Evaluation of 2014 Refresher Training: Descriptive Analysis

Wendy Hicks and Andrew Mercer and Doris C. Lefkowitz and Steven R. Machlin and Lisa Mirel

Agency for Healthcare Research and Quality Working Paper No. 14003

December 2014

Suggested citation: Hicks W, Mercer A, Lefkowitz DC, Machlin SR, Mirel L. Evaluation of 2014 Refresher Training: Descriptive Analysis. 14003 December 2014, http://gold.ahrq.gov.

Evaluation of 2014 Refresher Training: Descriptive Analysis

Wendy Hicks, Andrew Mercer, Doris Lefkowitz, Steve Machlin, and Lisa Mirel

ABSTRACT

In 2014 Westat held an in-person refresher training designed explicitly for experienced interviewers that focused primarily on techniques for enhancing the quality and completeness of respondent recall and reporting of health care events. Throughout all stages of the Data Quality Initiative, training emphasized interviewer behaviors and skills, and interviewing procedural changes geared towards minimizing recall error, without making significant changes to the computer assisted personal interview (CAPI) instrument. The bulk of the changes to the training program addressed underreporting of health care events by focusing on interviewer skills and behaviors that can facilitate better recall.

Observed changes in the descriptive measures are consistent with our hypothesis that interviewers are in fact engaging in the desired behaviors targeted with the Data Quality Initiative and the refresher training more specifically. However, the Medical Expenditure Panel Survey did not implement any phase of the Data Quality Initiative as a randomized experiment and thus cannot attribute the observed changes in interviewer behaviors, nor observed increases in the utilization estimates directly to the changes to the training program. Analysis indicated:

- Data show a higher proportion of completed cases reported as using records to aid recall, and a higher proportion of interviews using "key" records (i.e., records with more complete information about health care events) as recall aids during the interview.
- Interviewers also seem to engage in more effective probing as measured by increased time spent in the sections of the interview that require more probing to elicit better reporting.
- Comparisons indicated that mean annualized estimates of office-based provider visits, physician visits and prescription medicines were higher than for prior estimates.

Though not conclusive, these data suggest that the Data Quality initiative and especially the in-person refresher training provided interviewers with the skills and tools needed to facilitate better respondent recall and reporting of health care events within the current CAPI instrument design.

Westat
1600 Research Boulevard
Rockville, MD 20850-3129
(301)251-1500

Refresher Training and the Data Quality Initiative

In the Spring of 2013 the Medical Expenditure Panel Survey (MEPS) began a new initiative to improve the accuracy and completeness of household respondents' reports of the study's key variables -- medical events experienced by members of the participating households. This initiative, the Data Quality (DQ) initiative, focused on improving those interviewer skills and behaviors that facilitate respondent recall and reporting of health care events, such as the appropriate use of memory aids, effective probing and active listening.

MEPS implemented the Data Quality initiative in several phases. In the first stage, Westat adapted the existing in-person training program developed for newly hired interviewers to place greater emphasis on skills and behaviors that can elicit better respondent reporting. Elements of the revised in-person training were also incorporated into the home study materials used by experienced interviewers. A data quality monitoring program was put in place in the latter part of the Spring 2013 data collection in which field supervisory staff received monthly reports on selected data quality metrics such as the interviewers' use of records in completed interviews, appropriate use of the instrument's alternative paths for collecting event utilization data, and interview length. Supervisors used the reports to provide additional coaching for their interviewers.

MEPS launched the second stage of the initiative in early 2014 with a new, three-day in-person 'refresher' training for all experienced interviewers. Traditionally, any additional training provided to interviewers after the initial new-hire training occurs via a remote, self-paced training package. The three-day in-person training designed explicitly for experienced interviewers as part of the Data Quality Initiative focused primarily on techniques for enhancing the quality and completeness of respondent recall and reporting of health care events. This report provides a descriptive evaluation of the effect of the Data Quality Initiative and the addition of the

experienced interviewer Refresher Training on interviewer behaviors relating to the quality of respondent reported data and the resulting estimates of health care usage.

Summary of In-Person Refresher Training

Refresher Training Schedule

The three-day training for experienced interviewers was conducted in January 2014 in Los Angeles, California. All 328 of the experienced interviewers continuing with the project were expected to attend, with half attending a first session January 4-6 and half a second session January 7-9. When winter storms and poor traveling conditions prevented a substantial number (41) of interviewers from attending their scheduled sessions, MEPS project staff conducted a make-up in-person session February 18-20 at a hotel near Westat's Rockville offices to ensure that all active interviewers received the new training.

Training Content

Refresher training focused primarily on interviewing procedures and specific behaviors expected to facilitate better respondent recall and reporting of health care use. The content drew largely from literature on memory, recall error, and measurement error. Attachment A shows the agenda for the three-day training.

Trainers presented three categories of content across the three days:

- Conceptual or theory-based sessions addressing the reasons for and correlates of underreporting error;
- Applied sessions that focused on specific behaviors and project-supplied materials and procedures that encourage more complete respondent reporting; and
- Sessions targeting specific sections of the MEPS interview identified by field staff or through home office review as susceptible to error.

Across the three days, sessions oscillated between the three categories of content, with most of the theory-based sessions presented the first day. Training formats varied across sessions using lecture, interactive exercises, and hands-on practice with the MEPS CAPI instrument.

One of the main goals of the more theory-based sessions was to help interviewers understand why respondent underreporting occurs in a survey like MEPS, and to point out features of the MEPS overall design that can hinder full and accurate reporting. In these sessions, trainers covered basic interviewing skills necessary for identifying situations when respondents might underreport and introduced the concept of "Respondent-Centered Interviewing" (RCI). With RCI, interviewers learn to tailor their approach – what they say and what materials they use --to the respondent and the particular interviewing situation. The tailoring requires the interviewer to assess the respondent and the situation in order to determine the most effective way to encourage full reporting without compromising willingness to participate. In particular, this part of the training focused on teaching interviewers to identify situations in which they could comfortably decide not to start an interview or to break an interview in progress in order to get better records, or to get the participation of another household member when the family's primary respondent could not report fully for certain household members.

The applied sessions covered the new tools created for 2014 to help respondents become better, more motivated reporters as well as the related interviewer skills:

- Coaching respondents on record keeping and on identifying the types of records most accessible for a given respondent and most likely to contain detailed utilization information

- Probing within the CAPI question-answer sequence and probing outside of the strict question-answer flow.

The sessions targeting specific components of the CAPI questionnaire used a hands-on format. In these sessions, trainers used training cases pre-loaded on the laptops to take interviewers through parts of the interview, re-teaching them how to interact with the CAPI instrument to better capture respondent reported information (both health care use and other types of information) and to take greater advantage of the full set of navigational features available in the current application.

Data

The evaluation of the data quality initiative is based on descriptive metrics for Rounds 1 and 3 of the Panels interviewed since 2010 (Panels 15-19 for Round 1 and Panels 14-18 for Round 3). As described above, much of the training covered interviewer behaviors geared towards helping respondents reduce recall error. Given the focus on minimizing recall error, the analysis keeps the relative recall periods across panels consistent. However Panel 19 Round 1 shifted the start and end dates for data collection two weeks later relative to Panels 16, 17 and 18, resulting in a different distribution of reference period days across the full data collection period (see Figure 1, Technical Appendix A). As such, we restrict the eligible cases to those completed between January 24 and June 30 for all panels so that all panels reflect a consistent set of calendar days relative to January 1. (See Technical Appendix A for additional analysis and discussion based on full Round 1 data, without truncation.) Round 3 data collection used comparable dates across all panels allowing the use of all cases in the analysis. The analysis for both Round 1 and Round 3 includes all interviews completed by any interviewer, new or experienced.

Calculations are weighted using the MEPS Monitoring Weight. The Monitoring Weight incorporates a household's probability of selection into the National Health Interview Survey

(NHIS) and the probability of selection into MEPS conditional on completing the NHIS. The analysis excludes any nonresponse adjustment for either the NHIS or MEPS, as nonresponse adjustments are not yet available for MEPS Panel 19. Panel 19 Round 1 cases included one additional adjustment so that the weighted number of completed interviews in each two-week interval between January 24 and June 30 matched the distribution observed for Panel 18 Round 1. This last weight adjustment further standardized differences in the fielding of cases resulting from the later start of Panel 19 Round 1 data collection. For this descriptive analysis, we did not do any statistical testing. All discussion of increases or decreases are based on the observed percentage distributions.

Evaluation Metrics

The evaluation of the Data Quality initiative and specifically refresher training involves two primary components. First, the analysis assesses the extent to which interviewers can apply the behaviors covered in training, such as coaching respondents on better record keeping and probing to elicit complete responses. Second, the analysis looks for observable increases in reported medical utilization consistent with reduced underreporting for those types of medical events most susceptible to recall error.

Interviewer Behaviors

MEPS does not currently have a method for directly measuring interviewer behaviors that occur outside of the CAPI instrument. As such, there is no direct measure of how well an interviewer "coaches" respondents on record keeping. Nor is there available a direct way to observe or measure probing skills or other verbal interactions not directed or recorded by CAPI. For this reason, the evaluation uses several indirect metrics as proxy indicators of the interviewer behaviors targeted in the refresher training. Though indirect, we anticipate that the measures should change in the aggregate as a result of the training emphasis. Table 1 identifies the specific

measures used in the analysis as they relate to training priorities and the a priori hypothesized direction of change. Because these data are descriptive in nature and not the product of a randomized experiment, we cannot definitively isolate the effects of training from other external changes in the population that may have affected fieldwork. However, the absence of change, or change in the direction opposite to that expected, would provide evidence that the training was not effective. Although not definitive, change in the expected direction across many of these measures would suggest that interviewers are in fact engaging in the desired behaviors targeted with the DQ Initiative and the refresher training more specifically.

Record Use and Key Record Use

The first row shown in Table 1 addresses interviewer behaviors related to increasing respondent use of medical records in reporting health care events in order to reduce the likelihood of recall error. The second row addresses interviewers' ability to teach respondents to use a more complete type of record, referred to as "key" records in MEPS. "Key" records are those records that contain the most complete information about a health care event, such as an Explanation of Benefits (EOB) or similar insurance or medical provider statements, accessing a patient portal account, or a personal calendar with notes regarding the provider and the health care visit.

As seen in the first row of Table 2a, the data suggest that interviewers are successful in applying these particular training points. For Round 1, the proportion of completed reporting units (RU's) that did not use records of any kind increased from 20% in P15 to just under 28% in Panel 17. With Panel 18 at the start of the Data Quality Initiative, the data show a drop to 17% of households not using any records. Following the Panel 19 refresher training, we observe an additional reduction in the proportion of Round 1 interviews completed without

any memory aids, down to 13.5%.

Table 2a also shows a complimentary pattern in regards to the proportion of households using key records during the interview. Between P15 and P17, the data show a decline in the proportion of households using key records to report. However, that trend reverses with the start of the Data Quality Initiative in P17, jumping up by 15 percentage points between P16 and P17, and then increasing another seven percentage points after the refresher training.

Table 2b shows that for Round 3, the pattern is similar although less dramatic for both record use in general and key record use. By Round 3, interviewers have had two prior interview opportunities to help respondents be better record keepers, so we anticipated smaller changes with the Round 3 data. Prior to the start of the Data Quality Initiative, Table 2b shows a trend of increasing higher proportions of households not using records during the interview. However, with the start of the Data Quality initiative in P17R3 we see a drop of about six percentage points in the proportion of households completing an interview without records. After refresher training, we can see an additional reduction of close to six percentage points for Panel 18 Round 3.

As with general record use, the observed proportion of completed cases using key records before and after the Data Quality initiative and Refresher training differ less for Round 3 as compared to Round 1. The proportion of RU's using some form of key record prior to 2013 was about 70% in P14R3 dropping down to 66% in P16R3, but increasing in P17R3 to 74% and a further increase to 84% in P18R3.

The MEPS instrument captures only that one or more of a particular type of record was used during an interview, as entered by the interviewer, so this metric does not indicate whether

more key records are being used in RU's that were already using some key records. Rather, an increase in this metric tells us that the pool of RU's that are using at least some key records is growing.

More Selective use of the Calendar Path

The "calendar path" in the MEPS CAPI instrument was designed to reduce the burden of the interview for respondents who had kept a calendar of health care events or had records available for reference. The design of the calendar path assumes that respondents have full and complete medical records available for reference when reporting health care events for most if not all of the household members. With this assumption, the calendar path uses slightly less structure and includes fewer prompts to stimulate recall of medical care. In contrast, the non-calendar path assumes that the respondent has few if any records for any household member and will report the family's health care events largely based on memory. The non-calendar path uses a structured set of probes to stimulate recall, looping through each household member one at a time. The person- specific probing allows the respondent to focus on one household member at a time when trying to recall or reconstruct health care events, simplifying the cognitive task relative to recall for the entire household at once. Because of the person-level probing, the non-calendar path can take more time than the calendar path to administer.

At refresher training, interviewers reviewed the assumptions guiding the selection of the calendar versus non-calendar paths. The training points identified the two situations in which the selection of the calendar path or non-calendar path are most clear - using the calendar path when full and complete records are available for most or all RU members, and conversely, using the non-calendar path if no key records are available. However, if the respondent has fewer than all records, but does have some, the selection of the calendar path or non-calendar path is less clear-cut and can depend on the level of cooperation, the size of the household and the round. In

training interviewers on the assumptions behind the design of the two paths, our goal was to have interviewers select the non-calendar path even if the respondent might have key records for him/herself, but not for other household members. Using the non-calendar path in these situations allows the respondent to think about the other household members individually when reporting their health care. The longer series of probes associated with the non-calendar path can be particularly helpful in Round 1, when the probes acquaint respondents with the broad range of events included in the study's definition of health care.

The data in tables 3a and 3b suggest that interviewers changed their behavior in accordance with the training regarding record usage. For Round 1 cases that used the calendar path, the proportion using key records showed a decreasing trend from 78% in P15R1 to 72% in P17R1, but reversed that trend at the start of the Data Quality Initiative. In P18R1, 86% of the interviews following the calendar path used key records, and continuing to rise to close to 92% in P19R1. For Round 3 calendar path cases, the proportion using key records rose from 86% in P14R3 to 96% in P18R3.

Perhaps more telling is the fact that among non-calendar path cases, the proportion with key records rose from 46% in P15R1 to 58% in P19R1 and in Round 3, rose from 41% in P14R3 to 58% in P18R3. This might suggest that interviewers are not simply using the presence of any key records as a reason to use the calendar path, but apply more discrimination in the choice of path.

Better Probing Skills

Training materials covered two probing approaches for improving respondent reporting, as well as specific application of each approach to the MEPS interview. Trainers discussed that the probing approach used to improve respondent reporting differs depending on the specific

goal –

- To get a clear and valid answer to a CAPI question, or
- To get the respondent to think differently or more deeply about a topic – in particular to reduce the chance of under-reporting - but not directly for the purpose of answering a specific CAPI question

This evaluation focuses on the second probing skill – probing to get the respondent to think more deeply or in a different way about a topic. This type of probing applies well as a method for improving health care reporting, specifically within certain portions of the event enumeration section of the interview. In the Calendar (CA) screens within the instrument, interviewers worked on probing skills that could help them to better identify the types of records available, barriers to record use as well as the breadth of coverage of the health care records across household members. In addition, at the CA screens, probing techniques included methods for getting respondents to expand the range of health care events they consider appropriate for reporting during the interview.

Training also gave interviewers probing techniques to re-engage respondents and stimulate better recall of health care events during the Provider Probes (PP) section of the instrument for both the calendar and non-calendar path versions of the PP section.

The data shown in Table 4a suggest that interviewers did have some success in applying the new probing skills covered at Refresher Training in Panel 19 Round 1, relative to prior years. Most notable is the three percentage-point drop in the proportion of household respondents indicating that no one in the household had any health care events during the reference period (i.e., % volunteer no events), dropping from 9.3% in Panel 15 to 6.4% in Panel 19. This decrease suggests that in the aggregate, interviewers were able to get

households to think more expansively about health care events and realize that in fact, someone in the household had at least one health care event to report.

Table 4a also suggests improvements in interviewer probing skills for both the CA section and PP section of the instrument. For Round 1 of Panel 15, Panel 16 and Panel 17, the mean administration time for the CA section ranged from 40 seconds dropping to 34 seconds. This increased to 47 seconds for Panel 18 Round 1 and over a minute for Panel 19 Round 1, almost doubling the mean time from just prior to the Data Quality Initiative. Within the PP section, we calculated the mean administration time per RU member for the section separately for respondents using the calendar path and respondents using the non-calendar path. For Panel 15 Round 1 through Panel 17 Round 1, the mean administration time showed a decreasing pattern per person for the calendar path. At Panel 18 Round 1 the trend reversed for both the calendar and non-calendar path mean administration time per person. The mean per person time on the calendar path increased to just under five and a half minutes, and just under five minutes for the non-calendar path in Panel 18 Round 1. Mean administration time increased further in Panel 19 Round 1, with the calendar path increasing by a little over a minute on average per person, and the non-calendar path close to a minute longer in length per person.

The pattern is consistent for Round 3 for probing to help household respondents realize that they do have events to report, as well when probing within both the CA screens and the PP section. Between Panel 14 Round 3 and Panel 17 Round 3, the proportion of households volunteering no health care events to report remained fairly flat at a little over 6%. However, at Panel 18 Round 3, the proportion had dropped to 4.7%.

Mean administration time within the CA section was essentially flat at about 30 seconds between Panel 14 Round 3 and Panel 17 Round 3. However, the mean administration time in the Calendar Section increased to just under a minute in Panel 18 Round 3. Although these

times seem short, they represent growth of over 50% in the mean length by Panel 19 for Round 1 administration, and by Panel 18 for Round 3 administration.

Round 3 cases exhibited larger increases in administration time from 6.6 minutes per person for Panel 15 Round 3 to 9.1 minutes per person in Panel 18 Round 3 for the calendar path. For the non-calendar path, the timings grew from 5.6 minutes in Panel 14 Round 3 to 7.7 minutes in Panel 18 Round 3. All of these changes are consistent with more time spent on more thorough administration and probing.

Willingness to Break Interviews to get Better Data Quality

The refresher training also included methods for determining viable situations in which the interviewer could defer the start of an interview or decide to break an interview in progress with the expectation that the delay would facilitate better health care reporting. Specifically, interviewers received guidance on identifying when getting the participation of a more knowledgeable respondent might improve reporting, or when the current respondent would be willing to do more preparation gathering better health care records before continuing the interview. This could mean coming back on another occasion when the more knowledgeable respondent is available, or after the respondent assembled medical records for more accurate reporting. If an interviewer decides to defer the start of an interview for either of these reasons, we'd expect an increase in the average number of visits to respondent households. If the interviewer decides to break-off an interview already in progress and reschedule for another time for either of these reasons, we'd expect an increase in the proportion of multi-session cases. (However, the way CAPI captures the classification of a multi-session interview changed in 2013, so we can only evaluate the change between 2013 and 2014.)

To accurately measure interviewers' readiness to delay or break interviews with willing respondents, the metrics need to include cases from the start of fieldwork. Length of reference period is not a factor for either of these interviewer behaviors. As such, the metrics shown in Table 5a include all cases completed for the entire Round 1 data collection period for all panels prior to 2014, and do not use the weighting adjustment to standardize the length of the reference period.

For Panel 18 Round 1, 7.8% of the 7,699 completed interviews in multi-person households were conducted as a multi-session interview. This grew to 10.3% of the 7,475 completed interviews in multi-person households in Panel 19 Round 1. Additionally, Table 5a shows a small reduction for Round 1 cases completed at the first contact with a household member (e.g., first true contact), going from 15% in P15R1 to 13% in P19R1, perhaps a modest indication of willingness to defer the start of an interview with a willing respondent. The mean number of true contacts per completed interview grew slightly from 3.2 in Panel 18 Round 1 to 3.5 in Panel 19 Round 1.

For Round 3, we expected smaller changes in these metrics since, by the third interview, household respondents know how to prepare for the MEPS interview. However, Table 5b, shows a two percentage-point increase in the percentage of multi-session interviews in multi-person households between Panel 17 Round 3 and Panel 18 Round 3, just a little smaller than the increased observed in Round 1. The last row of Table 5b also shows about a two percentage-point decrease in the proportion of interviews completed at the first true contact between Panel 14 Round 3 (6%) and Panel 18 Round 3 (4%). Round 3 also showed a complimentary increase in the mean number of true contacts at Panel 18, up to 3.4 from the relatively stable mean of 3 true contacts per complete in prior years.

Medical Event Reporting

As is the case with interviewer behaviors, we cannot directly observe the effect of the Data Quality Initiative and Refresher Training on medical event reporting. However, we would expect to see changes in event reporting that are consistent with improved recall. In particular, we expect to see increased utilization reporting for events that are more likely to be forgotten or those less likely to be known by all members of a household. For instance, one household member may not be aware of all of the office-based doctor's visits or prescription medications used by other household members. Increased use of medical records and gaining the participation of the most knowledgeable person should result in increased reporting of these kinds of events. On the other hand, hospitalizations (HS) and emergency room (ER) visits are likely more salient and memorable, and more likely known by all household members. Reporting of these types of events should be less sensitive to the trained behaviors. As such we expect less change in HS and ER visit reporting as a result of the training.

For this evaluation, we looked at the following event types:

- Office based visits to any type of medical provider (MV)
- Physician office visits (MVE)
- Hospital inpatient stays (HS)
- Hospital outpatient department visits (OP)
- Emergency room visits (ER)
- Prescribed medicine purchases (PM)

Because the number of events reported by a household will vary based on the size of the RU and the length of the reference period, we use an annualized measure of events per person per day, calculated as follows:

$$\text{Utilization Rate} = \frac{\text{\# Events Reported in RU}}{\text{\# Person in RU} \times \text{\# Days in Reference Period}} \times 365$$

For each type of medical event, we calculated the mean utilization rate for that event type using the adjusted monitoring weights. For each event type, we also calculated a utilization rate among those households that reported one or more event of that same type. Finally, to confirm that more households are reporting events (as opposed to the same households reporting more events), we calculated the proportion of RU's that reported no events of any kind.

We expect that MV, MVE and PM events will increase the most following the Data Quality Initiative and the Refresher Training. HS and ER events generally represent significant medical events that are memorable and likely known to all household members. As such we expect less change in the reporting of these event types. We do not have a clear hypothesis for Outpatient visits (OP) since those can represent care similar in significance to an HS or ER event, but may also reflect an experience more like an office-based visit. Additionally, anecdotal reports from the field staff suggest that respondents have difficulty determining what types of places count as Outpatient visits.

Tables 6a and 6b show these measures for Rounds 1 and 3 respectively.

The results for both Round 1 (Table 6a) and Round 3 (Table 6b) largely correspond to the hypothesized changes. For Round 1, MV, MVE and PM events show a small increase in Panel 18 Round 1 followed by a larger increase after the Refresher Training in Panel 19 Round 1. Looking at the Round 3 comparisons, the primary increases for MV, MVE and PM occur with Panel 18 Round 3 after Refresher training. On the other hand, HS, OP and ER events do not show a discernable trend following the initiation of the Data Quality Initiative. For MV, MVE and PM events, the magnitude of the change is greater for Round 1 cases than Round 3 cases.

Similarly, the proportion of completed interviews with zero events reported dropped approximately 1.5 percentage points between Panel 17 Round 1 and the start of the Data Quality Initiative at Panel 18 R1 (18.9% and 17.2% respectively). The proportion of 0 event households dropped an additional three percentage points after the Refresher Training to 13.9% in Panel 19 Round 1. In Panel 18 Round 3, the reduction in the percentage of households reporting 0 events occurred after Refresher training only, dropping 2 percentage points from 11.5% in Panel 17 Round 3 to 9.4 % in Panel 18 Round 3.

Summary

Westat launched the Data Quality Initiative in Spring 2013 in response to AHRQ's concerns about declines observed in MEPS utilization estimates that were not consistently observed in other comparable data sources. The initiative started with a retooling of the annual new hire training to increase emphasis on interviewer skills and behaviors that can elicit better respondent reporting. The experienced interviewer home study materials received similar updates at the same time. Later in the Spring 2013 data collection, the home office introduced a set of interviewer level monitoring reports that field supervisors and managers used to identify interviewers needing further coaching and training on these same skills. In Spring 2014, Westat held an in-person refresher training designed explicitly for experienced interviewers that focused primarily on techniques for enhancing the quality and completeness of respondent recall and reporting of health care events. Throughout all stages of the Data Quality Initiative, including the in-person Refresher Training for experienced interviewers, training emphasized interviewer behaviors and skills, and interviewing procedural changes geared towards minimizing recall error, without making significant changes to the CAPI instrument. Prior to the refresher training, the last in-person training of experienced interviewers occurred in 2007 and focused primarily on

the conversion of the MEPS interview into a new Computer Assisted Personal Interviewing (CAPI) software application. Since then, experienced interviewers received only self-paced remote learning training packages.

MEPS did not implement any phase of the Data Quality Initiative as a randomized experiment and thus cannot attribute the observed changes in interviewer behaviors, nor observed increases in the utilization estimates directly to the changes to the training program. However, observed changes in the descriptive measures presented here consistent with our hypothesis suggests that interviewers are in fact engaging in the desired behaviors targeted with the DQ Initiative and the refresher training more specifically. Relative to Round 1 data collection in prior panels, the Panel 19 Round 1 data show a higher proportion of completed cases reported as using records to aid recall, and a higher proportion of interviews using "key" records (i.e., records with more complete information about health care events) as recall aids during the interview. Compared to prior panels, the Panel 19 Round 1 interviewers also seem to engage in more effective probing as measured by increased time spent in the sections of the interview that require more probing to elicit better reporting. Similarly, Panel 19 Round 1 had the lowest proportion of households indicating before the start of event enumeration that no one had health care events to report – a response that interviewers were told to probe further to help respondents understand the full breadth of the health care events included in MEPS. The Round 3 comparisons similarly suggested that interviewers did in fact engage in the behaviors targeted by the Data Quality initiative and refresher training.

As discussed, the bulk of the changes to the training program addressed underreporting of health care events by focusing on interviewer skills and behaviors that can facilitate better recall. Not all types of health care are as prone to recall error. Hospital stays and in many instances,

emergency room visits are salient events less likely to be forgotten or unknown by the adult household members. In contrast, office-based visits, dental visits, and prescription medicines may be forgotten more easily. In fact, the Panel 19 Round 1 mean annualized estimates of office-based provider visits, physician visits and prescription medicines were higher than the Panel 15, 16, 17 and 18 Round 1 estimates. Similar increases occurred with Panel 18 Round 3 mean annualized estimates of office based provider visits, physician visits and prescription medicines relative to Panel 14, 15, 16 and 17 Round 3 estimates.

Though not conclusive, these data suggest that the changes to the training program implemented as part of the Data Quality initiative, and especially the in-person refresher training provided interviewers with the skills and tools needed to facilitate better respondent recall and reporting of health care events within the current CAPI instrument design.

TECHNICAL APPENDIX A

Discussion of overall changes in Panel 19 targeting improved utilization reporting

With Panel 19, we introduced an additional change from prior years as part of the Data Quality Initiative, also with the goal of positively affecting utilization. Specifically, Panel 19 Round 1 started later in January than each of Panels 16, 17 and 18. Shifting the Round 1 start date later in the month allowed us to keep the minimum number of days in a reference period to 24, as compared to 10 in prior years. The increase in the minimum number of days in a reference period also increases the chance of health events occurring across households.

Shifting the start of the Round 1 fielding period meant that the other two spring rounds would already be in the field. Round 1 for Panels 16 – 18 had two-weeks of data collection without any simultaneous Round 3 or 5 case work. With all three rounds in the field, the productivity of working Round 1 cases also shifted, resulting in a more gradual build of completes relative to the years with Round 1 only in the field.

Figure 1 below shows the change in the relative distribution of completed Round 1 cases across reference period days for Panel 15 through Panel 19. Panel 15 which started February 1, and Panel 19 which started January 24, have the most similar distribution of completes across reference period days.

Figure 2 provides a visual comparison of mean annualized utilization rate, per week, for MV events, MVE events and PM events, across the Round 1 field period for Panel 15 through Panel 19. For all of the panels reviewed here, respondents interviewed in the first few weeks reported the highest rates of medical utilization. While the reporting in those early weeks may reflect a higher relative degree of telescoping (Zuvekas, 2011), the charts suggests that omitting the first two weeks from the comparison of Round 1 data across panels selectively excludes

households with the highest utilization for Panel 16, 17 and 18.

As a sensitivity analysis, we repeated the same utilization metrics shown in the main analysis but without truncation and without the additional weight adjustment to verify that the positive findings in the main analysis still applied with the data from the full field period. Table 7 shows the same utilization using the full field period for each panel and using only the base monitoring weight without the additional adjustment to standardize distribution of completes across the reference period days. Without comparable reference periods, differences in utilization rates do not account for the potential for differential recall error between the panels with earlier start and end dates, and panels with a later start and end date. Since Panel 15 and Panel 19 had similar start dates (January 24 and February 1, respectively) and end dates (July 16 and July 13 respectively) we rearranged the order of the panels in the tables so that the two panels with similar distribution of reference periods display in consecutive columns. This allows us to compare panels with similar reference period distributions without excluding the high utilization cases from the early weeks of Panels 16 - 18.

Comparing the annualized means between Panel 15 and Panel 19, the table shows that the event types we expected to show the biggest increase following the Data Quality Initiative and the Refresher Training (MV, MVE and PM), are in fact higher in Panel 19 relative to Panel 15 with both panels having similar field period start and end dates. Panels 17 and 18 had similar field period start and end dates, but Panel 18 data reflect the initial stages of the Data Quality initiative. As anticipated, the mean MV, MVE and PM per person estimates are all higher in Panel 18 relative to Panel 17, prior to the Data Quality Initiative.

Similarly, the proportion of households reporting 0 events is also slightly lower in Panel 19 as compared to Panel 15. (The larger relative proportion of households with 0 events reported in Panels 16, 17 and 18 likely reflects the earlier start date which results in less

opportunity for events to actually occur.)

The increase in the annualized mean events per person is less dramatic between the full Panel 18 and Panel 19 Round 1 data for both MV and MVE events, and in fact decreases between Panel 18 and Panel 19 for PM events. However, Figure 3 shows the cumulative total events by event type for Round 1 of Panels 15 through Panel 19. For all event types, the cumulative total is larger in Panel 19 as compared to all previous panels, including Panel 15 with a similar distribution of reference period days. The Panel 19 cumulative total is also higher than the Panel 18 cumulative total for all event types, suggesting even further gain from the in-person refresher training that followed the start of the Data Quality Initiative in Panel 18.

Looking towards the later rounds of data collection for Panel 19, we anticipate the need to continue accounting for differences in distribution of reference period days across panels to evaluate the long-term effect, if any, of the refresher training on utilization reporting. While the cost of repeating in-person training for all interviewers may be prohibitive in the next few years, we hope to have more frequent and more rigorous automated, self-paced remote trainings to extend the effectiveness of the refresher training.

Reference:

Zuvekas, Samuel H. (2011) “The effects of recall length and reporting aids on household reporting of health care events in the medical expenditure panel survey” Journal of Economic and Social Measurement 36, 321–343.

Table 1. Metrics Assessing Interviewer Behaviors Covered in Refresher Training

Training point	Measure	Expected Direction of Change
1) Coaching respondent to use records	Proportion using no records	Lower
2) Coaching respondent to get most informative records	Proportion using one or more key records (e.g., *calendar, insurance statement, provider statement, appointment card/reminder, electronic records)*	Higher
3) Increasing use of key records	% of all interviews using key records - Overall - Calendar path - Non-calendar path - Volunteer no events in calendar section	Higher
4) More selective use of the calendar path by interviewer (Calendar path is the flow through CAPI that assumes the respondent refers to complete health care records for all household members in response to the health care use questions)	Distribution of interviews by path/use of records (sum=100%) - Calendar path/key records - Calendar path/no records - Non-calendar path/key records - Non-calendar path/no records - Volunteer no event/key records - Volunteer no event/no records	Increasing % in top category
5) Better probing skills	Proportion volunteering no events in calendar section Administration time of the calendar section (2 screens) Administration time of the provider probes (PP) per person, by calendar path or not	Lower Higher Higher
6) Willingness to break interviews if need another/different RU member, or records aren't available	Proportion multi-session interviews among multi-person RUs (*Limited to P18 vs P19*) Mean number of true contacts per complete Proportion of first-time-final completes (complete at first true contact)	Higher Higher Lower

Table 2a: Interviewer behaviors to encourage record use, Round 1

Training Point	Metric	P15R1	P16R1	P17R1	P18R1	P19R1
	Count of Completes	6,332	6,481	6,301	6,225	7,114
Coaching respondents to use records	% using no records	20.3%	22.8%	27.6%	17.0%	13.5%
Coaching respondents to get most informative (key) records	% using one or more "key" records	60.8%	59.8%	54.9%	67.9%	74.8%

Table 2b: Interviewer behaviors to encourage record use, Round 3

Training Point	Metric	P14R3	P15R3	P16R3	P17R3	P18R3
	Count of Completes	6,970	6,257	8,047	7,655	7,197
Coaching respondents to use records	% using no records	18.2%	19.9%	21.1%	15.0%	9.4%
Coaching respondents to get most informative records	% using one or more "key" records	70.2%	67.3%	66.3%	74.5%	84.2%

Table 3a: Interviewer behaviors to encourage use of more complete records, Round 1

Training Point	Metric	P15R1	P16R1	P17R1	P18R1	P19R1
Increasing use of "key" records	Count of Calendar Path Completes	2,815	2,951	2,670	2,816	3,410
	% using Key records	78.2%	75.5%	72.0%	86.4%	91.5%
	Count of Non-Cal Path Completes	2,988	2,910	2,955	2,866	3,256
	% using Key records	45.6%	45.5%	40.9%	50.2%	57.9%
	Count of Volunteer No Events	529	620	676	543	448
	% using Key records	45.8%	44.1%	35.7%	51.0%	62.7%
More selective use of the calendar path	Calendar path w/ key records	36.5%	36.3%	33.6%	42.1%	45.2%
	Calendar path w/ no records	10.1%	11.8%	13.1%	6.6%	4.2%
	Non-Cal path w/ key records	20.1%	19.5%	17.7%	21.4%	25.6%
	Non-Cal path w/ no records	24.0%	23.4%	25.6%	21.3%	18.6%
	Vol. No Events w/ key records	4.2%	4.0%	3.5%	4.4%	4.0%
	Vol. No Events w/ no records	5.0%	5.1%	6.4%	4.2%	2.4%

Table 3b: Interviewer behaviors to encourage use of more complete records, Round 3

Training Point	Metric	P14R3	P15R3	P16R3	P17R3	P18R3
Increasing use of "key" records	Count of Calendar Path Completes	4,189	3,657	4,773	4,458	4,807
	% using Key records	86.0%	83.6%	82.4%	88.4%	95.9%
	Count of Non-Cal Path Completes	2,307	2,186	2,775	2,707	2,049
	% using Key records	40.9%	40.2%	39.8%	51.3%	57.2%
	Count of Volunteer No Events	474	414	499	490	341
	% using Key records	51.5%	46.7%	38.9%	53.3%	62.2%
More selective use of the calendar path	Calendar path w/ key records	54.5%	51.5%	51.5%	55.1%	66.4%
	Calendar path w/ no records	8.9%	10.1%	11.0%	7.3%	2.8%
	Non-Cal path w/ key records	12.4%	12.9%	12.5%	16.2%	14.9%
	Non-Cal path w/ no records	17.9%	19.2%	18.9%	15.3%	11.2%
	Vol. No Events w/ key records	3.2%	3.0%	2.4%	3.3%	2.9%
	Vol. No Events w/ no records	3.0%	3.4%	3.8%	2.9%	1.8%

Table 4a: Probing skills, Round 1

Training Point	Metric	P15R1	P16R1	P17R1	P18R1	P19R1
Better probing skills	% Volunteer No Events	9.26%	9.15%	9.93%	8.59%	6.39%
	Mean administration time in Calendar Section *(minutes)*	0.67	0.63	0.57	0.78	1.14
	Mean administration time per person, Provider Probes/ Calendar Path *(minutes)*	4.59	4.47	4.41	5.24	6.41
	Mean administration time per person, Provider Probes/ Non-Cal Path *(minutes)*	4.53	4.50	4.15	4.55	5.49

Table 4b: Probing Skills, Round 3

Training Point	Metric	P14R3	P15R3	P16R3	P17R3	P18R3
Better probing skills	% Volunteer No Events	6.23%	6.36%	6.20%	6.13%	4.73%
	Mean administration time in Calendar Section *(minutes)*	0.51	0.51	0.53	0.65	0.93
	Mean administration time per person, Provider Probes/Calendar Path *(minutes)*	6.60	6.33	6.50	7.23	9.05
	Mean administration time per person, Provider Probes/ Non-Cal Path *(minutes)*	5.55	5.36	5.15	5.90	7.67

Table 5a: Balancing Cooperation and Data Quality, Round 1

Training Point	Metric	P15R1	P16R1	P17R1	P18R1	P19R1
Willingness to break interviews to get better data quality (e.g., more records, another respondent)	Count of Completes (includes all cases from start of data collection)	6,819	8,362	8,116	7,699	7,475
	Count of multi-person completes	4,863	5,909	5,806	5,461	5,191
	% multi-session of multi-person	7.9%	4.8%	6.1%	7.8%	10.3%
	Mean # true contacts per complete	3.4	3.3	3.3	3.2	3.5
	% completes at first true contact	15.4%	15.2%	15.5%	14.7%	13.3%

Table 5b: Balancing Quality and Cooperation, Round 3

Training Point	Metric	P14R3	P15R3	P16R3	P17R3	P18R3
Willingness to break interviews to get better data quality (e.g., more records, another respondent)	Count of Completes (includes all cases from start of data collection)	6,982	6,267	8,058	7,668	7,213
	Count of multi-person completes	4,875	4,355	5,546	5,356	4,922
	% multi-session of multi-person	8.1%	8.6%	8.6%	11.9%	13.9%
	Mean # true contacts per complete	3.1	3.2	3.1	3.1	3.4
	% completes at first true contact	5.6%	6.5%	6.7%	5.0%	4.1%

Table 6a. Medical Utilization Measures, Round 1

General Measure	Event Type	P15R1	P16R1	P17R1	P18R1	P19R1
Annualized mean number of events per person	Office visit to any type of medical provider (MV)	5.77	5.76	5.78	6.28	8.00
	Physician office visits (MVE)	4.49	4.54	4.39	4.79	6.03
	Hospital inpatient stays (HS)	0.11	0.12	0.14	0.11	0.13
	Hospital outpatient visits (OP)	0.51	0.59	0.52	0.62	0.60
	Emergency Room visits (ER)	0.24	0.27	0.25	0.25	0.27
	Prescribed Medicine purchases (PM)	9.41	8.82	7.76	8.21	9.96
Annualized mean number of events per person, of persons w/ at least 1 event	Office visit to any type of medical provider (MV)	8.76	8.78	9.14	9.52	11.35
	Physician office visits (MVE)	7.21	7.29	7.38	7.76	9.13
	Hospital inpatient stays (HS)	2.67	2.63	2.85	2.73	2.59
	Hospital outpatient visits (OP)	5.09	5.96	5.62	6.18	5.25
	Emergency Room visits (ER)	2.38	2.76	2.75	2.73	2.57
	Prescribed Medicine purchases (PM)	13.43	12.68	12.03	12.38	14.23
Reporting of zero-events	Count of Completed RU's	6,332	6,481	6,301	6,225	7,114
	% with 0 events reported	15.4%	15.7%	18.9%	17.2%	13.9%

Table 6b. Medical Utilization Measure, Round 3

General Measure	Event Type	P14R3	P15R3	P16R3	P17R3	P18R3
Annualized mean number of events per person	Office visit to any type of medical provider (MV)	5.85	5.30	5.25	5.53	6.69
	Physician office visits (MVE)	4.55	4.03	4.11	4.27	4.88
	Hospital inpatient stays (HS)	0.13	0.11	0.10	0.11	0.11
	Hospital outpatient visits (OP)	0.59	0.47	0.53	0.43	0.63
	Emergency Room visits (ER)	0.21	0.20	0.20	0.22	0.24
	Prescribed Medicine purchases (PM)	4.92	4.61	4.36	4.30	4.59
Annualized mean number of events per person, of persons w/ at least 1 event	Office visit to any type of medical provider (MV)	7.60	6.94	6.95	7.23	8.37
	Physician office visits (MVE)	6.12	5.51	5.63	5.84	6.39
	Hospital inpatient stays (HS)	1.53	1.30	1.24	1.26	1.24
	Hospital outpatient visits (OP)	3.65	2.87	3.31	2.85	3.51
	Emergency Room visits (ER)	1.36	1.29	1.33	1.41	1.34
	Prescribed Medicine purchases (PM)	6.55	6.08	5.91	5.85	6.11
Reporting of zero-events	Count of Completed RU's	6,970	6,257	8,047	7,655	7,197
	% with 0 events reported	10.7%	10.2%	11.1%	11.5%	9.4%

Table 7. Medical Utilization Measures, Round 1, full reference period, no adjustment

General Measure	Event Type	P16R1	P17R1	P18R1	P15R1	P19R1
Annualized mean number of events per person	Office visit to any type of medical provider (MV)	6.28	6.13	6.74	5.63	7.49
	Physician office visits (MVE)	5.01	4.69	5.14	4.34	5.65
	Hospital inpatient stays (HS)	0.11	0.15	0.12	0.11	0.13
	Hospital outpatient visits (OP)	0.59	0.58	0.65	0.49	0.57
	Emergency Room visits (ER)	0.26	0.26	0.29	0.24	0.27
	Prescribed Medicine purchases (PM)	11.42	10.03	10.43	8.08	8.80
Annualized mean number of events per person, of persons w/ at least 1 event	Office visit to any type of medical provider (MV)	10.24	10.43	10.92	8.32	10.56
	Physician office visits (MVE)	8.68	8.60	8.98	6.78	8.48
	Hospital inpatient stays (HS)	2.98	3.49	3.23	2.41	2.41
	Hospital outpatient visits (OP)	6.74	6.77	7.07	4.80	4.79
	Emergency Room visits (ER)	3.17	3.23	3.44	2.25	2.39
	Prescribed Medicine purchases (PM)	17.20	16.45	16.59	11.43	12.55
Reporting of zero- events	Count of Completed RU's	8,362	8,116	7,699	6,819	7,475
	% with 0 events reported	18.6%	22.0%	20.5%	14.6%	13.9%

Figure 1: Reference Period Distributions

Panels 15-19, Round 1

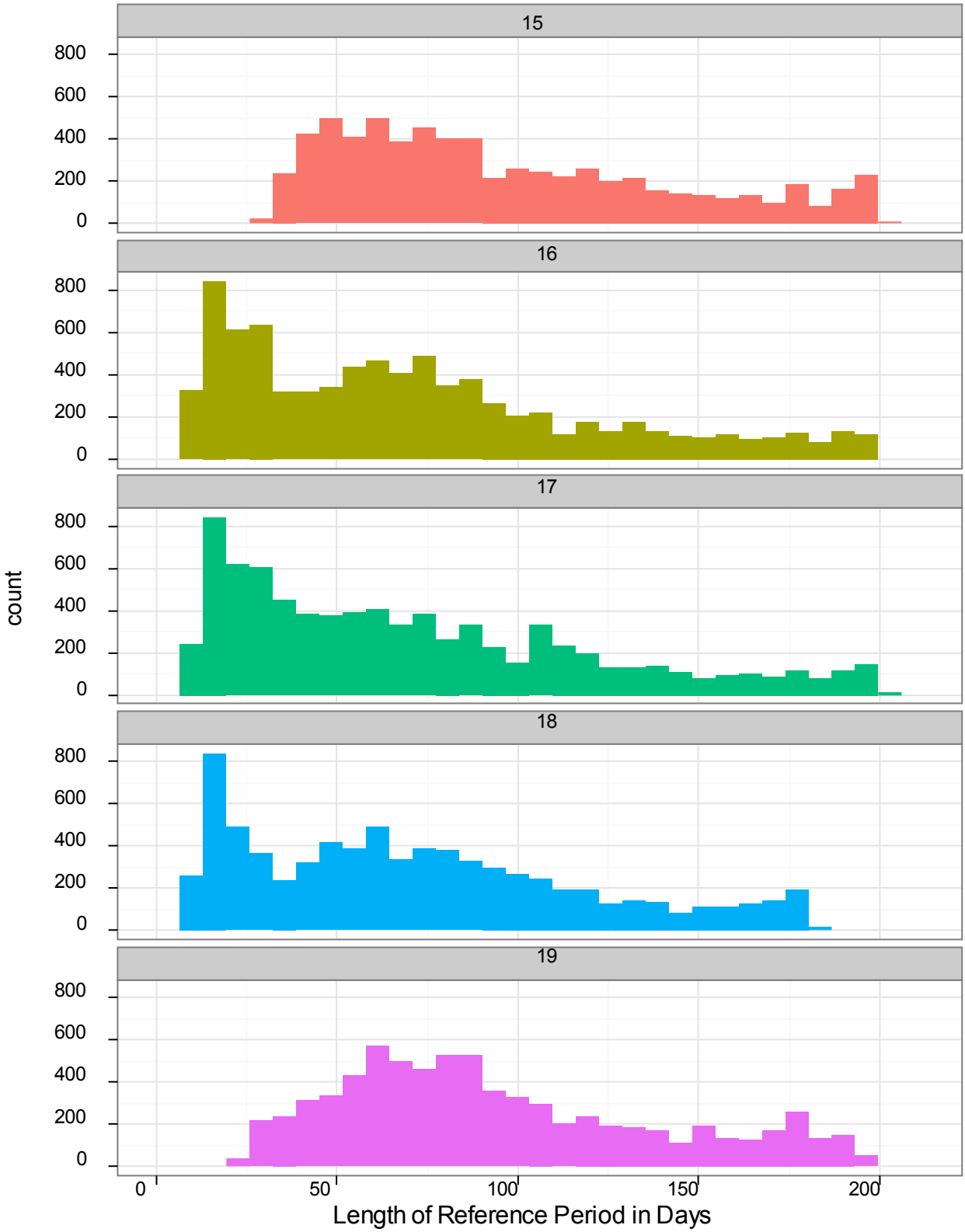

Figure 2: Mean Annualized Utilization Rate by Completion Week
(Non-cumulative)
Panels 15-19, Round 1

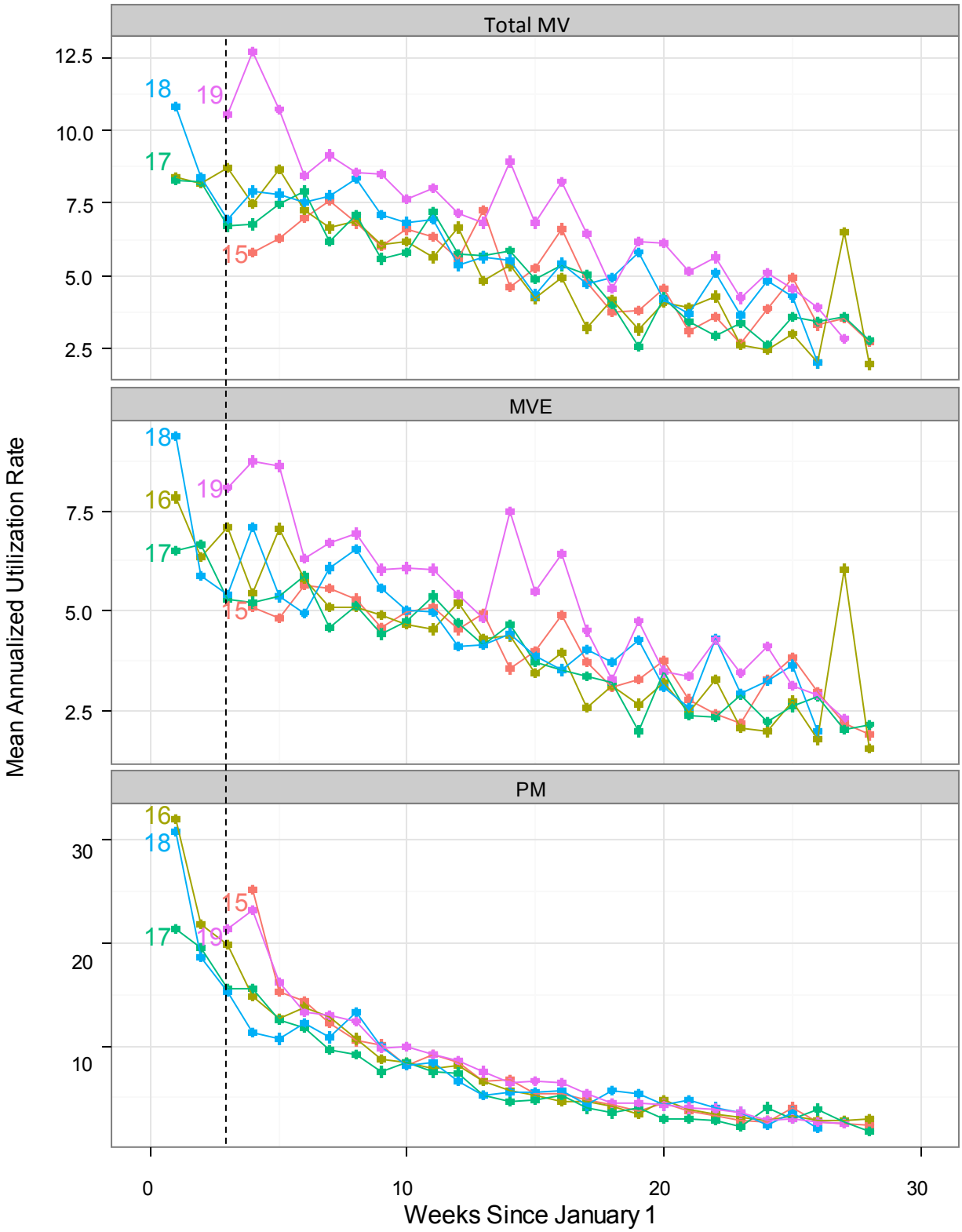

Figure 3: Total Reported Events Cumulative by Reference Week

Panels 15-19, Round 1