Risk-Benefit Analysis of Sampling Methods for Fine-Needle Aspiration Cytology

A Mathematical Modeling Approach

Robert L. Schmidt, MD, PhD, MBA,[1] Michal A. Kordy, MS,[2] Kirsten Howard, PhD,[3] Lester J. Layfield, MD,[4] Brian J. Hall, MD,[5] and Douglas G. Adler, MD[6]*

Key Words: Fine-needle aspiration cytology; Risk-benefit analysis; Modeling; Adverse events

DOI: 10.1309/AJCPEAKR4MO2GQBO

Abstract

The effectiveness of fine-needle aspiration (FNA) increases with the number of needle passes, but needle passes are also associated with increased risk of adverse events. The trade-off between needle passes and adequacy has not been well characterized. Clinical studies are limited because of their inherent variability and limited sample size. We developed mathematical models to compare the performance of a variety of sampling protocols under a wide range of conditions. Specifically, we compared the performance of sampling methods using a fixed number of needle passes with sampling methods using a rapid onsite evaluation (ROSE) with a variable number of needle passes. Variable sampling with ROSE generally required fewer needle passes than fixed sample size policies to achieve a desired adequacy rate. Variable sampling policies using ROSE achieve greater per-case adequacy with fewer needle passes than sampling policies using a fixed number of passes if assessor accuracy is high.

Fine-needle aspiration cytology (FNAC) is a widely used sampling technique in clinical medicine. FNAC is a complex multistep process, and sampling can be performed by a number of different methods. Despite the long-standing role FNAC has played in the diagnosis of many disorders, no gold standard approach to the process exists, and institutional practices vary widely, with individual physicians within a single institution not always adhering to identical protocols.

Physicians performing FNAC at some institutions perform a fixed number of needle passes and then terminate the procedure to await a formal pathologic evaluation. This approach is most commonly employed at sites where rapid onsite evaluation (ROSE) of cytology specimens is not available. At sites that employ ROSE, a variable number of needle passes are often obtained; after each needle pass, the tissue is evaluated by a cytologist for a variety of findings, including cellular adequacy and the presence of diagnostic tissue. At sites that use ROSE, there is often a dynamic interaction and interplay between the physician performing the FNA and the cytologist during the procedure, with the former supplying information about the location and appearance of the lesion in question and the latter supplying information about the appearance of aspirated specimens, suggestions regarding the site of successive FNAs based on the appearance of aspirated tissue obtained, and the need (or lack thereof) for further samples to reach a diagnosis.

Sample adequacy is an important component of FNA performance and is frequently reported in FNA studies. In general, sample adequacy is expected to increase with the number of needle passes, and the impact of the number of passes has been investigated in a small number of studies.[1-9] However, the risk of adverse events, although small overall,

increases with the number of needle passes, and the benefits of an adequate sample must be balanced against the potential risk of an adverse event (both in terms of patient harm and monetary costs). The trade-off between adverse events and successful FNA can be modified by the use of ROSE, but the relationship between the per-case adequacy rate and needle passes has not been demonstrated in clinical studies. It is unlikely that a single study could obtain a sufficient number of cases to fully reveal these relationships.

Sampling can be viewed as a process with binary outcomes (success vs failure) that are governed by laws of probability. Each needle pass is associated with a probability of success. Because it is governed by the laws of probability, the sampling process is amenable to mathematical modeling. Modeling has the advantage of not being subject to the site-to-site variation that commonly complicates clinical studies. Modeling studies can reveal relationships that could not otherwise be examined because it is possible to compare the impact of several different variables (per-pass probability, ROSE accuracy, and number of needle passes) in a single study. We therefore developed mathematical models to compare the trade-off between adequacy and needle passes for several different methods of endoscopic ultrasound FNA sampling. Specifically, we compared outcomes when either a fixed number of passes or a variable number of passes using ROSE were utilized during FNA.

Materials and Methods

Conceptual Model

We compared 2 categories of sampling policies that we designated as fixed and variable. In both policy types, sampling continues until a stopping point is reached. For a fixed policy, samples are not evaluated for adequacy, and sampling is stopped when the number of required samples is reached or an adverse event occurs (whichever comes first). In a variable policy using ROSE, each sample is sequentially evaluated for adequacy by a pathologist or cytotechnologist as tissue is obtained from the patient. Sampling is stopped after the cytologist observes the required number of adequate samples, after the occurrence of an adverse event, or after reaching the maximum number of passes. A sampling policy is determined by the policy type (fixed vs variable), the number of required samples (number of passes vs number of observed adequate samples), and the use of ROSE.

Model Details and Notation

In a fixed sampling method, a set number, n_f, of samples are taken and the samples are not assessed for adequacy. In a variable sampling method, samples are assessed for adequacy, and sampling is discontinued after a designated number, n_o, of samples is observed to be adequate. Variable sampling corresponds to ROSE. In a variable policy with sample limit, sampling is discontinued after a designated number, n_o, of samples are observed to be adequate or after a maximum number, M, of samples are obtained, whichever occurs first. This type of sampling could occur if the aspirator is concerned about limiting the number of needle passes or the total procedure time.

We designate sampling policies as follows. Let $\pi(C, n)$ designate a sampling policy, where C indicates the policy category (F = fixed, V = variable), and n indicates the number of required samples (nf for a fixed policy, n_o for a variable policy). We use the symbol, π, to designate a generic policy. The above notation is useful for describing categories of policies, but we sometimes use a simpler notation to describe a particular sampling method. We will indicate a particular policy using the notation Cn. For example, F2 indicates a fixed policy with 2 passes, and V1 indicates a variable policy that stops after 1 adequate sample is observed with no limit on the number of needle passes.

Each sample has a per-pass probability of adequacy, p. The per-pass probability depends on a variety of factors such as the anatomic site (eg, pancreas vs thyroid), lesion characteristics (size, cystic vs solid), operator experience, and technical factors such as needle size and the use of guidance. Our model does not attempt to determine the value of the per-pass probability of adequacy but takes this as an input **Figure 1**. In principle, the per-pass probability can be estimated from published studies or by retrospective review at a particular site. Our model allows p to vary from case to case (depending on the factors listed above) but assumes that the per-pass probability remains constant within a particular case. This is a simplifying assumption and is justified because most factors affecting p (eg, anatomic site, aspirator experience, tumor size) remain constant within a single case. It is possible that the per-pass adequacy could depend on the number of needle passes (eg, prior needle passes could disturb the tissue and decrease the probability of success of subsequent passes), but we are unaware of any data showing that the per-pass probability of success depends on the number of needle passes.

The samples are evaluated by an assessor. The assessor may be a pathologist, cytotechnologist, or a nonpathologist clinician. The assessor reviews samples in real time, and the accuracy of the assessor is compared with a processed sample that has been reviewed by a pathologist, which is considered the reference standard (gold standard). The accuracy of the assessor is characterized by the sensitivity, θ_1, and specificity, θ_2, of the assessment of adequacy relative to the gold standard. We make no presumption regarding the relative accuracy of various personnel. The assessor accuracy can be determined by case review.

DOI: 10.1309/AJCPEAKR4MO2GQBO

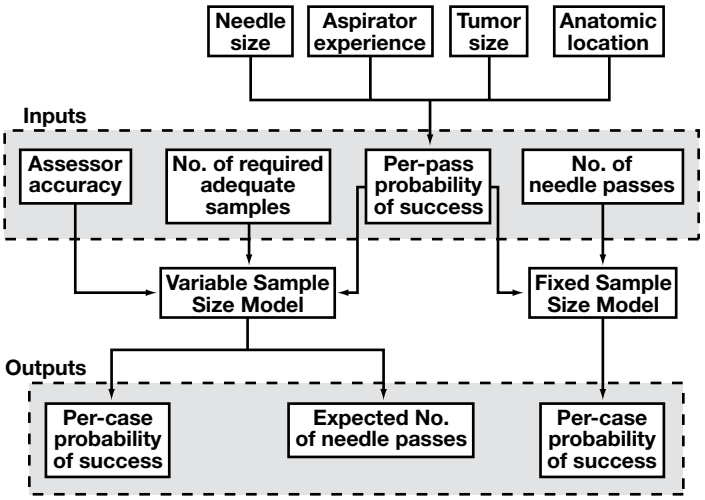

Figure 1 Comparison of inputs and outputs of fixed and variable sampling models. The inputs to the variable model are assessor accuracy (θ_1 = sensitivity, θ_2 = specificity); the required number of observed adequate samples, n_o; and the per-pass probability of an adequate sample, p. In the fixed model, the input parameters are the per-pass probability of an adequate sample, p, and the number of needle passes, nf. The per-pass adequacy rate, p, is an input that is common to both models. The factors that determine p are external to the model.

Two outcomes are associated with each sample: the actual adequacy, A_a, and the observed adequacy, A_o. The actual adequacy is the final determination (gold standard). The observed adequacy is the assessment made during ROSE. Each outcome has 2 possible values: adequate or not adequate. We will use the following notation:

Equation 1

$$P(A_a) = P(\text{sample is actually adequate}) = p$$

Equation 2

$$P(\bar{A}_a) = P(\text{sample is actually not adequate}) = (1-p)$$

Equation 3

$$P(A_o) = P(\text{sample is observed as adequate}) =$$
$$P(A_o|A_a)P(A_a) + P(A_o|\bar{A}_a)P(\bar{A}_a) = \theta_1 p + (1-\theta_2)(1-p)$$

Equation 4

$$P(\bar{A}_o) = P(\text{sample is observed as not adequate})$$
$$= P(\bar{A}_o|A_a)P(A_a) + P(\bar{A}_o|\bar{A}_a)P(\bar{A}_a) = (1-\theta_1)p + \theta_2(1-p)$$

The positive predictive value, PPV, and negative predictive value, NPV, of the assessments are

Equation 5

$$PPV = P(A_a|A_o) = \frac{P(A_a \cap A_o)}{P(A_o)} = \frac{\theta_1 p}{\theta_1 p + (1-\theta_2)(1-p)}$$

Equation 6

$$NPV = P(\bar{A}_a|\bar{A}_o) = \frac{P(\bar{A}_a \cap \bar{A}_o)}{P(\bar{A}_o)} = \frac{\theta_2(1-p)}{\theta_2(1-p) + (1-\theta_1)p}$$

For each sampling method, we compare the following statistics: the probability of success, $P(S)$, and the number of needle passes, N. We define sampling as successful if at least one of the samples is truly adequate:

Equation 7

$$P(S) = P(N_a \geq 1)$$

where N_a is the number of actually adequate samples. We define this as the per-case probability of success as opposed to the per-pass probability of success. The probability of success can be conditioned on a particular sampling method, π:

Equation 8

$$P(S|\pi) = P(N_a \geq 1|\pi)$$

We also calculate the probability of a particular number of needle passes associated with a particular policy as $P(N|\pi)$ and the expected number of needle passes as $E(N|\pi)$, as well as the variance of the number of needle passes given a particular policy, $\text{Var}(N|\pi)$.

Mathematical Model for FNA Sampling

Fixed-Sampling Method

The probability of obtaining a truly adequate sample follows binomial distribution with per-pass probability of success, p, and n_f trials.

The per-case probability of success (obtaining at least 1 adequate sample) is

Equation 9

$$P\left(S|\pi(F,n_f)\right) = 1-(1-p)^{n_f}$$

The expected number of needle passes per case is

Equation 10

$$E\left(N|\pi(F,n_f)\right) = \bar{N}_f = n_f$$

DOI: 10.1309/AJCPEAKR4MO2GQBO

The variance of the number of needle passes per case is

Equation 11

$$Var\left(N|\pi(F,n_f)\right)=0$$

A flow diagram showing the inputs and outputs of the fixed model is shown in Figure 1.

Variable Sampling Method (ROSE)

The variable sampling method follows a negative binomial distribution with probability of success, $P(A_o)$, and number of successes, n_o.

The per-case probability of success (obtaining at least 1 adequate sample) is

Equation 12

$$P\left(S|\pi(V,n_o)\right)=1-P(failure)=\ 1-P(N_a=0)$$

Suppose n_o apparently adequate samples are observed by the ROSE assessor after N trials. Given $N = n$ trials, the probability of failure (ie, that none of the samples are actually adequate) is

Equation 13

$$P(failure|N=n)=\ [P(\bar{A}_a|A_o)]^{n_0}[P(\bar{A}_a|\bar{A}_o)]^{n-n_0}$$

Moreover, probability that there will be exactly n trials is given by a negative binomial probability mass function, where the probability of success is the probability of the sample being assessed as adequate, $P(A_o)$:

Equation 14

$$P(N=n)=\ \binom{n-1}{n_0-1}P(\bar{A}_o)^{n-n_0}P(A_o)^{n_0}$$

Equation 15

$$P(failure)=\sum_{n=n_o}^{\infty}P(N=n)\ P(failure|N=n)$$

Equation 15.1

$$=\sum_{n=n_o}^{\infty}\binom{n-1}{n_0-1}P(\bar{A}_o)^{n-n_0}P(A_o)^{n_0}[P(\bar{A}_a|\bar{A}_o)]^{n-n_0}[P(\bar{A}_a|A_o)]^{n_0}$$

Equation 15.2

$$=\sum_{n=n_o}^{\infty}\binom{n-1}{n_0-1}P(\bar{A}_a\cap\bar{A}_o)^{n-n_0}P(\bar{A}_a\cap A_o)^{n_0}$$

Equation 15.3

$$=\ P(\bar{A}_a\cap A_o)^{n_0}\sum_{n=n_o}^{\infty}\binom{n-1}{n_0-1}P(\bar{A}_a\cap\bar{A}_o)^{n-n_0}$$

Equation 15.4

$$=\frac{P(\bar{A}_a\cap A_o)^{n_0}}{(n_0-1)!}\sum_{n=n_o}^{\infty}(n-1)(n-2)\ldots(n-n_0+1)P(\bar{A}_a\cap\bar{A}_o)^{n-n_0}$$

Equation 15.5

$$\overset{m}{=}\frac{P(\bar{A}_a\cap A_o)^{n_0}}{(n_0-1)!}\left(1-P(\bar{A}_a\cap\bar{A}_o)\right)^{-n_0}(n_0-1)!$$

Equation 15.6

$$=\left(\frac{P(\bar{A}_a\cap A_o)}{1-P(\bar{A}_a\cap\bar{A}_o)}\right)^{n_0}$$

The step denoted with *m* is a consequence of the fact that for $|x| < 1$,

Equation 16.1

$$\sum_{n=n_o}^{\infty}(n-1)(n-2)\ldots(n-n_0+1)x^{n-n_0}=\frac{d^{n_0-1}}{dx^{n_0-1}}\left(\sum_{n=0}^{\infty}x^n\right)$$

Equation 16.2

$$=(1-x)^{-n_0}(n_0-1)!$$

The overall probability of success is given by

Equation 17.1

$$P\left(S|\pi(V,n_o)\right)=1-P(failure)=\ 1-\left(\frac{P(\bar{A}_a\cap A_o)}{1-P(\bar{A}_a\cap\bar{A}_o)}\right)^{n_0}$$

Equation 17.2

$$=1-\left(\frac{(1-\theta_2)(1-p)}{1-\theta_2(1-p)}\right)^{n_0}$$

The expected number of needle passes per case is

Equation 18

$$E(N|\pi(V,n_o))=\bar{N}_v=\frac{n_o}{P(A_o)}$$

The variance of the number of needle passes per case is

Equation 19

$$Var(N|\ \pi(V,n_o))=\ n_o\frac{P(\bar{A}_o)}{P(A_o)^2}$$

A flow diagram showing the inputs and outputs of the variable model is presented in Figure 1.

Bayesian Updating of the Per-Pass Probability

In our model, an estimate of the per-pass probability of success, p, is provided as an input. This value, along with the accuracy of the assessor, is used to estimate the number of needle passes that would be required to observe an adequate sample when using a variable (ROSE) sampling policy. The model also estimates the probability that an adequate sample could be obtained in n_f more needle passes. These estimates are based on the assumption that the per-pass probability of success is constant and ignores the information obtained from previous attempts. In addition, the estimate of p is imprecise. The per-pass probability of success depends on many factors that vary from case to case (eg, size of tumor, experience of the aspirator). The imprecision of the initial estimate of p can be represented by providing a probability distribution for p. It turns out that the β distribution

Equation 20.1

$$g(p|a,b)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}p^{a-1}(1-p)^{b-1}$$

is a flexible prior distribution for p and is the conjugate prior distribution of the binomial distribution. The average of the β distribution is given by

DOI: 10.1309/AJCPEAKR4MO2GQBO

Equation 20.2

$$\bar{p} = \frac{a}{a+b}$$

Given y successes out of n needle passes, the revised distribution for p (ie, the posterior probability distribution) is given by

Equation 20.3

$$g(p|a,b,y,n) = \frac{\Gamma(n+a+b)}{\Gamma(y+a)\Gamma(n-y+b)} p^{y+a-1}(1-p)^{n-y+b-1}$$

and the revised estimate for p, p', is given by

Equation 20.4

$$p' = \frac{y+a-1}{y+a-1+n-y+b-1} = \frac{y+a-1}{n+a+b-2}$$

For example, suppose the initial estimate for the per-pass probability for a particular case is 0.5. This can be represented by a β distribution with parameters $a = 5$, $b = 5$. Now suppose that n_o adequate samples have been observed after 3 needle passes ($y = 0$, $n = 3$). The per-pass probability would be revised downward to

$$p' = \frac{y+a-1}{n+a+b-2} = \frac{0+3-1}{3+3+3-2} = 0.29$$

Given this new estimate of p, the probability of obtaining an adequate sample in n_f additional attempts is given by Equation 9, and the expected number of additional needle passes required to observe n_o adequate samples is given by Equation 18.

Risk-Benefit Analysis

This study is a risk-benefit analysis in which risk is expressed in terms of needle passes and benefit is expressed in terms of the per-case adequacy rate (ie, the probability of obtaining at least 1 adequate sample from the set of collected samples). We use economic terminology to describe the relationships between different policies.[10]

Policies are compared along 2 dimensions: risk (needle passes) and benefit (per-case probability of success). A policy is *absolutely dominated* if it provides both higher risk and lower benefit when compared with another sampling policy. In other words, a policy is absolutely dominated if another policy is superior in both dimensions (risk and benefit). When plotted, the risk-benefit combinations of strategies that are not dominated form an *efficiency frontier*. The efficiency frontier represents the sampling alternatives with the most favorable risk-benefit ratios (ie, fewest needle passes for a given per-case adequacy rate). The slope of the line connecting any 2 policies corresponds to the incremental risk-benefit (or efficiency) of moving from one policy to an alternative policy.

Results

The per-case probability of success increases as the number of needle passes increases in both fixed and variable sampling policies **Figure 2**. Variable sampling policies

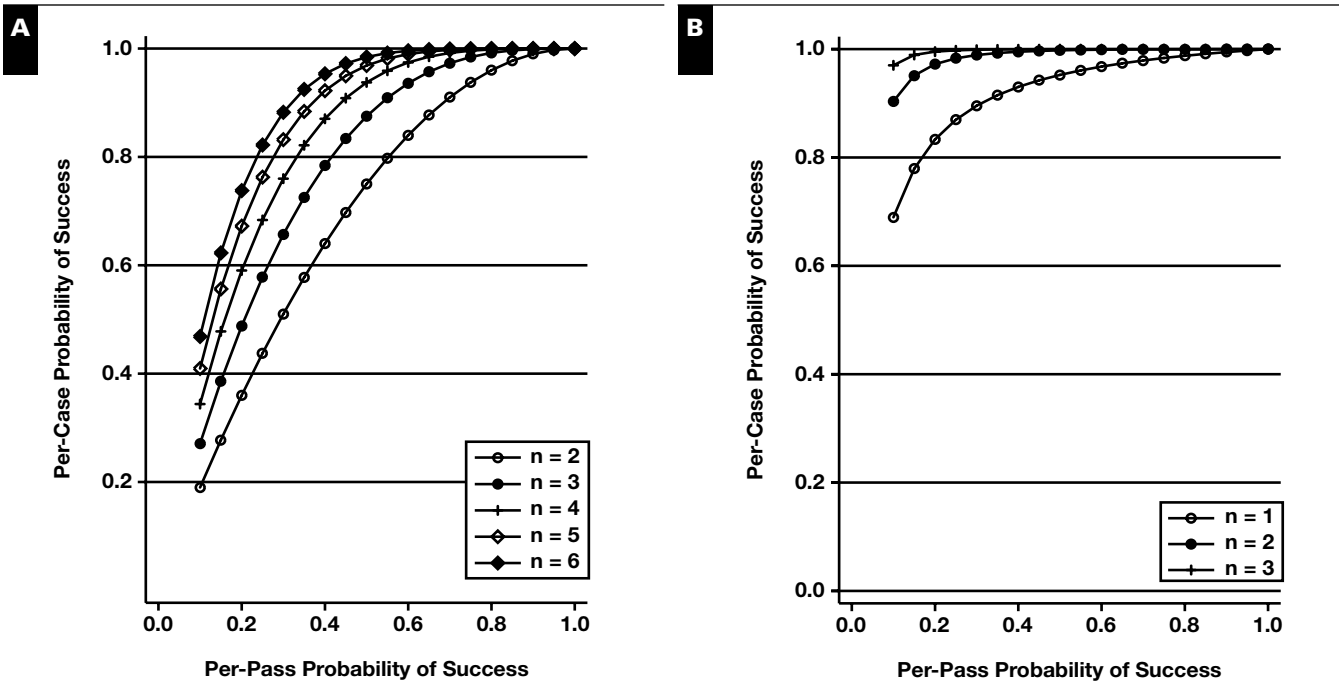

Figure 2 Effect of per-pass probability on per-case adequacy. The per-case probability of success (ie, obtaining at least 1 adequate sample) varies with the per-pass adequacy rate. **A** shows the results for fixed-sampling policies, and **B** shows the results for variable sampling using rapid onsite evaluation. The number, n, corresponds to the number of needle passes (fixed policy) or the required number of observed as adequate samples (variable policy).

DOI: 10.1309/AJCPEAKR4MO2GQBO

with n_o greater than 1 achieve high levels of adequacy even when the per-pass adequacy rate is low. The number of required needle passes increases at an increasing rate as the per-pass adequacy rate decreases **Figure 3**. The variability in the number of needle passes also increases as the per-pass adequacy decreases **Figure 4**.

FNA is successful if at least 1 adequate sample is obtained. In general, a greater number of needle passes are required to achieve a higher per-case success rate **Figure 5**. The difference between policies depends on the per-pass success rate. As shown in Figure 5, there were significant differences in the sampling policy performance when the per-pass

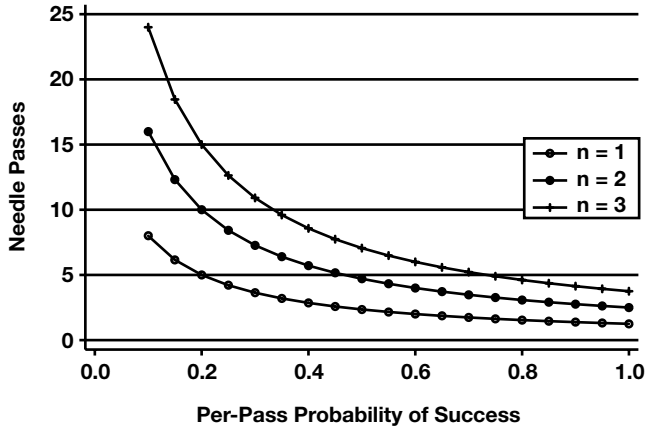

Figure 3 Expected number of needle passes in variable sampling. The figure shows the average number of needle passes that would be required to observe n_o adequate samples in a variable sampling policy.

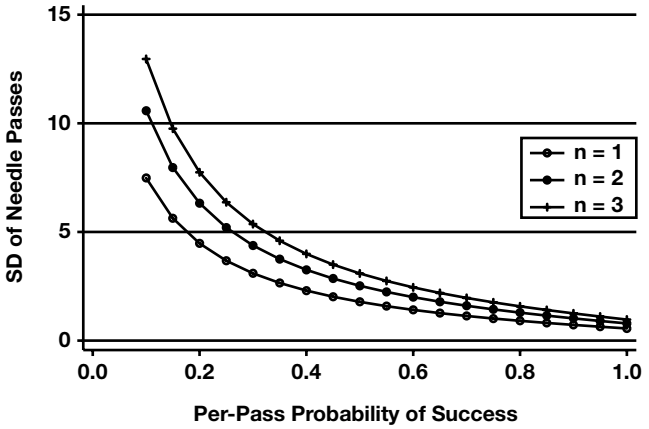

Figure 4 Uncertainty in needle passes in variable sampling. The figure shows the standard deviation of the expected number of needle passes required to observe n adequate samples.

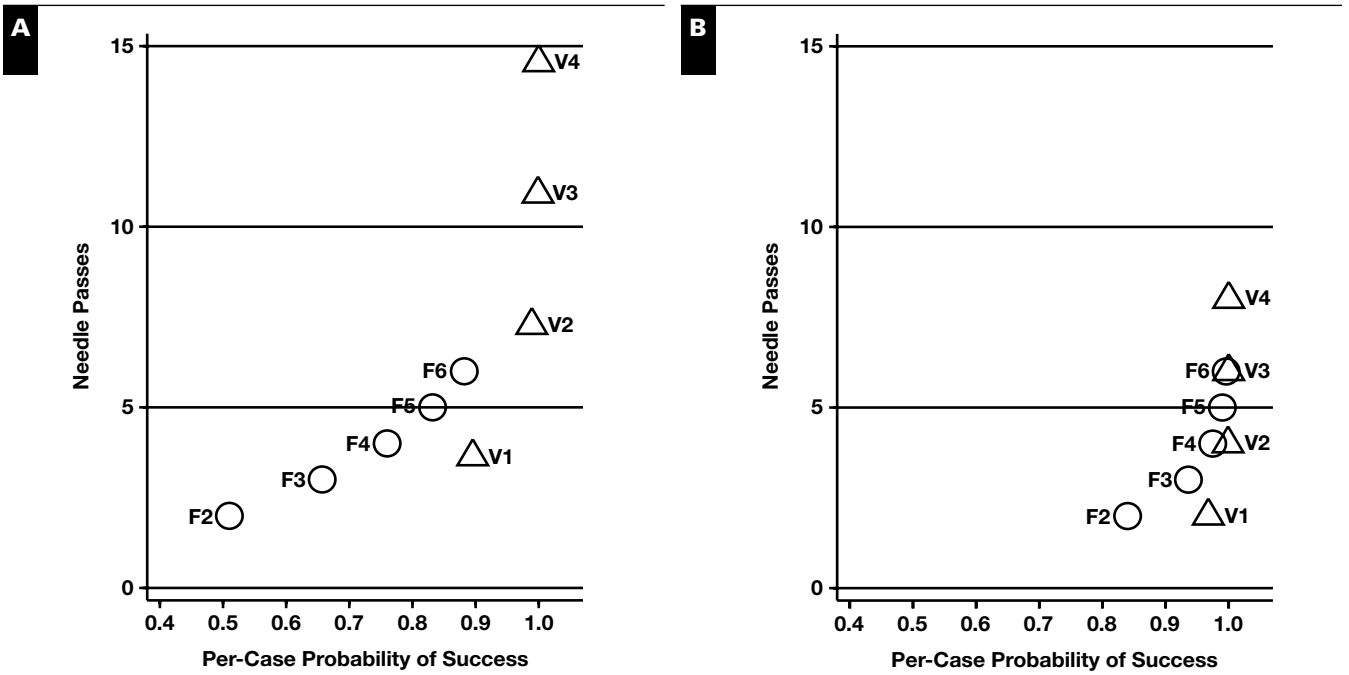

Figure 5 Sampling policy comparison. The figure shows the average performance for each policy as determined by the per-case probability of success and the expected number of needle passes per case. Each policy is given a 2-character code. The first character indicates the policy type (F = fixed, V = variable), and the second character indicates the number of needle passes (fixed policy) or the required number of observed adequate samples (variable policy). Sensitivity was set at 0.8 and specificity set at 0.95. The per-case probability of success was 0.3 (**A**) and 0.6 (**B**).

DOI: 10.1309/AJCPEAKR4MO2GQBO

adequacy was low (P = .3); however, the differences in performance were relatively small when the per-pass adequacy was high (P = .6).

A fixed policy with 2 needle passes had the lowest probability of success and the lowest number of needle passes (F2; Figure 5). When the per-pass probability of success was low (P = .3), variable sampling policies were superior to fixed policies. Policy V1 provided a higher probability of success with fewer needle passes than any fixed policy. The probability of success could be increased by selecting a variable policy with 2 adequate samples (V2); however, the incremental risk/return ratio was approximately 0.3 needle passes per percent gain in probability of success. A similar pattern was seen when the per-pass adequacy rate was high (P = .6).

Model Validation

Relatively few studies have published results showing how the adequacy rate varies with the number of needle passes. Diacon et al[6] published data for transbronchial FNA and found that an adequate sample was obtained on 64.5% of the initial needle passes. Using this value as an estimate of p, we compared the predictions of our fixed model with the results observed by Diacon et al[6] **Table 1** and found a statistically significant fit ($\chi^2 = 2.2 \times 10^{-7}$, $P < .001$).

Table 1
Comparison of Observed and Predicted Per-Case Adequacy Rate

No. of Passes	Observed	Predicted
1	0.645	0.645
2	0.874	0.874
3	0.955	0.955
4	0.984	0.994

Discussion

Our study provides mathematical formulas that show how the overall probability of success, expected number of needle passes, and the variability in the number of needle passes depend on the per-pass adequacy rate, the sampling policy, and the accuracy of the assessor in ROSE. These formulas make it possible to compare a wide range of sampling procedures.

No single sampling policy was strictly superior. The efficiency frontier (ie, the set of most efficient sampling methods) shows that increased per-case adequacy cannot be achieved without increasing risk. In addition, increasing adequacy required more risk (ie, needle passes) as the adequacy rate increased. With the exception of policy F2, we found that variable sampling policies were superior to fixed sampling policies. That is, variable or ROSE-based sampling methods achieve any desired adequacy rate with fewer needle passes than can be achieved using a fixed or non-ROSE sampling method.

The difference in sampling policy performance depends on the per-pass adequacy rate. When the per-pass adequacy rate was low, there were significant differences between methods (Figure 5, P = .3). When per-pass adequacy was high (P = .6), the differences in performance were smaller. This strongly suggests that the advantage of variable sampling methods depends on the underlying per-pass adequacy rate. The per-pass adequacy rate will depend on factors such as the tissue being sampled (thyroid vs pancreas), characteristics of the lesion (size), and FNA operator skill.

Studies have shown that ROSE improves FNA adequacy when sampling the pancreas,[11,12] lung,[13-15] thyroid,[16] soft tissue,[17] head and neck,[18] and breast[19]; however, the studies show considerable variability. Our study shows that the impact of ROSE is strongly dependent on the fixed number of needle passes in the comparator policy. ROSE might be expected to have a large impact in sampling environments with low per-pass adequacy when compared with a fixed sampling policy with a low number of passes (eg, F2 or F3). Otherwise, ROSE would not be expected to show a significant effect.

Our model predicts that FNAC per-case adequacy increases with the number of needle passes. This is consistent with the results of studies that have been conducted in a wide range of anatomic sites (parathyroid gland,[2] abdominal and pelvic organs,[3,5,7] breast,[1] pancreas,[4,8] liver,[20] and lung[6]) using a variety of FNAC techniques (percutaneous,[1,2] endoscopic ultrasound-guided,[3-5] transbronchial,[6] and fluoroscopic and computed tomography–guided[7] FNA). The optimal number of passes recommended from each of these studies has varied from greater than 2[3] to at least 7,[5] with 5 passes being most often recommended. Other studies have disputed the number of needle passes correlating with diagnostic yield specifically regarding renal FNAs.[9] However, it has been agreed upon by the experts at the National Cancer Institute's State of the Science conference in October 2007 that in every single study of thyroid FNA in which different numbers of passes were compared, the more passes performed, the higher the adequacy rate (up to 12 passes).[21] Their consensus was that 5 passes were optimal, because beyond that number the small increment in diagnostic material was outweighed by the potential for harm to the patient. It can be concluded from all of these studies that increasing the number of passes up to a point is directly associated with an increased diagnostic yield and therefore greater diagnostic accuracy with regard to FNA. Our study adds to this knowledge by comparing the performance of different sampling policies (the number of needle passes, use of ROSE) in different sampling environments. In

DOI: 10.1309/AJCPEAKR4MO2GQBO

particular, our study shows the impact of the number of needle passes on FNAC effectiveness over a wide range of per-pass adequacy rates.

Validation of our fixed-sample model requires data showing the per-case adequacy as a function of needle passes. Although studies often report adequacy rates, very few studies report the adequacy rate as a function of needle passes. We found only 1 study that reported the effect of needle passes on per-case adequacy using a fixed-sample policy.[6] The predictions of our model showed almost perfect correspondence (Table 1). Our fixed-sample model assumes that the per-pass adequacy rate does not change with the number of needle passes. It is possible that the per-pass needle adequacy rate decreases with increasing needle passes due to bleeding or other disruptions of the tissue; however, we are aware of no data showing such an effect. The correspondence of our model with clinical data suggests that the assumption of constant per-pass adequacy is correct in the case of transbronchial FNA.

Our model does not attempt to predict the effect of factors such as needle size, aspirator experience, anatomic location, or tumor size on the per-pass adequacy rate, p. Rather, our model takes the per-pass adequacy rate as an input and uses it to compare the performance of various sampling policies using the assumed per-pass adequacy rate. It is not necessary to have precise estimates of p to use our model. For example, one might use the model to compare the expected performance of ROSE against a non-ROSE fixed sampling policy in an environment in which the per-pass adequacy rate, p, is likely to be low. Figures 1 and 5 provide a broad comparison of the expected performance of ROSE vs non-ROSE sampling policies under such a scenario. Many of the factors that affect p (anatomic site, tumor size, aspirator experience) would be constant in such a comparison and would not be affected by the presence of an assessor. Thus, even if p is not precisely known, our model allows for valid comparisons between sampling policies. Our model is intended to compare the *relative* performance of sampling policies for a given value of the per-pass adequacy rate (Figure 5) rather than to predict the performance of a single sampling policy.

Although our model assumes a constant per-pass adequacy rate, p, this estimate can be updated by incorporating real-time data. For example, one might assume a value for p based on historical data sampling a particular type of tumor at a particular site (endoscopists sampling solid pancreatic tumors at a particular hospital). The value of p is based on an average of past history; however, each case is unique. Our model shows how to use Bayesian updating to revise the estimate of p in light of failed needle passes. Such updating can indicate when further sampling is likely to be futile.

Our study is limited because it only captures the trade-off between effectiveness and needle passes. A more complete study would include the monetary costs associated with failed sampling (cost of open biopsy, resampling, or surgery to obtain a biopsy specimen), the cost of adverse events, and the cost of service provision (eg, variable cost per needle pass, fixed cost of FNA session) in a cost-effectiveness study.

We focused our investigation on scenarios with relatively high assessor accuracy (sensitivity = 80%, specificity 95%). These assessor accuracy levels are similar to those commonly reported in the literature.[22-24]

Our approach has several strengths. Our modeling approach eliminates unwanted sources of variation (patient-to-patient variation, site-to-site variation, etc). This enabled us to demonstrate the trade-off curve between per-case FNA effectiveness and needle passes for a range of different sampling policies. It is unlikely that a clinical study would be able to obtain enough cases or control variation sufficiently to reveal these relationships. For example, a sample size of approximately 400 cases (200 in each arm) would be required to demonstrate a 10% difference in adequacy. Relatively few studies obtain sample sizes of this size.[25,26]

FNA adequacy can be analyzed with mathematical models. We found that variable sampling methods (ROSE) are generally superior to fixed sampling methods because they require fewer needle passes for a given level of per-case adequacy. The advantage of variable sampling is greatest when the per-pass adequacy rate is low.

From the [1]Department of Pathology and [4]Department of Surgical Pathology, University of Utah School of Medicine and ARUP Laboratories, Salt Lake City, UT; [2]Department of Mathematics, University of Utah, Salt Lake City, UT; [3]Department of Health Economics, Sydney School of Public Health, University of Sydney, Sydney, Australia; [5]Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX; and [6]Gastroenterology Division, University of Utah School of Medicine, Salt Lake City, UT.

[*]Dr Layfield is now with the Department of Pathology and Anatomical Sciences, University of Missouri, Columbia.

Address reprint requests to Dr Schmidt: Dept. of Pathology, University of Utah School of Medicine, 15 N Medical Dr E, Salt Lake City, UT 84112; e-mail: Robert.schmidt@hsc.utah.edu.

References

1. Pennes DR, Naylor B, Rebner M. Fine needle aspiration biopsy of the breast: influence of the number of passes and the sample size on the diagnostic yield. *Acta Cytologica*. 1990;34:673-676.
2. Guerrero MA, Suh I, Vriens MR, et al. The number of needle passes affects the accuracy of parathyroid hormone assay with intraoperative parathyroid aspiration. *Am J Surg*. 2010;200:701-705, discussion 705-706.
3. Thomas T, Kaye PV, Ragunath K, et al. Efficacy, safety, and predictive factors for a positive yield of EUS-guided Trucut biopsy: a large tertiary referral center experience. *Am J Gastroenterol*. 2009;104:584-591.

DOI: 10.1309/AJCPEAKR4MO2GQBO

4. Möller K, Papanikolaou IS, Toermer T, et al. EUS-guided FNA of solid pancreatic masses: high yield of 2 passes with combined histologic-cytologic analysis. *Gastrointest Endosc*. 2009;70:60-69.
5. LeBlanc JK, Ciaccia D, Al-Assi MT, et al. Optimal number of EUS-guided fine needle passes needed to obtain a correct diagnosis. *Gastrointest Endosc*. 2004;59:475-481.
6. Diacon AH, Schuurmans MM, Theron J, et al. Transbronchial needle aspirates: how many passes per target site? *Eur Respir J*. 2007;29:112-116.
7. Gothlin JH, Gadeholt G. Percutaneous fine needle biopsy of abdominal and pelvic lesions: passes necessary for secure diagnosis with fluoroscopy and CT-guidance. *Eur J Radiol*. 1986;6:288-290.
8. Cleveland P, Gill KRS, Coe SG, et al. An evaluation of risk factors for inadequate cytology in EUS-guided FNA of pancreatic tumors and lymph nodes. *Gastrointest Endosc*. 2010;71:1194-1199.
9. Andonian S, Okeke Z, Okeke DA, et al. Number of needle passes does not correlate with the diagnostic yield of renal fine needle aspiration cytology. *J Endourol*. 2008;22:2377-2380.
10. Gray AM, Clarke PM, Wolstenholme J, et al. *Applied Methods of Cost-Effectiveness Analysis in Healthcare*. Oxford, UK: Oxford University Press; 2011.
11. Iglesias-Garcia J, Dominguez-Munoz JE, Abdulkader I, et al. Influence of on-site cytopathology evaluation on the diagnostic accuracy of endoscopic ultrasound-guided fine needle aspiration (EUS-FNA) of solid pancreatic masses. *Am J Gastroenterol*. 2011;106:1705-1710.
12. Klapman JB, Logrono R, Dye CE, et al. Clinical impact of on-site cytopathology interpretation on endoscopic ultrasound-guided fine needle aspiration. *Am J Gastroenterol*. 2003;98:1289-1294.
13. Davenport RD. Rapid on-site evaluation of transbronchial aspirates. *Chest*. 1990;98:59-61.
14. Diette GB, White P Jr, Terry P, et al. Utility of on-site cytopathology assessment for bronchoscopic evaluation of lung masses and adenopathy. *Chest*. 2000;117:1186-1190.
15. Santambrogio L, Nosotti M, Bellaviti N, et al. CT-guided fine-needle aspiration cytology of solitary pulmonary nodules: a prospective, randomized study of immediate cytologic evaluation. *Chest*. 1997;112:423-425.
16. Raab SS, Grzybicki DM, Sudilovsky D, et al. Effectiveness of Toyota process redesign in reducing thyroid gland fine-needle aspiration error. *Am J Clin Pathol*. 2006;126:585-592.
17. Virayavanich W, Ringler MD, Chin CT, et al. CT-guided biopsy of bone and soft-tissue lesions: role of on-site immediate cytologic evaluation. *J Vasc Interv Radiol*. 2011;22:1024-1030.
18. Moberly AC, Vural E, Nahas B, et al. Ultrasound-guided needle aspiration: impact of immediate cytologic review. *Laryngoscope*. 2010;120:1979-1984.
19. Dray M, Mayall F, Darlington A. Improved fine needle aspiration (FNA) cytology results with a near patient diagnosis service for breast lesions. *Cytopathology*. 2000;11:32-37.
20. Skjoldbye B, Horn T, Torp-Pedersen S, et al. Ultrasound guided fine needle aspiration biopsies, from the liver: how many needle passes? *Eur J Ultrasound*. 1996;4:43-47.
21. Pitman MB, Abele J, Ali SZ, et al. Techniques for thyroid FNA: a synopsis of the National Cancer Institute Thyroid Fine-Needle Aspiration State of the Science Conference. *Diagn Cytopathol*. 2008;36:407-424.
22. Schmidt RL, Hall BJ, Wilson AR, et al. A systematic review and meta-analysis of the diagnostic accuracy of fine-needle aspiration cytology for parotid gland lesions. *Am J Clin Pathol*. 2011;136:45-59.
23. Houssami N, Ciatto S, Turner RM, et al. Preoperative ultrasound-guided needle biopsy of axillary nodes in invasive breast cancer: meta-analysis of its accuracy and utility in staging the axilla. *Ann Surg*. 2011;254:243-251.
24. Hewitt MJ, McPhail MJW, Possamai L, et al. EUS-guided FNA for diagnosis of solid pancreatic neoplasms: a meta-analysis. *Gastrointest Endosc*. 2012;75:319-331.
25. Ardengh JC, Lopes CV, de Lima LF, et al. Diagnosis of pancreatic tumors by endoscopic ultrasound-guided fine-needle aspiration. *World J Gastroenterol*. 2007;13:3112-3116.
26. Alsharif M, Carlo-Demovich J, Massey C, et al. Telecytopathology for immediate evaluation of fine-needle aspiration specimens. *Cancer Cytopathol*. 2010;118:119-126.

DOI: 10.1309/AJCPEAKR4MO2GQBO