Sampling Strategy to rule out the presence of unidentified samples in IRB-approved biorepositories

Purpose: To conduct a statistically sound strategy for sampling biospecimens collected and stored in NCI supported biorepositories to rule out with high certainty the presence of unanticipated "select" agents as defined by HHS/CDC.

Background: The NCI has maintained a major commitment to collection and storage of biospecimens under IRB approval. There are no known clinical or epidemiological studies that collect or analyze "select" agents in either epidemiological or clinical trials conducted by CCR, DCEG and DCTD. All studies have been conducted under the protection of NCI IRBs with informed consent. Biospecimens reside in established biorepositories and are regularly surveyed. Thus, the likelihood of discovery of unanticipated vials with "select" agents is extremely small.

Freezers or refrigerators with samples can be separated into two categories: those that contain only NCI samples or those shared with entities, separately contracted. The sampling of freezers can be stratified by these two categories to increase precision. Because of the risk related to unknown co-occupants in shared freezers could be slightly higher, oversampling of shared freezers will be considered.

Approach: The primary endpoint measured is the true proportion of unidentified samples as a fraction of the sample size. The expected upper (1-sided) 95% confidence interval for the true proportion of unidentified samples, π, provides an upper limit on the proportion of unidentified samples based on a survey of n samples. The premise is that IRB-approved samples have a very low risk for infiltration with unknown vials. To assure an estimation of the proportion of unidentified samples with sufficient precision, one can choose n such that the relative error, defined as the half-width of the confidence interval divided by the proportion, is less than 20%. This is more stringent than the criteria often used by the National Center for Health Statistics.

All vials within a chosen box would be visually inspected and matched to the electronic manifest kept by NCI and its contractors, thus confirming the accuracy of the manifests. This approach is based on the standard configuration of biorepositories, namely, vials are stored in boxes (containing between 80 and 96 vials), which reside in racks/shelves in individual freezers/refrigerators. Because unidentified vials may cluster within boxes, the goal is to inspect a large fraction of boxes to achieve increased precision in the presence of potential clustering.

With a sample of n=10,000 or more per each strata, the expected upper 95% confidence limit (CL) for various true underlying proportions of unidentified samples is presented in the **Table** below. Notable are the tight confidence intervals between the true unidentified proportion and 95% CL, underscoring the extremely small likelihood that the underlying proportion of unidentified samples would be substantially underestimated. The sampling strategy gains little after sampling more than 10,000 samples per strata (**Figure**). Moreover, the table demonstrates that with a sample size of 10,000 or more per strata, we can determine a small upper bound on the proportion of unidentified vials. A sample of size n=10,000 achieves this precision if the true proportion of unidentified samples is 0.01 or greater.

Table. Expected upper 95% confidence limits from a sample of n=10,000 vials for various true proportions of unidentified vials

True unidentified proportion	0.000	0.001	0.005	0.010	0.020
Expected upper 95% confidence limit	0.0002	0.0015	0.0060	0.0114	0.0219