Implementing an Effective Survey Operations for a Research and Development Survey in the Philippines

Ramoncito G. Cambel[1], Dalisay S. Maligalig, Maurice C. Borromeo, and Ronald R. Roldan Jr.
Institute of Statistics, University of the Philippines Los Baños

Clifford B. Lesmoras
Education Development Center

ABSTRACT

Studies have shown that research and development (R&D) is a good driver of economic growth. Policies and programs that are based on good quality data are expected to produce better results. Hence, to formulate and implement policies and programs in R&D, good quality data is vital. A good data support system is also essential in identifying critical areas that need intervention, formulating viable approaches in addressing these issues, and allocating limited resources. In the Philippines, the Department of Science and Technology (DOST) has been conducting the Survey on Research and Development Expenditures and Personnel (R&D Survey) since 2003 so that R&D data and indicators can be compiled. To ensure that good quality R&D data and indicators are achieved, the DOST granted a research fund to the Institute of Statistics (INSTAT) of the University of the Philippines Los Baños (UPLB) in 2018 to further improve the design, conduct and analysis of the R&D Survey. This paper describes the processes that were developed and implemented through this research grant in relation to the dimensions of data quality, namely, relevance, accuracy, timeliness, accessibility, coherence, and comparability. Based on the evaluation of these processes, the paper also recommends further improvement on the survey operations of future rounds of the R&D Survey.

Keywords: *Survey on Research and Development Expenditures and Personnel, data quality*

1. Introduction

Research and development (R&D) comprise creative and systematic work undertaken to increase the stock of knowledge and to devise new applications of available knowledge (OECD, 2015). It is an important driver of sustainable economic development (Khan, 2015). Research has shown that R&D expenditure is significantly correlated with economic growth (Bayarcelic and Tasel, 2012) in developing countries. Akcali and Sismanoglu (2015) also established that R&D expenditures have a positive effect on the economic growth of both developing and developed countries. Sokolov-Mladenovic et al. (2016) confirmed that R&D investment has a positive effect on the real economic growth rate. Based on the data of 28 countries of the European Union for the period 2002 to 2012, the time when a financial crisis hit Europe, a percentage increase in the gross domestic expenditure on R&D (GERD) increased the GDP growth by more than 2 percentage points. Blanco et al. (2016) stated that R&D has a

[1] **Address correspondence to Ramoncito G. Cambel**: rgcambel2@up.edu.ph

significant effect on both the states' outputs in the United States and their respective total factor productivity in the long term.

Policies and programs to promote R&D must be formulated and implemented to propel economic growth. They can be developed with better quality data. Good quality data is also needed to identify critical areas that need intervention, formulate viable approaches in addressing these issues, and allocate limited resources.

At the national level, data and indicators that are important to national issues like economic growth, inflation, poverty, and employment, are compiled and published by the Philippine Statistics Authority (PSA). Because R&D indicators are not yet part of the Philippines' official statistics, the Department of Science and Technology (DOST) has been conducting the Survey on Research and Development Expenditures and Personnel (R&D Survey) since 2003 so that R&D indicators can be compiled. This survey captures the R&D expenditure and personnel data from three different sectors of the economy, namely, government, higher education institutions (HEIs), and private non-profit institutions (PNPIs) while the Philippine Statistics Authority (PSA) gathers some R&D data from the business and industry sector through the Annual Survey of Philippine Business and Industry (ASPBI). DOST combines the data from the R&D survey and ASPBI to provide national and regional estimates of the R&D indicators.

In the interest of achieving better quality data, the DOST gave a research grant to the Institute of Statistics (INSTAT) of the University of the Philippines Los Baños (UPLB) to further improve the design, conduct and analysis of the R&D Survey in December 2018. This paper describes the processes that were developed and implemented through this research grant to achieve good quality R&D data and indicators. Based on the evaluation of these processes, the paper also recommends further improvement on the survey operations of future rounds of the R&D Survey.

2. Data Quality and Survey Operations

The quality of data collected can be assessed using the six dimensions of data quality or quality of statistical outputs (Astrologo et. al., 2019) namely: relevance, accuracy, timeliness, accessibility, coherence, and comparability. These six dimensions were applied in planning the R&D Survey. Because R&D data and indicators are already used for planning and monitoring, their relevance is already demonstrated. Usually, accuracy is measured using the mean squared error of important estimates, say $\mathrm{MSE}(\theta)$, where θ is an estimator. Since $\mathrm{MSE}(\theta) = \mathrm{Var}(\theta) + \mathrm{Bias}^2(\theta)$, where $\mathrm{Var}(\theta)$ is the variance and $\mathrm{Bias}(\theta)$ is its bias of the estimator θ, the approach that is usually taken in designing probability sample surveys is to ensure that adequate sample size is achieved at a tolerable margin of error and best practices in survey operations are implemented so that bias is kept at a minimum. While the variance of an estimator can be estimated from a probability sample survey, bias can only be measured if credible external data sources are available, or another survey is conducted specifically for this purpose. Timeliness of survey results, comparability of estimates, coherence and accessibility can also be achieved through effective survey operations.

As shown in Figure 1, conducting a sample survey involves three major stages: planning, operations, and evaluation. The objectives of the survey, its contents and procedures are developed in the planning stage and the sample design and data collection plans are implemented in the survey operation stage. The third and last stage is evaluation to assess the quality of the survey data which is important for planning the next survey rounds, if any.

Survey operations include the construction of the sampling frame, selection of the sample, data collection, processing, estimation and analysis, and dissemination of survey results. Bias is controlled and hence, accuracy could be improved, if these tasks are done well and efficiently. Timely dissemination and access to data and indicators will also be achieved in the process. Coherence is attained when the concepts and definitions and determination of the target population are uniformly applied across time and space. Thus, the development of survey instruments, the construction of sampling frame, and the training of data collectors and supervisors are important tasks that need to be done well to establish coherence. Interpretability is gained when users could easily use and properly analyze the survey data. This implies that the meta-data – the attributes of characteristics of interest to be measured like definitions of concepts, units of measures, target population, and the limitation of the data should be made available to users.

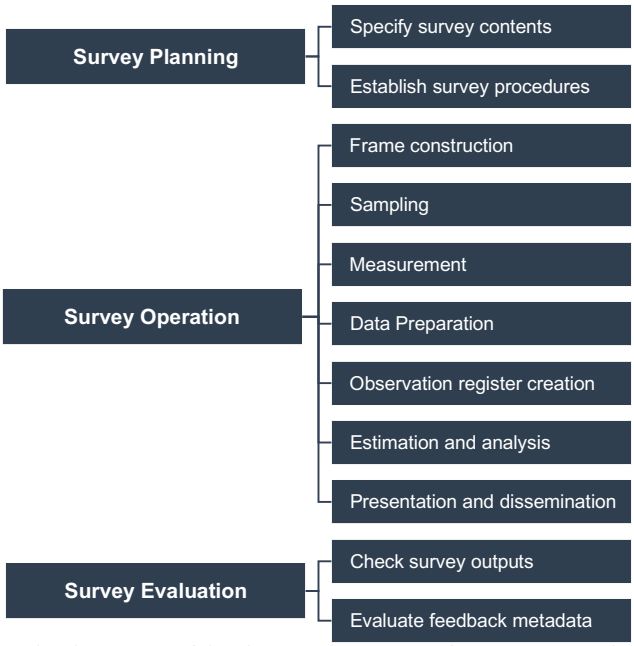

Figure 1. A visual summary of the phases and processes of a survey processing system. Source: Sundgren (1999)

2.1 Sampling Frame Construction

Survey operations begin with the construction of the sampling frame. This task is contingent upon the target population, which in the case of the R&D Survey, consists of all institutions that undertook in-house research in a given reference period, the calendar year of 2018. Since the R&D survey was designed to be a probability survey, each institution that performs R&D should have a chance of being represented in the survey. This is possible if the sampling frame that is used for selecting the sample includes all the target population units. However, unlike many surveys in which there is a readily available sampling frame, there is no comprehensive list of institutions that undertook in-house research for government, HEIs, and PNPIs. An initial sampling frame was developed from the following data sources:

For the government sector, online searches, and visits to websites of various government agencies and offices were done. An updated list of government institutions was constructed with information about their head agency, head of the office, address, telephone number, and email address.

For the HEIs, the Commission on Higher Education (CHED) provided a list of HEIs for the academic year 2017-2018. The list, however, needed updating since some contact information like the head of the institution and email address were outdated or missing for some HEIs especially for the satellite campuses of university and college systems. Moreover, new HEIs that were identified from online searches which were not found in the list were added and HEIs found to have ceased their operations were removed from the list. Updating the list was done by examining every HEI's website and contacting them through emails and telephone calls.

For the PNPIs, a list of non-stock, non-profit organizations that meet established minimum criteria for non-government organization (NGO) governance and accountability found in the Philippine Council for NGO Certification (PCNC) website was combined with the initial list of PNPIs who responded in the previous R&D surveys. Institutions known to conduct R&D but are not part of the initial list were also added to the list.

Because the web searches and other research that were done did not render a strong indication that all the institutions in the initial sampling frame have in-house R&D, it is quite likely that some of these institutions are not eligible to take part in the survey. If many ineligible institutions are selected in the sample, then the resulting survey estimates may not be as precise as planned. To improve the efficiency of the initial sampling frame, a two-phase survey was implemented, in which Phase 1 would screen out institutions that did not have any in-house research while the R&D Survey questionnaire would be administered in Phase 2 on the selected sample drawn from the improved sampling frame from Phase 1.

The Phase 1 survey was conducted online using SurveyMonkey, an online survey platform that allows the creation and sending of questionnaires in electronic format. The use of this online platform allowed respondents to complete the survey at their convenience. It resulted in a very low participation rate as shown in Table 1. It turned out that despite the preliminary activities described above, many sampled institutions did not have updated contact information. There were also changes in the management hierarchy of some institutions. Moreover, some of them also ignored the invitation to participate in the R&D Survey that was sent through email because they do not consider them as an official invitation. These cases are examples of cultural and technological practices of population units that should be considered when designing a survey so that costs and data quality are optimized (De Leeuw, Hox, and Dillman, 2008). In the context of the R&D survey, there should have been more analysis that was done on how the telephone or Internet surveys are viewed by the respondents from the various sectors – government, HEIs, and PNPIs.

Because of the low response rates in the Phase 1 of the survey, the desired improvement in the initial sampling frame did not materialize. Only 194 of the 3,378 institutions in the initial list responded, and of these, only 107 institutions have in-house R&D. To improve the sampling frame, imputations based on credible external sources were done. The 107 institutions with in-house R&D from those that responded form the initial list in the revised sampling frame. The respondents in the previous rounds of the R&D Survey were reviewed and those that did not respond to the Phase 1 survey were added in the sampling frame. Both CHED and DOST also provided lists of institutions that were given research grants. These were also added if they were not yet in the initial list. Web searches were also done to identify additional institutions that have published research or some other indications of R&D like seminars, workshops, news releases about research studies that they did.

Table 1 summarizes the resulting counts of institutions that may have R&D in the reference period in the different phases of sampling frame construction. Note that there was an increase in the number of institutions in the sampling frame at the end of Phase 2 for both government and PNPIs. The additional institutions were added when they were only identified during survey operations as having R&D. This addition to the sampling frame did not affect the sampling design since institutions in both government and PNPIs were selected with certainty.

Table 1. Number of institutions that undertake R&D by sector

Sector	Initial list	Phase 1 results		Imputations*			With R&D at the end of Phase 1	With R&D at the end of Phase 2
		Responded	With R&D	DOST R&D List	CHED/ DOST Projects	Web search		
GOV	670	73	32	88 (88)	21 (31)	14 (15)	155	263
HEI	2,354	100	66	269 (298)	43 (171)	487 (530)	865	865
PNPI	354	21	9	29 (32)	0 (0)	24 (24)	62	66

* The numbers enclosed in parentheses indicate the number of institutions found in the external data source, whereas the number above is the number of institutions in the external source that was captured in addition to the ones identified in Phase 1 or using the previously used external source, e.g., data sources from CHED and DOST listed 171 HEIs that were granted research projects, of which 43 HEIs were found to perform R&D but were not captured in Phase 1 and the DOST R&D list.

Sources: Challenges in Designing and Implementing Research and Development Surveys in the Philippines (Maligalig et al., 2019), 2018 R&D Survey Report (Maligalig et al., 2021)

The imputations did not guarantee that all institutions in the revised sampling frame have R&D and hence, the status of the sampled institutions must be identified so that adjustments for non-coverage error can be introduced after the survey. A system for monitoring the responses of the institutions and identifying their eligibility status was developed and implemented.

The approach of combining various data sources for constructing the sampling frame that was taken is similar to the study of Arora et al. (2021). The sampling frame of small and medium-sized enterprises (SMEs) was constructed by using patent, search engine, and website data. The resulting sampling frame of innovative SMEs did not have substantial coverage error. Similarly, the Italian National Statistical Institute (Istat), uses various data sources in creating a sampling frame for the Italian business R&D survey. These include previous Istat R&D surveys, other Istat business surveys with R&D-related questions, the Italian register of active enterprises, the Italian register of R&D performing institutions, data on national and European Union funding to research projects, patent databases, business reports, and data from the Italian Tax Agency (Istituto Nazionale Di Statistica, 2016).

2.2 Sampling

The planned sampling design was implemented after the revised sampling frame was finalized. For the government and PNPIs, a census of all institutions was employed since their population sizes are manageable. For the 865 HEIs in the sampling frame, proportionate stratified random sampling was employed. Since research has shown that HEIs with a higher number of graduate students are likely to have R&D, HEIs were stratified according to the graduate student size.

Those with at least 1,000 graduate students were grouped as "Large HEIs", and those with less than 1,000 were grouped in the "Small HEIs" stratum. Those that did not have data on the number of graduate students were classified as "Unknown". All the 55 HEIs in the "Large HEIs" stratum were included in the sample.

Sample sizes for the two strata (Small HEIs and Unknown HEIs) were computed using the formula:

$$n = \frac{Z_{\alpha/2}^{\ 2} PQ}{d^2 + \frac{Z_{\alpha/2}^{\ 2} PQ}{N}}$$

where $Z_{\alpha/2}$ is the abscissa of the standard normal distribution given $(1-\alpha)100\%$ confidence level; N is the population size; P is the proportion of a major characteristic of interest; $Q = 1 - P$; and d is the margin of error. Since P is usually unknown, it can be assumed based on prior information about its value from previous surveys or studies, but it can also be set to $P = 0.50$ to produce the sample size that would result to the most conservative estimate of the population variability. Considering several scenarios, the sample sizes were determined using a margin of error of 0.05 and a level of confidence of 0.95 to ensure a greater balance between the resources and the precision of estimates. Out of 355 Small HEIs, 193 were selected. Meanwhile, 218 out of 455 Unknown HEIs were selected. Regional allocations were done proportionately.

2.3 Measurement

2.3.1 Questionnaire Design

While the sampling frame was being constructed, the questionnaires were also being developed. A short questionnaire was developed for the Phase 1 survey. Questions included whether the institution performed in-house R&D during the reference year, the contact details of personnel who are most knowledgeable about their R&D activities, information on whether institutions implement centralized or decentralized reckoning systems to monitor their R&D activities, administrative units of institutions that independently conducted R&D and the incumbent heads of these units. The administrative unit is defined as the institution's constituent entity that has a certain degree of autonomy to perform its respective mandates. It is usually characterized by the appointment of an official designated to lead, supervise, and/or oversee the unit's activities. Identification of administrative units from the institutions is important so that appropriate estimates can be derived.

For the higher education sector, additional questions about the total number of graduate faculty and students were added to the Phase 1 questionnaire. As explained in the sampling part, this information was used in the stratification scheme in the sampling design of the said sector. For the government and private non-profit sectors, the Phase 1 survey collected information if the institution provided R&D funds to other institutions in 2018. In general, the information derived from Phase 1 and the records of the previous survey rounds served as the basis for identifying the target institutions in Phase 2.

In the case of the survey proper, the questionnaire for the previous survey rounds, the UNESCO Institute for Statistics (UIS) recommended template and the required data items to be collected were considered in redesigning the questionnaire. As per UIS (2014), the survey questionnaire must include a minimum number of basic questions on R&D activity. The questionnaire should

be simple and short, logically structured, and provide clear definitions and instructions including explanatory notes and hypothetical examples. Recommended guidelines indicated in the Frascati Manual (OECD, 2015) served as a reference of the definitions and classifications of R&D personnel and expenditures that were included in the questionnaires. This ensures that the indicators derived from the survey are of international standards.

A consultative workshop with key stakeholders was held in which the plan for the redesign of the questionnaire was discussed. After extensive consultations, it was decided that the questionnaire be streamlined to reduce the respondent's burden and increase the response rate. Instead of requiring the respondents to enumerate the research projects and personnel involved of institutions, summary data at the institution level became the data requirements in the survey. Although there were changes in the questions, the data collected could still measure the R&D indicators generated in the past surveys.

Definitions and key concepts were included in the questionnaires like basic concepts of research and development, the characteristics of activities that can be considered as R&D and types of activities that should be included or excluded as part of R&D, definitions of R&D personnel and personnel, types of personnel, percentage of time spent on R&D, research expenditures according to accounting categories, type of research, fields of science, and socio-economic objectives. They helped improve the accuracy of responses and minimize the risk of committing measurement error.

The revised questionnaire was pre-tested on some institutions in government, private sector, and HEIs. It was found that the questionnaire requires data from different sources within the institutions and hence, completing it requires substantial time and effort, and coordination among different units of an institution like the human resources and personnel, accounting, and planning units.

2.3.2 Data Collection

The modes of data collection of these surveys are primarily contingent on the organizational structure of sectors. Commonly, data on R&D are collected through multiple R&D surveys, which can be consolidated in coming up with national estimates. As per the UNESCO Institute of Statistics (UIS) in 2014, prevailing norms that govern information exchange, showing an understanding of the way that organizations may guard their information assets are of consideration in conducting R&D surveys.

For instance, in the annual U.S. Higher Education Research and Development (HERD) Survey, respondents may answer through a paper survey or using the web-based data collection system. For both methods, telephone and email follow-ups were employed to increase the participation rate. For the Survey of Industry Research and Development (SIRD), which is an annual sample survey that intends to include or represent all for-profit R&D-performing companies, either publicly or privately held, respondents are mailed with the survey. After a certain number of days, letters and telephone follow-ups are made. Lastly, for the annual Survey of State Government Research and Development (SGRD), respondents respond thru a self-administered questionnaire. Same with HERD and SIRD, telephone and e-mail follow-up with survey respondents are done.

In Canada, the Annual Survey of Research and Development in Canada Industry (RDCI) collects data primarily through an electronic questionnaire while providing respondents with the option of receiving a paper questionnaire, replying by telephone interview, or using other electronic filing methods. Data collected in this survey are supplemented with information that is extracted from administrative files. In addition to RDCI, the Survey on Research and Development in the Higher Education Sector (RDHES) and the Annual Survey of Research

and Development of Canadian Private Non-profit Organizations (RDNP) are also carried out to gather R&D data from higher HEIs and PNPIs, respectively, which are not covered in the RDCI. In RDNP, survey questionnaires are mailed out to target respondents and followed up with a phone call to verify receipt. Institutions who have not yet responded to the survey are followed up thru telephone up to five times, with effort devoted to organizations that are believed to perform R&D

In the 2018 Philippine R&D Survey, the questionnaires were intended to be administered online with telephone and personal follow-ups. This approach would reduce data processing and validation because online survey applications allow for self-administered questionnaire and automated validation of responses. Online probability surveys, however, require that sampled units must have updated email addresses and access to the Internet. Phase 1 of the survey was launched in the last week of May 2019. Weblinks to the questionnaires, the endorsements from the DOST Secretary and the CHED Chair for HEIs, were included in the invitation to participate in the survey that was sent to respondents through email. To increase the participation in the survey, telephone follow-ups were conducted two weeks after the launching of the online survey. These follow-ups were also used to update the contact information of the sampled institutions. Several rounds of telephone follow-ups were carried out but as shown in Table 1, the response rate was quite low, and thus, the planned two-phase sampling design had to be modified

The objective of the Phase 1 survey to eliminate the ineligible institutions in the initial sampling frame was not achieved and hence, imputations on the eligibility status of institutions were done. Because the likelihood of getting ineligible sampled units has not decreased, a monitoring system that identifies and records ineligible units was established. This procedure is presented in Figure 2. Each reply that is received is examined and its eligibility status is determined. If the respondent sent a fully filled-out questionnaire, with both expenditure and personnel data, then it is classified as eligible. If critical expenditure data is missing, a telephone follow-up by the designated supervisor or survey manager is done to check whether the responding institution has undertaken in-house R&D or not. If the institution is deemed ineligible, a letter stating its ineligibility status is requested for documentation purposes. There were also HEIs that would have several campuses in different locations that were included in the sample, but they share a common research fund with the main campus managing the R&D budget, coordinating the research activities, and outputs. In this case, the main campus is deemed eligible while the other campuses were declared ineligible. When a sampled institution conveys that it did not undertake in-house R&D, a telephone follow-up is also done to make sure that the respondent understands the definition of R&D and the reference year before the final eligibility status is determined.

The Phase 2 survey questionnaire was rolled out via SurveyMonkey in the first week of September 2019. Separate questionnaires were sent to administrative units of institutions that advised in the Phase 1 that they have independent administrative units doing R&D. Telephone follow-ups were done one month after the launch of the survey. At most three attempts were made by the telephone enumerator to each non-responding sampled institution.

Because of the lower-than-expected response rate in the last quarter of 2019, the online survey mode had to be adjusted. Based on the telephone follow-ups that were done, it turned out that many institutions do not recognize email correspondence as official, and hence, they ignored the email invitations to participate in the R&D Survey. In January 2020, formal letters enclosed with a printed copy of the questionnaire, a link to the SurveyMonkey questionnaire and, the option of using a digital copy of the questionnaire to reply were sent to the institutions through the post. Endorsement letters from the DOST Secretary, and the CHED Chairperson, in the case of HEIs, were also included. The endorsements were important in getting the trust and cooperation of the sampled institutions. Just as more responses were beginning to come in, the

COVID-19 pandemic struck which slowed down not only the planned training of data collectors that were intended to supplement the telephone follow-ups.

In addition to the telephone follow-up by INSTAT, selected DOST regional staff were also requested to help in following up and collecting the completed questionnaires from sampled institutions in their respective area. A training program was conducted to orient the data collectors about the R&D Survey, their responsibilities, and the role of INSTAT as technical support to facilitate efficient collaboration. In the training, quality practices in collecting, validating, and encoding responses in the R&D survey were discussed. Information on the follow-ups that INSTAT had already made with the institutions were forwarded to the assigned data collectors so as not to disrupt the flow of communication. Illustrations of common errors and inconsistencies in the accomplished questionnaires were also given to guide them in validating the collected responses. Because of the implementation of the community quarantine restrictions in almost all the regions beginning 16 March 2020, in-person training was conducted only for CALABARZON and the Zamboanga Peninsula, while virtual training was organized for the other regions.

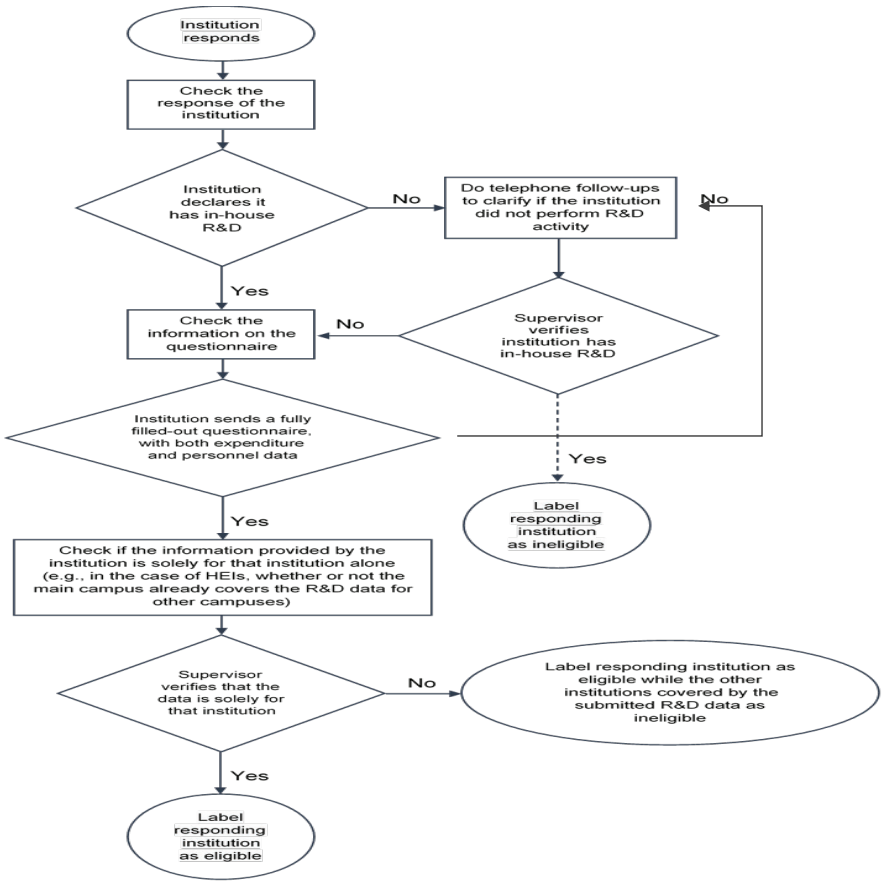

Figure 2. Process flowchart for distinguishing and validating eligibility status of responding institution.

An INSTAT survey supervisor was designated for each region to oversee the data collectors. These survey supervisors, in close collaboration with the lead data collectors for each region, served as the focal persons in providing updates to the INSTAT survey team. Regular team meetings were held to discuss and address critical issues that were encountered in the field.

Regular updates on the status of the survey were communicated to the data collectors to ensure that all responses were accounted for in the survey. The field follow-ups closed in November 2020 but a grace period on the submission of accomplished questionnaires was set until January 2021. By then, higher than targeted response rates were achieved per region and at the national level. A national-level response rate of 81% (computed as the non-weighted percentages of institutions that responded either as eligible or ineligible over the sample size) was achieved at the end of the data collection (Figure 3).

As discussed, adjustments in survey mode were adapted in the duration of the survey operation as needed. These resulted in improved response rates across the months of the implementation of Phase 2 of the survey. For instance, a very low constant series of response rates were observed from September 2019 until February 2020. An increase in the response rate for March 2020 can be attributed to the sending of questionnaires to the identified institutions through the post. For April and May, a dismal increase in the response rate is due to the adjustments of institutions due to the COVID-19 pandemic. Though most of the regional staff of DOST were already engaged as data collectors in March 2020, the effect of the various community lockdowns implemented in the country can be seen on the response rates. The response rate started to consistently increase beginning June 2020, when all the data collectors have been trained and deployed.

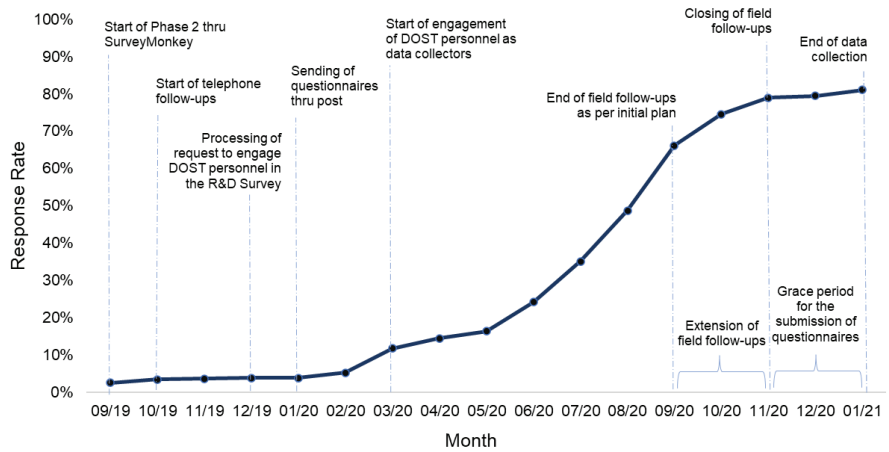

Figure 3. Response rates of Phase 2 of the 2018 R&D Survey.

2.4 Data Preparation and Observation Register Creation

To ensure that data quality is maintained, appropriate data processing and validation procedures were implemented. A coding system was developed to ensure that each institution in the sampling frame is uniquely identified, and all meta-data are properly recorded. The monitoring system that was described above was also automated so that weekly reports can be generated.

Data from the questionnaires were entered through SurveyMonkey. Scanned copies of the questionnaires were also systematically stored so that suspect data can be easily verified. Guidelines for encoding completed questionnaires were given to the data collectors to minimize coding errors. Periodic data validation was conducted with each round of data processing and preliminary data analysis, gaps between the identified eligible institutions in the monitoring form and the encoded responses in the database were flagged for the appropriate action of the designated supervisors. Institutions found to have submitted their accomplished questionnaires but were still not encoded in the database were identified. Responses recorded in the database with suspicious eligibility status in the monitoring form were clarified. Multiple responses of institutions in the database were individually cross-checked with the completed questionnaires to identify the correct entry. When all collected responses were accounted for in the database and data cleaning was completed, responses of reporting administrative units of institutions with decentralized management system were consolidated into institution-level data. Figure 4 shows the total number of eligible and ineligible institutions and nonresponding institutions.

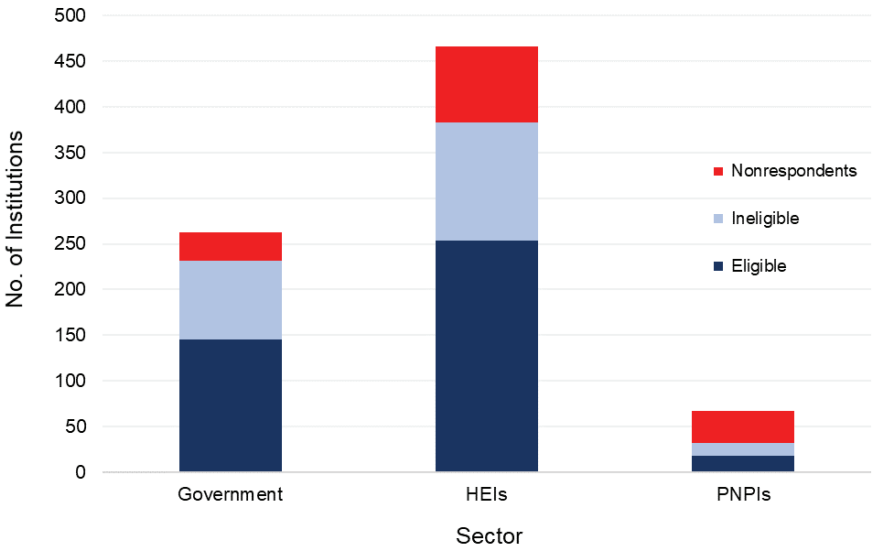

Figure 4. Distribution of the eligibility status of sampled institutions across sectors.

As shown in Figure 4, more than half of the PNPIs did not respond to the R&D survey, while less than 20% of the institutions in the government and HEIs did not respond. In terms of eligibility status, about a third of the responding institutions were deemed ineligible. These statistics were used in the adjustments for nonresponse and coverage errors.

Looking closely at the data, the management of research activities and financial reckoning systems differed widely across sampled institutions. In general, large institutions with very decentralized research management systems are likely not able to complete the R&D Survey.

At the end of the data validation process, institution-level responses were produced. Finally, design variables, like region, stratum, stratum sizes for finite population correction, and survey weights, were incorporated into the data to perform the analysis.

2.5 Estimation and Analysis

Final survey weights were determined as the product of the base weights that were computed as planned and adjustments to compensate for nonresponse and coverage errors. The R&D indicators that were computed were mostly totals and proportions, like total R&D expenditure and percent share of government R&D expenditure. To provide a measure of the precision of the estimates, the corresponding standard errors were also computed using Taylor Series Linearization (TSL), which is a variance estimation method that renders robust results even for nonlinear estimators, like subpopulations means and totals. TSL is the default variance estimation procedure in many reputable statistics software like R, Stata, and SAS. All computations in this survey report were executed using the 'survey' package in R. The complete methodological notes on the weighting adjustments and estimation implemented in the survey can be found in the 2018 R&D Survey Report that was disseminated to stakeholders (Maligalig et al., 2021).

Statistical tables were generated for all the R&D indicators at the national level, by sector, and by region. Indicators for the HEIs were also disaggregated by public and private HEIs. The corresponding standard errors of all estimates were also incorporated in the statistical tables in the survey report. Graphs were generated for the most important survey results. The estimates were also compared with those from previous R&D survey rounds. A thorough review of the estimates was undertaken before writing the survey report.

2.6 Presentation and Dissemination

The survey report was submitted to the National Research Council of the Philippines (NRCP), the monitoring agency for this research project, and DOST for review. All comments were discussed, and the survey report was finalized accordingly. The results of the survey were presented in two webinars – for the DOST staff and management and another for the respondents, UPLB constituents, and the data collectors. The survey report can also be downloaded from the INSTAT website (instat.uplb.edu.ph). Printed copies of the survey report were also sent to the responding institutions and data collectors.

A complete set of data files were turned over to NRCP and DOST. The set included an anonymized survey data file that can be turned into a public utility file that can be shared with researchers, the sampling frame, and the data dictionary.

3. Summary and Conclusions

The most challenging task in the survey operations is the construction of the sampling frame because there is no central database from which data on all the institutions that undertake R&D can be extracted. Data and information from many sources were combined to construct the sampling frame. Because there is no guarantee that all the institutions in the sampling frame are eligible, a monitoring system that could distinguish eligible and ineligible respondents was incorporated in the workflow so that appropriate weighting adjustments to compensate for coverage errors can be derived. When the response rate was low at the beginning months of the survey, a more flexible combination of survey modes was adapted. These innovations led to an effective survey operation for the R&D Survey and data of good quality was obtained.

All the statistical tables in the survey report contained the corresponding measure of the precision of the estimates being presented so that users can do their evaluation. Bias due to non-response and coverage errors were mitigated through appropriate weighting adjustments. Measurement errors were controlled by applying proven good practices. The changes

implemented in the questionnaire also helped in controlling for the non-sampling errors in the survey.

To ensure the accessibility and clarity of the information in the R&D survey, dissemination of the results through two webinars, distribution of survey report, and provision of a downloadable version of the survey report were done. Moreover, anonymized survey data that can be turned into a public utility file was turned over to DOST for possible sharing with researchers.

In terms of timeliness, there had been delays in the proposed survey activities. Though the implementation of the mixed survey modes was successful, still it has room for improvement. For instance, instead of starting the data collection approach using the online mode, it may be better to send by post the sampled institutions the questionnaire, invitation to participate and endorsement letters since most institutions have yet to consider emails and online surveys as official communication.

For the coherence and comparability dimensions, data across time and space were achieved by employing the same set of concepts and definitions. Streamlining the questionnaires based on the Frascati manual and based on the review of various R&D questionnaires ensured the coherence and comparability of the R&D survey.

4. Recommendations

The R&D questionnaire can be expanded to enable aggregates by sex and personnel characteristics such as age group, highest educational attainment, and the field of specialization. Moreover, the number of graduate students, especially the number of Ph.D. candidates can be collected to indicate the potential sources of future researchers. The size of the graduate class can also be included in the questionnaire to help sharpen the estimation strategy. The inclusions of questions on other forms of dissemination of completed research and outputs must also be explored to enable more in-depth analysis.

The construction of sampling frames for government, HEIs, and PNPIs were one of the most challenging activities that were undertaken to ensure that the 2018 R&D Survey is a probability sample survey that can render robust estimates. To reduce this difficult burden for the next survey rounds, a centralized R&D institution database that stores information on the institutions with in-house research must be maintained. Updating of this database can be done at the point of entry – when DOST or other government agencies or PNPIs provide a grant to a research project. In this regard, the collaboration with other research-granting institutions such as CHED, Department of Health (DOH), Department of Education (DepEd), Department of Agriculture (DA), Department of Environment and Natural Resources (DENR) should be strengthened. Following the endorsement of DOST, CHED, and DOH that were given to the 2018 R&D Survey, other government agencies can also encourage their respective research units and grantees to cooperate and participate in the next R&D Survey rounds.

While the 2018 R&D Survey paved for some methodological innovations in terms of data collection and analysis, there are still areas in the survey design and operations that need further improvement so that better quality data can be achieved. Of the three sectors covered by the R&D Survey, PNPIs have the lowest response rate. Despite the low response rate, both the total number of R&D personnel and expenditures have considerably increased compared to 2015 indicating that coverage errors may have been reduced. There were 12 regions that do not have PNPIs. While there could really be no PNPIs in all these regions that undertake in-house research, it is still worth exploring possible approaches that could improve the sampling frame for PNPIs. PNPIs that have R&D activities must be engaged by DOST through regular consultations so that they can also be included in the database. There must also be an

information campaign to let institutions know the importance of R&D in our economy and to encourage them to participate in the R&D Survey

The mixed survey mode that was implemented turned out to be successful. Because of this approach, and with a streamlined questionnaire and the cooperation of the DOST regional offices staff that played a critical role in implementing the mixed-mode approach, the data collection for the 2018 R&D Survey was completed at about 81% response rate. Instead of starting the data collection approach using the online mode, it may be better to send by post the sampled institutions the questionnaire, invitation to participate and endorsement letters since most institutions have yet to consider emails and online surveys as official communication. The link to the online survey can be given in the letter to offer the sampled institutions an alternative way of responding to the questionnaire.

The DOST regional offices, as well as regional development councils, must also be engaged in updating the R&D institution database. Given the help of the DOST regional staff in improving the response rates of the 2018 R&D Survey, they must be engaged at the onset of the survey to help update the database mentioned above as well as undertake field operations for the R&D Survey.

DOST can develop application software that can be used for managing research activities in government, HEIs, and PNPIs. The software should enable institutions to store and manage all the information regarding their research activities. The software can also generate reports that can be used by the institutions to manage their research agenda as well as, for completing the R&D Survey. Large institutions with decentralized research management systems can use this application software to consolidate the R&D information from different reporting administrative units and consequently, manage their research agenda more effectively. The creation of the application software will further improve the response rate and reduce measurement errors of future R&D Survey rounds.

Literature Cited

AKCALI, B.Y., and SISMANOGLU, E. 2015. Innovation and the Effect of Research and Development (R&D) Expenditure on Growth in Some Developing and Developed Countries. Procedia – Social and Behavioral Sciences, Vol. 195, pp. 768-775.

ARORA, S.K, KELLEY, S., and MADHAVAN, S. 2021. Building a Sample Frame of SMEs Using Patent, Search Engine, and Website Data. Journal of Official Statistics, Vol. 37, No. 1, 2021, pp. 1–30, http://dx.doi.org/10.2478/JOS-2021-0001

ASTROLOGO, C.J., BAÑARES, J.S., GARRAEZ, J.A.H., ROBLES, M.P., and BANTANG, J.A.O. 2019. Evaluating the Quality of Statistics Produced by the Philippine Statistics Authority Using the UN NQAF Assessment Tool. Paper presented at the 14th National Convention on Statistics, Manila, Philippines. October 1-3, 2019.

BAYARCELIK, E.B., and TASEL, F. 2012. Research and Development: Source of Economic Growth. Procedia - Social and Behavioral Sciences 58, pp. 744 – 753.

BLANCO, L.R., GU, T., and PRIEGER, J.E. 2016. The Impact of Research and Development on Economic Growth and Productivity in the U.S. States. Southern Economic Journal, Vol. 82, No. 3, pp. 914-934. Southern Economic Association.

DE LEEUW, E.D., HOX, J.J., and DILLMAN, D.A. 2008. In E.D. De Leeuw, J.J. Hox, and D.A. Dillman (Ed). International Handbook of Survey Methodology, pp. 1-17.

ISTITUTO NAZIONALE DI STATISTICA. 2016. The Behaviour of Respondents While Filling in a Web Questionnaire: The Case of the Italian Business R&D Survey. M. Masselli, A. Nuccitelli, and A.L. Palma. Istat Working Papers.

KHAN, J. 2015. The Role of Research and Development in Economic Growth: A Review. Journal of Economics Bibliography, Vol. 2, Issue 3, pp. 128-133.

MALIGALIG, D.S., BORROMEO, M.C., CAMBEL, R. G., ROLDAN, JR., R. R. and LESMORAS, C. B. 2019. Challenges in Designing and Implementing Research and Development Surveys in the Philippines. Paper presented at the 14th National Convention on Statistics, Manila, Philippines. October 1-3, 2019.

MALIGALIG, D.S., BORROMEO, M.C., CAMBEL, R. G., ROLDAN, JR., R. R. and LESMORAS, C. B. 2021. 2018 R&D Survey Report. Department of Science and Technology, and Institute of Statistics.

NATIONAL RESEARCH COUNCIL. 2015. Measuring Research and Development Expenditures in the U.S. Nonprofit Sector: Conceptual and Design Issues, Summary of a Workshop. C. House, H. Rhodes, and E. Sinha, Rapporteurs. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

NATIONAL SCIENCE FOUNDATION. Survey of Industrial Research and Development. Retrieved from https://www.nsf.gov/statistics/srvyberd/prior-descriptions/overview-sird.cfm

OECD. 2015. Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264239012-en

SOKOLOV-MLADENOVIĆ, S., CVETANOVIĆ, S., and MLADENOVIĆ, I. 2016. R&D expenditure and economic growth: EU28 evidence for the period 2002–2012, Economic Research-Ekonomska Istraživanja, 29:1, 1005-1020, DOI: https://doi.org/10.1080/1331677X.2016.1211948

SUNDGREN, B. 1999. Information Systems Architecture for National and International Statistics Offices - Guidelines and Recommendations, United Nations Statistical Commission and Economic Commission for Europe, Geneva.

THE AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH. 2015. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 8th edition. AAPOR.

UNESCO INSTITUTE FOR STATISTICS. 2014. Guide to Conducting an R&D Survey: For countries starting to measure research and experimental development. Downloaded from: http://dx.doi.org/10.15220/978-92-9189-151-1-en

UNIVERSITY OF READING. 2009. International Household Survey Network Survey Quality Assessment Framework SQAF.