

Supporting Statement for the ERS - Census of Users of the National Plant Germplasm System

December 7, 2016

Review Conducted by Mingshan Zheng (Mathematical Statistician), at NASS Methods Division

General Comments:

This docket is well described and documented. Its direct impact on efficient government will be very beneficial in our current budget situation.

Part B: Collections of Information Employing Statistical Methods Part1.

Part1.

First, I found an inconsistency with regards to the percentage of U.S. users between the first two paragraphs. In the first paragraph it is stated that the population is 6,009 individuals, of which 78% are U.S. users. In the second paragraph there is a breakdown of the exact number of each group amongst the U.S. users, but when summed up it only comes out to be 76% of the total users. Clarification is needed as to where the extra 2% is coming from.

ERS response: The reviewer is correct. A typographical error has been rectified.

Secondly, since the expected response rate for the entire Census is 73 percent, and the demand-weighted response rate is 78 percent, in that case, even if you can get 80 percent response rate for U.S. users, it will still only be 62.4 percent of the entire Census. Due to a limited budget for telephone follow-up with non-U.S. users, this could cause non-response bias. Additionally, I believe the stated response rate of 100 percent for U.S. non-respondents to be over-estimated, as it seems likely that a non-respondent may also not respond to additional follow-ups.

My suggestion is to possibly make the additional follow-up version of the questionnaire an online survey of both U.S. and non-U.S. non-respondents. This would help to decrease the potential non-response bias of both response groups and potentially save time on phone surveys, assuming that some non-respondents would do the online survey instead.

ERS response: The census findings will focus on the demand for NPGS resources, which concerns the number of accessions distributed rather than the census population. More detailed analysis of U.S. users is required for several reasons. First, the geospatial analysis of germplasm use is directed to U.S. requestors only. Second, U.S. users receive the majority of the samples distributed and meeting the needs of U.S. users (now and in the future) is the primary policy objective for the NPGS. Thus, the non-response bias study will focus on U.S. users and will provide additional information about whether non-response bias exists for this group. The NPGS is nonetheless interested in the ways germplasm is being used by international requestors, in part because of the goodwill generated by the two-way exchange of germplasm. For the benefit of NPGS, asking questions of international users will allow comparisons to be made with results of the earlier study of respondents who demanded germplasm during 1995-1999 that was

mentioned in Statements A and B (“Demand for genetic resources and the U.S. National Plant Germplasm System,” *Crop Sci.*, 2006).

We have adopted the reviewer’s suggestion that the sample of 100 non-respondents for the bias study should first receive an invitation by email with a link an online survey. If there is no response to that, the requestor will be asked the same questions via a telephone call. The 97% response rate is based on Frey’s *National Plant Breeding Study I* (1996), which addresses a similar body of U.S. plant breeders. Frey used telephone communications to contact non-respondents, and achieved an overall response rate of 97.5%. *The National Plant Breeding Study* relied on professional relationships similar to those that the census organizers share with NPGS users. In addition, we note that Attachment D shows the assumption of 70% response rate from all non-U.S. users, unweighted. This implies a 73% unweighted response rate from all users, including non-U.S. users, and a 77% demand-weighted response rate for all users.

## Part2.

Regarding the paragraph discussing the degree of accuracy needed, there doesn’t appear to be a clear statement on the accuracy required. The paragraph mainly discusses the reasoning and importance of the accuracy, but doesn’t go into further detail. In addition to detailing how this is not a sampling survey and is instead a one-time collection, there should be more elaboration on what level of accuracy is required and how that will be pursued, possibly an approximate sampling error anticipated based on response rate that you are expecting.

ERS response: Because this information collection is a census, sampling error is not an issue. Our primary concern will be unit non-response. Various methods can be used to adjust for non-response, for example the use of known information about the non-respondents to model likely response. In some, but not all, cases, assumptions about the nature of the non-response process can be used to develop statistical properties of these models. Sensitivity analyses can also be used to illustrate the effects of model assumptions (Little, 1988; Schaffer and Graham, 2002; NASS, 2012). These methods can be applied to both U.S. and international users. However, the planned geospatial analysis, applied to data from U.S. users only, will require greater attention to non-response error and content error in the U.S. sub-population. We plan to examine the nature of non-response for U.S. users if their demand-weighted response rate falls below 80%. The non-response study is designed to determine in particular whether non-respondents are more likely than respondents to expect decreasing use of NPGS germplasm. (See the note of caution described in the response to Part 3.)

## Part3.

See suggestion for non-response study in the first section above. Additionally, it is mentioned that if non-response bias is found to be an issue, caution will be noted in the reporting of the result. How will this caution be noted, and if it is noted, will the survey still be of the required quality given that non-response bias would have been found to be an issue?

ERS response: If non-response bias is found, in particular a bias suggesting non-respondents are more likely to decrease their use of the NPGS, caution will be noted, specifically that results may apply only to respondents of the census. In that case, results from the population of respondents can still be compared with the population of respondents who demanded germplasm during

1995-1999 (“Demand for genetic resources and the U.S. National Plant Germplasm System,” *Crop Sci.*, 2006). (We have no reason to suppose that the structure of non-response bias would be any different in these two groups.) Other observations from this study (e.g. regarding the traits sought by respondents) will still provide valuable information for policy makers and NPGS managers. For managers of genetic resources, using this information will be preferable to having no information about these variables. Also, these results will still enhance ERS’ understanding of changes in the role of genetic resources in the agricultural R&D process.

Part4.

No comment.

Part5.

No comment.

Frey, K. J. 1996. *National Plant Breeding Study-I: Human and Financial Resources Devoted to Plant Breeding Research and Development in the United States in 1994*, Special Report 98. Iowa Agricultural and Home Economics Experiment Station.

Little, R. J. A., 1988. “Missing-Data Adjustments in Large Surveys,” *Journal of Business & Economic Statistics*, Vol. 6, No. 3 (Jul., 1988), pp. 287-296.

National Agricultural Statistics Service. 2014. “Census of Agriculture Methodology,” *2012 Census of Agriculture*, Appendix A, (May 2014), USDA.

Rubenstein, K. D., Smale, M., & Widrlechner, M. P. 2006. Demand for genetic resources and the U.S. National Plant Germplasm System,” *Crop Science*, 46(3), 1021-1031.

Schafer, J and J. W. Graham. 2002. “Missing Data: Our View of the State of the Art,” *Psychological Methods*, Vol. 7, No. 2, 147–177.