{final}}$	Estimate of the expectation of $\overline{E_{final}}$	(4)
$u(\overline{n_{dil}})$	Symmetric uncertainty of $\overline{n_{dil}}$	(4)
$u(c_{final})$	Symmetric uncertainty of c_{final}	(4)
$u(p_{final})$	Symmetric uncertainty of p_{final}	(4)
$u(\overline{e_{final}})$	Symmetric uncertainty of $\overline{e_{final}}$	(4)
c_5	Sensitivity coefficient of C_{final}	(4)
c_6	Sensitivity coefficient of P_{final}	(4)
c_7	Sensitivity coefficient of $\overline{E_{final}}$	(4)
N_{rw}	Effective number of wells containing the same specific calibrator	
A_{c_after}, B_{c_after}	Left- and right-side confidence intervals for the expectation of C_{after}	(5) and (6)
$A_{c\_before}, B_{c\_before}$	Left- and right-side confidence intervals for the expectation of C_{before}	(5) and (6)
A_p, B_p	Left- and right-side confidence intervals for the expectation of P	(5) and (6)
A_q, B_q	Left- and right-side confidence intervals for the expectation of Q	(5) and (6)
A_e, B_e	Left and right intervals for representing the asymmetric variety of E	(5) and (6)
$A_{\overline{n\_dil}}, B_{\overline{n\_dil}}$	Left- and right-side confidence intervals for the expectation of $\overline{n_{dil}}$	(7) and (8)
A_{c_final}, B_{c_final}	Left- and right-side confidence intervals for the expectation of C_{final}	(7) and (8)
A_{p_final}, B_{p_final}	Left- and right-side confidence intervals for the expectation of P_{final}	(7) and (8)
$A_{\overline{e\_final}}, B_{\overline{e\_final}}$	Left- and right-side confidence intervals for the expectation of $\overline{E_{final}}$	(7) and (8)
P_{agg}	Cell aggregation rate	
N_{agg}	Number of masses of cell aggregate containing ≥ 2 cells when measuring P_{agg}	
N_{sgl}	Number of independent cell when measuring P_{agg}	
P_D	Detection rate of the APDs corresponding to detect a fluorescent particle (either a cell of a fluorescent contaminant) in a droplet	(9) and (11)
P_{FD}	False detection rate of the APDs corresponding to detect a fluorescent contaminant in a droplet	(10)
Continued		

Variable	Definition	Equation no.
N_1	Number of droplets containing one cell aggregate (both a independent cell and a cell aggregate containing 2 cells are referred to as a cell aggregate) when measuring P_D and P_{FD}	(9) and (10)
N_{APD_0}	Number of times the APDs detected droplets containing 0 cell aggregates	(9) and (10)
N_{APD_1}	Number of times the APDs detected droplets containing 1 cell aggregate	(9) and (10)
λ_{pa}	Mean fluorescent particle number in each droplet	(11)
P_{pa_0}	Proportion of droplets containing no fluorescent particle	(11)
N_{pos}	Number of droplets in which APD detected particles when measuring λ_{pa}	(11)
N_{all}	Total number of droplets when measuring λ_{pa}	(11)
P_{con}	Probability of well contamination by DNA	
P_{intra_2}	Proportion of single cells with an intracellular DNA copy number of two	
$\widehat{P}_{pa\_0}$	Worst value of P_{pa_0}	
$N_{pa\_droplet}$	Number of fluorescent particles in a droplet	(12)
N_{pa}	Number of particles dispensed between two consecutive APD detections	(12)
n_{pa}	Specific value of N_{pa}	(12)
$\widehat{P}_D$	Worst value of P_D	(12)
$\widehat{\lambda}_{pa}$	Worst value of λ_{pa}	
N_{new}	DNA copy number of a calibrator prepared with the new reference material in each well	
$\overline{N_{new}}$	Mean of N_{new} in each well	(13)
N_{APD}	Predetermined number of APD detections in each well set in the inkjet system	(13)
$\overline{N_{pa}}$	Mean of N_{pa}	(13)
$\overline{N_{agg\_pa}}$	Mean number of cell aggregates in a fluorescent particle	(13)
$\overline{N_{sgl\_agg}}$	Mean number of cells in a cell aggregate	(13)
$\overline{N_{intra}}$	Mean intracellular DNA copy number	(13)
$\overline{N_{con}}$	Mean contaminant DNA copy number in a well from reagents and environment	(13)
$\overline{n_{new}}$	Estimate of the expectation of $\overline{N_{new}}$	(14)
$u(\overline{n_{new}})$	Symmetric uncertainty of $\overline{n_{new}}$	(14)
σ_{N_new}	Standard deviation of N_{new}	(14)
A_{n_new}, B_{n_new}	Left and right intervals for representing the asymmetric variety of N_{new}	(15) and (16)
$A_{\overline{n\_new}}, B_{\overline{n\_new}}$	Left and right confidence intervals for the expectation of $\overline{N_{new}}$	(15) and (16)
$u(e_{curve})$	Symmetric uncertainty of a calibration curve in real-time PCR process	(17)
e_{curve}	Estimation of the error of the mean quantification result sourced from the calibration curve	(17)
$\overline{n_Q}$	Mean PCR quantification result of replicates of target samples	(17)
$N_{rw,i}$	Effective number of wells containing calibrators with the same i-th representative DNA copy number	(17)
$\overline{n_{r,i}}$	Estimate of mean DNA copy number of calibrators with the same i-th representative DNA copy number	(17)
$R_{\overline{n\_r,i}}$	Representative value of $\overline{n_{r,i}}$	(17)
$u\left(\overline{n_{r,i}}\right)$	Symmetric uncertainty of $\overline{n_{r,i}}$	(17)
$s_{n\_rQ,i}$	Corrected sample standard deviation of PCR quantification results of replicates of calibrators with the same i-th representative DNA copy number	(17)
$\overline{n_{rQ,i}}$	Mean of PCR quantification results of replicates of calibrators with the same i-th representative DNA copy number	(17)
U_{conv}	Symmetric expanded uncertainty of the mean DNA copy number of target samples in each well	(18)
s_{n_Q}	Corrected sample standard deviation of real-time PCR quantification results of replicates of target samples	(18)
N_w	Number of wells containing replicates of target samples including those failed in amplification (undetermined)	(18)
$t_{0.05,N_w-1}$	Two-tailed t critical value	(18)
X	Lognormal distributed variable, corresponding to the PCR quantification result of either a calibrator or a target sample	
$\overline{X}$	Mean of X	(19) and (20)
s_X	Corrected sample standard deviation of X	(19) and (20)
Y	Natural logarithm of X	
s_Y	Corrected sample standard deviation of Y	(19)
$\overline{Y}$	Mean of Y	(20)
$A_{\overline{X}}, B_{\overline{X}}$	Left- and right-side confidence intervals for $\overline{X}$	(21) and (22)
M_X	Median of X	(21–23)
A_{e_curve}, B_{e_curve}	Left- and right-side confidence intervals for the expectation of e_{curve}	(24) and (25)
M_Q	Estimated median of PCR quantification results of replicates of target samples	(24) and (25)
Continued		

Variable	Definition	Equation no.
$A_{\overline{n\_r,i}}, B_{\overline{n\_r,i}}$	Left- and right-side confidence intervals for the expectation of $\overline{n_{r,i}}$	(24) and (25)
$M_{rQ,i}$	Estimated median of PCR quantification results of replicates of calibrators with the same i-th representative DNA copy number	(24) and (25)
$A_{\overline{n\_rQ,i}}, B_{\overline{n\_rQ,i}}$	Left- and right-side confidence intervals for the expectation of $\overline{n_{rQ,i}}$	(24) and (25)
A_{prop}, B_{prop}	Left and right confidence intervals for the expectation of the mean DNA copy number of target samples	(26) and (27)
A_Q, B_Q	Left and right confidence intervals of the mean PCR quantification result of replicates of target samples	(26) and (27)

Table 1. List of variables and parameters.

At each dilution step, $u(c_{before})$ equals $u(c_{after})$ calculated in the previous step. The volume of the original solution was dispensed once at each step. Therefore, $u(e)$ was the standard deviation of the Poisson distribution $\sigma_{Poisson} = \sqrt{p \cdot C_{before}}$. Note that $u(c_{before})$ at the first step is the uncertainty of the certified reference material, which is 9.75×10^7 μL^{-1} according to the manufacturer's certificate.

We used these calculations to determine the concentration uncertainties of a series of diluted DNA solutions.

A conventional method to calculate the uncertainty of DNA copy number prepared by dispensing a dilution series. We calculated the symmetric uncertainty of the mean DNA copy number in calibrators prepared by dispensing a diluted solution of a specific concentration into each well. After preparing the dilution series, 4 μL of each diluted solution was dispensed into a 96-well plate with a pipette. Therefore, the mean final copy number N_{dil} in each well, $\overline{N_{dil}}$, is described using the following equation, corresponding to f in equation (S1):

$$\overline{N_{dil}} = P_{final} \cdot C_{final} + \overline{E_{final}} \tag{3}$$

In this equation, C_{final} (unit: μL^{-1}) is the concentration of the diluted solution in a container to be dispensed into each well; P_{final} (unit: μL) is the volume of solution dispensed in each well, set to 4 μL; and $\overline{E_{final}}$ is the mean error of DNA copy number E_{final} in each well. The probability mass function of E_{final} was defined by $P_{E\_final}(e_{final})$ $= \Pr(E_{final} = e_{final}) = \Pr(N_{dil} = e_{final} + \lambda_{final})$, where e_{final} is a specific value of E_{final}, and N_{dil} follows the Poisson distribution with an expectation $\lambda_{final} = p_{final} \cdot c_{final}$. Here, p_{final} and c_{final} are estimates of the expectations of P_{final} and C_{final}, respectively. Moreover, p_{final} was much smaller than the remaining volume of the diluted solution in the container. Thus, c_{final} was assumed to remain the same throughout dispensing. Furthermore, dispensing was performed independently for each well. Therefore, N_{dil} in each well was assumed to independently and identically follow the Poisson distribution with the same λ_{final}. Note that c_{final} corresponds to c_{after} calculated at a certain dilution step.

Based on equation (S1), the uncertainty of the mean of the final copy number in each well $u(\overline{n_{dil}})$ was calculated using the following equation:

$$u(\overline{n_{dil}}) = \sqrt{\left[c_5 \cdot u\left(c_{final}\right)\right]^2 + \left[c_6 \cdot u\left(p_{final}\right)\right]^2 + \left[c_7 \cdot u\left(\overline{e_{final}}\right)\right]^2}$$
$$= \sqrt{\left[p_{final} \cdot u\left(c_{final}\right)\right]^2 + \left[c_{final} \cdot u\left(p_{final}\right)\right]^2 + u^2\left(\overline{e_{final}}\right)} \tag{4}$$

In this equation, $\overline{n_{dil}}$ and $\overline{e_{final}}$ are estimates of expectations of $\overline{N_{dil}}$ and $\overline{E_{final}}$, respectively; $u(c_{final})$, $u(p_{final})$, and $u(\overline{e_{final}})$ are uncertainties of their respective variables, respectively; and c_5, c_6, and c_7 are sensitivity coefficients of C_{final}, P_{final}, and E_{final}, respectively. The only component of $u(p_{final})$ was the uncertainty of the pipette, which can be obtained directly from the calibration certificate. Moreover, the standard error of the mean $\overline{e_{final}}$, $\sigma_{e\_final} = \sqrt{(p_{final} \cdot c_{final})/N_{rw}}$, corresponds to $u(\overline{e_{final}})$. Here, N_{rw} is the effective number of wells containing the same specific calibrator. The number of wells that failed to amplify is excluded. An alternative model for calculating $u(\overline{n_{dil}})$ is introduced in Supplementary Information File 2 and assumes that c_{final} and p_{final} were correlated in all wells.

We used these calculations to determine the uncertainties of the mean DNA copy number of calibrators prepared by dispensing 4 μL of diluted solution at different concentrations into each well, $u(\overline{n_{dil}})$. A representative value is defined in this paper to represent the measurement results of a variable, generally expressed as a mean. The representative value of the mean copy number $\overline{n_{dil}}$ was $p_{final} \cdot c_{final}$.

A proposed method to calculate the uncertainty of a serial dilution. We calculated asymmetric uncertainties of concentration in serial dilutions and the mean DNA copy number of calibrators prepared by dispensing diluted solution into independent wells, compared to the symmetric uncertainty calculated above.

First, we used an asymmetric 95% confidence interval for each component to estimate the uncertainty of C_{after}. We assumed an asymmetric 95% confidence interval could be denoted by $[R - A, R + B]$. R denotes the representative value. A and B denote the left-side confidence interval (the interval between the representative value and the lower confidence limit) and the right-side confidence interval (the interval between the representative value and the upper confidence limit). Therefore, A and B represent the asymmetric expanded uncertainty of the variable. Because it was not considered proper to calculate expanded uncertainty using a cover factor when a variable is not normally distributed, we estimated the expanded uncertainty of each variable separately.

The left- and right-side confidence intervals for the expectation of C_{after}, A_{c_after} and B_{c_after}, were calculated separately based on Eq. (2), using the following equations:

$$A_{c\_after} = \sqrt{\left(c_1 \cdot A_{c\_before}\right)^2 + \left(c_2 \cdot A_p\right)^2 + \left(c_3 \cdot A_q\right)^2 + \left(c_4 \cdot A_e\right)^2} \tag{5}$$

$$B_{c\_after} = \sqrt{\left(c_1 \cdot B_{c\_before}\right)^2 + \left(c_2 \cdot B_p\right)^2 + \left(c_3 \cdot B_q\right)^2 + \left(c_4 \cdot B_e\right)^2} \tag{6}$$

In these equations, A_{c_before} and B_{c_before} denote left- and right-side confidence intervals for the expectation of C_{before}; A_p and B_p denote left- and right-side confidence intervals for the expectation of P; A_q and B_q denote left- and right-side confidence intervals for the expectation of Q; and A_e and B_e denote left- and right-side intervals representing the asymmetric variety of E. At each dilution step, A_{c_before} and B_{c_before} are A_{c_after} and B_{c_after} of the previous step. The concentration of the certified reference material was assumed to follow a normal distribution. Therefore, A_{c_before} and B_{c_before} at the first dilution step were approximated to be twice the uncertainty obtained from the manufacturer's certificate. In addition, the volume of the solution measured and dispensed by a pipette is also generally assumed to follow a normal distribution. Thus, A_p and B_p, and A_q and B_q were approximated with $2u(p)$ and $2u(q)$, respectively. Furthermore, A_e and B_e were asymmetrically evaluated as shown in Figure S2 in Supplementary Information File 1. The probability mass function of E was $P_E(e)$ defined above. The representative value of e was the expectation 0, and the initial value of A_e and B_e defined in Figure S2 were both 0. We then compared the probability that e would fall within the interval $[-A_e, B_e]$ when A_e was increased by a step value of 1 (because e is an integer) with that when B_e was increased by 1 and updated the interval to the one with a higher probability. This process was repeated until the probability reached or exceeded 95%. Note that the representative value of c_{after} was the estimate of the expectation, which was equal to $\frac{p \cdot c_{before}}{p+q}$.

As the asymmetric expanded uncertainty of each serially diluted solution was calculated, we then calculated the asymmetric expanded uncertainty for the expectation of the mean DNA copy number of calibrators in each well prepared by dispensing a specific diluted solution. The left- and right-side confidence intervals for the expectation of $\overline{N_{dil}}$, $A_{\overline{n\_dil}}$ and $B_{\overline{n\_dil}}$, were calculated separately based on Eq. (4) using the following equations:

$$A_{\overline{n\_dil}} = \sqrt{\left(p_{final} \cdot A_{c\_final}\right)^2 + \left(c_{final} \cdot A_{p\_final}\right)^2 + A^2_{e\_final}} \tag{7}$$

$$B_{\overline{n\_dil}} = \sqrt{\left(p_{final} \cdot A_{c\_final}\right)^2 + \left(c_{final} \cdot B_{p\_final}\right)^2 + B^2_{e\_final}} \tag{8}$$

In these equations, A_{c_final} and B_{c_final} denote left- and right-side confidence intervals for the expectation of C_{final}; A_{p_final} and B_{p_final} denote left- and right-side confidence intervals for the expectation of P_{final}; and $A_{\overline{e\_final}}$ and $B_{\overline{e\_final}}$ denote left- and right-side confidence intervals for the expectation of $\overline{E_{final}}$. A_{c_final} and B_{c_final} correspond to A_{c_after} and B_{c_after} of the specific diluted solution. Moreover, p_{final} is generally assumed to follow a normal distribution. Therefore, A_{p_final} and B_{p_final} were approximated with $2u(p_{final})$ and $A_{\overline{e\_final}}$ and $B_{\overline{e\_final}}$ were approximated with $A_e/\sqrt{N_{rw}}$ and $B_e/\sqrt{N_{rw}}$. The approximation of $A_{\overline{e\_final}}$ and $B_{\overline{e\_final}}$ was similar to the calculation for the standard error of the mean; however, the right and left sides were calculated separately. Note that the representative value $\overline{n_{dil}}$ is the mean DNA copy number in each well equal to $p_{final} \cdot C_{final}$.

We thus obtained asymmetric expanded uncertainties of mean DNA copy number of calibrators prepared by dispensing 4 μL of different concentrations into each well separately. The asymmetric expanded uncertainties were evaluated as $A_{\overline{n\_dil}}$ and $B_{\overline{n\_dil}}$, compared to symmetric uncertainties evaluated as $u_{\overline{n\_dil}}$. Note that symmetric uncertainties were converted to expanded uncertainties in the later calculations of uncertainties in the PCR process.

Measurement of the production variability of the new reference material.

Compared with a calibrator prepared by serial dilution, we calculated the uncertainty when the new reference material was used as a calibrator. To obtain the new reference material's production variability and uncertainty more precisely, we measured distributions of different influential factors of the production variability. We then used these distributions to estimate the DNA copy number distribution of the calibrator. First, we will describe the measurement of distributions of factors influencing production variability.

The cell aggregation rate in the cell suspension to be dispensed by inkjet was measured using Countess Cell Counting Chamber Slides (Thermo Fisher Scientific) and a microscope (Axio Observer.D1, Carl Zeiss). A 10-μL-sized cell suspension was dispensed into one of the chambers of a slide, and cells in the chamber were imaged using a microscope. Then, the number of masses of aggregated cells, N_{agg}, and the number of independent cells, N_{sgl}, in all images were counted by visual inspection. The aggregation rate P_{agg} was defined as the ratio of N_{agg} to the sum of N_{agg} and N_{sgl}.

The intracellular DNA copy number distribution was measured in the same manner as in the previous study[13], although a flow cytometer (MA900, Sony) was used for the measurement. In flow cytometry, a cell containing two nuclei cannot be distinguished from an aggregate of two cells containing one nucleus per cell. Therefore, the aggregation rate was subtracted from the proportion of cells with two copies of DNA measured using flow cytometry. This aggregation rate had a different value from the above-mentioned aggregation rate because different processes were used to prepare cell suspensions.

The APD's detection rate and false detection rate were obtained as follows. A cell suspension was ejected onto an SD00011 slide glass (Matsunami Glass Ind.) by inkjet, and flying droplets were measured using two

APDs. Then, the droplets that landed were observed using a microscope to count cell aggregates in each droplet (for convenience of explanation, hereinafter an independent cell is also referred to as an aggregate). To acquire a sufficient number of data, the concentration of particles in a droplet was adjusted to 0.566 (equivalent to 1.48×10^6 mL^{-1}). For actual production, the concentration was lower (≤ 0.1); thus, the probability of 2 or more cell aggregates detected in a droplet was very low ($\leq 5\%$). The total fluorescence intensity increased when the droplet contains two or more cell aggregates, and thus the detection rate of APD increases. Therefore, only droplets containing ≤ 1 cell were evaluated in this experiment. The detection rate P_D and the false detection rate P_{FD} were defined using the following equations:

$$P_D = \frac{N_{APD\_1} + N_{APD\_0}}{N_1 + N_{APD\_0}} \tag{9}$$

$$P_{FD} = \frac{N_{APD\_0}}{N_1 + N_{APD\_0}} \tag{10}$$

In these equations, N_1 was the number of droplets containing one cell aggregate, and N_{APD_0} and N_{APD_1} are the number of times the APDs detected droplets containing 0 and 1 cell aggregate, respectively.

False detection events occur primarily because of contamination by fluorescent substances during cell suspension preparation. Therefore, $1 - P_{FD}$ represents the proportion of cell aggregates among all fluorescent particles in the cell suspension.

The mean particle number in each droplet was calculated based on the APD measurements of droplets ejected by the inkjet system. The number of particles in each droplet $N_{pa\_droplet}$ (both cell aggregate or fluorescent contaminant count) is generally assumed to follow the Poisson distribution, so the expectation λ_{pa} is equivalent to the mean particle number. Thus, the proportion of droplets containing no particle P_{pa_0} was an estimate of the Poisson distribution probability when $N_{pa\_droplet} = 0$. The probability is equal to $\exp(-\lambda_{pa})$ based on the definition of the Poisson distribution[19]. Therefore, λ_{pa} was calculated using the following equation:

$$\lambda_{pa} = -\ln P_{pa\_0} = -\ln\left(1 - \frac{N_{pos}}{N_{all}} \cdot \frac{1}{P_D}\right) \tag{11}$$

In this equation, N_{pos} is the number of droplets in which APD detected particles, and N_{all} is the total number of droplets. The detection rate of APD was also considered in the calculation of P_{pa_0}.

Moreover, we measured contamination by DNA containing the target sequence from reagents and the environment. First, 16 wells of an empty 96-well plate were covered with a seal as negative controls. The plate was then left in the inkjet device without dispensing droplets for the same length of time as normal production. After the seal was removed, cell wall lysis solution was added to all wells as in normal production, and the amplification reagent was added to measure the DNA copy number using real-time PCR. The probability of well contamination by DNA, P_{con}, was then calculated as the ratio of the number of positive wells out of 80 wells that were not covered using a seal.

Moreover, to consider each component's measurement uncertainty in the calculation, the mean level for each component was set to the worst value in which the uncertainty was considered as follows. First, the measurement of the three components, the intracellular DNA copy number, and the detection and false detection rates of APD, are Bernoulli trials, which follow the Bernoulli distribution[20,21]. Experientially, the aggregates of cells used were mostly composed of two cells, and the contaminant DNA copy number was ≤ 1 in most cases; therefore, the Bernoulli trial was used to approximate these components. Measuring the mean particle number in each droplet follows the Bernoulli distribution and depends on the binary question of whether or not a droplet contains any particles. The parameters of the six Bernoulli distributions described above were P_{intra_2} denoting the proportion of single cells with an intracellular DNA copy number of 2, the detection rate P_D, the false detection rate P_{FD}, the aggregation rate P_{agg}, the probability of DNA contamination P_{con}, and the proportion of droplets with no particles P_{pa_0}. The confidence interval of each of the six parameters was estimated using the Wilson score interval with continuity correction[22]. Finally, the worst values of P_{intra_2}, P_{FD}, P_{agg}, and P_{con} were set to the corresponding upper confidence limits. The worst values of P_D and P_{pa_0} were set to the corresponding lower confidence limits. Therefore, the worst value of λ_{pa} was calculated as $-\ln \widehat{P}_{pa\_0}$ based on Eq. (11), where $\widehat{P}_{pa\_0}$ denote the worst value of P_{pa_0}.

Next, we calculated the distribution of the number of particles dispensed between two consecutive APD detections. After each APD detection, several droplets are usually dispensed before another droplet containing fluorescent particles are detected. Because the detection rate is not 100%, there is a small probability that some particles are present in the undetected droplets. The following equation gives the probability mass function of the total number of particles N_{pa} in the detected droplet and undetected droplets $P_{N\_pa}(n_{pa})$:

$$P_{N\_pa}(n_{pa}) = Pr(N_{pa} = n_{pa}) = \sum_{i=1}^{n_{pa}} \left[\left(1 - \widehat{P}_D\right)^{i-1} \cdot \widehat{P}_D \cdot \frac{Pr(N_{pa\_droplet} = n_{pa} - (i-1))}{1 - Pr(N_{pa\_droplet} = 0)} \right] \tag{12}$$

In this equation, n_{pa} is a specific value of N_{pa}, and $\widehat{P}_D$ is the worst value of P_D calculated above. $N_{pa\_droplet}$ was assumed to follow the Poisson distribution with an expectation $\widehat{\lambda}_{pa}$, where $\widehat{\lambda}_{pa}$ is the worst value of λ_{pa} calculated by using Eq. (11). Experientially, the droplets not detected by APD never contained ≥ 2 particles. Therefore, the number of particles in these droplets was assumed to be 1. Thus, if $n_{pa} - (i-1)$ particles were contained in the detected droplet, then $(i-1)$ droplets with a single particle were undetected. The item $\frac{Pr(N_{pa\_droplet} = n_{pa} - (i-1))}{1 - Pr(N_{pa\_droplet} = 0)}$ corresponds to the probability that $n_{pa} - (i-1)$ particles exist in the detected droplet. Note that the number of fluorescent contaminants was included in n_{pa}. The influence of fluorescent contaminants is considered next.

Using the above measurement and calculations, we obtained the distributions of the number of particles dispensed between two consecutive APD detections N_{pa}, the number of cell aggregates in a fluorescent particle determined by P_{FD}, the number of cells in an aggregate determined by P_{agg}, the copy number of intracellular DNA in a cell, and the number of DNA contaminants from reagents and the environment determined by P_{con}.

Uncertainty of the new reference material. We combined distributions of factors influencing production variability to calculate the uncertainty of the new reference material.

First, we calculated the probability mass function of the DNA copy number of the calibrator prepared with the new reference material, N_{new}. The mean value of N_{new} in each well, $\overline{N_{new}}$, is given by the following equation:

$$\overline{N_{new}} = N_{APD} \cdot \left[\left(\overline{N_{pa}} \cdot \overline{N_{agg\_pa}} \right) \cdot \left(\overline{N_{sgl\_agg}} \cdot \overline{N_{intra}} \right) \right] + \overline{N_{con}} \tag{13}$$

In this equation, the constant N_{APD} is the predetermined number of APD detections in each well set in the inkjet system, $\overline{N_{pa}}$ is the mean of N_{pa}, $\overline{N_{agg\_pa}}$ is mean number of cell aggregates in a particle, $\overline{N_{sgl\_agg}}$ is the mean number of cells in a cell aggregate, $\overline{N_{intra}}$ is the mean intracellular DNA copy number, and $\overline{N_{con}}$ is the mean contaminant DNA copy number in a well from reagents and environment. Based on the relationship between N_{new} and factors influencing production variability indicated in Eq. (13), we combined the distributions of these factors and obtained the probability mass function of N_{new}. The method for this calculation is detailed in Supplementary Information File 2.

Next, we calculated the symmetric uncertainty of the mean DNA copy number of the calibrator in each well. The symmetric uncertainty $u(\overline{n_{new}})$ is the standard error of the mean given by the following equation:

$$u(\overline{n_{new}}) = \sigma_{N\_new} / \sqrt{N_{rw}} \tag{14}$$

In this equation, $\overline{n_{new}}$ is an estimate of the expectation of $\overline{N_{new}}$, σ_{N_new} is the standard deviation of N_{new}, from equation (S12) in Supplementary Information File 2, and N_{rw} is the effective number of wells containing the same calibrator.

Finally, as a comparison, we calculated asymmetric expanded uncertainties of the mean DNA copy number of the calibrator in each well. First, left and right intervals representing the asymmetric variety of N_{new}, A_{n_new} and B_{n_new}, were asymmetrically calculated (Figure S2). The representative value used in the calculation was the median of N_{new}. We considered the median to be a more appropriate representation of an asymmetrically distributed variable, because the median is closer to the peak of the distribution. Note that this interval was an integer because N_{new} is an integer in principle. Therefore, the interval was symmetric even when the distribution of N_{new} was asymmetric. Finally, the asymmetric expanded uncertainties for $\overline{N_{new}}$, represented by the left and right confidence intervals $A_{\overline{n\_new}}$ and $B_{\overline{n\_new}}$, were calculated separately (Eq. 14), using the following equations:

$$A_{\overline{n\_new}} = A_{n\_new} / \sqrt{N_{rw}} \tag{15}$$

$$B_{\overline{n\_new}} = B_{n\_new} / \sqrt{N_{rw}} \tag{16}$$

We thus obtained the symmetric uncertainty $u(\overline{n_{new}})$, and asymmetric expanded uncertainties $A_{\overline{N\_new}}$ and $B_{\overline{N\_new}}$ of the mean DNA copy number of the calibrator prepared with new reference material in each well. These uncertainties were compared with those of calibrators prepared by serial dilution. Note that the median of N_{new} was selected to be the representative value of $\overline{N_{new}}$.

Uncertainty of real-time PCR measurement results. Here we describe the calculation of symmetric and asymmetric expanded uncertainties of the mean DNA copy number of target samples in each well quantified by real-time PCR. Relative quantification was performed by plotting calibration curves using calibrators with different representative DNA copy numbers, prepared by serial dilution or new reference material. Quantification with a calibration curve is described in Supplementary Information File 2.

The combined uncertainty was calculated from the mean quantification result and pooled relative uncertainties of several components in a previous study[14]. Here we used the same approach to combine the relative uncertainties of several components. However, the uncertainty of the mean quantification result of DNA copy number of target samples to be quantified was considered to have two major components. The first is the uncertainty of the calibration curve, and the second is the variance between target sample replicates. The calibration curve uncertainty was proposed to have two components: the uncertainty of calibrators prepared by serial dilution or new reference material and the variance in the results of calibrators with the same representative DNA copy number quantified by PCR. Note that DNA copy number of replicates of calibrators were quantified by using the calibration curve plotted by themselves to evaluate the PCR amplification variability of calibrators.

The symmetric expanded uncertainty of the mean DNA copy number of target samples in each well was calculated as follows. First, based on the calculation of pooled relative uncertainty introduced in Eqs. (3) and (4) in the previous study[14], the following equation gives the uncertainty of a calibration curve $u(e_{curve})$:

$$u(e_{curve}) = \overline{n_Q} \sqrt{ \frac{ \sum_{i=1}^{k} (N_{rw,i}-1) \left[\left(\frac{u(\overline{n_{r,i}})}{\overline{n}_{n\_r,i}} \right)^2 + \left(\frac{s_{n\_rQ,i} / \sqrt{N_{rw,i}}}{\overline{n}_{rQ,i}} \right)^2 \right] }{ \sum_{i=1}^{k} (N_{rw,i}-1) } } \tag{17}$$

In this equation, e_{curve} denotes an estimation of the error of the mean quantification result sourced from the calibration curve, $\overline{n_Q}$ is the mean PCR quantification result of target sample replicates; $N_{rw,i}$ is the effective number

of wells containing calibrators with the same i-th representative DNA copy number; $R_{\overline{n\_r,i}}$ is the representative value of an estimate of the mean DNA copy number of these calibrators $\overline{n_{r,i}}$ defined in previous sub-sections; $u(\overline{n_{r,i}})$ is the symmetric uncertainty of $\overline{n_{r,i}}$ calculated in previous sub-sections; and $s_{n\_rQ,i}$ and $\overline{n_{rQ,i}}$ are the corrected sample standard deviation and mean PCR quantification results for calibrator replicates. Finally, based on equations (S1) and (S3) in Supplementary Information File 2, the symmetric expanded uncertainty of the mean DNA copy number of target samples in each well, U_{conv}, was calculated with the following equation:

$$U_{conv} = \sqrt{\left(t_{0.05,N_w-1}\frac{s_{n\_Q}}{\sqrt{N_w}}\right)^2 + [2u(e_{curve})]^2} \tag{18}$$

In this equation, s_{n_Q} is the corrected sample standard deviation of real-time PCR quantification results of target sample replicates, N_w is the number of wells containing target sample replicates including those that failed to amplify (undetermined), and $t_{0.05,N_w-1}$ is the two-tailed critical t-value introduced in Supplementary Information File 2. We considered the uncertainty of the calibration curve was irrelevant to the sample size. Therefore, the corresponding coverage factor was fixed at 2. The representative value of the mean quantification result of target samples in this method was $\overline{n_Q}$.

Wells with no amplification were excluded from the calculation of $\overline{n_{rQ,i}}$ and $s_{n\_rQ,i}$ because a calibration curve is a fitted line between calibrator Cq values and copy numbers. Failure to amplify indicates a copy number of 0 and should be included in the quantification results. Therefore, the quantification results for all wells were used to calculate $\overline{n_Q}$ and s_{n_Q}. We compared the result with undetermined replicates defined as 0 to the results with these wells excluded.

Next, we evaluated asymmetric expanded uncertainties of the mean quantification result of target samples by calculating asymmetric confidence intervals. First, we calculated asymmetric expanded uncertainties of a mean PCR quantification result. DNA copy numbers converted from Cq values were assumed to follow the log-normal distribution[23–25]. In general, the confidence interval of a variable X that follows the log-normal distribution is calculated using the mean and corrected sample standard deviation of $Y = \ln(X)$[26]. However, no Cq values exist for wells in which the copy number was 0. This indicates that PCR quantification results are not strictly log-normal distributed when zero-copy samples are included, and the mean and corrected sample standard deviation of Y are unavailable. However, most of the results with positive values are log-normal distributed. Thus, we concluded that the approximation based on the log-normal distribution was still a practical and useful method compared to other distributions. Therefore, the mean and corrected sample standard deviation of Y when zero-copy samples are included, $\overline{Y}$ and s_Y, were estimated according to the relationship shown in Eqs. (11) and (12) in a previous study[27], using the following equations:

$$s_Y = \sqrt{\ln\left(\frac{s_X^2}{\overline{X}^2} + 1\right)} \tag{19}$$

$$\overline{Y} = \ln(\overline{X}) - \frac{s_Y^2}{2} \tag{20}$$

In these equations, $\overline{X}$ and s_X are the mean and corrected sample standard deviation of X, respectively. Based on confidence limits calculated with the modified Cox method[26], $A_{\overline{X}}$ and $B_{\overline{X}}$, left- and right-side confidence intervals for $\overline{X}$ were calculated with the following equations:

$$A_{\overline{X}} = M_X - \exp\left(\overline{Y} + \frac{s_Y^2}{2} - t_{0.05,N_X-1}\sqrt{\frac{s_Y^2}{N_X} + \frac{s_Y^4}{2(N_X-1)}}\right) \tag{21}$$

$$B_{\overline{X}} = \exp\left(\overline{Y} + \frac{s_Y^2}{2} + t_{0.05,N_X-1}\sqrt{\frac{s_Y^2}{N_X} + \frac{s_Y^4}{2(N_X-1)}}\right) - M_X \tag{22}$$

In these equations, N_X is the sample size of X; and M_X is the median of X, which is given by the following equation (see also equation 1.4 in reference[28]):

$$M_X = \exp(\overline{Y}) \tag{23}$$

In this study, the median was used as a representative real-time PCR quantification result because the median is closer to the peak of an asymmetric distribution than the mean. Note that X here corresponds to the PCR quantification result of DNA copy number of either a calibrator or a target sample.

Subsequently, left- and right-side confidence intervals for the expectation of the error of the mean result sourced from a calibration curve, A_{e_curve} and B_{e_curve}, were calculated separately based on Eq. (17):

$$A_{e\_curve} = M_Q\sqrt{\frac{\sum_{i=1}^{k}(N_{rw,i}-1)\left[\left(\frac{A_{\overline{n\_r,i}}}{R_{n\_r,i}}\right)^2 + \left(\frac{A_{\overline{n\_rQ,i}}}{M_{rQi}}\right)^2\right]}{\sum_{i=1}^{n}(N_{rw,i}-1)}} \tag{24}$$

$$B_{e\_curve} = M_Q\sqrt{\frac{\sum_{i=1}^{k}(N_{rw,i}-1)\left[\left(\frac{B_{\overline{n\_r,i}}}{R_{n\_r,i}}\right)^2 + \left(\frac{B_{\overline{n\_rQ,i}}}{M_{rQ,i}}\right)^2\right]}{\sum_{i=1}^{n}(N_{rw,i}-1)}} \tag{25}$$

Representative DNA copy number R_r	Symmetric uncertainty of mean DNA copy number of replicates $u_r(\overline{n_r})$	Symmetric expanded uncertainty of mean DNA copy number of replicates U_r	Left-side confidence interval for mean DNA copy number of replicates A_r	Right-side confidence interval for mean DNA copy number of replicates B_r
Dilution series				
80	5.14	10.3	10.0	10.3
40	3.22	6.43	6.21	6.21
20	2.11	4.22	3.88	4.23
10	1.43	2.86	2.73	2.74
5	1.00	2.00	1.81	1.82
1	0.726	1.45	0.769	1.45
New reference material				
79	1.04	2.07	1.63	2.04
39	0.730	1.46	1.22	1.22
20	0.527	1.05	0.817	1.22
9	0.369	0.739	0.408	0.816
5	0.282	0.565	0.408	0.816
1	0.186	0.373	0.000	0.500

Table 2. Results of measurement uncertainty of calibrators.

In these equations, M_Q is the estimated median of PCR quantification results of replicates of target samples calculated with Eq. (23); $A_{\overline{n\_r,i}}$ and $B_{\overline{n\_r,i}}$ are asymmetric expanded uncertainties of $\overline{n_{r,i}}$ calculated in previous sub-sections; $M_{rQ,i}$ is the estimated median of PCR quantification results of replicates of corresponding calibrators calculated with Eq. (23); and $A_{\overline{n\_rQ,i}}$ and $B_{\overline{n\_rQ,i}}$ are the left- and right-side confidence intervals for the mean of the PCR replicates of the calibrators from Eqs. (21) and (22).

Finally, asymmetric expanded uncertainty of the mean DNA copy number of target samples, represented by left and right confidence intervals A_{prop} and B_{prop}, is given by the following equations based on equation (S1):

$$A_{prop} = \sqrt{A_Q^2 + A_{e\_curve}^2} \qquad (26)$$

$$B_{prop} = \sqrt{B_Q^2 + B_{e\_curve}^2} \qquad (27)$$

In these equations, A_Q and B_Q, are left- and right-side confidence intervals of the mean PCR quantification result of target sample replicates based on Eqs. (21) and (22). The representative value of the mean DNA copy number of target samples here was the median M_Q.
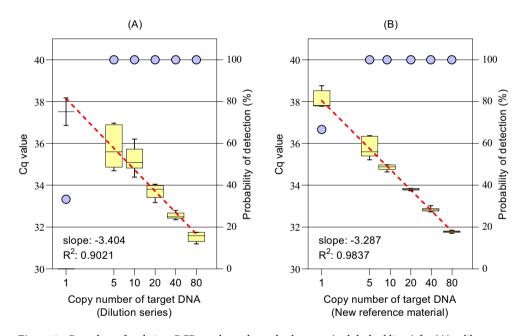
We thus obtained symmetric and asymmetric expanded uncertainties of the mean DNA copy number of target samples in each well quantified by real-time PCR, U_{conv}, and A_{prop} and B_{prop}. The corresponding 95% confidence intervals were $[\overline{n_Q} - U_{conv}, \overline{n_Q} + U_{conv}]$ and $[M_Q - A_{prop}, M_Q + B_{prop}]$.

Calculation results for the uncertainty of reference materials. The upper half of Table 2 shows the uncertainty calculation results using the conventional method and interval widths on both sides of the representative value of 95% confidence interval calculated by the proposed method for the calibrator prepared by the serial dilution method (the number of wells in which the calibrator at each concentration amplified is shown in Table 3). Note that the symmetric expanded uncertainties of mean DNA copy number of calibrators shown in Table 2, U_r, are simply two times of the corresponding symmetric uncertainties. Table S1-1 in Supplementary Information File 1 shows parameters and results, such as uncertainty of the pipette, the uncertainty of concentration of the dilution series and interval widths on both sides of the representative value of 95% confidence interval, as well as interval widths on both sides of the representative value of confidence interval arising from deviation because of the Poisson distribution.

Measurement results for factors that affect the uncertainty of the new reference material were as follows. The flow cytometry measurement results of the intracellular DNA copy number revealed that the proportion of cell aggregates with a DNA copy number of 2 was 1.29%. The aggregation rate of the cell suspension was 1.17%. Therefore, the measured proportion of single cells with a DNA copy number of 2 was 0.12%, and the worst value was 0.144%. Regarding the aggregation rate of the cell suspension for inkjet dispensing, the number of aggregated masses was 45 of 2553 cell masses; therefore, the worst aggregation rate was 2.37%. The measurement results of APD detection and false detection rates revealed that 1043 of 1047 droplets contained one cell aggregate. Moreover, seven droplets had no cell aggregates; therefore, the worst detection and false detection rates were 99.0% and 1.43%, respectively. The APD measurement results for inkjet ejection of the cell suspension showed that particles were detected in 876 of 14,479 droplets, and thus the worst particle concentration in droplets was 0.0639. Further, in contamination evaluation by real-time PCR, all negative controls were undetermined; moreover, among 80 remaining wells, amplification occurred in one well, but the Cq value was 45.0. Because the mean Cq value + 6s for the sample with a copy number of 1 was ≤ 40 in this experiment; therefore, we concluded

Representative DNA copy number R_r	Mean Cq value of replicates rCq	Mean DNA copy number of replicates quantified by the calibration curve $\overline{n_{rQ}}$	Corrected standard deviation of DNA copy number of replicates quantifed by the calibration curve s_{rQ}	Estimated median of DNA copy number of replicates quantified by the calibration curve M_{rQ}	Effective numbers of well excluded undetermined wells N_{rw}
Dilution series					
80	31.5	88.2	14.6	87.0	6
40	32.5	44.6	4.57	44.4	6
20	33.7	20.4	4.95	19.8	6
10	35.2	7.75	3.04	7.22	6
5	35.8	5.84	3.33	5.07	6
1	37.5	1.69	0.998	1.45	2
New reference material					
79	31.8	80.5	3.01	80.4	6
39	32.8	38.5	2.66	38.4	6
20	33.8	19.5	0.585	19.5	6
9	34.9	9.345	0.883	9.30	6
5	35.8	5.16	1.60	4.93	6
1	38.0	1.04	0.291	1.00	4

Table 3. Real-time PCR results of calibrators.

Figure 1. Box plots of real-time PCR results and standard curves (red dashed lines) for (**A**) calibrator prepared by the serial dilution method and (**B**) calibrator prepared using the new reference material. A whisker in the box plots shows the 10th and 90th percentiles. A purple circle indicates the percentage of wells in which the calibrator of each copy number amplified (probability of detection).

that specific amplification did not occur, and the number of DNA contaminants was 0. In other words, the worst probability of DNA contamination from reagents and the environment was 0.0571%. Finally, the distribution of the DNA copy number of the new reference material was combined from these components. The lower half of Table 2 shows the uncertainty calculation results using the conventional method and interval widths on both sides of the representative value of a 95% confidence interval calculated using the proposed method. Overall, the uncertainty was lower than that for the calibrator prepared by serial dilution. Moreover, strong asymmetry was reported when the copy number was 1, 5, 9, 20, and 79. When the copy number was 39, the asymmetry was rounded during the calculation process of the minimum integer interval of production variability.

Real-time PCR quantification results and measurement uncertainty estimation results. Table 3 and Fig. 1 show real-time PCR results for two different calibrators. All negative controls did not show amplification. The slope, intercept, and coefficient of determination (R^2) of the calibration curve prepared by the

Mean DNA copy number of replicates $\overline{n_Q}$	Corrected sample deviation of DNA copy number of replicates s_Q	Estimated median of DNA copy number of replicates M_Q	Left-side confidence interval for mean DNA copy number of replicates A_Q	Right-side confidence interval for mean DNA copy number of replicates B_Q
Calibrated by the dilution series				
1.28	1.16	0.951	0.453	2.36
Calibrated by the new reference material				
1.23	1.12	0.906	0.434	2.28

Table 4. Real-time PCR results of target samples to be quantified including undetermined wells as 0 copy.

Mean DNA copy number of replicates $\overline{n_Q}$	Corrected sample deviation of DNA copy number of replicates s_Q	Estimated median of DNA copy number of replicates M_Q	Left-side confidence interval for mean DNA copy number of replicates A_Q	Right-side confidence interval for mean DNA copy number of replicates B_Q
Calibrated by the dilution series				
1.93	0.777	1.79	0.779	1.90
Calibrated by the new reference material				
1.84	0.768	1.70	0.809	1.99

Table 5. Real-time PCR results of target samples to be quantified excluding undetermined wells.

serial dilution method were -3.40, 38.1, and 0.902, respectively; therefore, the amplification efficiency was 96.7%. Moreover, the slope, intercept, and R^2 of the calibration curve prepared with the new reference material were -3.29, 38.0, and 0.984, respectively. Therefore, the amplification efficiency was 101%. The relative uncertainty $u(e_{curve})/\overline{n_Q}$ of the calibration curve prepared with dilution series, A_{e_curve}/M_Q, and B_{e_curve}/M_Q were 0.247, 0.414, and 62.3, respectively. The relative uncertainty $u(e_{curve})/\overline{n_Q}$ of the calibration curve prepared with the new reference material, A_{e_curve}/M_Q, and B_{e_curve}/M_Q were 0.101, 0.166, and 0.340, respectively. For the calibration curve prepared with the dilution series, the one-copy sample amplified in only two wells, and thus the two-tailed t critical value increased to 12.7. Therefore, the relative uncertainty on the right side became very large.

Next, Tables 4 and 5 shows quantification results for the target sample with two different calibration curves. Among the six wells, the result was undetermined for two wells and the probability of detection (POD) of the target sample was 66.7%. Cq values for the wells in which amplification occurred were 37.8, 36.66, 37.8, and 36.8. The copy number quantification results with the calibration curve prepared with the dilution series were 1.24, 2.74, 1.28, and 2.44, whereas the copy number quantification results with the calibration curve prepared with the new reference material were 1.17, 2.64, 1.20, and 2.35, respectively. Quantification results for undetermined wells were treated as zero in Table 4.

The final measurement uncertainty was calculated with Eqs. (17)–(27). When the calibration curve prepared with the dilution series was used for quantification, the quantification result with the expanded uncertainty for the target sample calculated by the conventional method was 1.28 ± 1.38. Therefore, the 95% confidence interval was $[-0.092, 2.66]$. Moreover, the quantification result calculated by the proposed method was 0.951, and the 95% confidence interval was $[0.350, 60.3]$. When the calibration curve prepared with the new reference material was used for quantification, the quantification result with the expanded uncertainty for the target sample calculated by the conventional method was 1.23 ± 1.20. Therefore, the 95% confidence interval was $[0.025, 2.43]$. Moreover, the quantification result calculated by the proposed method was 0.906, and the 95% confidence interval was $[0.447, 3.21]$.

For comparison, we also calculated the results after excluding wells in which the target failed to amplify. When the calibration curve prepared with the dilution series was used for quantification, the result with the expanded uncertainty for the target sample calculated by the conventional method was 1.93 ± 1.56. Therefore, the 95% confidence interval was $[0.364, 3.49]$. The quantification result calculated by the proposed method was 1.79, and the 95% confidence interval was $[0.711, 113]$. When the calibration curve prepared with new reference material was used, the result with the expanded uncertainty for the target sample calculated by the conventional method was 1.84 ± 1.28. Therefore, the 95% confidence interval was $[0.564, 3.12]$. The quantification result calculated by the proposed method was 1.70, and the 95% confidence interval was $[0.891, 3.69]$.

Discussions

In this paper, we describe models for calculating asymmetric uncertainties in measurements obtained by real-time PCR quantification. We separately evaluated left- and right-side uncertainties of two kinds of serially diluted and new reference material produced by an inkjet system and real-time PCR. We considered the influence of the Poisson distribution as a significant component in the dilution series model and proposed a new model for calculating the production variability of new reference material. We also introduced several factors for consideration in modeling real-time PCR quantification. The dilution series models were based on the general models

given in Guide to the Expression of Uncertainty in Measurement[18]. We also used the concept to combine the relative uncertainties of several components described previously[14].

The uncertainty evaluation results for the one-copy calibrator by the proposed method showed similar tendencies with the dilution series and the new reference material. When the dilution series were used, the sample's confidence interval with a copy number of 1 was strongly asymmetric with respect to the representative value because the Poisson distribution with $\lambda = 1$ is strongly asymmetric. When the new reference material was used, the distribution was strongly asymmetric for the one-copy sample because it follows a peculiar distribution pattern. When the copy number was ≥ 5, the calibrator's uncertainty prepared by serial dilution was generally symmetric because the Poisson distribution's asymmetry was very weak. However, a certain level of asymmetry was seen for the calibrator prepared with the new reference material even when the copy number was ≥ 5 because of the characteristics of the distribution. Such a measurement result with asymmetric uncertainty can be separately expressed with the representative value N and the confidence interval $[L, H]$ as "N, $[L, H]$". Moreover, it can be expressed with the representative value and distances from the representative value to the lower and upper limits of the confidence interval, A and B, respectively, like N_{-A}^{+B}.

The mean copy number's relative uncertainty was considerable for the dilution series because the Poisson distribution has a dominant effect when the copy number is ≤ 80. Similarly, the Poisson distribution has certain impact on the new reference material depending on the concentration of the cell suspension used. Nevertheless, the effect was minor because the cell suspension concentration was very low. Overall, the new reference material had low uncertainty.

Furthermore, the POD results for the one-copy calibrators showed an enormous difference. While the POD for the new reference material was 66.7%, the POD for the dilution series was 33.3%. These results suggest that the calibrator prepared with the dilution series does not contain DNA at a high probability. Note that the one-copy calibrator was prepared with the same diluted solution used to prepare target samples. However, the sample size was small and the value of POD varied widely. Moreover, the variability of Cq values was relatively high for the dilution series. This arises from the high uncertainty of the calibrator itself and the operator's work variability; however, the new reference material is associated with relatively low uncertainty and is less susceptible to the operator effect. Therefore, relatively highly precise calibration curves can be obtained with the new reference material.

The amplification efficiency for the new reference material's calibration curve exceeded 100%, presumably because the confidence interval for the low-copy-number calibrator was skewed toward the right. The copy number of a low-copy-number calibrator is possibly larger than the representative value. Moreover, the calibration curve was an approximated plot of data with representative values. Therefore, the left side of the calibration curve was lowered, and the slope's absolute value became smaller. The lower limit of the confidence interval estimated by the conventional method was a negative value for quantifying the target sample with both reference materials. However, the estimation results with the proposed method were always ≥ 0. As the DNA copy number cannot be a negative value, in reality, the negative part of the confidence interval has no physical meaning. Therefore, the proposed method can allow for more realistic uncertainty estimation. Moreover, for the dilution series, the asymmetry of the Poisson distribution has relatively strong effects when the copy number is low. Therefore, the actual copy number is possibly larger than the calculated value. When the new reference material is used, the actual copy number is likely to be larger than the calculated value because of the effects of cell aggregation and false-negative results of APD detection. Therefore, the calibration curve is likely to shift to the right from the approximation result. In other words, even when the Cq value was the same, it is likely to be larger than the calculated copy number of the calibrated sample to be quantified. This is another reason why uncertainty estimation with the proposed method is more appropriate.

Moreover, when the dilution series was used for quantification, the relative uncertainty on the right side was substantial because the POD of the one-copy calibrator was low. Therefore, the uncertainty on the right side estimated by the new method was very large, suggesting that copy number quantification with low POD is inappropriate.

Subsequently, when excluding wells in which the target failed to amplify, the mean DNA copy number increased to nearly 2, although the expectation was 1. One possible reason for this discrepancy is that the concentration of the diluted solution drifted, creating a large uncertainty in the one-copy solution. However, it is also possible that a one-copy sample, once diluted, would contain no DNA. Therefore, it is appropriate to include the zero-copy wells.

Furthermore, the uncertainty of the calibrator used is added to the measurement uncertainty. The use of reference material with smaller uncertainty improved the measurement precision of real-time PCR. Assay optimization, such as internal control and an increased number of samples, is necessary to reduce the uncertainty further. Moreover, the quantifiable range of copy numbers may be expandable using a calibrator that is excellent in terms of uncertainty in low copy numbers, such as 1 copy and 5 copies, and POD.

Currently, the coefficient of variation (CV) of quantification results is used to evaluate the LOQ of real-time PCR[29,30]. However, the calibrator uncertainty and bias of the uncertainty of real-time PCR in a low-copy-number region have effects on measurement results. Therefore, these factors should be considered. Moreover, criteria for evaluation of LOD include the percentage of positive replicates (equivalent to POD)[29,30]; however, considerations are required for copy number results of zero, which are included in measurement results, and the variability of the sample itself. Therefore, how to reflect these effects in the definition of LOD is considered to be an important issue.

Concentration after diluted C_{after} (μL^{-1})	Volume of previous dilution p (μL)	Volume of buffer q (μL)
2.24×10^9		
2.00×10^7	4.00	444
2.00×10^5	4.00	396
2.00×10^3	4.00	396
20.0	4.00	396
10.0	200	200
5.00	200	200
2.50	200	200
1.25	150	150
0.250	40.0	160

Table 6. Protocol of serial dilutions.

Methods

We prepared low-copy-number calibrators by the serial dilution of certified reference material (6205-a DNA600-G, National Metrology Institute of Japan) with an initial DNA copy number concentration of $2.24 \times 10^9 \, \mu L^{-1}$. The dilution protocol is shown in Table 6. The dilution buffer was prepared by mixing 1480 µL of 1× TE buffer (TE, pH 7.0, RNase-free, Thermo Fisher Scientific), 1480 µL of UltraPure DNase/RNase-Free Distilled Water (Thermo Fisher Scientific), and 40 µL of ColE1 DNA (318-00436, 450 ng μL^{-1}, NIPPON GENE). ColE1 DNA was added to prevent DNA containing the target sequence from adhering to the inner wall of wells.

The forward and reverse primers and probes used for real-time PCR experiments were 5′-TCGAAGGGTGAT TGGATCGG-3′, 5′-TGGCTAGCTAAGTGCCATCC-3′, and 5′-6-FAM-TGCATTCTGGCTTCGATTGTCCCT AC-TAMRA-3′, respectively. The size of the amplicon produced by the primers was 100 bp. A 100 µM primer solution (100 µL) was diluted to 10 µM primer by mixing with 1× TE buffer (900 µL). A 100 µM probe solution (20 µL) was diluted to 2 µM probe with 1× TE buffer (980 µL). The amplification reagent was prepared by mixing 1050 µL of GenCheck qPCR Probe Master (dUTP, FASMAC), 105 µL of 10 µM forward primer, 105 µL of 10 µM reverse primer, 210 µL of 2 µM probe, and 210 µL of UltraPure DNase/RNase-Free Distilled Water. Finally, 16 µL of the amplification reagent and 4 µL of a template DNA solution were added to each well. To construct a calibration curve, we dispensed two different calibrators, the dilution series, and the new reference material, into a 96-well plate. The solutions of each calibrator at six different concentrations (1, 5, 10, 20, 40, and 80 copies) were added to five wells each in rows B to F. As a negative control, the amplification reagent alone was added to 6 wells in row A of the plate. In addition, calibrators with copy numbers of 1, prepared by the serial dilution method, were added to six wells in row H as quantification targets.

QuantStudio 12K Flex (Applied Biosystems) was used to carry out real-time PCR measurements. The thermal cycling condition was: reaction at 50 °C for 2 min and at 95 °C for 10 min, followed by 50 cycles of reaction at 95 °C for 30 s and at 61 °C for 1 min. The threshold was fixed to 0.2. For target sample wells, the DNA copy number was set to zero when the real-time PCR result was undetermined.

The preparations of cell suspensions for preparation of the new reference material and treatments after dispensing, such as cell wall lysis, were performed as described in the previous study[13], except that *Synthetic construct DNA 6203-a-G* (600-G, GenBank registration number AB610938.1), instead of *hmg-Le1*, was incorporated into yeast.

Conclusions

This risk of false-negative results in real-time PCR detection of environmental or viral DNA is great, and the demand for improved measurement precision is significant. However, the spread and development of genetic testing remain unsatisfactory because of inadequate concentration precision of low-copy-number calibrators and the unavailability of methods for appropriate evaluation of the uncertainty of test results. With the advent of reference material with high precision in a low-copy-number region, genetic testing has come one step closer to a new stage. Furthermore, in this study, we showed a method to reflect the asymmetric character of the uncertainty of low-copy-number calibrators in the calculation and evaluate the uncertainty of real-time PCR measurements more appropriately. Combining these, real-time PCR quantification of samples from one copy, which should be theoretically possible, will become possible in the near future. Moreover, the conclusion that the confidence intervals of real-time PCR measurement results are skewed to the right is applicable to qualitative tests. It may contribute to improving the reliability of qualitative results in various applications such as tests to rule out viruses or mycoplasma.

Data availability

All data generated and/or analyzed during current study are available within the manuscript and the supplementary information file.

References

1. Mullis, K. B. & Faloona, F. A. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* **155**, 335–350 (1987).
2. Holst-Jensen, A., Rønning, S. B., Løvseth, A. & Berdal, K. G. PCR technology for screening and quantification of genetically modified organisms (GMOs). *Anal. Bioanal. Chem.* **375**, 985–993 (2003).
3. ISO 21570:2005. Foodstuffs—Methods of analysis for the detection of genetically modified organisms and derived products—Qualitative nucleic acid based methods. *Int. Organ. Stand.* (2005).
4. ISO 21571:2005. Foodstuffs—Methods of analysis for the detection of genetically modified organisms and derived products—Nucleic acid extraction. *Int. Organ. Stand.* (2015).
5. Kuribara, H. *et al.* Novel reference molecules for quantitation of genetically modified maize and soybean. *J. AOAC Int.* **85**, 1077–1089 (2002).
6. Notification No.201; Consumer Affairs Agency, Government of Japan: Tokyo, Japan. (2012).
7. Yang, S. & Rothman, R. E. PCR-based diagnostics for infectious diseases: Uses, limitations, and future applications in acute-care settings. *Lancet Infect. Dis.* **4**, 337–348 (2004).
8. Morling, N. PCR in forensic genetics. *Biochem. Soc. Trans.* **37**, 438–440 (2009).
9. ISO 20395:2019. Biotechnology—Requirements for evaluating the performance of quantification methods for nucleic acid target sequences—qPCR and dPCR. *Int. Organ. Stand.* (2019).
10. Higuchi, R., Fockler, C., Dollinger, G. & Watson, R. Kinetic PCR analysis: Real-time monitoring of DNA amplification reactions. *Nat. Biotechnol.* **11**, 1026–1030 (1993).
11. Heid, C. A., Stevens, J., Livak, K. J. & Williams, P. M. Real time quantitative PCR. *Genome Res.* **6**, 986–994 (1996).
12. Svec, D., Tichopad, A., Novosadova, V., Pfaffl, M. W. & Kubista, M. How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomol. Detect. Quantif.* **3**, 9–16 (2015).
13. Seo, M. *et al.* Novel bioprinting application for the production of reference material containing a defined copy number of target DNA. *Anal. Chem.* **91**, 12733–12740 (2019).
14. Griffiths, K. R., Burke, D. G. & Emslie, K. R. Quantitative polymerase chain reaction: A framework for improving the quality of results and estimating uncertainty of measurement. *Anal. Methods* **3**, 2201 (2011).
15. Trapman, S. *et al. Guidance Document on Measurement Uncertainty for GMO Testing Laboratories.* (European Commission Joint Research Centre, 2009). https://doi.org/10.2787/18988.
16. Haight, F. A. *Handbook of the Poisson Distribution* (Wiley, 1967).
17. Hedges, A. J. Estimating the precision of serial dilutions and viable bacterial counts. *Int. J. Food Microbiol.* **76**, 207–214 (2002).
18. JCGM 100:2008 Evaluation of measurement data—Guide to the expression of uncertainty in measurement.
19. Johnson, N. L., W. Kemp, A. & Kotz, S. Poisson Distribution. In *Univariate Discrete Distributions: Third Edition* 156 (Wiley, 2005). https://doi.org/10.1002/0471715816.
20. Papoulis, A. Bernoulli Trials. In *Probability, Random Variables, and Stochastic Processes* 57–63 (McGraw-Hill, 1984).
21. Bertsekas, D. P. & Tsitsiklis, J. N. *Introduction to Probability* (Athena Scientific, 2002).
22. Newcombe, R. G. Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Stat. Med.* **17**, 857–872 (1998).
23. Thompson, M. GMO Proficiency testing: Interpreting z-scores derived from log-transformed data. *AMC Tech. Br.* (2004).
24. Johnson, N. L., Kotz, S. & Balakrishnan, N. Lognormal distributions. In *Continuous Univariate Distributions*, Vol.1 (ed. 2nd) xvi+387 (Wiley, 1994).
25. Bland, J. M. & Altman, D. G. Statistics notes: Transformations, means, and confidence intervals. *BMJ* **312**, 1079–1079 (1996).
26. Olsson, U. Confidence intervals for the mean of a log-normal distribution. *J. Stat. Educ.* **13**, (2005).
27. Weisstein, E. Log Normal Distribution. *Wolfram Mathworld* https://mathworld.wolfram.com/LogNormalDistribution.html (2021).
28. Daly, L. E. & Bourke, G. J. Data transformations. In *Interpretation and Uses of Medical Statistics* 27–31 (Wiley, 2000) https://doi.org/10.1002/9780470696750.
29. Forootan, A. *et al.* Methods to determine limit of detection and limit of quantification in quantitative real-time PCR (qPCR). *Biomol. Detect. Quantif.* **12**, 1–6 (2017).
30. Kralik, P. & Ricchi, M. A basic guide to real time PCR in microbial diagnostics: Definitions, parameters, and everything. *Front. Microbiol.* **8**, 108 (2017).

Acknowledgements

Author contributions

U.K. developed the model, wrote the main manuscript text and prepared the equations, tables and figures. T.S. wrote the introduction and revised the manuscript. S.N., Y.Y. and K.W. prepared the samples and acquired the data. M.H. prepared the samples. S.H. and H.U. revised the manuscript. All authors reviewed the manuscript.

Competing interests

All authors are employees of Ricoh Company, Ltd. who are manufacturers of the RICOH Standard DNA Series. All authors do not have any competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-90959-0.

Correspondence and requests for materials should be addressed to U.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.