Nonsampling errors and their implication for estimates of current cancer treatment using the Medical Expenditure Panel Survey

Survey Methods: Insights from the Field

Jeffrey M. Gonzalez | Lisa B. Mirel | Nina Verevkina

How to cite this article : Gonzalez J. M., Mirel L. B. & Verevkina N. (2016). Nonsampling errors and their implication for estimates of current cancer treatment using the Medical Expenditure Panel Survey. Survey Methods: Insights from the Field. Retrieved from https://surveyinsights.org/?p=7860

DOI : 10.13094/SMIF-2016-00007

Copyright : © the authors 2016. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

Abstract : Survey nonsampling errors refer to the components of total survey error (TSE) that result from failures in data collection and processing procedures. Evaluating nonsampling errors can lead to a better understanding of their sources, which in turn, can inform survey inference and assist in the design of future surveys. Data collected via supplemental questionnaires can provide a means for evaluating nonsampling errors because it may provide additional information on survey nonrespondents and/or measurements of the same concept over repeated trials on the same sampling unit. We used a supplemental questionnaire administered to cancer survivors to explore potential nonsampling errors, focusing primarily on nonresponse and measurement/specification errors. We discuss the implications of our findings in the context of the TSE paradigm and identify areas for future research.

1. Introduction

A simple survey question such as “Are you currently being treated for cancer?” can be subject to different interpretations. For persons who have never been diagnosed with cancer or a malignancy of any kind, the response process should be straightforward. Cancer survivors, defined as a person still living after their initial cancer diagnosis, however, may experience difficulty in responding regardless of their “true” current treatment status because of the cognitive processes employed. Flaws in these processes, such as misinterpreting key words or misjudging items or events to include in the response, may contribute to incorrect responses and result in nonsampling error.

Regardless of steps taken by survey designers to mitigate nonsampling errors, they can still arise when there are weaknesses in the data collection and processing procedures [1], [2], [3]. Specification, measurement, nonresponse, and processing errors are all types of nonsampling errors [4]; not controlling or compensating for them can lead to biased inferences.

Survey error evaluation helps survey practitioners understand how best to control and compensate for these errors [4]. Analyzing data collected via supplemental questionnaires or follow-up surveys is one method survey researchers can utilize to evaluate survey errors [5]. Supplemental questionnaires can

provide additional information on survey nonrespondents or can yield repeated measurements on the same unit for use in nonresponse or measurement error analyses, respectively.

One opportunity for using a supplemental questionnaire to assess nonsampling errors arose when the 2011 Medical Expenditure Panel Survey-Household Component (MEPS) fielded a follow-up survey to cancer survivors. Each survey utilized a different method, later described, to elicit information on whether participants were currently receiving treatment for cancer; therefore, we were able to explore potential measurement/specification errors by comparing these two surveys. We also assessed nonresponse error in the supplemental questionnaire by analyzing differences between respondents and nonrespondents' data provided in MEPS.

2. Data sources

The Medical Expenditure Panel Survey is the most complete data source on the cost and use of health care and insurance for the civilian noninstitutionalized US population [6]. It is a multi-purpose survey that can be used by public health researchers to characterize health care costs for and utilization of health care services by certain health conditions, such as cancer [7]. It has three main components – the household, medical provider, and insurance components – in this paper we focus on the household component, henceforth referred to as MEPS.

Recognizing the need for a comprehensive data source on health care costs and utilization and other public health issues relevant to cancer survivors and their families, in 2011, the National Cancer Institute collaborated with the Agency for Healthcare Research and Quality, the Centers for Disease Control and Prevention, LIVESTRONG, and the American Cancer Society to develop a supplemental paper-and-pencil, self-administered questionnaire for cancer survivors identified in MEPS. This self-administered questionnaire is formally referred to as the Experiences with Cancer Survivorship Supplement [8], henceforth referred to as the CSAQ. The purpose of the CSAQ was to collect information from cancer survivors on their experiences with cancer, including whether treatment is currently being sought, long-lasting effects of the disease, and financial impacts [8]. Together, MEPS and the CSAQ provide the only information system that allows public health researchers to assess both direct medical costs and opportunity costs for cancer-related events and to characterize associations between costs and experiences with cancer for a representative sample of cancer survivors from the US civilian noninstitutionalized population [9].

2.1. MEPS

MEPS is a complex, multi-stage, nationally representative sample of the US civilian noninstitutionalized population. It has been an annual survey since 1996, supporting national estimates of, including but not limited to, health care use, expenditures, and disease prevalence. Each year a new sample is drawn as a subsample of households that participated in the prior year's National Health Interview Survey, a survey conducted by the National Center for Health Statistics. It is a household-level sample meaning that data are collected for all target population members in the household.

MEPS employs an overlapping panel design [10] in which data are collected through a series of five personal visit interviews via CAPI. During each calendar year data are collected simultaneously for two MEPS panels. One panel is in its first year of data collection (e.g., in 2011, Rounds 1, 2, and 3 of Panel 16), while the prior year's panel is in its second year of data collection (e.g., in 2011, Rounds 3, 4, and 5 of Panel 15). In our analysis, we focused on Round 5 of Panel 15 and Round 3 of Panel 16 because that is

when the CSAQ was administered to eligible MEPS participants.

MEPS is typically completed by one household member who responds for themselves and all other eligible members of the family. The respondent is asked to report all health care events and descriptions of those events for themselves and family members during the reference period. By default, for multi-person families the MEPS respondent serves as a proxy respondent for everyone else in the family.

Events are reported by type of health care service (e.g., office based physician visit, outpatient hospital visit, and prescription medicine acquisition) and as part of event reporting, the respondent is asked what condition(s) was (were) associated with the event. These conditions are then coded into the Clinical Classifications Software (CCS) categories, based on ICD-9-CM codes, making cancer-related health care service events easily identifiable in MEPS databases. [11].

2.2. CSAQ

The CSAQ was fielded during the Panel 15, Round 5 and Panel 16, Round 3 MEPS interviews and was intended to be administered upon completion of those two interviews. The CSAQ went through two bouts of cognitive testing to assess respondents' answers to and understanding of survey questions [8]. Members, age 18+, of households responding to MEPS were identified as being eligible to participate in the CSAQ via an affirmative response to the following question in MEPS: "(Have/Has) (PERSON) ever been told by a doctor or other health professional that (PERSON) had cancer or a malignancy of any kind?" These participants were then classified as cancer survivors, using the definition noted above (e.g., a person still living after their first diagnosis of cancer).

3. Methods

3.1. Identifying those currently being treated for cancer

Identifying CSAQ respondents who were currently being treated for cancer was straight-forward as it was based on an explicit question in the CSAQ. We classified respondents as currently being treated for cancer via an affirmative response to the question: "Are you currently being treated for cancer-that is are you planning or recovering from cancer surgery, or receiving chemotherapy, radiation therapy, or hormonal therapy for your cancer?" All other CSAQ respondents were classified as not currently being treated.

Since no explicit question exists in MEPS that captures current treatment status for cancer, we created an indicator of this concept from other MEPS data. We used utilization reports of office-based medical provider visits, outpatient visits, and prescription medicines collected in MEPS to create the indicator. These reports contained information that corresponded to the procedures and treatments explicitly mentioned in the CSAQ current cancer treatment question text (e.g., "surgery, or receiving chemotherapy, radiation therapy, or hormonal therapy").

These reports also contain detailed information on the type of MEPS reported events (e.g., a visit to an office-based physician) for 2011, including date of medical visit/event or purchase/acquisition, types of treatments and services received, specialty of the medical practitioner, condition codes associated with the event/prescription, expenditures, sources of payment and coverage, and for the prescription medicine file, the National Drug Code for the prescribed medicine obtained.

We identified events from the medical provider and outpatient reports that occurred during the reference

period of the Panel 15, Round 5 or Panel 16, Round 3 interviews for which the household member identified the event as being associated with cancer (using the CCS codes) and either saw a medical practitioner with any of the following specialties – oncology, radiology, or surgery – or received chemotherapy or radiation therapy at the visit. These events were flagged as current cancer treatment events.

Next, we identified individuals with prescription medicine reports who acquired a prescription medicine associated with a cancer condition (using the CCS codes) or were treated with a prescription medicine typically used to treat cancer during the appropriate reference period. Cancer prescription drugs were identified by the set of National Drug Codes contained in the National Cancer Institute's drug tables [12].

We then created a person-level dichotomous indicator for current cancer treatment status from MEPS based on the individual having at least one event meeting any of the above criteria.

3.2. Analysis

Using only information collected in MEPS, we compared the socio-demographic characteristics and mean annual health care expenditures of MEPS participants, age 18+, who were eligible to participate in and completed the CSAQ (n = 1,419) to (1) the 2011 MEPS participant sample, age 18+ (n = 24,236); (2) all CSAQ-eligible persons (n = 1,764); and, (3) all CSAQ-eligible, but nonrespondents to the CSAQ (n = 345).

For MEPS participants who were eligible to participate in the CSAQ, we compared MEPS reports of cancer-related conditions among CSAQ respondents and nonrespondents. We note that comparing CSAQ respondents to CSAQ nonrespondents may provide insight into the potential nonresponse errors in the CSAQ.

Finally, for CSAQ participants, we assessed the agreement between their responses to the explicit current cancer treatment status question in the CSAQ to the corresponding MEPS indicator. Agreement was assessed using the Kappa statistic, a statistic that corrects the percentage of agreement between raters, by taking into account the proportion of agreement expected by chance [13]. Because this statistic accounts for chance agreement, we believe this is a more robust measure agreement than simply percent agreement. We note that comparing the agreement between the two current treatment status measures provides insight into potential measurement errors.

We also evaluated domain-specific rates of agreement in which domains were defined by each of respondent proxy status of MEPS, completion timeliness of the CSAQ, cancer type reported in MEPS. We considered self/proxy-respondent status as a domain because we hypothesized that there should be higher agreement between the two data sources when the CSAQ and MEPS respondents were identical. The rationale for including completion timeliness of the CSAQ was two-fold. First, CSAQ respondents who responded for themselves in MEPS may have been primed by event reporting, thus they may have been thinking about the cancer-related health care events when responding to the CSAQ. Second, the lag time between completing the CSAQ after MEPS may affect the cognitive processes involved in survey response. If the CSAQ was completed immediately after the MEPS, then any ambiguity regarding what was implied by the word "currently" (in the CSAQ question text) may have been mitigated for those persons. Finally, since skin cancer is the most common type of cancer diagnosed in the US, with non-melanoma skin cancers being the most common type, and a large number seek treatment [14], we also wanted to investigate whether there were differential rates in agreement in current treatment status between non-melanoma skin and other cancers.

All analyses were weighted to account for the complex sample design and to adjust for nonresponse to the MEPS [10], [15]. All statistical analyses were performed using either SAS survey procedures (Version 9.3 SAS Institute Inc., Cary, N.C.) or R (Version i386 3.0.1, The Comprehensive R Archive Network). R was used to calculate the standard error for the Kappa statistics following the method outlined in Lin et al. [16]. Variances for estimated means, proportions, and Kappas were estimated, using Taylor series linearization [17]. t-tests were used to assess whether there were significant differences between the CSAQ respondents and nonrespondents on continuous variables and between domain-level Kappa statistics while Rao-Scott design-adjusted Pearson $\chi 2$ tests were used for categorical variables.

4. Results

4.1. Indications of nonresponse error

Table 1 displays the socio-demographic characteristics of the four groups. We note that estimates from MEPS 18+ participants and CSAQ-eligibles are presented solely for reference and completeness as no statistical tests were conducted to compare these groups.

Table 1. Descriptive statistics from the 2011 MEPS

	MEPS, 18+ participants (n=24,236)			CSAQ eligible (n=1,764)			CSAQ respondents (n=1,419)			CSAQ nonrespondents (n=345)		
	n	%	SE	n	%	SE	N	%	SE	n	%	SE
Age												
18-44*	12,100	46.9	0.6	195	9.4	0.8	122	7.9	0.8	73	15.4	1.7
45-64	8,080	34.8	0.5	658	38.0	1.4	544	38.7	1.5	114	36.3	2.8
65+	4,056	18.4	0.5	911	52.6	1.6	753	53.3	1.8	158	48.3	2.7
Gender												
Male	11,223	48.2	0.3	730	44.6	1.4	571	44.2	1.4	159	49.7	2.8
Female	13,013	51.8	0.3	1,034	55.4	1.4	848	55.8	1.4	186	50.3	2.8
Race/ethnicity												
Hispanic *	6,196	14.8	0.8	154	5.0	0.6	105	4.4	0.6	49	8.4	0.9
Black, Non-Hispanic	4,743	11.4	0.7	211	5.4	0.5	154	5.4	0.6	57	7.7	1.0
Asian, Non-Hispanic	1,793	5.1	0.5	38	1.1	0.3	25	1.0	0.3	13	2.0	0.6
Other Race, Non-Hispanic *	11,504	68.7	1.0	1,361	88.5	0.8	1,135	89.1	0.9	226	81.9	1.3
Health Insurance Status												
Any Private*	13,775	67.5	0.7	1,081	68.3	1.4	904	70.6	1.4	177	58.4	3.4
Public Only*	5,406	17.8	0.5	578	27.3	1.4	445	25.7	1.3	133	35.0	3.7
Uninsured*	5,055	14.7	0.4	105	4.3	0.6	70	3.7	0.6	35	6.7	1.1
Perceived Health Status												
Excellent/ very good*	13,464	58.5	0.6	723	44.4	1.5	603	46.4	1.7	120	34.8	2.9
Good	7,083	27.6	0.5	578	32.1	1.2	465	31.9	1.3	113	32.6	2.3
Fair/ Poor*	3,372	12.4	0.3	457	22.7	1.3	351	21.6	1.3	106	28.7	2.8
Unknown	317	1.4	0.1	6	0.7	0.4	-	-	-	6	3.9	1.8
MEPS-based current treatment status												
Yes	460	2.4	0.2	330	19.2	1.3	280	19.9	1.3	50	15.9	2.3
No	23,776	97.6	0.2	1,434	80.8	1.3	1,139	80.1	1.3	295	84.1	2.3

*p-value<0.05 for CSAQ respondents vs. nonrespondents

CSAQ respondents and nonrespondents differ significantly with respect to age, race/ethnicity, health insurance status, and perceived health status (Table 1). CSAQ nonrespondents tended to be younger (15.4% vs. 7.9% in the 18-44 age category, $p < 0.01$), more likely to be Hispanic (8.4% vs. 4.4%, $p < 0.01$), less likely to be covered by any source of private insurance (58.4% vs. 70.6%, $p < 0.01$), and had lower perceptions of their health status (28.7% vs. 21.6% in the fair/poor category, $p < 0.05$) than CSAQ respondents. Though not presented, we did observe that CSAQ nonrespondents had lower mean annual outpatient health care expenditures than CSAQ respondents ($843 vs. $1,319, $p < 0.05$), but there were no other statistically significant differences in expenditures, even by treatment status.

We summarize the distributions for cancer type, years from first cancer diagnosis, years from last treatment, and time of CSAQ completion separately for CSAQ respondents and nonrespondents (Table 2).

Table 2. Characteristics of CSAQ-eligible persons

	CSAQ respondents (n=1,419)			CSAQ nonrespondents (n=345)		
	n	%	SE	n	%	SE
Cancer type from MEPS						
Breast	235	14.5	0.9	48	13.4	2.2
Prostate	159	11.6	1.1	41	9.9	1.9
Colorectal	59	3.9	0.5	17	3.4	0.7
Non-melanoma skin	216	17.9	1.4	42	14.6	2
Other *	750	52.1	1.5	197	58.6	2.5
Years from first cancer diagnosis from MEPS						
<5	455	32.3	1.5	126	35.9	2.7
>=5 *	932	65.6	1.5	182	54.2	3
Missing *	32	2.1	0.6	37	9.9	1.3
Current treatment status from MEPS						
Yes	280	20	1.4	50	15.9	2.3
No	1,139	80	1.4	295	84.1	2.3
MEPS respondent status						
Proxy *	410	29.5	1.2	148	47.6	2.5
Self *	1,009	70.5	1.2	197	52.4	2.5
Years from last treatment from CSAQ						
Currently Treated	196	14.2	1.1	-	-	-
<1	140	10.5	1.1	-	-	-
1-<5	336	24	1.5	-	-	-
>=5	572	39.4	1.5	-	-	-
Have not been treated/ Unknown	175	11.8	1	-	-	-
CSAQ completion timeliness						
At MEPS interview	880	61.2	1.8	-	-	-
Other	539	38.8	1.8	-	-	-

*p-value<0.05 for Cancer SAQ respondents vs. nonrespondents

The percentage distribution of the three main types of cancer – breast, prostate, and colorectal – did not differ between respondents and nonrespondents; however, there was a higher percentage of nonrespondents having a MEPS report of some other type of cancer (58.6% vs. 52.1%, $p < 0.05$). There was a higher percentage of respondents who had received their first cancer diagnosis more than five years ago compared to nonrespondents (65.6% vs. 54.2%, $p < 0.01$). There was a greater proportion of item nonresponse for years since first cancer diagnosis among CSAQ nonrespondents (9.9% vs. 2.1%, p <

0.05). There was a higher percentage of MEPS proxy respondents among CSAQ nonrespondents when compared to respondents (47.6% vs. 29.5%, $p < 0.05$). Finally, we note that there was not a statistically significant difference in current treatment status, based on the MEPS constructed indicator, between CSAQ respondents and nonrespondents.

4.2. Indications of measurement error

In Table 3 we summarize the overall concordance between CSAQ respondents' responses to the explicit current treatment status question and the corresponding indicator constructed from MEPS. We observe that 196 CSAQ respondents (or 14.2%) answered affirmatively to the CSAQ question.

Table 3. Overall concordance between MEPS and the CSAQ

	Current treatment (CSAQ)		No current treatment (CSAQ)		Total	
	n	%	n	%	N	%
Current treatment (MEPS)	112	8.06	168	11.84	280	19.9
No current treatment (MEPS)	84	6.15	1,055	73.96	1,139	80.11
Total	196	14.21	1,223	85.8	1,419	100
Kappa	0.37					
(SE)	0.04					

Of the 1,419 CSAQ respondents, 280 (or 19.9%) of them had evidence of being treated for cancer based MEPS data. The overall concordance between the MEPS and CSAQ was 82.0%. This is largely due to the high percentage of no evidence of treatment from either survey (74.0% of CSAQ respondents had no evidence of treatment from both surveys). The Kappa statistic was 0.37 which represents fair agreement [18].

Concordance between MEPS and the CSAQ for the domains defined by each of whether the CSAQ respondent was also the MEPS respondent, time of completion of the CSAQ and, type of cancer is presented in Tables 4-6.

Table 4. Concordance between MEPS and the CSAQ by proxy respondent status

	MEPS Respondent: Self				MEPS Respondent: Proxy			
	Current treatment (CSAQ)		No current treatment (CSAQ)		Current treatment (CSAQ)		No current treatment (CSAQ)	
	n	%	n	%	n	%	n	%
Current treatment (MEPS)	78	7. 9	125	12.5	34	8.4	43	10.4
No current treatment (MEPS)	53	5.1	753	74.5	31	8.5	302	73.7
Total	131	13.0	878	87.0	65	16.9	345	83.1
Kappa	0.41				0.44			
(SE)	0.04				0.07			

Table 5. Concordance between MEPS and the CSAQ by completion timeliness

	CSAQ completed at MEPS interview				CSAQ completed at a later date			
	Current treatment (CSAQ)		No current treatment (CSAQ)		Current treatment (CSAQ)		No current treatment (CSAQ)	
	n	%	n	%	n	%	n	%
Current treatment (MEPS)	79	9.3	114	13.3	33	6.2	54	9.5
No current treatment (MEPS)	48	5.5	639	72.0	36	7.2	416	77.1
Total	127	14.8	753	85.3	69	13.4	470	86.7
Kappa	0.44				0.39			
(SE)	0.05				0.06			

Table 6. Concordance between MEPS and the CSAQ by type of cancer

	Other cancer				Non-melanoma skin cancer			
	Current treatment (CSAQ)		No current treatment (CSAQ)		Current treatment (CSAQ)		No current treatment (CSAQ)	
	n	%	n	%	n	%	n	%
Current treatment (MEPS)	105	8.6	157	13.1	7	5.5	11	6.2
No current treatment (MEPS)	73	6.7	868	71.6	11	3.5	187	84.8
Total	178	15.3	1,025	84.7	18	9.0	198	91.0
Kappa	0.37				0.52			
(SE)	0.04				0.12			

Across each level of each domains, the concordance between the measures of current cancer treatment status was at least 80%. As with the overall concordance, this is likely driven by the high percentage of no evidence of treatment from either data source. In addition, we do not observe any statistically significant differences in the rates of agreement for any domain, but all level-specific Kappa statistics for each domain represented fair to moderate agreement [18].

5. Discussion and conclusions

We presented an analysis of a supplemental questionnaire administered to cancer survivors to explore potential measurement and nonresponse errors. Overall, we found a fair level of agreement between the two sources for current treatment status, but no statistically significant differences among any of the domains considered. There were, however, differences between respondents and nonrespondents.

A priori, we did not make any judgments about whether MEPS or the CSAQ would provide the better method for measuring current treatment status. Simple questions, with few probes or examples, might be easy and less burdensome to answer [19]; however, they may not align well with the underlying construct (i.e., there may be a specification error). Alternatively, questions requiring complicated response tasks, e.g., enumerating events that occurred during some time period, might be difficult and burdensome to answer [20], [21], but these question might align well with the desired construct. As such, survey practitioners must balance parsimony with burden to mitigate measurement and specification errors.

Since we found more evidence of currently receiving treatment among CSAQ respondents using the MEPS indicator (19.9% in MEPS vs. 14.2% in the CSAQ), then perhaps constructing this concept from a survey whose response tasks involve enumerating specific treatment events is preferred. Specifically, if we operate under the “more is better” premise [22], as is often the case in surveys that require event and expenditure reporting, then MEPS might be “fit enough” for identifying cancer survivors currently

receiving treatment and less susceptible to specification and measurement errors; acknowledging that, the CSAQ is still necessary for assessing other concepts like the relationship among health care costs and experiences with cancer since the latter are not collected in MEPS.

We note that since we constructed current treatment status from only three types of MEPS utilization reports and did not incorporate other types of health care service reports, we may have induced a specification error in the constructed indicator. For example, a breast cancer survivor responding to MEPS might have received a mastectomy but, the details of this surgical procedure, if part of a hospital stay, would have been included in an inpatient hospital stay report. At first glance, it may seem that our method would not classify this cancer survivor as currently receiving treatment for cancer, but it is likely that prior to receiving the mastectomy, the cancer survivor also consulted an oncologist or surgeon in an office-based physician setting. The physician visit would likely have been identified in our method of identifying individuals currently receiving treatment. Thus, provided that both events were reported and occurred in the reference period, the MEPS participant would have been correctly classified in the MEPS indicator.

In light of this example, one area of future research might include broadening the criteria used to classify a MEPS participant as currently receiving treatment. It would, however, also increase the discordance between the two data sources (because the set of CSAQ respondents identified as receiving treatment is fixed). Researchers may consider conducting sensitivity analyses with varying inclusion criteria.

As noted, the MEPS indicator was based on respondent reports of health care service events. There are at least two additional sources of nonsampling errors in MEPS that could affect the identification of those currently receiving treatment for cancer. These are underreporting and coding error. Failing to report events can be due memory problems like retrieval failures and encoding differences [3] and survey fatigue or burden [20]. Survey burden can also lead to satisficing behaviors and is particularly relevant to this research [20]. By the time cancer survivors are identified in MEPS and asked to participate in the CSAQ, they have already potentially participated in the National Health Interview Survey, at least two round of MEPS, and other MEPS supplemental questionnaires. As a consequence, these respondents have had several opportunities to learn poor responding behaviors, e.g., learning to say “no” so they are not asked a subsequent battery of questions [23], which could lead to underreports of cancer-related events within in an interview. Furthermore, event characteristics, such as the medical condition associated with the report, are provided in the form of open-ended responses. Errors arising from systematic differences in conditions being coded will result in biased estimates of those currently receiving treatment from the MEPS indicator.

Even though a nonresponse adjustment was made to account for MEPS nonresponse, we still observed statistically significant differences between CSAQ respondents and nonrespondents with respect to some characteristics. We did not, however, observe a statistically significant difference in the MEPS indicator of current treatment status, suggesting little, if any, evidence of nonresponse bias in this estimate. However, as Groves [24] cautions, estimates within a survey are likely to vary in terms of their nonresponse bias properties. If there is a correlation between some other substantive variable of interest and nonresponse propensity, then nonresponse biases may exist [25]. For example, we observed that there were statistically significant differences in the percentage of respondents and nonrespondents who had their first cancer diagnosis more than five years ago (65.6% and 54.2% of respondents and nonrespondents, respectively). It seems plausible that this characteristic may be correlated with treatment status; thus, inferences about this substantive area might be subject to nonresponse bias.

We note that our analyses were hampered by small sample sizes. This prohibited us from conducting multivariate analyses with sufficient power to investigate whether there were interactions among any of the domains considered. For example, we were not able to explore whether there were interactions between proxy/self-response and CSAQ completion timeliness. For self-respondents to MEPS, it is possible that MEPS could provide retrieval cues to help facilitate the recall of information relevant to data collected in the CSAQ. With larger sample sizes via pooling data from possible future administrations of the CSAQ, these multivariate analyses could potentially be conducted.

One final area of research is to explore the potential sources of coverage error associated with the screening methods employed in MEPS to determine eligibility for the CSAQ. To investigate potential coverage errors, using the MEPS indicator, we examined the subset of MEPS sample members that were identified as currently receiving treatment, but were ineligible to participate in the CSAQ. 460 MEPS participants were identified as currently receiving treatment based, but only 330 of them were eligible for participation in the CSAQ. The majority of these discrepant cases (106 out of 130) had their first cancer diagnosis prior to age 18, so they were, in fact, ineligible to participate in the CSAQ. Of the remaining 24, nine became out-of-scope during that round of data collection (e.g., died or institutionalized), but there were an additional 15 that were still ineligible for CSAQ participation. When administered the CSAQ, these 15 reported that their cancer diagnosis report from MEPS was an error. Ten of these 15 respondents were also the MEPS respondent while five had MEPS proxy respondents. Additional research pertaining to coverage error could explore these types of erroneous reports, i.e., were they the result of satisficing behaviors or misinterpreting the survey questions?

Finally it is important to acknowledge that survey practitioners can learn a great deal from supplemental questionnaires or follow-up surveys. These types of data sources offer unique opportunities for assessing survey errors since analyses can be conducted within the same unit and often provide additional information about the sample unit. Understanding these errors can lead to enhancements for and improvements of future surveys. Furthermore, supplemental questionnaires that piggy-back onto existing surveys are useful for targeting areas of research and researchers should not feel limited to using only one source. To the extent possible, they should consider combining the data sources to maximize the information available for analysis.

References

1. Lessler, J. & Kalsbeek, W. (1992). *Nonsampling Errors in Surveys.* John Wiley & Sons, Inc. Hoboken, New Jersey.

2. Groves, R. M. (1989). *Survey Errors and Survey Costs*. John Wiley & Sons, Inc. Hoboken, New Jersey.

3. Groves, R. M., Fowler, J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey Methodology*. New York: Wiley.

4. Biemer, P. B. & Lyber, L. E. (2003). *Introduction to Survey Quality*. John Wiley & Sons, Inc. Hoboken, New Jersey.

5. Zajacova, A., Dowd, J. B., Schoeni, R. F., & Wallace, R. B. (2010). Consistency and precision of cancer reporting in a multiwave national panel survey. *Population Health Metrics*, *8*(20), 1-11.

6. MEPS HC-147. 2011 Full Year Consolidated Data File, September 2013. Agency for Healthcare Research and Quality, Rockville, MD. Retrieved September 19, 2014 from http://meps.ahrq.gov/mepsweb/data_stats/download_data/pufs/h147/h147doc.pdf

7. Short, P. F., Moran, J. R., & Punekar, R. (2011). Medical expenditures of adult cancer survivors aged <65 years in the United States. *Cancer*, *117*(12), 2791-2800.

8. Yabroff, K. R., Dowling, E., Rodriguez, J., Ekwueme, D. U., Meissner, H., Soni, A., Lerro, C., Willis, G., Forsythe, L.P., Borowski, L., & Virgo, K. S. (2012). The Medical Expenditure Panel Survey (MEPS) experiences with cancer survivorship supplement. *The Journal of Cancer Survivorship*, *6*(4), 407-419.

9. Lerro, C. C., Stein, K. D., Smith, T., & Virgo, K. S. (2012). A systematic review of large-scale surveys of cancer survivors conducted in North America, 2000—2011. *The Journal of Cancer Survivorship*, *6*, 115-145.

10. Ezzati-Rice, T. M., Rohde, F., & Greenblatt, J. (2008). *Sample design of the medical expenditure panel survey household component, 1998–2007* (Methodology Rep. No. 22). Agency for Healthcare Research and Quality. Rockville, MD.

11. HCUP CCS. Healthcare Cost and Utilization Project (HCUP). July 2014. Agency for Healthcare Research and Quality, Rockville, MD. Retrieved September 19, 2014 from http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp.

12. Cancer Research Network (U19 CA79689). *Cancer Therapy Look-up Tables: Pharmacy, Procedure & Diagnosis Codes.* http://crn.cancer.gov/resources/codes.html.

13. Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37-46.

14. Machlin S., Carper, K., & Kashihara, D. (2011) *Health care expenditures for non-melanoma skin cancer among adults, 2005–2008* (Statistical Brief No. 345). Agency for Healthcare Research and Quality, Rockville, MD.

15. Machlin, S. R., Chowdhury, S. R., Ezzati-Rice, T., DiGaetano, R., Goksel, H., Wun, L. -M., Yu, W., & Kashihara, D. (2010). *Estimation procedures for the Medical Expenditure Panel Survey Household*

Component (Methodology Rep. No. 24). Agency for Healthcare Research and Quality, Rockville, MD.

16. Lin, H-M., Kim, H-Y, Williamson, J. M., & Lesser, V. M. (2012). Estimating agreement coefficients from sample survey data. *Survey Methodology*, *38*(1), 63-72.

17. Wolter, K. (1985). *Introduction to variance estimation.* New York: Springer-Verlag.

18. Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174.

19. Redline, C. D. (2011). *Clarifying Survey Questions*. Unpublished doctoral dissertation, University of Maryland, College Park.

20. Bradburn, N. M. (1978). Respondent burden. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 35-40.

21. Gonzalez, J. M. (2012). *The use of responsive split questionnaires in a panel survey*. Unpublished doctoral dissertation, University of Maryland, College Park.

22. Goldenberg, K. L., McGrath, D., & Tan, L. (2009). The effects of incentives on the Consumer Expenditure Interview Survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 5985-5999.

23. Shields, J. & To, N. (2005). Learning to Say No: Conditioned Underreporting in an Expenditure Survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 3963-3968.

24. Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*, 646-675.

25. Bethlehem, J. (2002). Weighting nonresponse adjustments based on auxiliary information. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 275-288). New York: Wiley.