DEPARTMENT OF HEALTH

Olympia, Washington 98504

DOH 200-001 September 2010

Healthcare Associated Infections Program
Group Administrator Instructions for Validation of Surveillance Programs

Version 1.0
September 2010

Table of Contents

Section I: Introduction

The purpose of this instruction manual is to document all aspects of the methods used by the Washington State Health Department's Healthcare Associated Infections Program to validate the accuracy of hospital infection surveillance reporting. This first section introduces the general concepts underlying established methods we've considered. Sections that follow explain the theory and practice in greater detail for the methods as actually adopted.

The approach we are taking is consistent with the industrial concept of acceptance sampling. Acceptance sampling is a well-established process for using a sample to judge the quality of a batch of material or product. It has its origins in World War II and military standards (MIL STD 105A issued in 1950) for manufacturing. MIL STD 105A was eventually replaced by updated military standards documents, which in 1991 was replaced as the reference source when adopted by the International Standards Organization as ISO 2859.

There are several variations within industrial acceptance sampling, starting with whether one counts "defectives" (or "defects") as a present versus absent ("binary") decision or takes measurements and records measured values. The former is called "sampling by attributes"; the latter "sampling by variables". Our program employs sampling by attributes.

Another variation is whether there is: (1) a single fixed sample size for all inspections, (2) a preliminary smaller sample followed by a larger quota only if defects are found in the initial small sample, or (3) even more sampling is done. These are called single sampling plans, double sampling plans and multiple (or sequential) sampling plans. Our internal validation component uses a single sampling plan approach, while our external validation verification visit component uses a double sampling plan approach.

A third variation is whether the inspection is at a "normal", a "relaxed, or a "tightened" level. In the MIL STD and ISO documents, a dynamic process is described that starts with normal sampling levels but permits a lowered quota when all inspections confirm absence of defects and raises the quota if successive inspections do discover defects. These documents also offer a choice among four "special inspection" levels (S1-S4) and three ""general inspection" levels (I-III). S1-S4 are appropriate when sample size must be kept as small as possible and larger sampling risks can be tolerated. I, II and III are appropriate where the level of discrimination can be less than normal, normal or greater than normal. Our validation verification visit component employs general inspection level III ("normal" for spot-checks, "tightened" for follow-up after any indication of problems from an annual internal validation or external validation verification result).

The ability of any particular industrial acceptance sampling plan to detect substandard batches is summarized in an "operating characteristic" curve for that particular plan. Operating characteristic curves for the plans used in our program are shown and explained elsewhere in this document.

This explains the industrial manufacturing heritage of acceptance sampling, which eventually was found equally applicable in service industries (where transactions rather than raw materials or manufactured goods are subject to inspection). There is an analogous counterpart in clinical trials study design.

Clinical trials study design requires specification of desired target values, precision of estimate and statistical power required in order to compute sample size – just like the industrial acceptance sampling plan design process. The classic design was a single-stage procedure, in which fixed numbers of subjects were recruited into treatment and control arms of the study. In the 1960s, Edmund Gehan introduced a two-stage alternative design. This is similar to the double sampling plan of industrial acceptance sampling. In Gehan's method a smaller initial sample is drawn and if no beneficial response from the new treatment is found the study is terminated; if a beneficial response is detected, then the study continues into a second stage by expanding with a larger number of subjects.

In both industrial acceptance sampling and clinical trial design, required sample size calculations are based on statistical properties and the amount of precision required in estimating a population parameter from a sample. The statistical properties are α-error (also called Type I error and also known as Producer's Risk) and β-error (also called Type II error and also known as Consumer's Risk). The first refers to the fact that a batch of perfectly fine product might fail a consumer's quality screening test due to random chance alone (more technically, due to variation among all possible samples that could be drawn from that batch) and be rejected; the latter refers to the fact that an unsuitably imperfect batch might slip through a consumer's quality screening and be accepted (again due to random chance sampling variability). Minimizing α- and β-error is the goal, but it comes at a price (higher sampling rates). Industrial acceptance double sampling schemes and Gehan's two-stage clinical trial method start by selecting an initial sample size that is as small as possible while still keeping β-error at a specified low probability – if no target events are detected in that initial sample, then in acceptance sampling one concludes that the batch is acceptable and in clinical trials one concludes that the new treatment offers no appreciable benefit. If target events are detected in that initial sample, then acceptance sampling would call for rejection (if a large enough number of defects is detected), or for testing a second sample before deciding to accept or reject; similarly, Gehan's method would call for enlisting more subjects into a second stage of the clinical trial. For our purposes, the target events are missed and misclassified cases of healthcare-associated infection, and quality of surveillance is measured as sensitivity and specificity.

Gehan's first stage and an ISO 2859 double sampling plan first sample stage are similar in purpose. Both consider a specified β-error (or a related measure, "statistical power" which is equal to 1- β) as the basis for calculating sample size. Gehan's approach differs from ISO 2859 in the second stage: his second stage requires larger samples so that the actual treatment response rate can be estimated with a given level of precision and risk of α-error. The second sample stage of the ISO 2859 double sampling approach does not attempt to estimate the defect rate in this manner, only to improve the accuracy of accept/reject decisions about the batch conforming or not to specified requirements. This requires a smaller sample, and is more consistent with the aims of our validation of hospital infection surveillance program reporting accuracy.

The importance of validating hospital infection surveillance program accuracy is well documented in the literature and demonstrated in our own experience. From the beginning, hospital members on our Advisory Committee emphasized the need to ensure a fair playing field – to know that each hospital is doing a credible job of surveillance behind any numbers it reports. Experience elsewhere certainly underscores the fact that not all hospitals put the same effort into ensuring the accuracy of their rate reports. We therefore developed a statistically and epidemiologically sound, feasible and practical approach to validating CLABSI surveillance. We worked with our program's Advisory Committee on it for over a year, then rolled it out throughout the state in 2010 with minimal difficulty, and subsequently worked with those hospitals that needed to improve on what it revealed.

There are three components to our validation program:
1. Whether the cases that should be reviewed are being reviewed (case-finding, laboratory notification… measured by each hospital using our instructions for annual internal validation of sensitivity)
2. Whether the infection control professional (ICP) is capable of applying NHSN definitions correctly (case coding self-evaluation practice through "newsletter" scenario our program e-mails to all ICPs, and measured by each hospital using our instructions for annual internal validation of specificity)
3. Whether cases are detected and if NHSN definitions are applied correctly (confirmed by external validation verification visit following poor annual internal validation result, or in random spot-check visits)

In the external validation verification visit process, we also are introducing use of a wildlife biology technique ("capture-recapture") that also has been used to estimate the total burden of disease in a community and completeness of coverage in the associated public health reporting program. An overview of this technique and its application to public health is provided in Stephen C, Capture-Recapture Methods in Epidemiological Studies, INFECTION CONTROL &

HOSPITAL EPIDEMIOLOGY 1996;17(4):262-6. Each hospital's infection surveillance program's records and that hospital's discharge abstracts records will be the two information sources used in this portion of the validation process (not to compare one set of records against the other directly, but to incorporate both into the formulae of capture-recapture to estimate the total number of cases from two imperfect sources of information about those cases).

Section II: Internal Validation Component

Theory

The epidemiologic concepts of sensitivity and specificity may cause confusion for those who have not studied epidemiology formally. This section provides an introduction to what they represent, how to calculate them, and how large a sample size might be needed. It is based on three educational articles published in the Statistics for Epidemiology column of a leading hospital epidemiology journal (Birnbaum D, Sheps SB, Validation of New Tests, INFECTION CONTROL & HOSPITAL EPIDEMIOLOGY 1991;12(10):622-624. Birnbaum D, Who Is at Risk of What? INFEECTION CONTROL & HOSPITAL EPIDEMIOLOGY 1999;20(10):706-707. Birnbaum D, What's the Score? INFECTION CONTROL & HOSPITAL EPIDEMIOLOGY 2000;21(1):57-58.)

Look at Figure 1 reproduced here from the first article, "Validation of New Tests" – true condition refers to the reality of whether a condition is present or not in a given patient, test result refers to whether a diagnostic test for that condition in that patient is positive or not.

	True Condition	
Test Result	+	-
+	a	b
-	c	d

This leads to four cells in a so-called 2x2 table:
cell "a" represents true positives (condition really is present and test for it is positive);
cell "b" represents false positives (condition is absent but test for it is positive anyway);
cell "c" represents false negatives (condition really is present but test for it is negative);
and cell "d" represents true negatives (condition is absent, test is negative).
We are concerned with hospital-associated infection as the condition, and whether a surveillance system detects such infections as the diagnostic test to be validated.

There are two ways of examining these 2x2 tables for test validation: epidemiologists tend to look down the columns, clinicians tend to look across the rows. Epidemiologists want to know capabilities of the test itself, namely what proportion of real cases will be detected as positive (the sensitivity) and what proportion of those who are not real cases will have negative test results (the specificity). Clinicians want to know the probability their patient does or does not have the condition when a test is positive (positive predictive value) or negative (negative

predictive value). These test validation statistics are computed as pairs (sensitivity & specificity, or positive predictive value & negative predictive value). Sensitivity is computed as $a/(a+c)$, specificity as $d/(d+b)$; positive and negative predictive values, of less interest in our validation program, are computed as $a/(a+b)$ and $d/(c+d)$ respectively.   The reason positive predictive value is of less interest for our validation program is illustrated in this numeric example using the same table layout:

	True Condition	
Test Result	+	-
+	10	0
-	90	10

In this example, positive predictive value is 100%, which sounds very impressive. However, that obscures the fact that sensitivity is only 10% - this program would fail to detect 9 out of every 10 true positive cases. You should be aware that positive predictive value has been used in the validation audit process of the Centers for Medicare and Medicaid Services, and understand that their objective is different from ours. CMS wants to verify that when it receives a bill for some condition treated, the bill is coded accurately; positive predictive value confirms that. We want to verify that cases are not missed or misclassified in hospital infection surveillance programs; sensitivity and specificity are more appropriate for that.

That explains what these statistics are and how to compute them. The next question is how large a sample is needed, and that depends on how precise an estimate is wanted. If you are not familiar with sample size calculation from the binomial distribution, read the next two articles listed above ("Who Is at Risk of What?" and "What's the Score?"). They explain and illustrate how sample size can be computed from a formula based on a mathematical function called the binomial distribution. This distribution is used widely in statistics and probability to estimate outcomes in situations where there are two mutually exclusive possibilities (e.g. heads vs. tails, positive vs. negative, etc.). These articles also introduce a concept called a Confidence Interval – this is a technical way of saying that if we did an experiment over and over (e.g. if we flipped a coin ten times and recorded the number of heads vs. tails, then repeated that experiment many times), we would expect the outcome (e.g. the number of heads) to vary somewhat from experiment to experiment by random chance alone. The more times we repeat that experiment, the more confident we would be that our estimate of the probability of that coin landing heads-up is accurate. This, too, can be put into mathematical terms (a 95% confidence interval means that we expect the true value to be captured inside the indicated range 95 times out of 100).

We want to know how reliably each infection surveillance program produces accurate results, not just how it did during the past month or months, so our problem is like example number 1 in the "Who Is at Risk of What?" reprint. This distinction is important, because it determines whether something called the "finite population correction factor" needs to be included in the calculation of sample size. The formula for our situation does not include the population size, so size of hospital doesn't change the required sample size for a given level of precision. Smaller hospitals will take longer to accumulate enough cases, larger hospitals will take less time, but both would need the same sample size for a given level of precision in estimating sensitivity and specificity.

The actual number of patients with infection (to compute sensitivity) and without infection (to compute specificity) needed can be computed from a formula or can be read from tables (for example, Table 6.1 in the book *sample size tables for clinical studies, 2nd edition*). Table 6.1 shows a range of sample sizes based on three variables to specify. Those three are π which represents the expected proportion around which a confidence interval is wanted, α which when subtracted from 1.0 gives the confidence level, and ε which represents how wide an interval is wanted. For example, from the literature, it is reasonable to expect that an infection surveillance program will have 95% sensitivity for detecting bloodstream infections, so look for 0.95 in the π column. It is common to want a 95% confidence interval, so look for 0.05 (which is just 1-0.95, as explained above) in the adjacent α column. Now look across that row (which shows sample sizes ranging from 2,100 to 16). The last step is to decide how wide an interval (95% sensitivity ± what) we can tolerate. In this particular table, that ± value is represented as a fraction (ε=0.01, for example, indicating determining sensitivity to within 1% of its true value, so requiring a sample of 2,100 observations to produce a very narrow 95% confidence interval of 0.95±0.0095). At the right-hand extreme, a sample of just 16 observations is needed if we can tolerate a 95% confidence interval of 0.95±0.19. If we took just 16 observations and 15 out of 16 were positive, the real rate could range anywhere from 76% to 100%! Even more extreme, if we took just 3 observations and all were positive, the real rate could range anywhere from 0-100%! If we want a 95% confidence level that sensitivity is at least 90% given an expectation of π =95%, then sample size needed is about 110 observations according to this table.

We explained these concepts and steps to our program's Advisory Committee in 2008. First, define sample size based on precision wanted around an expected value. Next, review that many charts of patients who meet the definition of a hospital-acquired infection and check how many of those were known to the surveillance program (sensitivity). Also review that many charts of patients who don't meet the definition of a hospital-acquired infection, and check to see how many were documented by the surveillance program as having such infection (specificity). This process would need to be repeated for each infection site, to document sensitivity & specificity for bloodstream infections, pneumonias, surgical site infections, etc. The pace at which charts are reviewed, and whether done by an ICP alone or by interdisciplinary teams as an internal cross-

check and educational opportunity, are options open to discussion.   This led to decisions and agreement that 85% sensitivity and 98% specificity would be expected for central-line associated bloodstream infection surveillance, ±15 percentage points is an acceptable level of precision, and from that a determination that the required sample size for these specifications is 22 true positives and 22 true negatives for annual internal validation. An instructional resource "toolkit" booklet was created, and the process pilot tested to confirm practicality and feasibility (pilot test results are reported in poster #842 at the 5th Decennial International Conference on Healthcare-Associated Infections, and poster #4391 at the 2010 Council of State & Territorial Epidemiologists Annual Conference).

Practice

Instructions for hospitals are provided in the validation toolkit resource booklet stored in S:\EHSPHL\OAS\HAI\DataValidationToolKit. The only thing required beyond what is in that resource booklet is an annual reminder sent by e-mail to all hospitals each January to conduct their internal validation process within the first quarter of the current year for their previous year.

CLABSI scoring criteria for the returned report forms are based on the following tables and graphs. Hospitals pass if their sensitivity determination shows 17 or more of 22 actual cases as known to their infection surveillance program as true-positives; fail if only 15 or fewer are known; aren't failed but will need to be verified by a site visit if they recognize 16 of the 22 cases. Similarly, hospitals pass if their specificity determination shows no more than 1 patient misidentified as a false-positive; fail if 3 or more are misidentified; aren't failed but need to be verified by a site visit if they misidentify 2 as false-positives.

Sensitivity statistical parameters:

Table 1: Individual probabilities for different sensitivity levels

Cases	Sensitivity 85%		Sensitivity 70%		Sensitivity 60%		Sensitivity 50%		Sensitivity 30%	
	Individual Probability	Cumulative Probability	Individual Probability	Cumulative Probability	Individual Probability	Cumulative Probability	Individual Probability	Cumulative Probability	Individual Probability	Cumulative Probability
0	0.00000	0.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00039	1.00000
1	0.00000	0.00000	0.00000	1.00000	0.00000	1.00000	0.00001	1.00000	0.00369	0.99961
2	0.00000	0.00000	0.00000	1.00000	0.00000	1.00000	0.00006	0.99999	0.01659	0.99592
3	0.00000	0.00000	0.00000	1.00000	0.00001	1.00000	0.00037	0.99994	0.04740	0.97933
4	0.00000	0.00000	0.00000	1.00000	0.00007	0.99999	0.00174	0.99957	0.09649	0.93194
5	0.00000	0.00000	0.00001	1.00000	0.00035	0.99992	0.00628	0.99783	0.14886	0.83545
6	0.00000	0.00000	0.00004	0.99999	0.00150	0.99957	0.01779	0.99155	0.18076	0.68659
7	0.00000	0.00000	0.00020	0.99996	0.00513	0.99808	0.04066	0.97376	0.17707	0.50582
8	0.00000	0.00000	0.00088	0.99975	0.01442	0.99295	0.07624	0.93310	0.14229	0.32875
9	0.00000	0.00000	0.00320	0.99887	0.03364	0.97853	0.11859	0.85686	0.09486	0.18646
10	0.00002	0.00002	0.00971	0.99567	0.06560	0.94489	0.15417	0.73827	0.05285	0.09160
11	0.00010	0.00012	0.02471	0.98596	0.10734	0.87929	0.16819	0.58409	0.02471	0.03874
12	0.00053	0.00065	0.05285	0.96126	0.14760	0.77195	0.15417	0.41591	0.00971	0.01404
13	0.00231	0.00296	0.09486	0.90840	0.17031	0.62435	0.11859	0.26173	0.00320	0.00433
14	0.00842	0.01139	0.14229	0.81354	0.16422	0.45405	0.07624	0.14314	0.00088	0.00113
15	0.02545	0.03684	0.17707	0.67125	0.13138	0.28982	0.04066	0.06690	0.00020	0.00025
16	0.06311	0.09995	0.18076	0.49418	0.08622	0.15844	0.01779	0.02624	0.00004	0.00004
17	0.12621	0.22616	0.14886	0.31341	0.04564	0.07223	0.00628	0.00845	0.00001	0.00001
18	0.19866	0.42482	0.09649	0.16455	0.01902	0.02658	0.00174	0.00217	0.00000	0.00000
19	0.23700	0.66182	0.04740	0.06806	0.00601	0.00756	0.00037	0.00043	0.00000	0.00000
20	0.20145	0.86328	0.01659	0.02067	0.00135	0.00156	0.00006	0.00006	0.00000	0.00000
21	0.10872	0.97200	0.00369	0.00408	0.00019	0.00021	0.00001	0.00001	0.00000	0.00000
22	0.02800	1.00000	0.00039	0.00039	0.00001	0.00001	0.00000	0.00000	0.00000	0.00000

Note: Cumulative probability for 85% sensitivity is used to assess Type I error (a hospital with good sensitivity drawing a sample that suggests less than passing performance. Cumulative probability for lower sensitivity levels is used to assess Type II error (a hospital with poor sensitivity drawing a sample that suggests passing performance).

Figure 1: Probability distribution for different sensitivity levels
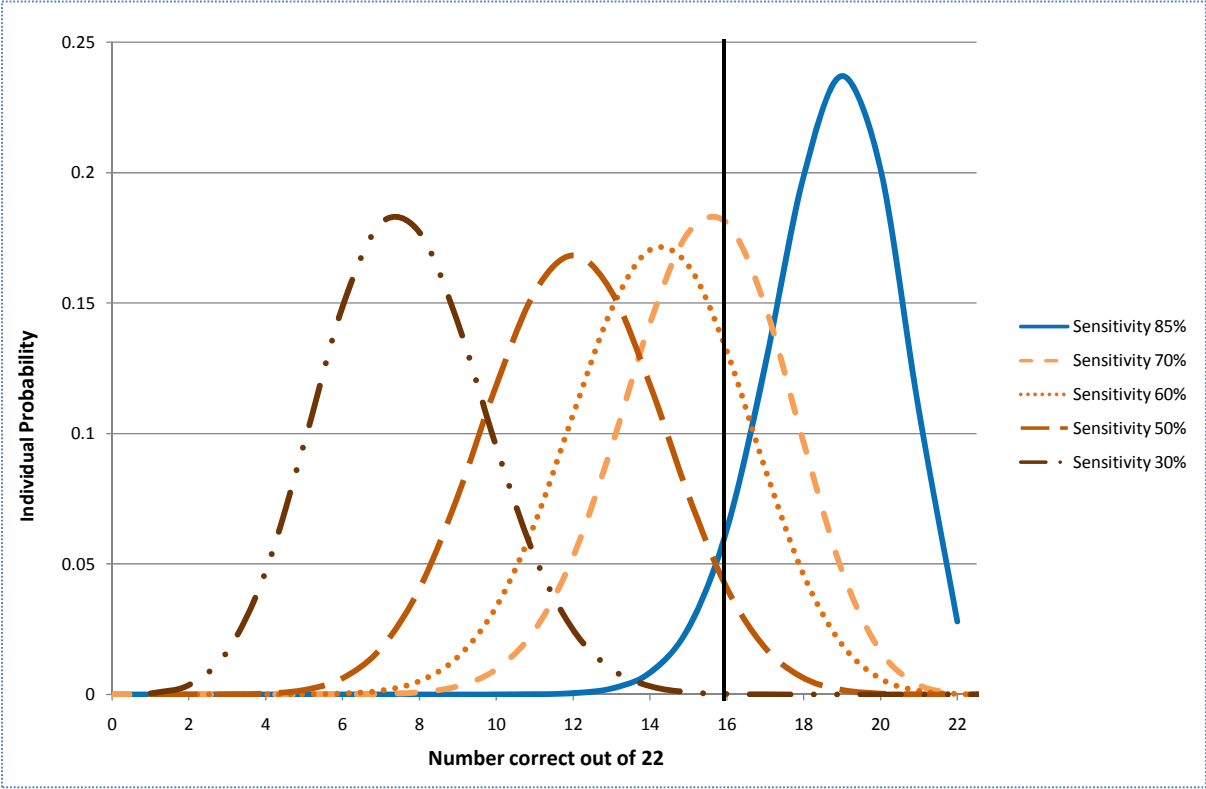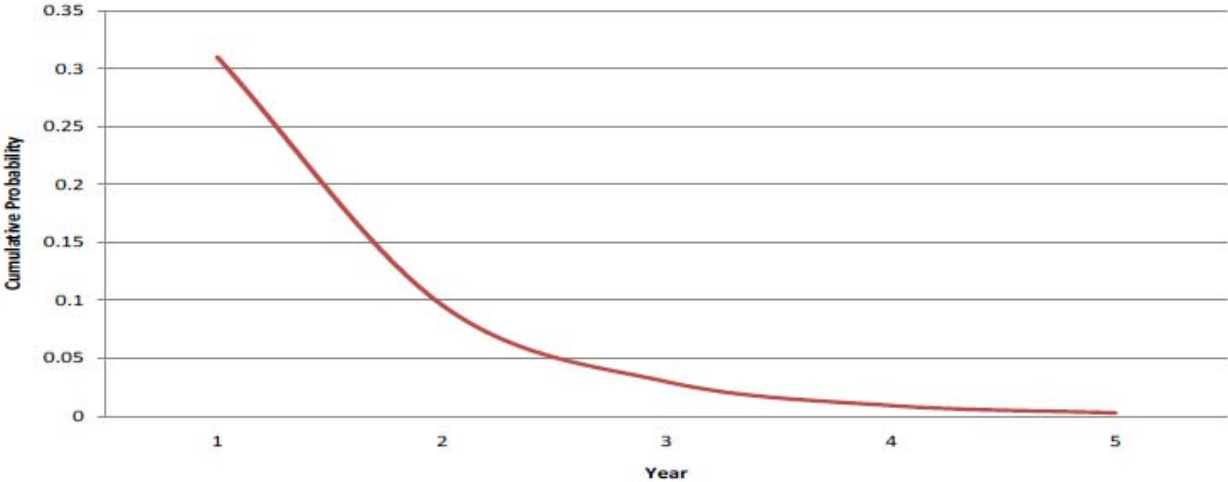
Figure 2: Probability of passing, year after year, with 70% sensitivity

Specificity statistical parameters:

Table 1: Individual probabilities for different specificity levels

Cases	Specificity 98%		Specificity 90%		Specificity 80%		Specificifty 70%		Specificifty 60%	
	Individual Probability	Cumulative Probability	Individual Probability	Cumulative Probability	Individual Probability	Cumulative Probability	Individual Probability	Cumulative Probability	Individual Probability	Cumulative Probability
0	0.00000	0.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000
1	0.00000	0.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000
2	0.00000	0.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000
3	0.00000	0.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00001	1.00000
4	0.00000	0.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00007	0.99999
5	0.00000	0.00000	0.00000	1.00000	0.00000	1.00000	0.00001	1.00000	0.00035	0.99992
6	0.00000	0.00000	0.00000	1.00000	0.00000	1.00000	0.00004	0.99999	0.00150	0.99957
7	0.00000	0.00000	0.00000	1.00000	0.00000	1.00000	0.00020	0.99996	0.00513	0.99808
8	0.00000	0.00000	0.00000	1.00000	0.00001	1.00000	0.00088	0.99975	0.01442	0.99295
9	0.00000	0.00000	0.00000	1.00000	0.00005	0.99999	0.00320	0.99887	0.03364	0.97853
10	0.00000	0.00000	0.00000	1.00000	0.00028	0.99994	0.00971	0.99567	0.06560	0.94489
11	0.00000	0.00000	0.00000	1.00000	0.00124	0.99965	0.02471	0.98596	0.10734	0.87929
12	0.00000	0.00000	0.00002	1.00000	0.00455	0.99841	0.05285	0.96126	0.14760	0.77195
13	0.00000	0.00000	0.00013	0.99998	0.01400	0.99386	0.09486	0.90840	0.17031	0.62435
14	0.00000	0.00000	0.00073	0.99985	0.03600	0.97986	0.14229	0.81354	0.16422	0.45405
15	0.00000	0.00000	0.00351	0.99912	0.07681	0.94386	0.17707	0.67125	0.13138	0.28982
16	0.00000	0.00000	0.01383	0.99561	0.13441	0.86705	0.18076	0.49418	0.08622	0.15844
17	0.00006	0.00006	0.04392	0.98178	0.18976	0.73264	0.14886	0.31341	0.04564	0.07223
18	0.00081	0.00088	0.10979	0.93787	0.21084	0.54288	0.09649	0.16455	0.01902	0.02658
19	0.00839	0.00927	0.20803	0.82807	0.17755	0.33204	0.04740	0.06806	0.00601	0.00756
20	0.06169	0.07096	0.28084	0.62004	0.10653	0.15449	0.01659	0.02067	0.00135	0.00156
21	0.28787	0.35883	0.24072	0.33920	0.04058	0.04796	0.00369	0.00408	0.00019	0.00021
22	0.64117	1.00000	0.09848	0.09848	0.00738	0.00738	0.00039	0.00039	0.00001	0.00001

Note: Cumulative probability for 98% specificity is used to assess Type I error (a hospital with good specificity drawing a sample that suggests less than passing performance. Cumulative probability for lower specificity levels is used to assess Type II error (a hospital with poor specificity drawing a sample that suggests passing performance).

Figure 1: Probability distribution for different specificity levels
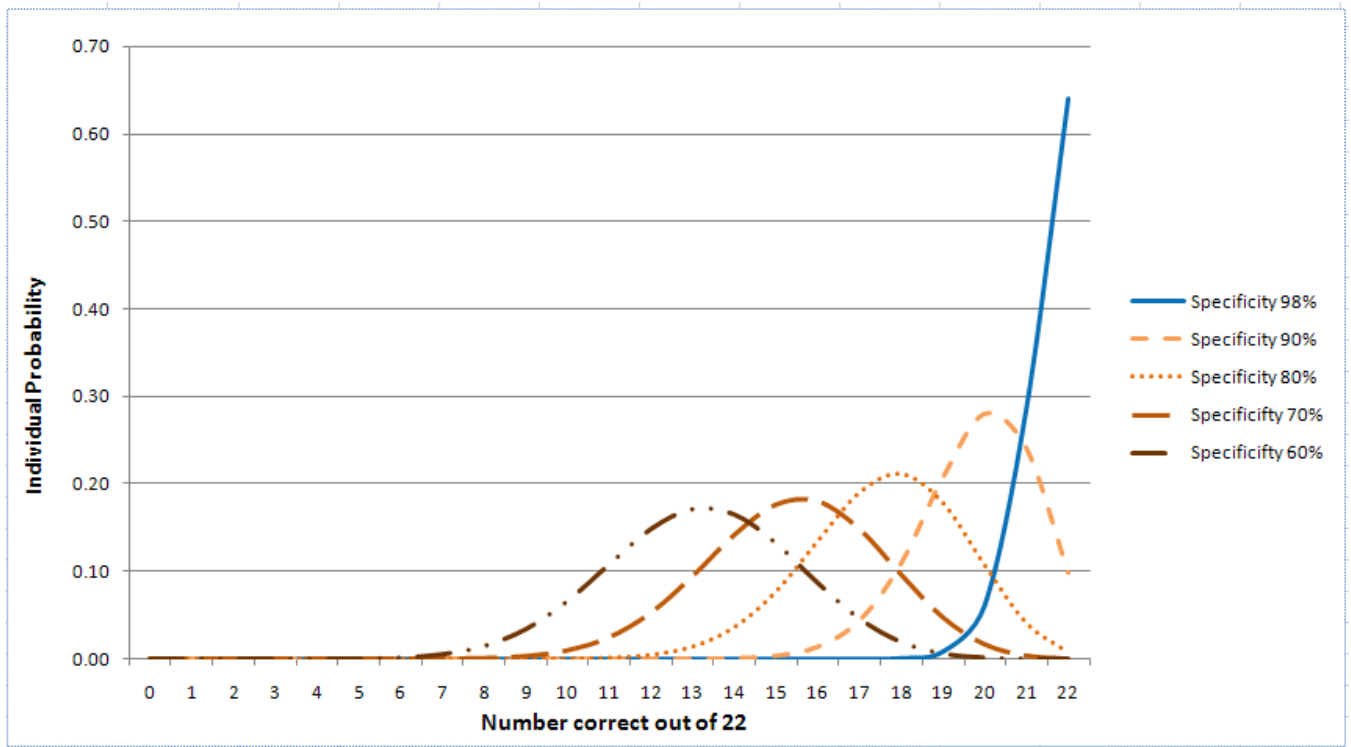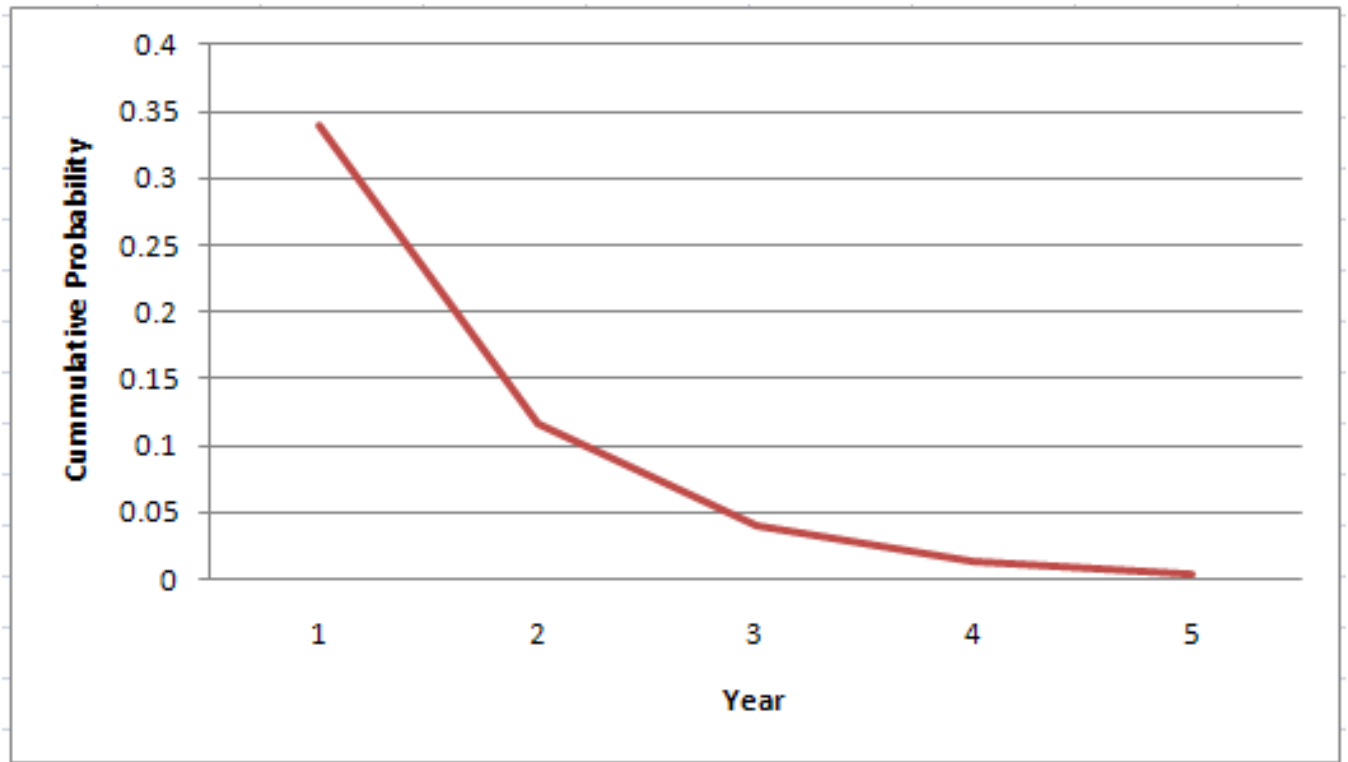
Figure 2: Probability of passing, year after year, if specificity is 90%

Section III: External Validation Verification Visit Component

We will use the ISO 2859 approach to verify whether a hospital's infection surveillance program is performing or not with the expected sensitivity and specificity. As described in Section I, this approach offers three general inspection levels; the greater discrimination afforded by Level III does not require a prohibitively large increase in sample size over Level II, so we will use Level III. There are two situations to consider for the external validation verification visit component. Some hospitals will have been picked at random; others will be chosen because of borderline or lower performance on their annual internal validation. For hospitals selected at random without prior discovery of problems in the annual internal validation component, "normal" inspection can be considered. For hospitals selected due to discovery of problems in the annual internal validation component, "tightened" inspection can be applied. ISO 2859 incorporates "switching rules" to determine when to shift back and forth between relaxed, normal and tightened inspection.

We will use the capture-recapture approach both to enrich the sample for detection of any flaws in sensitivity, as well as to explore the extent and cause of discrepancies between clinical versus surveillance definition of cases in each facility visited.

We discuss Gehan's Phase II approach here as a method held in reserve in case it ever becomes necessary to estimate the actual sensitivity or specificity rather than just verify whether it is consistent with expected levels. The primary reason for holding it in reserve rather than using it routinely is the much larger sample required (as described below, ISO 2859 is likely to require a maximum of 40 sampled items and an average closer to 30; Gehan's approach is likely to require at least as many but potentially up to 100).

Theory – Sampling by Attribute

Reference publications, tables and other tools are available from several sources to compute sample size and decision rules for ISO 2859 Sampling by Attributes plans. See, for example, the on-line calculator at http://www.sqconline.com). They start by requiring selection of "lot" or "batch" size. This is easier to define in manufacturing industries (e.g. as the number if items in each incoming shipment of materials). Sampling plans for smaller lots tend to produce operating characteristic curves showing greater likelihood of accepting non-conforming lots (β-error). Since any definition of a "lot" or "batch" size for infection surveillance reports would be arbitrary, we've selected the range that optimizes statistical power within manageable sample size (which is a Level III normal inspection plan for lots of 91-150 items). It might take large hospitals just a few weeks to have that many positive blood cultures, while for the smallest of hospitals it might take a couple of years to accumulate that many, so this affects the frequency which we might need to revisit hospitals (more frequent spot-checks among larger facilities)

among other factors that could trigger a visit (e.g. low score on annual internal validation, turn-over in infection control program staffing, etc.).

On-line calculators produce all required sampling plan design documents in one step (enter the lot size, acceptable quality level [AQL] wanted, level of inspection stringency, and they produce not only the sample size along with the decision points for accepting or rejecting the lot but also the operating characteristic curves). Solving the same design problem with tables requires use of at least two ISO 2859 (MIL STD 105) tables. On the first, you select one of the lot size ranges and the level of inspection stringency to find a code letter. That code letter is taken to a second table on which it is aligned with the desired AQL number to find the sample size and decision points. Operating characteristic curves then need to be constructed separately from the exact binomial probability formula:

$$P_a = P(k \le c) = \sum_{k=0}^{c} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

in which P_a is probability of accepting a lot of k items in which no more than c are defective based on samples of size n when the AQL is set at p.

Here, for example, are sections reproduced from two of the MIL STD 105 tables illustrating our sampling plan derivation (note that only a few rows from each table are shown here, not the complete tables, and only from those two tables pertinent to our plan rather than all of the tables in the set):

Lot or batch size	Special Inspection Level				General Inspection Level		
	S-1	S-2	S-3	S-4	I	II	III
51 to 90	B	B	C	C	C	E	F
91 to 150	B	B	C	D	D	F	G
151 to 280	B	C	D	E	E	G	H

Code Letter	Sample	Sample Size	Cumulative Size	AQL (normal inspection)									
				2.5		4.0		6.5		10		15	
				Accept	Reject	Accept	Reject	Accept	Reject	Accept	Reject	Accept	Reject
G	First	20	20	0	3	1	3	2	5	3	6	5	9
G	Second	20	40	3	4	4	5	6	7	9	10	12	13

From the first table, for a lot size of 91-150 and general inspection level III, the required code letter is G. Taking that to the second table, for code G the sample size required would be 20 for each stage; for AQL of 15% (the level of sensitivity deemed acceptable), the result would be acceptable if 5 or fewer defects are found among the first stage sample of 20, not acceptable if 9
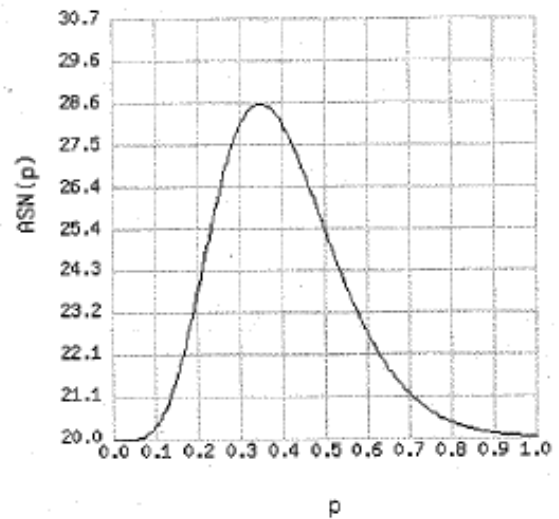
or more, otherwise continue on to the second stage. If a second stage sample of 20 more is examined, then the result would be acceptable if 12 or fewer defects are noted, unacceptable if 13 or more. Note that these Accept and Reject numbers are for "normal inspection"; for the same specifications but at "tightened" inspection" the first stage Accept number becomes 3 or less to accept and 7 or more for Reject, while the second stage Accept number becomes 11 or less to Accept and 12 or more to Reject. The operating characteristics curve for tightened inspection is steeper than the curve for normal inspection, as shown below (produced with the on-line calculator at www.sqconline.com), since it is more likely to detect non-conforming lots.

There are formal switching rules guiding when to move between "relaxed", "normal" and "tightened inspection". They generally permit moving down one level after five consecutive batches are accepted, and call for moving from "reduced" to "normal" after any one batch is rejected or from "normal" to "tightened" after any two of five consecutive batches are rejected.

The following pair of figures shows our Level III normal inspection sampling plan operating characteristics for assessing CLABSI sensitivity:
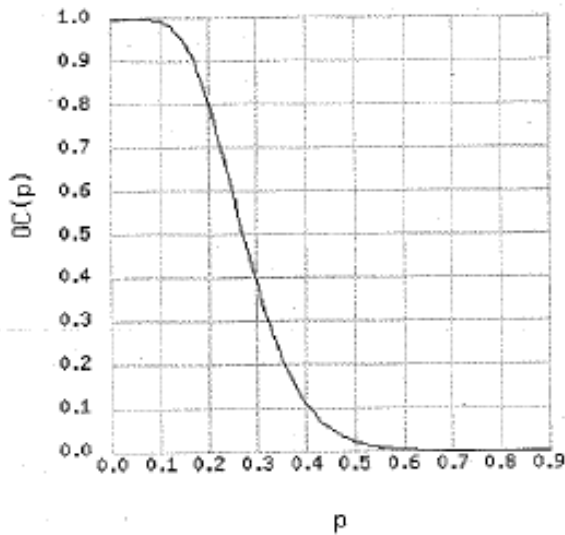

The OC curve describes the probability of accepting a lot, OC(p), as a function of the proportion non-conforming (p), for the single and double sampling plans described above.
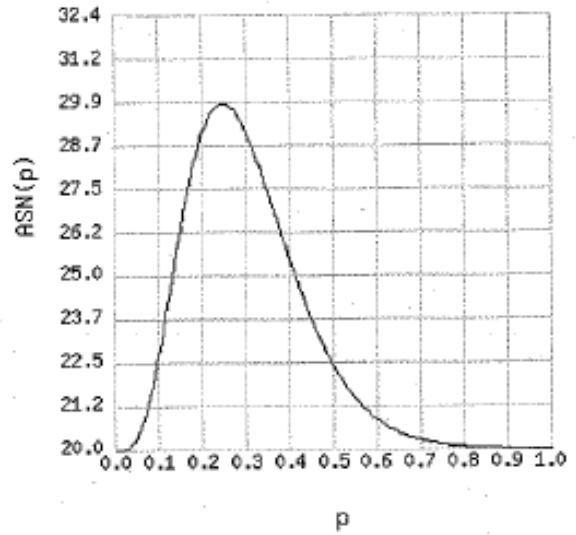
The ASN curve describes the average sample size, when employing the double sampling, as a function of the proportion of non-conforming items (p).

The following pair of figures shows our Level III tightened inspection sampling plan operating characteristics for assessing CLABSI sensitivity:
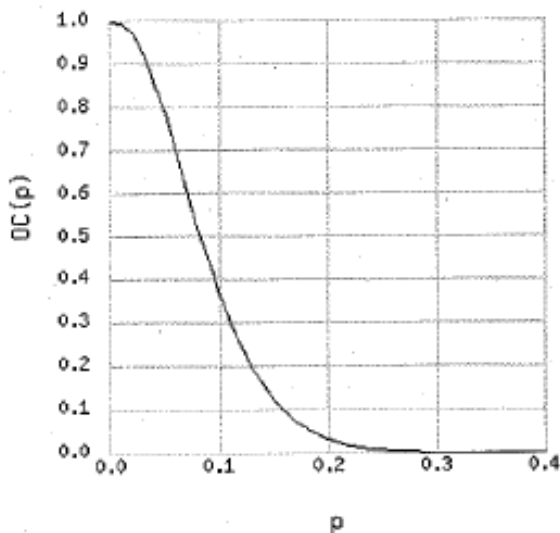
The OC curve describes the probability of accepting a lot, OC(p), as a function of the proportion non-conforming (p), for the single and double sampling plans described above.

The ASN curve describes the average sample size, when employing the double sampling, as a function of the proportion of non-conforming items (p).

The following pair of figures shows our Level III normal inspection sampling plan operating characteristics for assessing CLABSI specificity:

The OC curve describes the probability of accepting a lot, OC(p), as a function of the proportion non-conforming (p), for the single and double sampling plans described above.
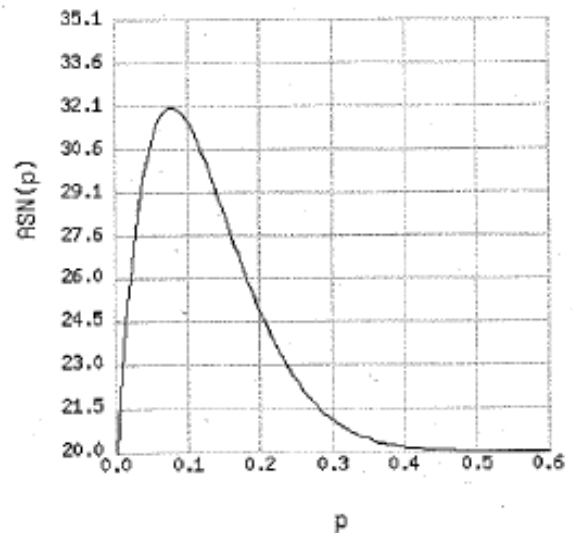
The ASN curve describes the average sample size, when employing the double sampling, as a function of the proportion of non-conforming items (p).

The following pair of figures shows our Level III tightened inspection sampling plan operating characteristics for assessing CLABSI specificity:
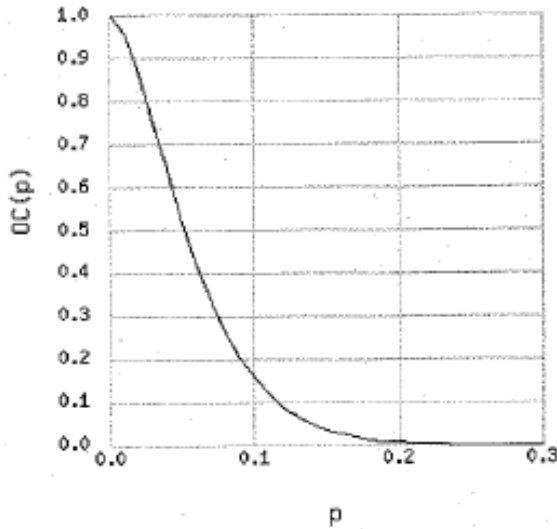

The OC curve describes the probability of accepting a lot, OC(p), as a function of the proportion non-conforming (p), for the single and double sampling plans described above.
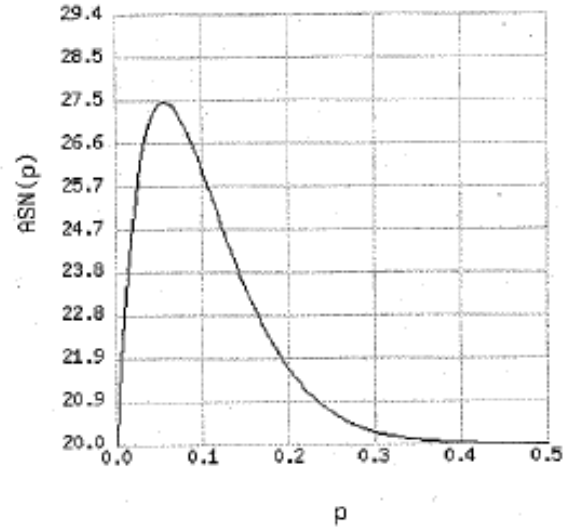
The ASN curve describes the average sample size, when employing the double sampling, as a function of the proportion of non-conforming items (p).

Theory – Capture-Recapture

A common problem in field biology is to estimate the number of mice in a field, fish in a lake, etc. The method developed for that purpose subsequently found many other applications (e.g. how many words did Shakespeare know, how complete is the reporting for a particular disease of public health importance, etc.). Essentially, the method consists of using some means (e.g. baited traps, a net, etc.) to capture a sample, tag those captured (and, in field biology, released); then using a second means to capture another sample independently (the recapture step) see not only how many were caught in a second sample but also how many of them were tagged. The basic Lincoln-Petersen method formula for this is:

$N = MC/R$

where:

N = Estimated total population number

M = Total number caught and tagged in the first capture

C = Total number caught in the second capture (the recapture)

R = Number in the second capture (the recapture) that were tagged.

This formula only works if M, C and R are all greater than zero. If M, C and R are not all greater than zero, then there also is a more refined formula that reportedly reduces bias in estimates from small samples. The refined formula uses the same notation and is:

$$N = [(M+1)(C+1)/(R+1)] – 1$$

And its variance is:

$$Var_{(N)} = [(M+1)(C+1)(M-R)(C-R)]/[(R+1)(R+1)(R+2)]$$

Theory – Gehan's Phase II Trial Design

Tables for the first and second stage of Gehan's Phase II trial design are available in chapter 10 of Machin D, Campbell M, Fayers P, Pinol A, Sample Size Tables for Clinical Studies 2nd edition, Blackwell Science Ltd, 1997. Note that precision in this book is defined differently than in the binomial distribution formulae used to compute internal validation sample size values (ε in this book expresses precision as the standard error, not as the absolute difference between an upper or lower confidence interval boundary and the central point estimate. ε=0.10 would correspond to a 95% confidence interval of ± 20%age points, ε=0.05 to ± 10%age points).

Table 10.3 in that book requires specification of "therapeutic efficacy" and statistical power (the probability of detecting a true effect if it exists). The basis for this table is to define the number of patients in whom no effect of a treatment is observed before one would decide to terminate the clinical trial. For our purposes, select in the left column ("Therapeutic efficacy") the tolerated false-positive rate upper boundary (15% ±15%age points in the internal sensitivity validation component gives an upper boundary of 30%) and look for any errors (case not known to infection surveillance program, or case known but miscoded) in the number of charts indicated under the 95% or 99% power columns (sample sizes of 9 or 13 respectively).
- ✓ If no errors are found, stop at that sample size. Conclude that the hospital's performance is satisfactory.
- ✓ If errors are detected in one or more of those charts, then continue on to the second stage of sampling (Table 10.4 in Machin et al.). For a 95% confidence interval of 30% ± 10% detected with 90% statistical power (meaning 9 times out of 10, we would expect our sample to detect a 30% error rate if that is really the hospital's performance level), this would require review of an additional 71 charts if 1 error had been found in the first stage (or 87 more if 2, 91more if 3, 87 more if 4, 74 more if 5, and 54 more if 6 errors had been found in the first stage). Compute the error rate and confidence interval from the total sample size (first and second stage combined).

Practice

1. Contact each hospital's lead ICP prior to the validation verification visit. Explain why they have been selected (either random spot-check or follow-up of internal validation result). Confirm a date for your visit, and explain what they will need to have ready for your review:
   a. A list from their laboratory of every consecutive patient who had any blood culture positive for any organism, for a total of 40 patients fitting that hospital's surveillance program profile (could be positive blood cultures from all patients, all in-patients, all ICU patients… depends on each individual hospital's surveillance practices – use the same filter criteria that the hospital applied to define the group of patients from which they pulled their sample for the internal validation process), starting with two months ago and moving backward in time. If the requisite number of cases cannot be achieved within a 24 month span, limit the list to 24 months and include every patient with positive blood culture.
   b. If a line list is maintained by an ICP, have a copy for that time period available for inspection. If other means are used to indicate the cases for which infection surveillance was notified and case review documented, have that available for inspection.
   c. A list from medical records of all patients with central-line-associated bloodstream infection noted in their discharge abstract, starting with patients discharged two months ago and moving backward in time as long as necessary to match the time period defined for (a) and (b) above. This corresponds to ICD9 code 999.31 ("infection due to central venous catheter" which by coding definition includes Hickman, PICC, Portacath, Triple lumen and umbilical venous catheters).
   d. If clinical records are paper based, have the charts of all those patients listed in (a) and (c) above available on the day of the verification visit. If clinical records are computer based, have a work station available for the reviewer on the day of the verification visit.
   e. Indicate that you will want to meet with a representative of their infection control program for approximately 30 minutes at the beginning of the visit (to be oriented to the way records are maintained there), and again at the end of your visit (to share your findings and reconcile any points they might dispute). Advise them that they may want to have the administrator to whom they report present at the end-of-visit meeting, and that person would be welcome to attend both meetings.
2. On the day of the verification visit:
   a. Bring a copy of our NHSN data listing that hospital's central-line associated bloodstream infection events.
   b. Introduce yourself, explain the purpose of the visit, and ensure that all required items are available as requested.

i. If all are not available, indicate that the visit result will be recorded as unsatisfactory, that a second verification visit will be rescheduled and if all required items are not available at that time then the matter will be reported to licensing & inspection and may be noted on the HAI website. Leave at that point.

ii. If all required items are available as requested, and only if all are available, proceed to (2)(c).

c. Review the clinical record for any patient records in which the discharge abstract indicated central-line associated bloodstream infection (ICD9 code 999.31). Make a list of those that would satisfy NHSN criteria versus those that would not (see worksheet for this purpose in Section IV). That list may include cases outside the patient care areas for which the hospital has been reporting numbers to NHSN – do not limit the list to only certain patient care areas! We will take patient care areas and reporting scope into consideration in assessing completeness of reporting, but we need the wider picture to understand magnitude of gaps between clinical versus surveillance defined cases as well as extent to which our current surveillance mandate coverage might be missing cases that occur elsewhere.

d. Review the clinical record for the 40 patients the laboratory has identified as having a positive blood culture. Decide which of those records satisfy NHSN criteria to code as a central-line associated bloodstream infection. Create a list (see worksheet in Section IV), coding each case reviewed as to primary source of infection. As specified in (1)(a) above, this set of positive cultures should correspond to the patient care areas for which the hospital has been reporting to NHSN.

i. If this is a spot-check and there have been no indications of problems, then start with 20 randomly selected charts, initially drawing from among the records in (2)(c); if no specificity (false-positive misclassification) errors and no more than 5 sensitivity (false-negative misclassification) are revealed upon comparing your list with the infection surveillance program's line-list, stop there and record the result as acceptable (the program appears to achieve the 85% sensitivity & 98% specificity expected). If 3 or more specificity errors, or 9 or more sensitivity errors were detected, stop and record the result as unacceptable. For anything in between, continue with another 20 records. If 3 or fewer specificity and 12 or fewer sensitivity errors are found among the 40 records, record the result as acceptable; if 4 or more specificity or 13 or more sensitivity errors are found among the 40, record the result as unacceptable. This corresponds to the normal level specifications explained in Section III.

ii. If this is a follow-up due to results of an internal validation result or prior validation verification visit, then start with 20 randomly selected charts, initially drawing from among the records in (2)(c); if no specificity (false-positive misclassification) errors and no more than 3 sensitivity (false-negative misclassification) are revealed upon comparing your list with the infection surveillance program's line-list, stop there and record the result as acceptable (the program appears to achieve the 85% sensitivity & 98% specificity expected). If 2 or more specificity errors, or 7 or more sensitivity errors were detected, stop and record the result as unacceptable. For anything in between continue with another 20 records. If 1 or fewer specificity and 11 or fewer sensitivity errors are found among the 40 records, record the result as acceptable; if 2 or more specificity or 12 or more sensitivity errors are found among the 40, record the result as unacceptable. This corresponds to the tightened level specifications explained in Section III.

e. Compare the lists created in (2)(c) and (2)(d) with the infection events that were reported through NHSN (reported according to our NHSN data listing). Calculate sensitivity and specificity based on the total number of charts read, then compare what you calculated to what the hospital calculated in its last internal validation documentation. Worksheets are provided in Section IV.

f. Compare the list created in (2)(c) with the list created in (2)(d). Determine why clinically defined cases fail to meet NHSN criteria: note if large numbers relate to blood culture practices (absence of cultures rather than negative cultures; high proportion of specimens drawn after initiation of antimicrobial therapy; etc.). If a reference to best practices in blood culturing is needed, refer to the most American Society for Microbiology "Cumitech" publication on blood culturing.

g. Use the capture-recapture formula to estimate the total number of central-line associated bloodstream infections during the period inspected for this hospital. The basic Lincoln-Petersen method formula applied to this purpose is:

$N = MC/R$

where:

N = Estimated total number of infections
M = Total number of infections noted in the NHSN data
C = Total number of infections noted in your chart review today
R = Number of infections noted in your chart review that were also identified in the NHSN data.

This formula only works if M, C and R are all greater than zero. If they are, go ahead to calculate first with C as only those cases meeting NHSN criteria, then a second time with C as all cases identified by clinicians whether they satisfy NHSN criteria or not. That will give us estimates for a minimum and maximum

number of infections; due to biases discussed in the Stephen reference, the first calculation might be biased toward underestimation, the latter toward overestimation. If M, C and R are not all greater than zero, then use the more refined formula below to calculate N instead.

There also is a more refined formula that reportedly reduces bias in estimates from small samples. Try it as well to help us see whether there is much of a difference. The refined formula uses the same notation and is:

$$N = [(M+1)(C+1)/(R+1)] - 1$$

And its variance is:

$$Var_{(N)} = [(M+1)(C+1)(M-R)(C-R)]/[(R+1)(R+1)(R+2)]$$

h. Discuss with the hospital's ICP your preliminary results on sensitivity, specificity, any coding misclassifications noticed and extent of discrepancies between clinical versus NHSN defined cases. If sensitivity or specificity is not satisfactory, then initiate discussion of the aspects that need to be addressed an improvement plan.

i. Retain all of the lists, worksheets and notes you created during the visit. Indicate that you will send a written report back to a designated individual at the hospital as soon as possible.

Section IV: Worksheets and Other Appended Material

A. Worksheet for Listing Clinically-Defined CLABSI Cases

Identifier for cases with an ICD-9 discharge diagnosis code indicating central-line associated bloodstream infection	Would this case meet NHSN criteria? *y = yes or n = no* *n1 = no blood cultures* *n2 = <2 blood cultures* *n3= Rx before neg culture* *n4 = other root cause (list)*	Was it reviewed by ICP? *y = yes, n = no, u = undocumented*	Was it entered in NHSN by ICP as a CLABSI event? *y = yes, n = no*

B. Worksheet for Assessing Review of Cases by ICP and Accuracy of CLABSI Coding Decision

Identifier for cases with a positive blood culture (so should have been reviewed by ICP)	Records indicate that the case actually was reviewed by ICP *y = yes*	Correctly coded by ICP as healthcare associated *y = yes;* *c = coded as CLABSI*	Incorrectly coded by ICP as healthcare associated *y = yes*
1-			
2-			
3-			
4-			
5-			
6-			
7-			
8-			
9-			
10-			
11-			
12-			
13-			
14-			
15-			
16-			
17-			
18-			
19-			
20-			
21-			
22-			
23-			
24-			
25-			
26-			
27-			
28-			
29-			
30-			
31-			

32-			
Identifier for cases with a positive blood culture (so should have been reviewed by ICP) [cont.]	Records indicate that the case actually was reviewed by ICP. [cont.]	Correctly coded by ICP as healthcare associated [cont.]	Incorrectly coded by ICP as healthcare associated [cont.]
33-			
34-			
35-			
36-			
37-			
38-			
39-			
40-			

C. Worksheet for Calculations

The decision as to whether sensitivity and specificity are sufficiently high is based on scoring criteria detailed in paragraph (2)(d) in Section III. The remaining calculations in this appendix further inform our ability to provide advice about areas to improve.

Record here whether this visit is a random spot check ☐ or a follow-up visit ☐

Record here whether normal inspection level ☐ or tightened inspection level ☐ criteria apply.

Record here the number of cases in the first column of Worksheet Table B ("Identifier for cases with a positive blood culture") that were not checked in the second column ("Records indicate that the case actually was reviewed by ICP") _____. This is the false-negative number.

Record here whether the false-negative number is ☐ or is not ☐ below the applicable unacceptable number limit specified in Section III paragraph (2)(d)(i or ii).

Record here the number of cases in the fourth column of Worksheet Table B ("Incorrectly coded by ICP as healthcare associated") _____. This is the false-positive number.

Record here whether the false-positive number is ☐ or is not ☐ below the applicable unacceptable number limit specified in Section III paragraph (2)(d)(i or ii).

If both values are below the corresponding limit, results are acceptable. Complete the calculations below, and no corrective action plan is required (although suggestions may be offered for the facility's consideration). If either value is above the corresponding limit, results are not acceptable. Complete the calculations below, and indicate that a corrective action plan will be required after our written report is provided to the facility.

<u>Sensitivity of Surveillance Program:</u>

SENSITIVITY FOR DETECTING BSI CASES

1. Total number of cases in first column of Worksheet Table B ("Identifier for cases with a positive blood culture…") =

2. Total number of "yes" check marks in second column of Worksheet Table B ("Records indicate that case actually was reviewed by ICP") =

3. Calculate sensitivity for BSI case detection by dividing the total in (2) by the total in (1). Multiply by 100 to express the result in percent here _____%.

SENSITIVITY FOR CLABSI CASE IDENTIFICATION

4. Total number of "yes" check marks in second column of Worksheet Table A ("Would this case meet NHSN criteria") =

5. Total number of "yes" check marks in corresponding rows of fourth column of Worksheet Table A ("Was it entered in NHSN by ICP as a CLABSI event? ") =

6. Calculate sensitivity for correctly coded BSI cases by dividing total in (5) by total in (4). Multiply by 100 to express the result in percent here _____%.

Specificity of Surveillance Program:

SPECIFICITY FOR CLABSI CASE IDENTIFICATION

7. Total number of cases checked "no" in second column of Worksheet Table A ("Would this case meet NHSN criteria") =

8. Total number of cases checked "no" in corresponding rows of fourth column of Worksheet Table A ("Was it entered in NHSN by ICP as a CLABSI event") =

9. Calculate specificity for correct coding by dividing total in (8) by total in (7). Multiply by 100 to express the result in percent here _____%.

Total Number of CLABSI Projected From Capture-Recapture Method:

10. Total number of cases listed in second column of Worksheet Table A ("Would this case meet NHSN criteria?") that also appear on our line list from NHSN =

11. Total number of CLABSI cases you confirmed in your chart review today (from second column in Worksheet Table A and third column in Worksheet Table B, being careful to not count duplicates as more than one case) =

12. Total number of CLABSI cases reported to NHSN in the same time period (from our NHSN line list for this time period) =

13. Calculate the minimum estimate for total number of CLABSI infections during this time period. If the totals in (10), (11) and (12) are all greater than zero, then use the formula N = MC/R where N is the minimum estimated total number of infections in the facility, M is the number from (12), C is the number from (11) and R is the number from (10). If the total in (10), (11) and (12) are not all greater than zero, then use the refined formula instead:

$$N = [(M+1)(C+1)/(R+1)] - 1$$

14. Total number of cases listed in first column of Worksheet Table A ("Identifier for cases with an ICD-9 discharge diagnosis code indicating central-line associated bloodstream infection") that also appear on our line list from NHSN =

15. Total number of CLABSI cases found in your chart review today including clinically defined cases (from <u>first</u> column in Worksheet Table A and third column in Worksheet Table B, being careful to not count duplicates as more than one case) =

16. Total number of CLABSI cases reported to NHSN in the same time period (from our NHSN line list for this time period) =

17. Calculate the maximum estimate for total number of CLABSI infections during this time period. If the totals in (14), (15) and (16) are all greater than zero, then use the formula $N = MC/R$ where N is the maximum estimated total number of infections in the facility, M is the number from (16), C is the number from (15) and R is the number from (14). If the total in (10), (11) and (12) are not all greater than zero, then use the refined formula:
$$N = [(M+1)(C+1)/(R+1)] - 1$$

Gap Analysis of Surveillance vs. Clinically Defined Cases:

18. Highlight cases that are identified clinically (first column of Worksheet Table A) but that would not fulfill NHSN criteria (see second column of Worksheet Table A). For each case, list any noteworthy findings from your chart review including whether blood cultures were not drawn, or too few were drawn in each set to have any chance of fulfilling the criteria, or not drawn until after antimicrobial therapy initiated. Also note if you discovered any cases that were missed by both physician's discharge summary and the infection surveillance program coding.

D. Form letter to notify hospital they will be visited

Dear …,

Your participation in our state's process to validate the accuracy of reported healthcare associated infection rates is very much appreciated. We are now following up the first step in that process, the annual internal validation of sensitivity and specificity, with the next step: validation verification visits. Your facility has been selected for a site visit, and we would like to confirm whether {INSERT DATE HERE} would be convenient.

Our visit will require us to be on site for one day, and take no more than a 30 minutes of your time at the start for an orientation plus 30 minutes at the end for a debriefing. If a member of your hospital's administration team would like to attend the debriefing, they would be most welcome to do so. We will need a work area where two people from the department's Healthcare Associated Infections Program can review various records during the site visit.

For the day of the visit, we would need you to have ready:

a. A list from your laboratory, starting with two months ago and moving backward in time, of every consecutive patient who had any blood culture positive for any organism, for a total of 40 patients fitting your hospital's surveillance program profile. This could be positive blood cultures from all patients, all in-patients, or all ICU patients. It depends on each individual hospital's surveillance practices – use the same filter criteria that you applied when you pulled your sample for the internal validation process. If the required number of cases cannot be achieved within a 24 month span, limit the list to 24 months and include every patient with positive blood culture.

b. If a line list is maintained by an ICP, have a copy for the same time period available. If other means are used to indicate the cases for which infection surveillance was notified and case review documented, have that available instead.

c. A list from medical records of all patients with central-line-associated bloodstream infection noted in their discharge abstract, starting with patients discharged two months ago and moving backward in time as long as necessary to match the time period defined for (a) and (b) above. This corresponds to ICD9 code 999.31 ("infection due to central venous catheter" which by coding definition includes Hickman, PICC, Portacath, Triple lumen and umbilical venous catheters).

d. If clinical records are paper based, have the charts of all those patients listed in (a) and (c) above available on the day of the verification visit. If clinical records are computer based, have a work station available for the reviewer on the day of the verification visit.

Please give Jeanne a call at (360) 236-4262 if you have question about this procedure.

Again, thank you for your assistance in ensuring that we all report accurate infection rates.

Sincerely,

David Birnbaum.

*****