30th May 2013

User Guide: Influenza Laboratory Testing Sample Size Calculators

Overview

Virologic surveillance is a key and complex component of the influenza surveillance system, informed by a variety of independent but related elements. Specific objectives of virologic surveillance include seasonal situational awareness and determination of virus strain prevalence, early detection of novel viruses or novel events, annual vaccine strain selection, and antiviral resistance monitoring. The amount of influenza testing performed both at CDC and in public health laboratories has largely been determined by the capacity of the laboratory. A statistical, systematic approach to determining the amount of testing needed to support disease response and control efforts and policy decisions has been lacking.

In this document, we provide a users guide for the web-based, sample size calculators.

Introduction to Sample size calculators

Abbreviations

MA-ILI+: number of medically attended patients diagnosed with influenza-like-illness

Flu+: number of medically attended patients diagnosed with influenza-like-illness that have influenza

Rare+: number of Flu+ patients that have a rare type of influenza

Usage Tips

1. Moving your mouse over any "blue text" will display a definition of that word/phrase as it pertains to the calculator.
2. Numbers or words in bold face will change as you alter the input values.
3. Moving your mouse over a plot will display the values for that point on the curve. Clicking on a curve will display a text description of how to interpret the information.

Calculator A. Sample size calculator for Flu+/MA-ILI+ (Situational Awareness)

Question

How many specimens from non-prescreened medically attended ILI patients do I need to test to determine that the prevalence of Flu+ specimens among medically attended ILI is X% at a specified confidence level and margin of error?

Assumptions

1. Each person in the US visits an emergency room or generalist physician 2.5 times per year.
2. A fixed percent of medically attended patients present with ILI.
   - The default is 2.2%.
3. The number of weekly MA-ILI+ patients is calculated as:
   - Population * 2.5 * 1/52 * 0.022 = Population * 1.0577x10^{-3}.
4. ILI+ patients submitted for testing are not pre-screened positive for influenza.
5. The sample size is large, i.e. greater than 20 individuals, and the prevalence of Flu+ among MA-ILI+ is not close to either 0 or 100 percent.
6. Sampling is performed from a finite population (national or state-level).

Equations

1. Sample size

$n = \frac{pz_{\alpha/2}^2 - p^2 z_{\alpha/2}^2}{\varepsilon^2}$ where;

n = required sample size

p = expected FLU+/MA-ILI+

α = confidence level. A value of α = 0.95 results in a 95% confidence interval.

$z_{\alpha/2}^2 = 1 - \frac{\alpha}{2}$ percentile of a standard normal distribution

ε = error level. A value of 0.05 results in a confidence interval that is +/- 5%

2. Finite size correction factor

$n^* = \frac{nN}{n + (N-1)}$ where;

n^* = corrected sample size

n = required sample size for an infinite population

N = total population size (e.g. the number of people living in the target state)

3. Sample power - error

$\varepsilon = z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$ where;

n = actual sample size

p = expected FLU+/MA-ILI+

α = confidence level. A value of α = 0.95 results in a 95% confidence interval.

$z^2_{\alpha/2} = 1 - \frac{\alpha}{2}$ percentile of a standard normal distribution

ε = error level. A value of 0.05 results in a confidence interval that is +/- 5%

4. Sample power - confidence

$z_{\alpha/2} = \frac{\varepsilon}{\sqrt{\frac{p(1-p)}{n}}}$ where;

n = actual sample size

p = expected FLU+/MA-ILI+

α = confidence level. A value of α = 0.95 results in a 95% confidence interval.

$z^2_{\alpha/2} = 1 - \frac{\alpha}{2}$ percentile of a standard normal distribution

ε = error level. A value of 0.05 results in a confidence interval that is +/- 5%

Tab 1: Sample Size

Uses Equations 1 and 2

This calculator is used to determine the minimum sample size of non-prescreened MA-ILI+ specimens needed to estimate the fraction of Flu+/MA-ILI+ with a specified confidence level and margin of error.

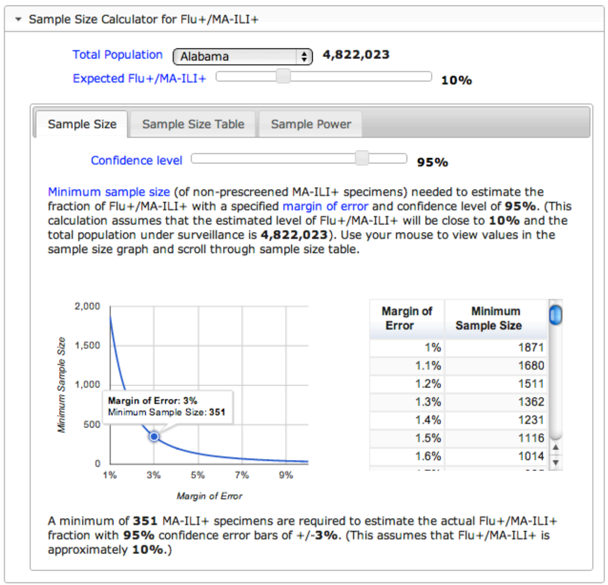


Figure 1.

Instructions for selecting input values:

1. Use the *Total Population* drop-down menu to select the population under surveillance. For labs collecting specimens from subsets of state populations or populations that cross multiple states, choose “Other” and enter the estimated size of the entire population under consideration. The calculator uses these numbers to estimate the weekly number of medically attended ILI cases in your jurisdiction (MA-ILI+). (See Figure 1, top dropdown menu.)
2. Set the expected Flu+/MA-ILI+ using the slider. This is your surveillance target: the level of Flu+/MA-ILI+ you would like to be able estimate accurately. For example, if you would like to detect when Flu+/MA-ILI+ crosses the 10% threshold at the beginning of the flu season, then move the slider to 10%. If, instead, you plan to use the data to estimate Flu+/MA-ILI+ later in the season, when it is closer to 30%, then move the slider closer to 30%. Although the actual fraction of Flu+ over MA-ILI+ may differ from the value you choose, this approximation still provides an important baseline for determining sample sizes. (See Figure 1, top slider bar.)
3. Set the confidence level using the slider. The desired confidence that the sample will yield an estimated level of Flu+/MA-ILI+ that is close to the true

value. When using laboratory samples to estimate Flu+/MA-ILI+, you will calculate an expected value plus or minus a margin of error. For example, you might calculate 10% plus or minus 2%, which means that you estimate Flu+/MA-ILI+ to fall somewhere between 8% to 12%). The higher the confidence level, the more confident you can be that the true level of Flu+/MA-ILI+ in your population falls within the estimated interval. Intuitively, high confidence levels and small margins of error require many samples, while low confidence levels or large. (See Figure 1, slider bar below main tabs.)

Instructions for interpreting output:

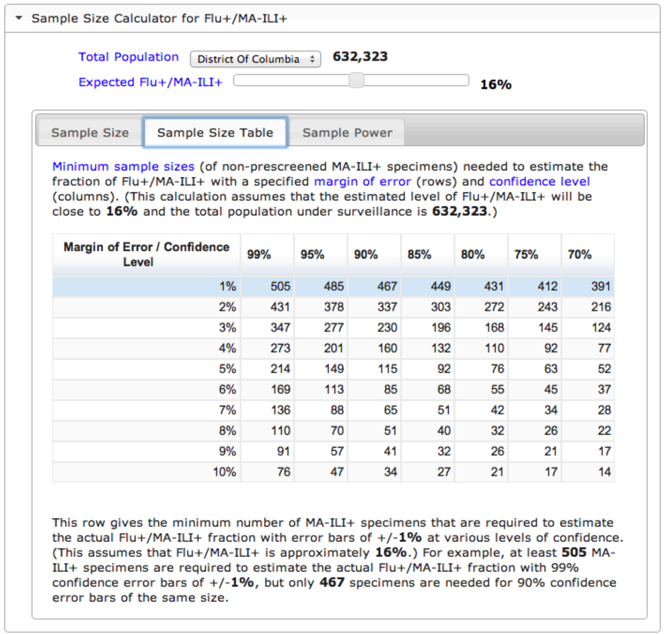
After setting the (1) population size, (2) expected Flu+/MA-ILI+, and (3) confidence level, a chart and table of minimum table sizes is generated. Each minimum sample size corresponds to a different margin of error, the width of the confidence interval around the estimated value of Flu+/ILI+. When using laboratory samples to estimate Flu+/MA-ILI+, you will calculate an expected value plus or minus a margin of error. For example, you might calculate 10% plus or minus 2%, which means that you estimate Flu+/MA-ILI+ to fall somewhere between 8% to 12%). The smaller the margin of error, the more precise your estimate of Flu+/MA-ILI+. Intuitively, high confidence levels and small margins of error require many samples, while low confidence levels or large margins of error require fewer samples.

By clicking on the curve, you will see the precise sample size for a given margin of error. The text below the plot will change as different margin of errors are selected. For example, if you are a conducting surveillance in Alabama, with an expected Flu+/MA-ILI+ of 10% and a confidence level of 95% and you chose a margin of error of 4.9% the output will read, “A minimum of **139** MA-ILI+ specimens are required to estimate the actual Flu+/MA-ILI+ fraction with **95%** confidence error bars of +/-**4.9%**. (This assumes that Flu+/MA-ILI+ is approximately **10%**.).” By decreasing the margin of error to 3%, the minimum sample size decreases and text now reads, “A minimum of **351** MA-ILI+ specimens are required to estimate the actual Flu+/MA-ILI+ fraction with **95%** confidence error bars of +/-**3%**. (This assumes that Flu+/MA-ILI+ is approximately **10%**.)”

<u>Tab 2: Sample Size Table</u>

Uses Equations 1 and 2

This calculator is used to generate a table of the minimum sample size of non-prescreened MA-ILI+ specimens needed to estimate the fraction of Flu+/MA-ILI+ for a range of margin of errors and confidence levels.



Margin of Error / Confidence Level	99%	95%	90%	85%	80%	75%	70%
1%	505	485	467	449	431	412	391
2%	431	378	337	303	272	243	216
3%	347	277	230	196	168	145	124
4%	273	201	160	132	110	92	77
5%	214	149	115	92	76	63	52
6%	169	113	85	68	55	45	37
7%	136	88	65	51	42	34	28
8%	110	70	51	40	32	26	22
9%	91	57	41	32	26	21	17
10%	76	47	34	27	21	17	14

Figure 2.

Instructions for selecting input values

1. Use the *Total Population* drop-down menu to select the population under surveillance. For labs collecting specimens from subsets of state populations or populations that cross multiple states, choose "Other" and enter the estimated size of the entire population under consideration. The calculator uses these numbers to estimate the weekly number of medically attended ILI cases in your jurisdiction (MA-ILI+). (See Figure 2, top dropdown menu.)
2. Set the expected Flu+/MA-ILI+ using the slider. This is your surveillance target: the level of Flu+/MA-ILI+ you would like to be able estimate accurately. For example, if you would like to detect when Flu+/MA-ILI+ crosses the 10% threshold at the beginning of the flu season, then move the slider to 10%. If, instead, you plan to use the data to estimate Flu+/MA-ILI+ later in the season, when it is closer to 30%, then move the slider closer to 30%. Although the actual fraction of Flu+ over MA-ILI+ may differ from the value you choose, this approximation still provides an important baseline for determining sample sizes. (See Figure 2, top slider bar.)

Instructions for interpreting output:

The output is a table of confidence levels and margin of errors corresponding to the population size and expected Flu_/MA-ILI+. If you highlight the table, you can copy and paste it into a tab-delimited text file. This file can be imported directly into an excel document.

Tab 3: Sample Power

Uses Equations 3 and 4

This calculator is used to determine the best combinations of margin of error and confidence level achievable for a pre-specified sample size. You might use this calculator to explore the benefits of increasing throughput or the cost of decreasing it.

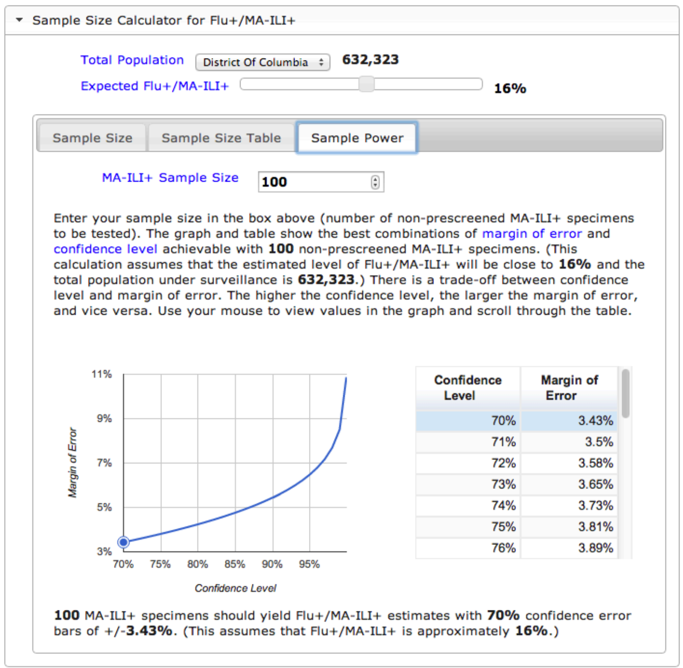


Figure 3.

Instructions for entering input values:

1. Use the *Total Population* drop-down menu to select the population under surveillance. For labs collecting specimens from subsets of state populations or populations that cross multiple states, choose “Other” and enter the estimated size of the entire population under consideration. The calculator uses these numbers to estimate the weekly number of medically attended ILI cases in your jurisdiction (MA-ILI+). (See Figure 3, top dropdown menu.)
2. Set the expected Flu+/MA-ILI+ using the slider. This is your surveillance target: the level of Flu+/MA-ILI+ you would like to be able estimate accurately. For example, if you would like to detect when Flu+/MA-ILI+ crosses the 10% threshold at the beginning of the flu season, then move the slider to 10%. If, instead, you plan to use the data to estimate Flu+/MA-ILI+ later in the season, when it is closer to 30%, then move the slider closer to 30%. Although the actual fraction of Flu+ over MA-ILI+ may differ from the value you choose, this approximation still provides an important baseline for determining sample sizes. (See Figure 3, top slider bar.)

3. Input the number of MA-ILI+ samples. This is the number of non-prescreened samples collected from the medically attended influenza-like-illness population (MA-ILI+). For example, this number might be the maximum throughput of your lab or the number of samples tested in the previous week. (See Figure 3, box below main tabs.)

Instructions for interpreting output:

After setting the (1) population size, (2) expected Flu+/MA-ILI+, and (3) sample size, a chart and table of confidence levels and margin of errors is generated. The graph and table show the best combinations of margin of error and confidence level achievable for your inputted sample size of non-prescreened MA-ILI+ specimens. There is a trade-off between confidence level and margin of error. The higher the confidence level, the larger the margin of error, and vice versa. Use your mouse to view values in the graph and scroll through the table.

By clicking on the curve, you will see the precise combination of confidence level and margin of error. The text below the plot will change as different margin of errors are selected. For example, if you are a conducting surveillance in Alabama, with an expected Flu+/MA-ILI+ of 10% and a sample size of 100 and you chose a margin of error of 3.8% the output will read, "**100** MA-ILI+ specimens should yield Flu+/MA-ILI+ estimates with **80%** confidence error bars of +/-**3.8%**. (This assumes that Flu+/MA-ILI+ is approximately **10%**.)" By increasing the margin of error to 5.02%, the output text now reads, "**100** MA-ILI+ specimens should yield Flu+/MA-ILI+ estimates with **91%** confidence error bars of +/-**5.02%**. (This assumes that Flu+/MA-ILI+ is approximately **10%**.)" Note how the confidence level increases as the margin of error increases.

Calculator B. Sample size calculator for Rare (Novel) Event Detection

Question

How many ILI specimens do I need to test in my state to allow the national surveillance system to detect a rare/novel virus at X% prevalence with Y% confidence?

Assumptions

1. ILI+ samples are collected uniformly randomly from ILI+ patients.
2. Flu+ samples are collected uniformly randomly from Flu+ patients.
3. Each Flu+ patient has a constant probability of being Rare+, and that probability is the same for all patients.
4. ILI+ patients submitted for testing are not pre-screened positive for influenza.
5. Flu+ samples are known to be influenza positive, but their Rare+/- status is unknown.
6. For national level surveillance, states are apportioned samples by population.
7. There is no correction for finite population size – this is a conservative assumption.

Why we do not correct for finite sample size

Correcting for finite sample size requires accurately characterizing the surveillance population. In the case of a novel event investigation, the size of the relevant population may be largely unknown. The sample size determined without correcting for a finite population size achieves a guaranteed confidence level and error, rendering it conservative. If a sample size correction factor is improperly applied, the target population will be under-sampled, resulting in an overestimate of the confidence level and underestimate of the error.

Equations

5. Sample size

$n = \frac{Log(1-\phi)}{Log(1-p)}$ where;

n = required sample size

ϕ = desired probability of observing at least one rare event

p = expected Rare+/FLU+

6. Sample power - confidence

$\phi = 1-(1-p)^n$ where;

n = actual sample size

ϕ = desired probability of observing at least one rare event

p = expected Rare+/FLU+

7. Sample power - threshold

$p = 1 - (1 - \phi)^{1/n}$ where;

n = actual sample size

ϕ = desired probability of observing at least one rare event

p = detection threshold for Rare+/FLU+

Tab 1: Flu+ Sample Size

Uses Equation 5

This calculator is used to determine the minimum sample size of Flu+ specimens needed to detect a rare type of influenza at specified detection threshold.

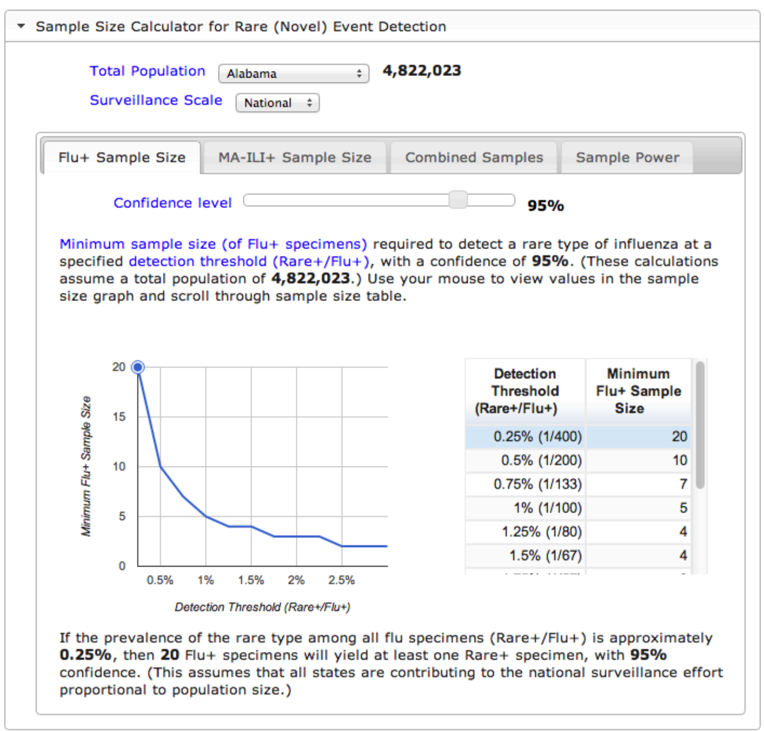


Figure 4.

Instructions for entering input values:

1. Use the *Total Population* drop-down menu to select the population under surveillance. For labs collecting specimens from subsets of state populations

or populations that cross multiple states, choose “Other” and enter the estimated size of the entire population under consideration. The calculator uses these numbers to estimate the weekly number of medically attended ILI cases in your jurisdiction (MA-ILI+). (See Figure 4, top dropdown menu.)

2. Indicated whether surveillance is being conducted at the national or state/regional level using the *surveillance scale* dropdown menu. A “National” surveillance scale means that all states are contributing to a national surveillance effort proportional to their population size. A state-based surveillance scale indicates a local surveillance effort for that state only. A “National” surveillance scale is usually the most appropriate selection. (See Figure 4, second dropdown menu.)
3. Set the confidence level using the slider. The desired confidence that the sample will contain at least one rare specimen (Rare+) when the prevalence of rare type reaches the specified limit of detection. Sample sizes can be calculated for non-prescreened MA-ILI+ or prescreened Flu+, or a combination of both types of specimens. For example, if you choose a confidence level of 95% and a detection threshold of 1/400, then the resulting minimum sample size should be sufficient to detect a rare type when it reaches a prevalence of 1/400 Rare+/Flu+, 95% of the time. Intuitively, a high confidence level and a low detection threshold requires many samples, while low confidence and a high detection threshold results in fewer samples. (See Figure 4, slider bar below main tabs.)

Instructions for interpreting output:

After setting the (1) population size, (2) surveillance scale, and (3) confidence level, a chart and table of minimum sample size and detection thresholds is generated. The graph and table show the minimum sample size (of Flu+ specimens) required to detect a rare type of influenza at a specified detection threshold (Rare+/Flu+), with a pre-specified confidence. Use your mouse to view values in the sample size graph and scroll through sample size table.

By clicking on the curve, you will see the precise combination of confidence level and detection threshold. The text below the plot will change as different detection thresholds are selected. For example, if you are a conducting surveillance in Alabama, at the national level, with a confidence of 95% and you select a detection threshold of 1.5%, the text will read, “If the prevalence of the rare type among all flu specimens (Rare+/Flu+) is approximately**1.5%**, then **4** Flu+ specimens will yield at least one Rare+ specimen, with **95%** confidence. (This assumes that all states are contributing to the national surveillance effort proportional to population size.)”

Tab 2: MA-ILI+ Sample Size

Uses Equation 5

This calculator is used to determine the minimum sample size of MA-ILI+ specimens needed to detect a rare type of influenza at specified detection threshold. The number of samples required for this form of sampling will be considerably higher than when only considering Flu+ specimens.

Instructions for entering input values:

1. Use the *Total Population* drop-down menu to select the population under surveillance. For labs collecting specimens from subsets of state populations or populations that cross multiple states, choose “Other” and enter the estimated size of the entire population under consideration. The calculator uses these numbers to estimate the weekly number of medically attended ILI cases in your jurisdiction (MA-ILI+). (See Figure 5, top dropdown menu.)

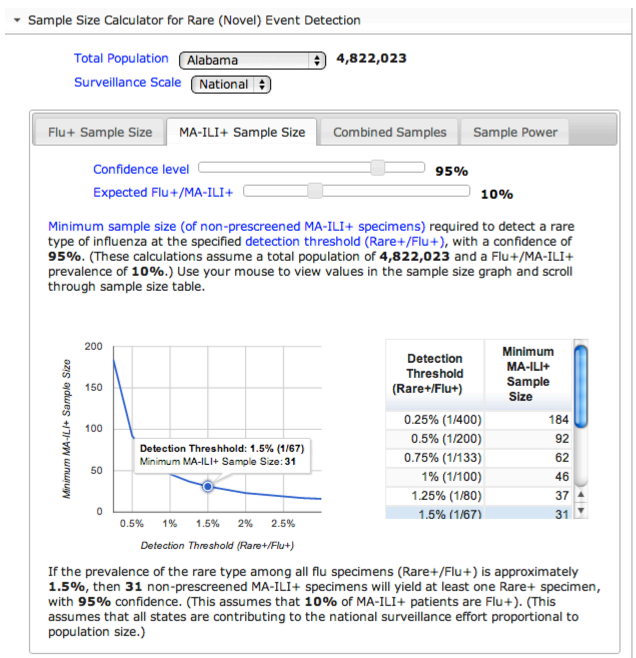


Figure 5.

2. Indicated whether surveillance is being conducted at the national or state/regional level using the *surveillance scale* dropdown menu. A “National” surveillance scale means that all states are contributing to a national surveillance effort proportional to their population size. A state-based surveillance scale indicates a local surveillance effort for that state only. A “National” surveillance scale is usually the most appropriate selection. (See Figure 5, second dropdown menu.)

3. Set the confidence level using the slider. The desired confidence that the sample will contain at least one rare specimen (Rare+) when the prevalence of rare type reaches the specified limit of detection. Sample sizes can be calculated for non-prescreened MA-ILI+ or prescreened Flu+, or a combination of both types of specimens. For example, if you choose a confidence level of 95% and a detection threshold of 1/400, then the resulting minimum sample size should be sufficient to detect a rare type when it reaches a prevalence of 1/400 Rare+/Flu+, 95% of the time. Intuitively, a high confidence level and a low detection threshold requires many samples, while low confidence and a high detection threshold results in fewer samples. (See Figure 5, top slider bar below main tabs.)
4. Set the expected Flu+/MA-ILI+ using the slider. This should represent your best estimate of Flu+/MA-ILI+. Although the actual fraction of Flu+ over MA-ILI+ may differ from the value you choose, this approximation still provides an important baseline for determining sample sizes. (See Figure 5, second slider bar below main tabs.)

Instructions for interpreting output:

After setting the (1) population size, (2) surveillance scale, (3) confidence level, and (4) expected Flu+/MA-ILI+, a chart and table of minimum sample size and detection thresholds is generated. The graph and table show the minimum sample size (of Flu+ specimens) required to detect a rare type of influenza at a specified detection threshold (Rare+/Flu+), with a pre-specified confidence. Use your mouse to view values in the sample size graph and scroll through sample size table.

By clicking on the curve, you will see the precise combination of confidence level and detection threshold. The text below the plot will change as different detection thresholds are selected. For example, if you are a conducting surveillance in Alabama, at the national level, with a confidence of 95% and you select a detection threshold of 1.5%, the text will read, "If the prevalence of the rare type among all flu specimens (Rare+/Flu+) is approximately**1.5%**, then **31** non-prescreened MA-ILI+ specimens will yield at least one Rare+ specimen, with **95%** confidence. (This assumes that **10%** of MA-ILI+ patients are Flu+). (This assumes that all states are contributing to the national surveillance effort proportional to population size.)" Note how this number is almost 8x higher than when only considering Flu+ samples.

<u>Tab 3: Combined samples</u>

Uses Equation 5

This calculator is used to determine the minimum sample size of Flu+ and MA-ILI+ specimens needed to detect a rare type of influenza at specified detection threshold. This models the scenario where Flu+ samples are requested from other labs for testing.

Instructions for entering input values:

1. Use the *Total Population* drop-down menu to select the population under surveillance. For labs collecting specimens from subsets of state populations or populations that cross multiple states, choose “Other” and enter the estimated size of the entire population under consideration. The calculator uses these numbers to estimate the weekly number of medically attended ILI cases in your jurisdiction (MA-ILI+). (See Figure 6, top dropdown menu.)
2. Indicated whether surveillance is being conducted at the national or state/regional level using the *surveillance scale* dropdown menu. A “National” surveillance scale means that all states are contributing to a national surveillance effort proportional to their population size. A state-based surveillance scale indicates a local surveillance effort for that state only. A “National” surveillance scale is usually the most appropriate selection. (See Figure 6, second dropdown menu.)

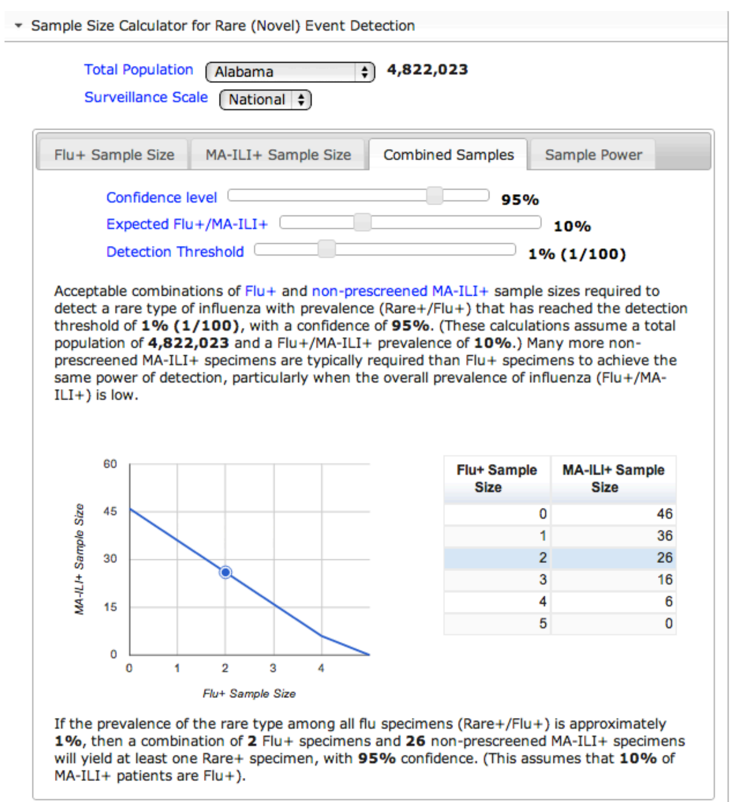


Figure 6.

3. Set the confidence level using the slider. The desired confidence that the sample will contain at least one rare specimen (Rare+) when the prevalence of rare type reaches the specified limit of detection. Sample sizes can be calculated for non-prescreened MA-ILI+ or prescreened Flu+, or a combination of both types of specimens. For example, if you choose a confidence level of 95% and a detection threshold of 1/400, then the resulting minimum sample size should be sufficient to detect a rare type when it reaches a prevalence of 1/400 Rare+/Flu+, 95% of the time. Intuitively, a high confidence level and a low detection threshold requires many samples, while low confidence and a high detection threshold results in fewer samples. (See Figure 6, top slider bar below main tabs.)
4. Set the expected Flu+/MA-ILI+ using the slider. This should represent your best estimate of Flu+/MA-ILI+. Although the actual fraction of Flu+ over MA-ILI+ may differ from the value you choose, this approximation still provides an important baseline for determining sample sizes. (See Figure 6, second slider bar below main tabs.)
5. Set the detection threshold. The detection threshold for a rare type of influenza is the prevalence of the rare type (out of all Flu+ cases) at which the first rare specimens are expected to appear in the lab. For example, a detection threshold of 0.25% (1/400) means that rare type should be detected by the lab when it rises to a prevalence of one out of every 400 cases of flu. (See Figure 6, third slider bar below main tabs.)

Instructions for interpreting output:

After setting the (1) population size, (2) surveillance scale, (3) confidence level (4) expected Flu+/MA-ILI+, and (5) detection threshold, a chart and table of minimum sample sizes for Flu+ and MA-ILI+ is generated. The graph and table show the minimum combination sample size of Flu+ and MA-ILI+ specimens required to detect a rare type of influenza at a specified detection threshold (Rare+/Flu+), with a specified confidence. Use your mouse to view values in the sample size graph and scroll through sample size table.

By clicking on the curve, you will see the precise combination of sample types and the corresponding detection threshold and confidence. The text below the plot will change as different combinations are selected. For example, if you are a conducting surveillance in Alabama, at the national level, with a confidence of 95% and you select 2 Flu+ and 26 non-prescreened MaILI+ samples the text will read, "If the prevalence of the rare type among all flu specimens (Rare+/Flu+) is approximately**1%**, then a combination of **2** Flu+ specimens and **26** non-prescreened MA-ILI+ specimens will yield at least one Rare+ specimen, with **95%** confidence. (This assumes that **10%** of MA-ILI+ patients are Flu+)."

Tab 4: Sample Power

Uses Equations 6 and 7

This calculator is used to determine the best combinations of margin of error and confidence level achievable for a pre-specified sample size. You might use this calculator to explore the benefits of increasing throughput or the cost of decreasing it.

Instructions for entering input values:

1. Use the *Total Population* drop-down menu to select the population under surveillance. For labs collecting specimens from subsets of state populations or populations that cross multiple states, choose "Other" and enter the estimated size of the entire population under consideration. The calculator uses these numbers to estimate the weekly number of medically attended ILI cases in your jurisdiction (MA-ILI+). (See Figure 7, top dropdown menu.)
2. Indicated whether surveillance is being conducted at the national or state/regional level using the *surveillance scale* dropdown menu. A "National" surveillance scale means that all states are contributing to a national surveillance effort proportional to their population size. A state-based surveillance scale indicates a local surveillance effort for that state only. A "National" surveillance scale is usually the most appropriate selection. (See Figure 7, second dropdown menu.)

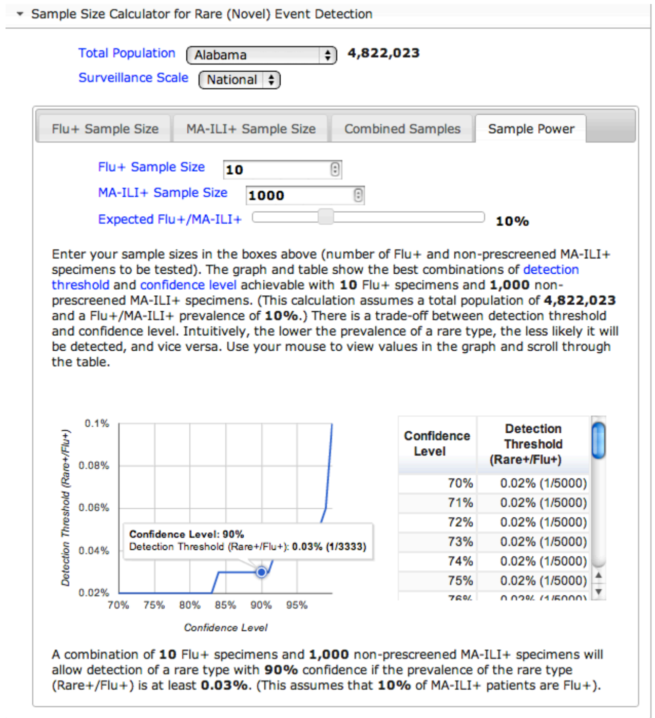


Figure 7.

3. Enter your Flu+ sample size. (See Figure 7, top box below main tabs.)
4. Enter your MA-ILI+ sample size. (See Figure 7, second box below main tabs.)
5. Set the expected Flu+/MA-ILI+ using the slider. This should represent your best estimate of Flu+/MA-ILI+. Although the actual fraction of Flu+ over MA-ILI+ may differ from the value you choose, this approximation still provides an important baseline for determining sample sizes. (See Figure 7, slider bar below main tabs.)

Instructions for interpreting output:

After setting the (1) population size, (2) surveillance scale, (3) number of Flu+ specimens (4) number of MA-ILI+ specimens, and (5) expected Flu+/MA-ILI+ a chart and table of detection threshold and confidence levels is generated. Use your mouse to view values in the graph and scroll through the table.

By clicking on the curve, you will see the precise combination of confidence level and detection threshold. The text below the plot will change as different confidence levels are selected. For example, if you are a conducting surveillance in Alabama, with an expected Flu+/MA-ILI+ of 10% and a sample size of 10 Flu+ and 1000 MA-ILI+ samples and you chose a confidence level of 90% the output will read, “A combination of **10** Flu+ specimens and **1,000** non-prescreened MA-ILI+ specimens will allow detection of a rare type with **90%** confidence if the prevalence of the rare type (Rare+/Flu+) is at least **0.03%**. (This assumes that **10%** of MA-ILI+ patients are Flu+).” Note how the confidence level increases as the detection threshold increases.

Calculator C. Sample Size calculator for Rare (Novel) Event Investigation

Question

Once a novel virus is detected, how many ILI specimens do I need to test to determine true prevalence the population with Y% confidence and Z% error?

Key Issue

When events are rare (p close to 0), the normal approximation to the binomial does not hold. Importantly, this results in an underestimate of the number of samples required. Here we have implemented an alternative approximation that holds when p is close to 0.

Assumptions

1. ILI+ samples are collected uniformly randomly from ILI+ patients.
2. Flu+ samples are collected uniformly randomly from Flu+ patients.
3. Each Flu+ patient has a constant probability of being Rare+, and that constant is the same for all patients.
4. The probability of a Flu+ patient also being Rare+ is close to zero.
5. The statistical test of interest is whether the prevalence exceeds a specified threshold and can be considered a one-tailed test.
6. ILI+ patients submitted for testing are not pre-screened positive for influenza.
7. Flu+ samples are known to be influenza positive, but their Rare+/- status is unknown.
8. For national level surveillance, states are apportioned samples by population.
9. There is no correction for finite population size – this is a conservative assumption.

Why we do not correct for finite sample size

Correcting for finite sample size requires accurately characterizing the surveillance population. In the case of a novel event investigation, the size of the relevant population may be largely unknown. The sample size determined without correcting for a finite population size achieves a guaranteed confidence level and error, rendering it conservative. If a sample size correction factor is improperly applied, the target population will be under-sampled, resulting in an overestimate of the confidence level and underestimate of the error.

Equation

8. Rare event investigations

$$p^* < p + \delta + \sqrt{z_\alpha^2 v(p) + \delta^2}$$

where;

$$\delta = \left(\frac{z_{\alpha}^{2}}{3} + \frac{1}{6} \right) \frac{1 - 2p}{n}$$

and

$$v(p) = \frac{p(1-p)}{n-1}$$

and

p^{*} = desired upper-bound on p.

n = required sample size

p = expected Rare+/Flu+

$z_{\alpha}^{2} = 1 - \alpha$ percentile of a standard normal distribution (one-sided test)

9. Sample size for estimating confidence interval of a small proportion. See Appendix.

<u>Tab 1: Flu+ Sample Size</u>

Uses Equations 8 and 9

This calculator is used to determine the minimum sample size of Flu+ specimens needed to estimate the prevalence of rare type of influenza at specified prevalence threshold.

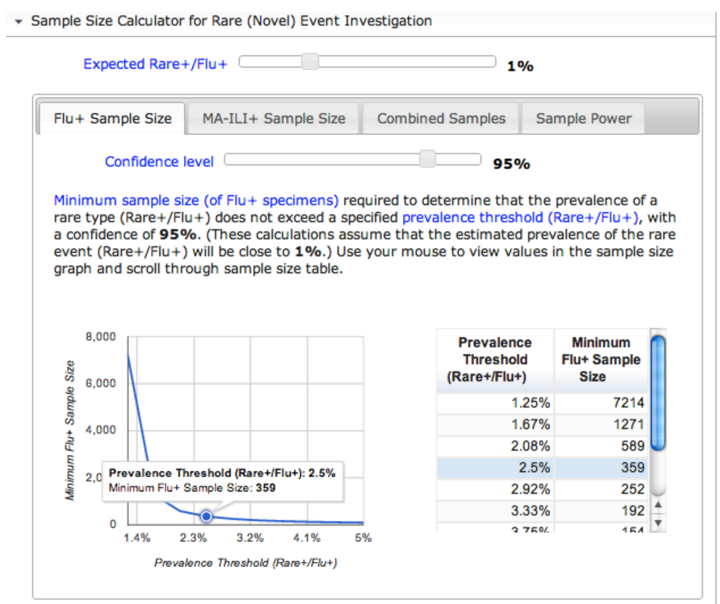


Figure 8.

Instructions for entering input values:

1. Set the expected Rare+/Flu+ using the slider. This is your surveillance target: the level of Rare+/Flu+ you would like to be able estimate accurately. (See Figure 8, top slider bar.)
2. Set the confidence level using the slider. The desired confidence that the sample will yield an estimated prevalence threshold level of Rare+/Flu+. The prevalence threshold is an upper bound (highest value) for the prevalence of the rare type (Rare+/Flu+) that you hope to establish with the specified level of confidence. The higher the confidence level, the more confident you can be that the true level of Rare+/Flu+ in your population falls below the prevalence threshold. Intuitively, high confidence levels and small prevalence thresholds require many samples, while low confidence levels or large prevalence thresholds require fewer samples. (See Figure 8, slider bar below main tabs.)

Instructions for interpreting output:

After setting the (1) expected Rare+/Flu+ and (2) confidence level, a chart and table of minimum sample sizes and prevalence thresholds is generated. Use your mouse to view values in the graph and scroll through the table.

By clicking on the curve, you will see the precise combination of minimum sample sizes and prevalence threshold. The text below the plot will change as different prevalence thresholds are selected. For example, if you are a conducting surveillance in Alabama, with an expected Rare+/Flu+ of 1% and a confidence level of 95% and you select a prevalence threshold of 2.5% the text will read, "A minimum of 359 Flu+ samples must be tested to ensure that the prevalence threshold does not exceed 2.5%"

<u>Tab 2: MA-ILI+ Sample Size</u>

Uses Equation 8 and Supplement

This calculator is used to determine the minimum sample size of MA-ILI+ specimens needed to estimate the prevalence of rare type of influenza at specified prevalence threshold.

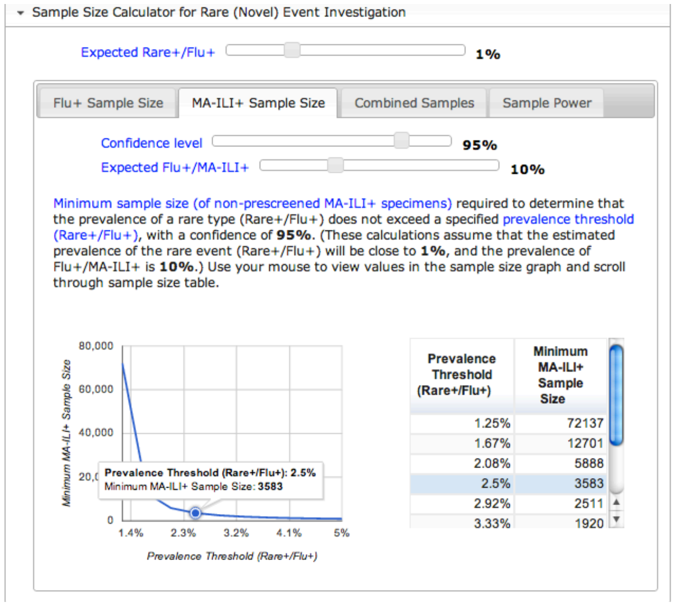


Figure 9.

Instructions for entering input values:

1. Set the expected Rare+/Flu+ using the slider. This is your surveillance target: the level of Rare+/Flu+ you would like to be able estimate accurately. (See Figure 9, top slider bar.)
2. Set the confidence level using the slider. The desired confidence that the sample will yield an estimated prevalence threshold level of Rare+/Flu+. The prevalence threshold is an upper bound (highest value) for the prevalence of the rare type (Rare+/Flu+) that you hope to establish with the specified level of confidence. The higher the confidence level, the more confident you can be that the true level of Rare+/Flu+ in your population falls below the prevalence threshold. Intuitively, high confidence levels and small prevalence thresholds require many samples, while low confidence levels or large prevalence thresholds require fewer samples. (See Figure 9, top slider bar below main tabs.)
3. Set the expected Flu+/MA-ILI+ using the slider. This should represent your best estimate of Flu+/MA-ILI+. Although the actual fraction of Flu+ over MA-ILI+ may differ from the value you choose, this approximation still provides an important baseline for determining sample sizes. (See Figure 9, second slider bar below main tabs.)

Instructions for interpreting output:

After setting the (1) expected Rare+/Flu+, (2) confidence level, and (3) expected Flu+/MA-ILI+, a chart and table of minimum sample sizes and prevalence thresholds is generated. Use your mouse to view values in the graph and scroll through the table.

By clicking on the curve, you will see the precise combination of minimum sample sizes and prevalence threshold. The text below the plot will change as different prevalence thresholds are selected. For example, if you are a conducting surveillance in Alabama, with an expected Rare+/Flu+ of 1%, a confidence level of 95%, and an expected Flu/MA-ILI+ of 10% and you select a prevalence threshold of 2.5% the text will read, "A minimum of 3583 Flu+ samples must be tested to ensure that the prevalence threshold does not exceed 2.5%." Note the substantial increase in sample size as compared to sampling only Flu+ specimens.

Tab 3: Combined Sample Size

Uses Equation 8 and Supplement

This calculator is used to determine the minimum combination sample size of MA-ILI+ and Flu+ specimens needed to estimate the prevalence of rare type of influenza at specified prevelence threshold.

Instructions for entering input values

1. Set the expected Rare+/Flu+ using the slider. This is your surveillance target: the level of Rare+/Flu+ you would like to be able estimate accurately. (See Figure 10, top slider bar.)

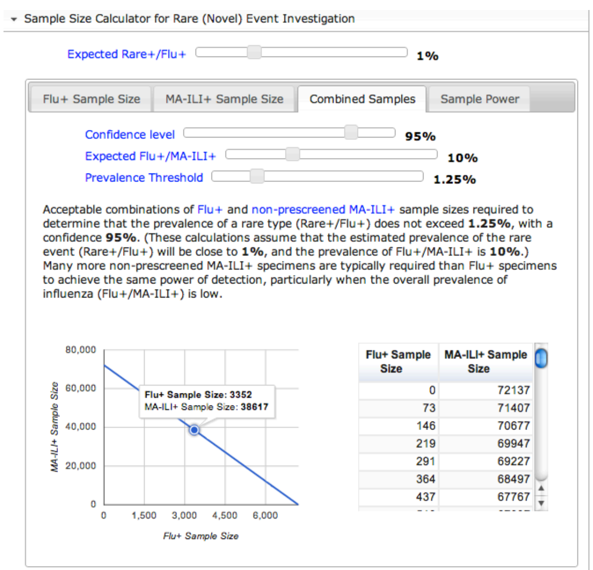


Figure 10.

2. Set the confidence level using the slider. The desired confidence that the sample will yield an estimated prevalence threshold level of Rare+/Flu+. The prevalence threshold is an upper bound (highest value) for the prevalence of

the rare type (Rare+/Flu+) that you hope to establish with the specified level of confidence. The higher the confidence level, the more confident you can be that the true level of Rare+/Flu+ in your population falls below the prevalence threshold. Intuitively, high confidence levels and small prevalence thresholds require many samples, while low confidence levels or large prevalence thresholds require fewer samples. (See Figure 10, top slider bar below main tabs.)

3. Set the expected Flu+/MA-ILI+ using the slider. This should represent your best estimate of Flu+/MA-ILI+. Although the actual fraction of Flu+ over MA-ILI+ may differ from the value you choose, this approximation still provides an important baseline for determining sample sizes. (See Figure 10, second slider bar below main tabs.)
4. Set the desired prevalence threshold. The prevalence threshold is an upper bound (highest value) for the prevalence of the rare type (Rare+/Flu+) that you hope to establish with the specified level of confidence. (See Figure 10, third slider bar below main tabs.)

Instructions for interpreting output:

After setting the (1) expected Rare+/Flu+, (2) confidence level, (3) expected Flu+/MA-ILI+, and (4) confidence threshold, a chart and table of minimum sample sizes of Flu+ and MA-ILI+ samples is generated. Use your mouse to view values in the graph and scroll through the table.

By clicking on the curve, you will see the precise combination of minimum sample sizes required. The text below the plot will change as different combinations of Flu+ and MA-ILI+ are selected. For example, if you are a conducting surveillance in Alabama, with an expected Rare+/Flu+ of 1%, a confidence level of 95%, an expected Flu/MA-ILI+ of 10%, and a prevalence threshold of 1.25% and you select a sample size of 3352 Flu+ samples the text will read, "A minimum of 3352 Flu+ samples and 38,617 MA-ILI+ must be tested to ensure that the prevalence threshold does not exceed 1.25%."

<u>Tab 4: Sample Power</u>:

-Equation 8 and Supplement

This calculator is used to determine the best combination of prevalence threshold and confidence level achievable for a pre-specified sample size. This sample size is a combination of Flu+ and MA-ILI+ samples. You might use this calculator to explore the benefit of requesting Flu+ samples from other public health labs to include in a rare-event investigation.

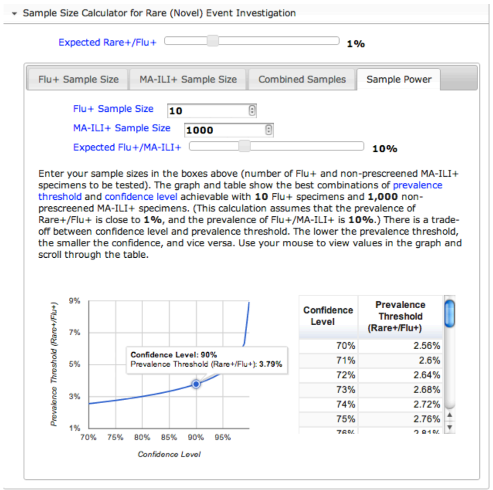


Figure 11.

Instructions for entering input values:

1. Set the expected Rare+/Flu+ using the slider. This is your surveillance target: the level of Rare+/Flu+ you would like to be able estimate accurately. (See Figure 11, top slider bar.)
2. Set the Flu+ sample size. (See Figure 11, top box below main tabs.)
3. Set the MA-ILI+ sample size. (See Figure 11, second box below main tabs.)
4. Set the expected Flu+/MA-ILI+ using the slider. This should represent your best estimate of Flu+/MA-ILI+. Although the actual fraction of Flu+ over MA-ILI+ may differ from the value you choose, this approximation still provides an important baseline for determining sample sizes. (See Figure 11, slider bar below sample size boxes.)

Instructions for interpreting output:

After setting the (1) expected Rare+/Flu+, (2) number of Flu+ samples, (3) number of MA-ILI+ samples, and (4) expected Flu+/MA-ILI+, a chart and table of prevalence threshold and confidence levels is generated. Use your mouse to view values in the graph and scroll through the table.

By clicking on the curve, you will see the precise combination of confidence level and prevalence threshold. The text below the plot will change as different confidence levels are selected. For example, if you are a conducting surveillance in Alabama, with an expected Flu+/MA-ILI+ of 10% and a sample size of 10 Flu+ and 1000 MA-ILI+ samples and you chose a confidence level of 90% the output will read, “A combination of **10** Flu+ specimens and **1,000** non-prescreened MA-ILI+ specimens will allow detection of a rare type with **90%** confidence if the prevalence of the rare type (Rare+/Flu+) does not exceed **3.79%**. (This assumes that **10%** of MA-ILI+ patients are Flu+).” Note how the confidence level increases as the prevalence threshold increases.