

Rapid On-Site Evaluation Reduces Needle Passes in Endoscopic Ultrasound-Guided Fine-Needle Aspiration for Solid Pancreatic Lesions: A Risk–Benefit Analysis

Robert L. Schmidt · Brandon S. Walker ·
Kirsten Howard · Lester J. Layfield ·
Douglas G. Adler

Received: 25 March 2013 / Accepted: 5 June 2013 / Published online: 4 July 2013
© Springer Science+Business Media New York 2013

Abstract

Background The effectiveness of endoscopic ultrasound-guided fine-needle aspiration 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. **Aims** The purpose of this study was to compare the risk–benefit tradeoff of different sampling protocols with and without rapid onsite evaluation (ROSE). **Patients and Methods** We used a discrete-event simulation model to compare eight different sampling protocols. Each sampling protocol was simulated 10,000 times to obtain the average performance for each scenario. The

per-pass adequacy rates, ROSE, accuracy of the assessor and sampling limits were varied to determine the impact of these factors on the number of needle passes and adequacy rates.

Results Increasing per-class adequacy can be achieved at a cost of increased needle passes. Sampling with ROSE achieved higher adequacy with fewer needle passes than policies using a fixed number of needle passes without ROSE.

Conclusions Variable sampling policies using ROSE generally achieve greater per-case adequacy with fewer needle passes than non-ROSE sampling policies using a fixed number of passes.

Keywords EUS · FNA · Pancreas · ROSE

R. L. Schmidt
Department of Pathology, University of Utah School
of Medicine, Salt Lake City, UT, USA

R. L. Schmidt · B. S. Walker
ARUP Laboratories, Salt Lake City, UT, USA

K. Howard
The Sydney School of Public Health, University of Sydney,
Sydney, Australia

L. J. Layfield
Department of Pathology and Anatomical Sciences, University
of Missouri, Columbia, MO, USA

D. G. Adler (✉)
Gastroenterology Division, Department of Gastroenterology,
University of Utah School of Medicine, 30 North 1900,
East SOM 4R118, Salt Lake City, UT 84132-2410, USA
e-mail: Douglas.Adler@hsc.utah.edu

D. G. Adler
Department of Internal Medicine, Gastroenterology Division,
University of Utah School of Medicine, Salt Lake City,
UT, USA

Introduction

Endoscopic ultrasound-guided fine-needle aspiration (EUS–FNA) is a widely performed procedure to acquire tissue from extraluminal organs and structures such as solid and cystic lesions of the pancreas, mediastinal and intra-abdominal lymph nodes, and other lesions. EUS–FNA is a complex multi-step process that can be performed using a variety of sampling methods [1].

Two EUS–FNA approaches are commonly utilized. In the first (fixed) method, endoscopists perform a fixed number of needle passes after which samples are formally evaluated. In the second (variable) method, endoscopists perform a variable number of needle passes in a setting where samples are evaluated after each needle pass by an onsite cytologist to ensure diagnostic adequacy.

In either approach, additional needle passes increase both the probability of obtaining an adequate sample and the probability of adverse events. The variable approach,

known as rapid onsite evaluation (ROSE), can increase sample adequacy [2]. ROSE can also reduce the number of needle passes and, potentially, reduce the risk of an adverse event [3, 4]. On the other hand, ROSE is associated with increased costs as it requires an onsite cytologist. Furthermore, although ROSE is becoming increasingly popular, cytologists are still not available at all locations where EUS–FNA is performed. Recent studies have shown that the benefits of ROSE are context dependent [2, 5]. ROSE is unlikely to be beneficial when the per-pass adequacy rate is high because there is little opportunity for improvement. As such, there is a need to further characterize the situations in which ROSE can benefit diagnosis.

Though EUS–FNA is commonly used, no gold standard approach exists. Institutional practices vary widely both between institutions and within single institutions. Published studies lack uniformity and have predominately focused on different needle sizes or technical issues of the specifics of needle passes rather than a fixed versus variable pass needle policy [6–12]. Although the tradeoff between the risk of adverse events and successful EUS–FNA can be modified by the use of ROSE, the relationship between the per-case adequacy rate and the number of needle passes has not been demonstrated in clinical studies.

Additionally, very few controlled studies of ROSE have been performed, and comparisons between institutions are complicated by factors such as sampling methodology (needle size, protocol variation) and case mix. Therefore, it is unlikely that a single study could obtain enough cases to fully reveal these relationships, and the number of clinical studies required to evaluate the comparative effectiveness of ROSE on EUS–FNA would be difficult to obtain.

Endoscopic ultrasound-guided fine-needle aspiration sampling can be viewed as a process with two outcomes: success or failure. Each needle pass is associated with a certain probability of success, and the overall success of a case depends on obtaining at least one successful needle pass among multiple trials.

Mathematical modeling is useful in understanding the probability of success in sampling processes. Modeling has several advantages; it is not subject to the site-to-site variation that commonly complicates clinical studies, it can reveal relationships that could not otherwise be examined, and it provides the opportunity to explore complex processes that cannot be expressed in simple mathematical equations. We have previously developed mathematical solutions for fixed sampling and for variable sampling with an unlimited number of trials [13]. In real life, a limit is often placed on the number of needle passes that can be performed (due to time, manpower issues, etc.), and a closed-form solution does not exist for this case.

The objective of this study was to compare the risk and benefits of ROSE sampling policies relative to non-ROSE

sampling in the specific context of EUS–FNA for solid pancreatic lesions. To that end, we used simulation modeling to evaluate both fixed and variable (ROSE) sampling policies (with and without a sampling limit) to assess the risk–benefit tradeoff between needle passes and sample adequacy for solid pancreatic lesions.

Methods

Model Overview

The model compared two categories of sampling policies which we designated as fixed and variable. For a fixed policy, samples are not evaluated for adequacy and sampling is stopped when the predetermined number of required samples is reached. In a variable policy using ROSE, each sample is evaluated for adequacy by a pathologist or cytotechnologist and sampling is stopped after the required number of adequate samples is observed or after reaching the maximum number of passes.

Adequacy can be defined on a per-pass or per-case basis. In our model, the per-pass adequacy rate is an input. In a fixed sampling policy, the per-case sample adequacy, a_a , is determined by the per-pass adequacy rate, p . In a variable sampling policy (ROSE), there are two potential outcomes associated with each sample: the *actual adequacy*, a_a and the *observed adequacy*, a_o . The actual per-case adequacy is the outcome of interest in this study. The assessor evaluates the sample for adequacy, but may fail to correctly categorize the sample. For example, the assessor may categorize an inadequate sample as adequate or vice versa. Given an adequate sample, the probability of actually observing an adequate sample is determined by the accuracy of the ROSE assessor. The accuracy of the assessor is expressed in terms of the sensitivity, S_n , and specificity, S_p .

$$S_n = P(a_o|a_a) \quad (1)$$

$$S_p = P(\bar{a}_o|\bar{a}_a) \quad (2)$$

An illustration of the ROSE component of the simulation with model probabilities p , S_n and S_p is shown in Fig. 1. As indicated, with each needle pass there is probability (p) that the obtained sample will be adequate (a_a); alternatively there is probability ($1 - p$) that the sample is not adequate (\bar{a}_a). After each sample is collected, it is evaluated by an onsite assessor and observed as either adequate (a_o) or not adequate (\bar{a}_o). The probabilities of an accurate assessment are represented by S_n and S_p while inaccurate assessments by $1 - S_n$ and $1 - S_p$. The process is then repeated until the number of required samples is observed or the maximum number of needle passes is reached.

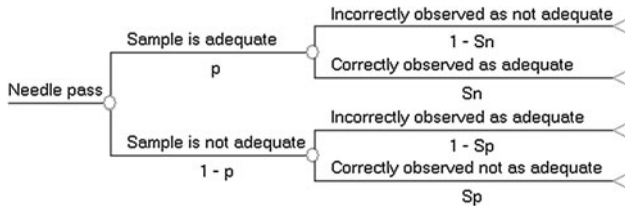


Fig. 1 Simplified decision tree of ROSE component of micro-simulation

Model Parameters

Model parameters are provided in Table 1. The per-pass adequacy rate was estimated to be 60 % using a binomial sampling model provided in Eq. 3, where $P(S)$ is the probability of success (the per-case adequacy rate), p is the per-pass adequacy rate, and n is the number of needle passes [4].

$$P(S) = 1 - (1 - p)^n \tag{3}$$

Based upon a survey of EUS–FNA studies, $P(S)$ and n were set to 93 % and 3, respectively [2]. Solving for p produced a per-pass adequacy rate of approximately 60 %. Based upon reported ranges of $P(S)$ and n a plausible range of 20–80 % was defined in which the parameter was allowed to vary during the simulation. Based upon assessor accuracy data in the literature, the baseline estimate for assessor sensitivity and specificity was to set 95 % [14].

Analysis

The simulated sampling performance (the per-case adequacy rate) is determined by the values of three input parameters: the average per-pass adequacy rate, the average assessor sensitivity, and the average assessor specificity. However, the true values of the input parameters may vary by case and institution. Thus, the relative performance of two sampling policies may be sensitive to variation in these parameters. We used sensitivity analysis to examine how parameter variation influenced model results. This analysis involved variation of one model parameter while holding all others constant at their baseline estimates. All simulations were conducted using the TreeAge Pro 2012 software (Williamstown, MA).

Table 1 Model parameters

Model parameter	Baseline estimate	Sensitivity range
Maximum number of needle passes for ROSE	10	n/a
Per-pass adequacy	60 %	20–80 %
Assessor accuracy	95 %	90–100 %

Results

The sampling policies are described here using a two-character code consisting of a letter followed by a number (e.g., F1). The first character indicates the type of policy; a fixed policy is indicated by “F” and a variable policy by “V.” The second character indicates the required number of samples. For a fixed policy, the number indicates the number of needle passes. For a variable policy, the number indicates the required number of samples that are required to be observed as adequate before sampling is stopped. For example, “F3” indicates a fixed policy that stops after three needle passes. “V1” indicates a variable policy (ROSE) that stops after the assessor observes one adequate sample. Therefore the eight sampling policies compared in the simulation can be summarized as F2, F3, F4, F5, F6, V1, V2 and V3. In most cases, variable policies would stop after observing the first adequate sample. There are rare circumstances in which one might consider policies V2 or V3 (e.g., low per-pass adequacy rate and inexperienced assessor). We included policies V2 and V3 for reference.

Baseline results are provided in Fig. 2. A fixed sampling policy with two, three, four, five, and six needle passes would have per-case adequacy rates of 84.2 % (F2), 93.9 % (F3), 97.4 % (F4), 99.0 % (F5), and 99.6 % (F6). ROSE sampling policies of V1, V2, and V3 have average per-pass adequacy rates of 83, 97.2, and 99.42 % with an average of 1.8, 3.8, and 5.5 needle passes, respectively. Here variable sampling policies are said to strictly dominate fixed polices since variable policies achieve higher per-case adequacy rates with fewer needle passes.

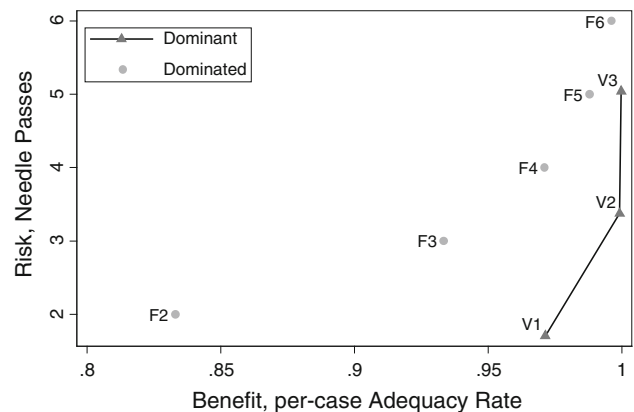


Fig. 2 Risk–benefit analysis fixed sampling versus variable sampling (ROSE). The figure indicates the average adequacy rate and average number of needle passes for different sampling policies. Each sampling policy is designated by a letter (F or V) indicating fixed or variable and a number. For fixed sampling policies, the number indicates the required number of needle passes. For variable sampling policies, the number indicates the required number of observed adequate samples. Each point represents the average outcome associated with a particular sampling policy

One-Way Sensitivity Analysis

The base case analysis (Fig. 2) assumed a per-pass adequacy rate of 60 %. As described in the methods, this value was derived from the average per-case adequacy rate (93 %) obtained in a recent meta-analysis [2]. The per-pass adequacy rate is likely to vary between institutions. The impact of the per-pass adequacy rate on the per-case adequacy is shown in Fig. 3. The figure shows that the per-case adequacy rate increases as the per-pass adequacy rate increases. At low per-pass adequacy rates ($p = 20\%$), F2 and F3 are no longer strictly dominated by variable policies. Although F2 and F3 have lower per-case adequacy rates than variable policies, these rates are achieved with fewer needle passes when the per-pass adequacy rate is low. At a per-pass adequacy rate of 40 % all fixed sampling policies except F2 are dominated by variable policies. When the per-pass adequacy rate is high ($p = 80\%$), V3 is dominated by V2 since the average per-case adequacy rate for both is 100 %; however, V2 averages 2.6 needle passes while V3 averages 3.9 needle passes.

The impact of assessor accuracy is demonstrated in Fig. 4. The per-case adequacy rate of policy V1 was quite sensitive to assessor accuracy when accuracy varied between 85 and 100 % but dominated all fixed sampling over this range of accuracy. Policies V2 and V3 were relatively insensitive to assessor accuracy.

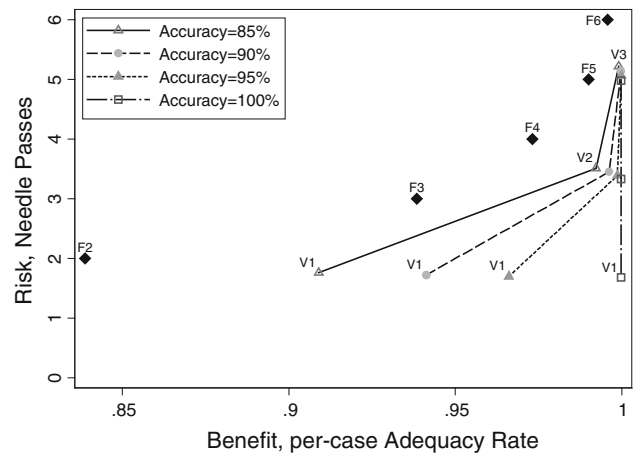


Fig. 4 The effect of assessor accuracy on outcomes. The figure shows the influence of assessor accuracy. Accuracy (sensitivity and specificity) varied from 85 to 100 %. Variable sample policies (using ROSE) are indicated by V followed by the number of observed adequate samples required to stop sampling (e.g., V1). Fixed sampling policies are indicated by F followed by the number of needle passes (e.g., F3)

The effect of the maximum limit of the number of needle passes (noted as M) is presented in Fig. 5. The average number of needle passes and per-case adequacy rates increase when M increases in variable policies. Variable policies dominate fixed policies at all levels of M ; however, the advantage of ROSE is small when the

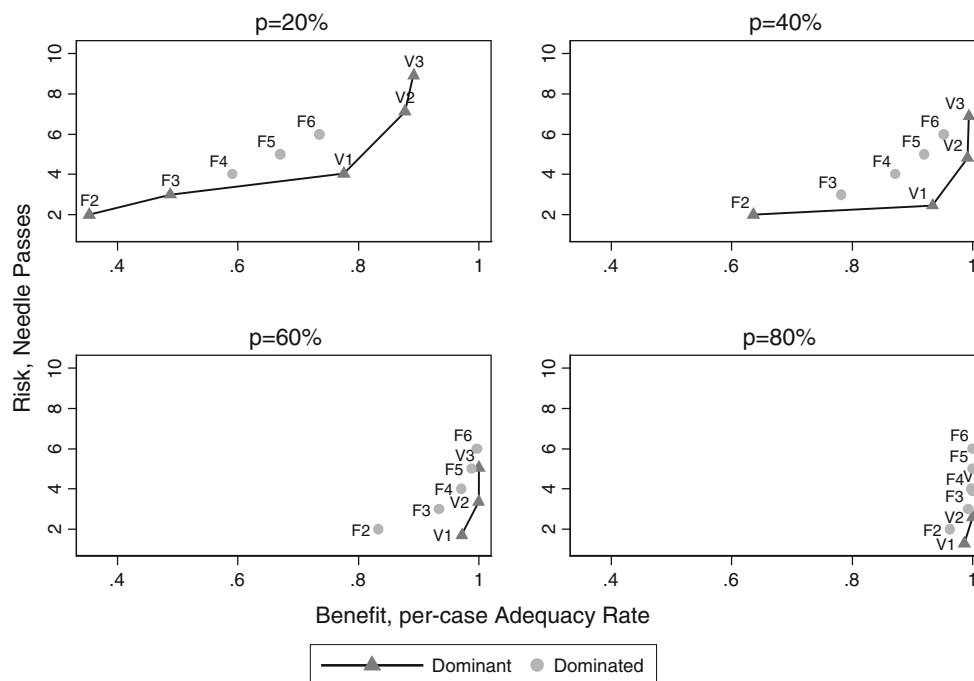


Fig. 3 The effect of the per-pass adequacy rate on the efficient frontier. The figure demonstrates the impact of the per-pass adequacy rate. The per-pass adequacy rate varied between 20 and 80 % as shown in the legend. Variable sample policies (using

ROSE) are indicated by V followed by the number of observed adequate samples required to stop sampling (e.g., V1). Fixed sampling policies are indicated by F followed by the number of needle passes (e.g., F3)

maximum number of needle passes is limited to two. V1 clearly dominates F2 when M increases to four. The sampling limit has little impact above six needle passes.

Discussion

Our results demonstrate variable sampling policies (ROSE) achieve higher per-case accuracy rates with fewer needle passes than non-ROSE sampling. As demonstrated by Fig. 2, a fixed policy of four needle passes (F4) would be needed to achieve approximately the same level of per-case accuracy (97 %) as policy V1 which averaged 2.2 fewer needle passes. More than six fixed needle passes would be required to achieve the per-case adequacy of V2. Our results contribute to the existing literature by quantifying the tradeoff between per-case adequacy and needle passes.

The relative advantage of ROSE depends on the context. As shown in Fig. 2, the relative advantage of ROSE decreases when the per-pass adequacy rate is high. The effect of the per-pass adequacy rate on the relative advantage of ROSE has also been observed in two recent meta-analyses [2, 5]. Thus, ROSE may only be cost-effective in situations where the per-pass adequacy rate is low. This might occur in situations with an inexperienced endoscopist, particular types of lesions that are difficult to sample, or at low volume centers. ROSE may confer additional benefits such as expediting patient care, reducing

Table 2 Settings where ROSE is likely to have the greatest impact

Settings and descriptions

Settings with low per-pass adequacy

Poor endoscopist skill (e.g. low volume center, trainee)

Lesion characteristics

Settings with a need to limit needle passes

Due to patient morbidity

Due to cost considerations

Settings with high assessor accuracy

endosonographer workload, and reducing costs associated with repeat procedures (Table 2). ROSE has also been shown to increase accuracy of EUS-FNA for pancreatic adenocarcinoma [15].

Sampling limits would be expected to reduce the advantage of ROSE; however, our results show that sampling limits have relatively low impact on ROSE performance if the limit, M , is greater than or equal to four. The impact of limits on needle passes has not been previously reported.

In general, the optimal policy will depend on the per-case adequacy rate. A decision-maker would use an institution-specific estimate of the per-pass adequacy rate as an input to the model to predict sampling policy performance and to select the best sampling policy at a particular institution. Bayesian methods (see Appendix) can be used to update the estimates of the per-pass sample rate [13].

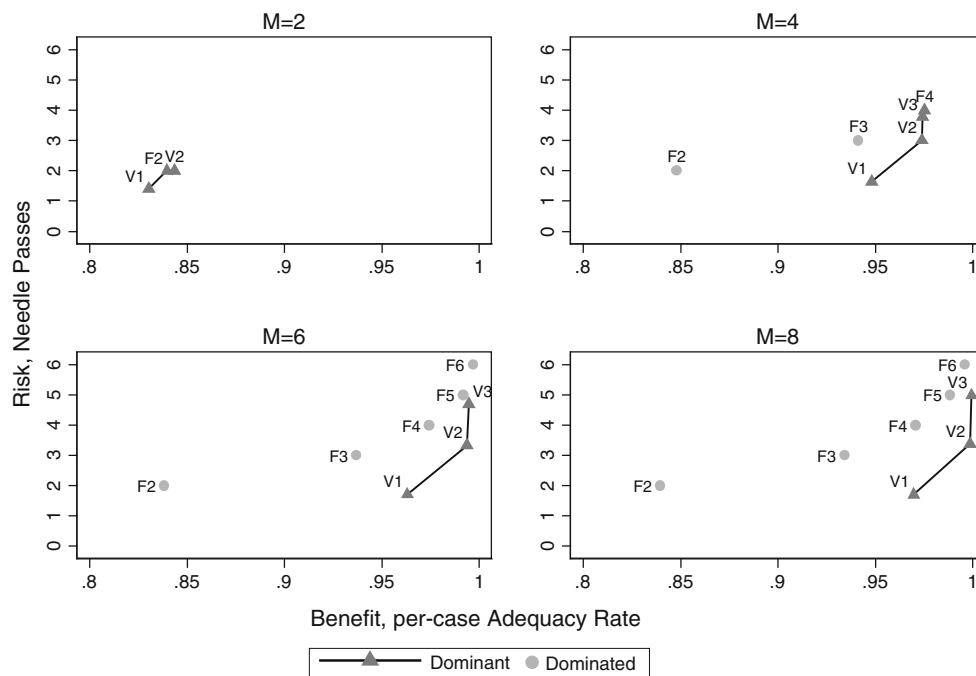


Fig. 5 The effect of a sampling limit on outcomes. The figure shows the results at different sampling limits (M). The limit varied between two and eight as shown in the legend. Variable sample policies (using ROSE) are indicated by V followed by the number of observed

adequate samples required to stop sampling (e.g., V1). Fixed sampling policies are indicated by F followed by the number of needle passes (e.g., F3)

Alternatively, one might correlate the per-pass adequacy rate with lesion characteristics to determine the best sampling policy for specific lesions. For example, the per-pass probability of success may depend on the lesion size [16, 17] and ROSE may be employed only for certain kinds of lesions.

Limitations and Strengths

Studies do not report per-pass adequacy rates and rarely adhere to a fixed sampling protocol over a set of cases. Thus, we based our model on the average performance obtained in a meta-analysis of sampling without ROSE [2]. Our model assumes a constant per-pass probability of success, p . We believe that this is a reasonable model; however, it is possible that the average per-pass probability changes with the number of needle passes. For example, each needle pass could damage the tissue, promote bleeding, and thereby decrease the probability of success on subsequent passes. We are not aware of any evidence showing that p varies with the number of needle passes and, lacking such evidence, we chose the simplest model. The impact of a pass-dependent success rate could be investigated in future work. This study is also limited because it only captures the tradeoff 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), adverse events, and service provision (e.g., variable cost per needle pass, fixed cost of FNA session) in a cost-effectiveness analysis.

Despite the mentioned limitations, the study has several strengths. By using simulation, we were able to eliminate unwanted sources of variation (patient-to-patient variation, site-to-site variation). This enabled us to demonstrate the tradeoff-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. Also, our study expresses the outcomes in terms of variables (per-case adequacy, needle passes) that are familiar to clinicians.

The validity of our model is supported by several qualitative and quantitative results. FNA sampling is known to yield diminishing returns with increasing needle passes. Our binomial sampling model predicts diminishing returns and shows a close fit to empirical data on the relationship between needle passes and adequacy for EUS–FNA in solid pancreatic lesions [18, 19]. The findings show that variable sampling (V1) will generally require fewer needle passes than fixed sampling as demonstrated in two previous studies [3, 20]. The model predicts that the relative advantage of ROSE increases as the per-pass adequacy rate decreases, as has been shown in pancreas [2] and a variety of other tissues [5, 21]. The model also predicts that

per-case adequacy will increase as a function of assessor accuracy as shown by Petrone et al. [14].

Conclusion

Our study demonstrates that modeling is a powerful approach for investigating questions related to EUS–FNA sampling. It also demonstrates that variable sampling policies with ROSE achieve higher adequacy rates with fewer needle passes than sampling without ROSE. The relative advantage of ROSE decreases when the per-pass adequacy rate is high or when a strict limit is placed on the number of needle passes.

Conflict of interest None.

Appendix: Bayesian Estimation of the Per-Pass Adequacy Rate

The per-pass adequacy rate is an important input to our model. We assume that the per-pass adequacy, p , is constant; however, the per-pass probability of success depends on many factors that vary from case to case (e.g., size of tumor, experience of the aspirator, etc.). The sampling experience in a single case can be used to revise the estimated per-pass probability [13].

The imprecision of the initial estimate of p , can be represented by providing a probability distribution for p . It turns out that the beta distribution

$$g(p|a, b) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} p^{a-1} (1 - p)^{b-1} \tag{4}$$

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

$$\bar{p} = \frac{a}{a + b} \tag{5}$$

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

$$g(p|a, b, y, n) = \frac{\Gamma(n + a + b)}{\Gamma(y + a)\Gamma(n - y + b)} p^{y+a-1} (1 - p)^{n-y+b-1} \tag{6}$$

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

$$p' = \frac{y + a - 1}{y + a - 1 + n - y + b - 1} = \frac{y + a - 1}{n + a + b - 2} \tag{7}$$

Thus, with every needle pass, the estimate is revised to become more case-specific. This information could be

used, for example, to estimate the number of additional passes that would be required to obtain an adequate sample.

References

- Erickson RA. EUS-guided FNA. *Gastrointest Endosc.* 2004;60:267–279.
- Schmidt RL, Witt BL, Matynia AP, et al. Rapid on-site evaluation increases endoscopic ultrasound-guided fine-needle aspiration adequacy for pancreatic lesions. *Dig Dis Sci.* 2013;58:872–882.
- 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.
- Schmidt RL, Adler DG. EUS-guided fine-needle aspiration of solid masses with or without on-site cytological evaluation: no paradox. *Am J Gastroenterol.* 2012;107:1446–1447.
- Schmidt RL, Witt BL, Lopez-Calderon LE, et al. The influence of rapid onsite evaluation on the adequacy rate of fine-needle aspiration cytology: a systematic review and meta-analysis. *Am J Clin Pathol.* 2013;139:300–308.
- Moller K, Papanikolaou IS, Toerner 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.
- Fabbri C, Polifemo AM, Luigiano C, et al. Endoscopic ultrasound-guided fine needle aspiration with 22- and 25-gauge needles in solid pancreatic masses: A prospective comparative study with randomisation of needle sequence. *Digestive and Liver Disease.* 2011;43:647–652.
- Song TJ, Kim JH, Lee SS, et al. The prospective randomized, controlled trial of endoscopic ultrasound-guided fine-needle aspiration using 22G and 19G aspiration needles for solid pancreatic or peripancreatic masses. *Am J Gastroenterol.* 2010;105:1739–1745.
- Puri R, Vilmann P, Săftoiu A, et al. Randomized controlled trial of endoscopic ultrasound-guided fine-needle sampling with or without suction for better cytological diagnosis. *Scand J Gastroenterol.* 2009;44:499–504.
- Lee JH, Stewart J, Ross WA, et al. Blinded prospective comparison of the performance of 22-gauge and 25-gauge needles in endoscopic ultrasound-guided fine needle aspiration of the pancreas and peri-pancreatic lesions. *Dig Dis Sci.* 2009;54:2274–2281.
- Harris MD, Buscaglia JM. How to do pancreatic mass FNA. *Gastrointest Endosc.* 2010;71:825–826.
- Larghi A, Verna EC, Ricci R, et al. EUS-guided fine-needle tissue acquisition by using a 19-gauge needle in a selected patient population: a prospective study. *Gastrointest Endosc.* 2011;74:504–510.
- Schmidt RL, Kordy MA, Howard K, et al. Risk-benefit analysis of sampling methods for fine-needle aspiration cytology: a mathematical modeling approach. *Am J Clin Pathol.* 2013;139:336–344.
- Petrone MC, Arcidiacono PG, Carrara S, et al. Does cytotechnician training influence the accuracy of EUS-guided fine-needle aspiration of pancreatic masses? *Digestive and Liver Disease.* 2012;44:311–314.
- Hebert-Magee S, Bae S, Varadarajulu S, et al. The presence of a cytopathologist increases the diagnostic accuracy of endoscopic ultrasound-guided fine needle aspiration cytology for pancreatic adenocarcinoma: a meta-analysis. *Cytopathology.* 2013;24:159–171.
- Hong SKS, Brown LJ, Draganova-Tacheva RA, et al. Relationship of pancreatic mass size and diagnostic yield of endoscopic ultrasound guided fine needle aspiration. *Gastrointest Endosc.* 2011;73:AB246–AB247.
- Siddiqui AA, Brown LJ, Hong SK, et al. Relationship of pancreatic mass size and diagnostic yield of endoscopic ultrasound-guided fine needle aspiration. *Dig Dis Sci.* 2011;56:3370–3375.
- Suzuki R, Irisawa A, Bhutani MS, et al. Prospective evaluation of the optimal number of 25-gauge needle passes for endoscopic ultrasound-guided fine-needle aspiration biopsy of solid pancreatic lesions in the absence of an onsite cytopathologist. *Dig Endosc.* 2012;24:452–456.
- Schmidt RL. Studies on the impact of onsite evaluation must account for the per-pass success rate. *Dig Endosc.* 2013. doi:[10.1111/den.12132](https://doi.org/10.1111/den.12132).
- 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.
- Witt BL, Schmidt RL. Rapid on-site evaluation (ROSE) improves the adequacy of fine needle aspiration for thyroid lesions: a systematic review and meta-analysis. *Thyroid* 2013. doi:[10.1089/thy.2012.0211](https://doi.org/10.1089/thy.2012.0211).