Considerations for Digital PCR as an Accurate Molecular Diagnostic Tool

Jim F. Huggett,1,2* Simon Cowen,1 and Carole A. Foy1

BACKGROUND: Digital PCR (dPCR) is an increasingly popular manifestation of PCR that offers a number of unique advantages when applied to preclinical research, particularly when used to detect rare mutations and in the precise quantification of nucleic acids. As is common with many new research methods, the application of dPCR to potential clinical scenarios is also being increasingly described.

CONTENT: This review addresses some of the factors that need to be considered in the application of dPCR. Compared to real-time quantitative PCR (qPCR), dPCR clearly has the potential to offer more sensitive and considerably more reproducible clinical methods that could lend themselves to diagnostic, prognostic, and predictive tests. But for this to be realized the technology will need to be further developed to reduce cost and simplify application. Concomitantly the preclinical research will need to be reported with a comprehensive understanding of the associated errors. dPCR benefits from a far more predictable variance than qPCR but is as susceptible to upstream errors associated with factors like sampling and extraction. dPCR can also suffer systematic bias, particularly leading to underestimation, and internal positive controls are likely to be as important for dPCR as they are for qPCR, especially when reporting the absence of a sequence.

SUMMARY: In this review we highlight some of the considerations that may be needed when applying dPCR and discuss sources of error. The factors discussed here aim to assist in the translation of dPCR to diagnostic, predictive, or prognostic applications.

© 2014 American Association for Clinical Chemistry

Digital PCR (dPCR)3 is an increasingly applied technique that offers a number of advantages for both detecting and quantifying nucleic acids. The method, which is performed by diluting target nucleic acid across a large number of reactions (termed partitions) so that some of the reactions contain no template, is gaining considerable traction. The concept behind dPCR has been long established (1), but it has been made practical in recent years by the development of microfluidics and/or emulsion chemistries to simplify and automate the process. The frequently cited benefits of improved precision affording finer fold-change measurements and deeper rare variant detection have been highlighted and increasingly applied to clinical challenges, including oncology (2–5) and infectious diseases (6–10) as well as fetal genetic screening (11, 12), and in predicting transplant rejection (13).

Despite positive press, dPCR is far from infallible. It depends on PCR, which, although it is possibly the most quintessential molecular method yet developed, requires fairly complex and accurate thermocycling instrumentation to perform. Real-time quantitative PCR (qPCR) revolutionized clinical application of PCR partly because it automated analysis by removing the need for postreaction manipulation. Most dPCR instruments have taken a step backward in this sense, requiring downstream manipulation to analyze the results, increasing time, and complicating the process. dPCR is also developing in the same primordial sea as next generation sequencing, which continues to advance with both hardware and chemistry to increase throughput as well as sequence read number and length. Advanced sequencing methods are also likely to eventually replace most molecular methods that are used to identify and quantify nucleic acids, including PCR in its many guises. Yet dPCR is here now and has much to offer. This review focuses on some of the key technical considerations in the performance of dPCR for a clinical application.

Routine Adoption of Quantitative Clinical Molecular Measurement

One undeniable truth about the clinical adoption of new diagnostic, predictive, or prognostic methods is it...
Clinical Chemistry 61:1 (2015)

Advantages

dPCR principally offers advantages when considering the identification of the presence and abundance of rare sequence mutations (19) and in the quantification of nucleic acids (20), while also enabling the cis/trans relationships (21, 22) to be determined. Quantitative and qualitative measurements offer different challenges when considering error, and many of the challenges that befall legacy PCR and qPCR when designing and optimizing an analytically sensitive and specific method directly apply to dPCR; i.e., a nonspecific and/or poorly optimized set of primers may bind the wrong target, leading to a false-positive signal. Consequently, when performing dPCR is performed it is vital to also perform control experiments using templates that are likely to complicate a result if they are present within a sample.

Because dPCR performs by identifying single partitions as positive or negative, it offers the potential to perform precise low-level quantification (23). However, studies that take advantage of this capability to investigate trace or low-abundance sequences should also measure sources of false-positive signal due to low-level contamination and/or nontarget amplification (caused by mis-priming/primer dimers). Because large numbers of trace experiments are performed, the chances of false positives increase, and these false-positive results must be captured when assessing limits of detection (24). Consequently, sufficient replication of appropriate controls is required to confidently attribute any low-level signal to the sample of interest and not to low-frequency false positives.

Rare Variant Detection

The term “digital” PCR was coined in a study that focused upon the detection of rare variant mutants of the ras oncogene in a sample that predominantly consisted of wild-type sequences (19). Designing a qPCR (or array)–based assay to a given single-nucleotide polymorphism (SNP) is not usually highly challenging because, as long as the target sequence is suitable, it is fairly simple to select a primer or probe to the desired SNP. The unwanted sequence will usually also be detected, but at a much lower (approximately 5%) detection efficiency (25); i.e., if the mutant and wild-type sequence are measured at the same concentration by qPCR, the wild-type sequence will amplify approximately 4 cycles later than the mutant. Where wild-type and mutant sequences are mixed there will be a proportion ($p_{eq}$) of the wild type for which an equivalent amount of signal comes from the wild-type and mutant sequence following equation 1:

\[ \text{Signal} = p_{eq} \times \text{Signal}_{\text{wild-type}} + (1 - p_{eq}) \times \text{Signal}_{\text{mutant}} \]
where $f$ is the efficiency (or frequency) by which the mutant probe binds the wild-type signal. So for a typical qPCR assay for which wild-type sequences are detected at a frequency of 0.05 compared to the mutant, the wild-type proportion at which 50% of the signal comes from each genotype would be 0.952. This means that when there are 20 wild-type sequences and 1 mutant sequence, 50% of the amplification plot will comprise wild-type sequence. The problem apparent from this equation is that the wild-type sequence does not have to be very abundant for it to be the predominant sequence detected, thus compromising the specificity of the assay. There have been several developments to further reduce the efficiency by which the wild-type sequence is detected using blocker probes (26, 27) that further reduce $f$. However, because the need to detect even rarer mutations is demanded, the specificity of the assay continues to be a potential problem.

dPCR sidesteps this problem because the partitioning that is its hallmark also reduces the wild-type/mutant ratio and the likelihood that $p_{eq}$ will be reached. This process can be taken to the extreme by partitioning the template at such a low concentration that each partition contains only a single molecule, either SNP or wild type (28). However, SNP detection by any PCR method is a balance between assay specificity and sensitivity in which design and optimization are paramount.

dPCR can be more sensitive than conventional approaches but remains susceptible to poor assay design, leading to cross-reactivity and false positives. Partitioning may not prevent assay cross-reactivity, however, and “wild-type only” partitions can still lead to a signal that may complicate analysis (Fig. 1). Consequently, robust application of such assays may require optimization and potentially redesign to determine the assay performance using both the desired and undesired genotypes to determine if the different genotypes can be discriminated by applying an amplitude threshold (Fig. 1C) or not (Fig. 1D).

Another unsung fact about dPCR is that, although the technique may be able to measure 1 SNP in a sea of 10 000 wild-type sequences, the practical fact is that this would require the measurement of approximately 50 000 gene copies to reach a 99% probability that the sample contains at least 1 mutant. This affords 2 challenges.

First, 50 000 gene copies of wild-type sequence will be too concentrated for accurate measurement with the use of all but one of the instruments that are currently available commercially. Consequently, if an accurate estimation of the abundance of mutant to wild type is needed, then the wild type must be measured at a concentration different from that of the mutant, rendering multiplexing useless and further complicating the measurement.

Second, depending on the biological sample, 50 000 copies of human genomic DNA could require a large amount of tissue. Although many sample specimens readily provide large amounts of DNA, this is not always the case. Cell-free nucleic acid measurement in plasma is an increasingly popular source of sample for rare-variant detection because it is less invasive than a tissue biopsy. To this end this sample has also been the focus of several studies using dPCR (5, 13). However, because cell-free DNA (cfDNA) is at a concentration of approximately 1000 copies per mL of plasma in healthy individuals (29) it would require about 50 mL of blood to measure 50 000 cfDNA copies. This is a very large sample for a potential routine clinical test and, in this case, 1 mutation in 10 000 wild types may not be possible for practical rather than technical reasons.

### Absolute Quantification

Quantification of nucleic acids is the other major strength of dPCR because it does not routinely require a calibration curve to provide a numeric value. This is further strengthened by the random nature of the distribution of the DNA molecules across the partitions that makes the precision of measurement both predictable and precise compared to qPCR (20), a fact that may also be applicable to RNA measurement (30) as well as other nucleic acid amplification methods like loop-mediated isothermal amplification (LAMP) (31) and recombinase polymerase amplification (32).

We describe in the Minimum Information for the Publication of Quantitative dPCR Experiments (dMIQE) guidelines that dPCR is uniquely able to detect and provide an “absolute” quantitative value of small amounts of specific target in a complex background without the need for a calibration curve (33). However, this accolade makes it very difficult to assess accuracy (if the results are actually correct) because there are few methods that are able to verify dPCR results. The vast majority of other molecular methods, including qPCR, are “relative” methods that must be calibrated using a similar template that must, in turn, be given a quantity (or value assigned) using another absolute quantification method.

Most value assignment for molecular standards is likely to be performed using spectrophotometry at an absorbance of $A_{260}$. Comparison of spectrophotometric methods for DNA measurements with dPCR has generally shown good agreement, differ-
To begin with this is very promising because when molecular quantification is routinely applied in the clinic, decisions are seldom made on changes of <10-fold. However, it must be remembered that although these studies show promise, because they suggest that the agreement between dPCR and A260 is fairly accurate, they could also be equally biased. This is particularly pertinent when we consider the fact that usually nucleic analysis of a clinical sample requires processing to purify and concentrate the nucleic acids.

**Error: Bias**

To better understand the accuracy of dPCR, which is influenced by bias and variance, we will initially con-
sider sources of bias (with which the given result is incorrect). As with any PCR assay, poor design or optimization may lead to overestimation due to nonspecific amplification, as will probe bleed-through during multiplexing (through which the signal from one fluorophore is detected during the reading of another). Accepting this, a well-designed and optimized, specific dPCR assay is limited in the ability for overestimation of the target sequence. This is because, unlike qPCR (which is dependent on factors like value assignment of the calibrator and PCR efficiency), theoretically, overestimation by dPCR can occur only if DNA molecules become denatured during partitioning and the resulting single strands frequent 2 different partitions. Consequently the worst case scenario will be a 2-fold overestimation.

Underestimation is a far greater potential source of bias when performing dPCR and is rooted in the fact that dPCR is only able to count positive partitions. A negative partition is routinely assumed to contain no DNA; however, it is known that when using methods like digital LAMP (31) and when detecting RNA by reverse transcriptase dPCR (30) it is possible for a large number of partitions to contain template that is not amplified. This “molecular dropout” is likely to be low in a dPCR reaction with a pure DNA template (36). However, template must be amplifiable and factors that damage DNA or that make DNA less accessible, like plasmid super-
coiling (23), are likely to also lead to molecular dropout.

Inhibition of dPCR appears to cause less negative bias than with qPCR (31, 37, 38); however, inhibitors are able to reduce the quantitative value measured by dPCR while maintaining good precision (31), and we strongly recommend including internal positive controls when measuring clinical samples, especially if negative results need to be reported. The popular routine use of internal positive controls when applying qPCR clinically is equally important for dPCR.

Other sources of underestimation will occur for which some of the assumptions that underpin dPCR are not met. The current method used for quantification by dPCR is totally dependent upon the random distribution of template. Situations in which this condition cannot be met, owing to factors like template linkage and/or sample inhomogeneity, will lead to an underestimation. A pertinent example is the measurement of genome instabilities like human epidermal growth factor-2 (HER-2) amplification, which manifests as cis amplification of the HER-2 oncogene on chromosome 17 (39). Unless the individual targets are separated they cannot frequent different partitions and will not be independently counted by dPCR. This effect, which does not have an impact on qPCR, may lead to underestimation of the number of copies when measured by dPCR (20).

Underestimation may also occur because the underlying model that is used to estimate the number of copies (Eq. 2) (33) describes a situation which is physically impossible:

\[ \lambda = -\ln\left(1 - \frac{k}{n}\right) \]  

(2)

This is because with Eq. 2, the number of positive partitions \( k \) is divided by the total number of partitions \( n \), assuming that there is no partition volume variation. Because the partition volume must have an associated error (34, 40), if this is not factored into the equation then there is potential for underestimation (Fig. 2). This has a greater effect in situations for which \( \lambda \) is high and will not have an impact on the results if experiments are performed at a \( \lambda \) that is low enough to ensure single-molecule occupancy of partitions. What is not clear is how this might really affect quantification by dPCR when the other associated errors are considered and how this variance could be factored into the equation (one possible solution is offered in Box 1 and the Supplementary Material that accompanies the online version of this article at http://www.clinchem.org/content/vol61/issue1). It is possible that the partition volume variance of a given instrument may be negligible compared to the error associated with other factors like pipetting, sample processing, and so on. However, what is needed is for this information to be made available for the different instruments so this can be tested, enabling users to be more confident in the bias of their measurements. Ideally, where run-to-run differences are substantial, instruments would provide these types of quality metrics for each run to enable these variations to be measured on a day to day basis.

**Error: Variance**

dPCR measurement is well documented as being based on the Poisson distribution (20, 41) which, for a given mean, has a constant variance. Although this may be true for the measurement of a given pure DNA extract, inhibitory substances copurified or introduced during the extraction may or may not affect the overall measurement variability (31). The other major factor that effects all molecular methods is the source of variance (and bias) contributed by the upstream steps involved in sampling, storing, and extracting the nucleic acid (42) to which dPCR is equally susceptible.

Consequently, when designing a dPCR experiment that aims to make quantitative clinical measurement, an appropriate estimation of the variance (and therefore potential confidence) must be captured by replicating upstream steps, including extraction procedure, to estimate all the potential sources of technical variance. This will also reduce the likelihood of the high technical precision of the dPCR step leading to significant but erroneous biased results.
Box 1. A New Method for Improving the Accuracy of Quantification by dPCR

dPCR is performed by distributing DNA across a large number of reactions (partitions) at such a concentration that a proportion of the partitions will contain no template at all, resulting in a negative signal. The proportion of positive signals is then used to estimate the average number of copies per partition (\(\lambda\)) using the Poisson distribution.

One of the key assumptions is that the DNA distribution is random and equal for each partition. The latter cannot be met if the volume varies among partitions; if partition volume variability is not considered when calculating \(\lambda\), the distribution of molecules among partitions may lead to a bias in the estimated number of copies per reaction (Fig. 2). This has implications for any research area in which accurate measurement is particularly important.

There is currently a range of commercially available instruments employing either chip or emulsion partitioning, with additional formats also being developed. With continuing developments meeting the challenges of reaction partitioning, differences in the associated volume variation are likely to lead to the possibility that this could have an impact on instrument comparability. To raise awareness of the existence of uncertainty in partition volume, the dMIQE guidelines included partition volume variance as one of the list of desirable pieces of information (“Desirables”) to include in a publication when available.

The problem with this Desirable is that instrument volume variability is not readily available, so it cannot be taken into account when estimating \(\lambda\). In particular, what has not been addressed is how such variation may have an impact on dPCR results if partition volume variation is significant, and \(\lambda\) is calculated using the conventional equation.

We have developed a potentially more accurate estimation of \(\lambda\), which considers volume variation for a given observed count. By removing the assumption that the partition volume is fixed (and \(\lambda\) constant), we use a model with a \(\gamma\)-distributed volume and consider \(\lambda\) as a mean over the volume variation. The resulting expressions for \(\lambda\) and its uncertainty remove the bias and provide a more accurate estimate. A detailed derivation is given in the online Supplementary Material, together with some example calculations.

When the volume variability is small, or the sample concentration is low, the effects described here are negligible, and so in some operating ranges, it is not relevant. However, the effect becomes more pronounced with increasing sample concentration and partition volume variability (see online Supplementary Fig. 1). In addition, there is associated deterioration in precision (see online Supplementary Fig. 2). These findings also highlight a difference in the value of \(\lambda\) for which dPCR is most precise (reportedly approximately 1.6 using the conventional equation and assuming no volume variation) between instruments of differing partition volume variability.

The volume effect we describe will initially be most relevant to measuring fine differences in fold changes or to assigning values to reference materials intended to support measurements in areas such as testing for genetic modification. However, as dPCR becomes more established with more instruments available, it may have broader implications when we consider interinstrument comparisons, for which the partition volume variability differs. It is important that manufacturers of dPCR instruments provide information on partition volume variation if absolute quantification is to be offered. Further, incorporating volume variability in their instrument analysis software using an approach such as the one we propose would be desirable for the most accurate dPCR measurement.

Another factor that is raised by our findings is the potential importance of knowing how such volume variability may differ between batches of chips, because this will affect the degree of bias present. In situations in which instruments generate partitions (as with the current droplet formats), it is important that they are routinely checked to define volume variability if run-to-run differences are to be avoided.

The impact of this phenomenon will ultimately increase as instruments capable of processing larger numbers of partitions are developed, enabling ever-increasing range in \(\lambda\). Our findings demonstrate that when \(\lambda\) is low and variation is small (less than about 10%), volumetric variability is unlikely to have a major impact. As \(\lambda\) increases, the bias becomes more severe, impacting measurement accuracy. We encourage instrument manufacturers to provide partition volume uncertainty information, and our model provides a mechanism for more accurate measurement when using instruments with differing partition volume variation.
Measuring Linkage

The problem for quantification by dPCR if sequences are linked and not randomly distributed can also be used advantageously. We used this in combination with multiplex analysis to estimate molecular dropout and investigate the impact of different template types on error (36). This application also has the potential to offer clinically valuable tools to investigate haplotypes in prenatal screening for diseases like \( \beta \)-thalassemia (43) and polymorphic gene complex arrangements encoding immunological cell receptors (21).

dPCR experiments that measure linkage depend on colocalization of targets within the partitions that do not follow the random distribution. As with quantification, other factors that disrupt random distribution of template, like sample inhomogeneity, will also cause colocalization and may cause artifacts. Conversely, negative associations will also be incorrectly measured if the linked molecules are separated. Consequently template quality and the physical distance of the target sequences of interest must be considered in data analysis.

Translation of dPCR to Routine Molecular Diagnostics

The earliest impact dPCR is likely to have on patient care will be through the value assigning of reference materials (44, 45) to support existing molecular diagnostics. Next will be the potential to use dPCR for routine calibration of the standards used in qPCR. Whether these are simple in-house standards or tied to other reference materials, for many of the reasons discussed above using dPCR is likely to improve the accuracy of assigning calibrators and thus the reproducibility of qPCR.

Many of the considerations discussed above will need to be addressed for dPCR to be routinely used as a diagnostic tool. There are, however, also some rudimentary issues that must be considered if the impact of dPCR on patient care is to be maximized. One of the key advantages of dPCR is that, unlike its main competitor next generation sequencing, most of the reagents have been developed by the demands of qPCR users and the response of some industrious biotechnology companies. qPCR has set the path and is increasingly accepted as a diagnostic tool; however, there are some key factors that we also predict will be necessary to implement for dPCR to maximize its clinical impact.

Greater Reaction Volumes and a Higher Dynamic Range

qPCR has started to define what is possible for molecular diagnostics, and this is particularly pertinent for clinical molecular quantification because it is versatile in many ways; one key advantage is that qPCR is readily scalable. Consequently, although dPCR has the potential to be more sensitive than qPCR when sample volumes are matched, qPCR will always have the edge if sensitivity can be improved by performing a larger-volume reaction (46, 47). With the use of most currently available instruments, dPCR is not readily scalable; a sample can be partitioned across multiple reactions to increase sensitivity (and dynamic range), but this adds complexity and cost.

qPCR is also capable of measuring a very wide dynamic range (\( >9 \) orders of magnitude), which means that essentially the same PCR setup can be used to quantify analytes that are present clinically at high, medium, and low abundance. This is not currently possible with the majority of the commercially available dPCR instruments, because a different dilution strategy is required for high and low analytes to ensure the instruments are not saturated, which complicates the protocol. To simplify this, what is needed are dPCR instruments with at least 100,000 partitions, which will begin to compete with what qPCR has to offer.

Cheaper, Faster, Simpler Instruments with Automated Analysis of Results

Although this requirement can be seen as stating the obvious, a specific point is noteworthy when dPCR is compared with qPCR. The automated reading associated with qPCR is a major advantage when applying this method clinically, such that it is often used for nonquantitative molecular analysis. Because dPCR does not need to be measured in real time, to measure the amplification curve the fluorescence reading need not be as complex. Consequently this offers the potential for dPCR instruments that offer automated reading while also being simpler than those required for qPCR.

Ideally this will lead to cheaper formats which will concomitantly reduce the turnaround time by reducing the time for instrument setup and analysis.

High Throughput vs Near Patient

qPCR has been applied in both high-throughput (supported by robotic automation and/or miniaturization) and lower-throughput “near-patient” formats. Although these developments arguably represent the pinnacle of quantitative clinical molecular measurement, when using qPCR to quantify, both are at the mercy of the amplification plot and \( C_q \) value. Furthermore, because more automation (for high throughput or near patient) results in the need for less expert input, the increased dependence on computer algorithms required to determine whether a PCR amplification
curve is suitable increases the challenge. The fact that dPCR is inherently simpler to measure, coupled with improved reproducibility, will lend itself far better to the analysis associated with automation than qPCR.

Terminology

One of the factors that is likely to hinder translation of any promising technology is the popular use of specific terms and/or acronyms among the early protagonists. One of the purposes of the dMIQE guidelines (33) was to use specific terminology to facilitate a common language and reduce confusion. The use of acronyms has its place, but they must be used sparingly and with a specific purpose. What is increasingly unclear is why there is a need to specify the fact that some instruments apply a “chip digital PCR” format (termed cdPCR) and others a “droplet digital PCR” format (termed ddPCR) when one of the increasingly key findings is the impressive concordance between the 2 formats.

When qPCR was developed it was not felt necessary to highlight whether the instrument used glass capillaries or a 96-well plate, which begs the question why this is necessary for dPCR. The preparation required to set up different dPCR formats can be very distinct, but this is not limited to chip and droplet formats and we advocate comprehensive description of experimental procedures in the Material and Methods or Supplemental Materials sections of any publication reporting the use of dPCR. To remove this confusion we would advocate using just the term “dPCR” (rather than cdPCR or ddPCR), with the format being apparent from the description of the manufacturer and make of the instrument. This can only assist in increasing the understanding and acceptance, and thus the likelihood, of eventual translation of such methods to the clinic.

Final Thoughts

If dPCR is developed into clinical tools it is likely to be initially most successful in the lower-throughput formats, but if the procedure is simplified to improve automation then there is no reason why high-throughput dPCR could not also be applied both in specialist areas where dPCR excels, like SNP detection, as well as routine quantification, like with viral load monitoring. When compared to qPCR the advantages of dPCR are many, whereas the only real disadvantages are those associated with any young technology. This an interesting time for molecular diagnostics; if those who are developing sequencing technology rapidly succeed then it is likely that PCR-based methods will be superseded for clinical applications. We predict, however, that routine clinical application of sequencing-based technologies is likely to take at least 10 years to realize. If this is correct, then PCR still has much to offer the clinic and dPCR is in a timely place to have a major impact on improving robust reproducible molecular diagnostics.

References

7. Last AR, Roberts C, Cassama E, Nabicass M, Molina-Gonzalez S, Burr SE, et al. Plasmid copy number and disease severity in naturally occur-