

# Considerations Regarding the Use of Global Survey Questions



Paul Beatty

National Center for Health Statistics

Prepared for the Consumer Expenditures Survey  
Method Workshop

December 8-9, 2010

# Example of a global question

---

This question is about moderate or strenuous physical activities you may have done at home or in your leisure time. By moderate or strenuous, we mean physical activities that lasted 10 minutes or longer, and caused at least some increase in heart rate or breathing. Please do not include physical activities done in any job for pay.

From [START DAY] to [END DAY], how much time did you spend doing moderate or strenuous physical activities, including yard work or other chores, walking for exercise or to get somewhere, or other exercise such as running, cycling, working out in a gym, or playing sports?

# Common problems from cognitive testing

---

- ❑ Too long and complicated
  - ❑ Probing revealed that some forgot, or never grasped, some elements
- ❑ Components not thought about or remembered in the same manner
  - ❑ Formal exercise often different than times when you happen to be physically active
- ❑ Response strategies often guesses or crude estimates
  - ❑ *Probing* revealed omissions and errors

# What's different about our challenge:

---

- ❑ Usually global questions are written by default, and the burden of proof is to show that smaller questions would lead to substantial improvements
- ❑ Here we are starting with smaller questions, and considering whether global questions would be just as good (or at least adequate given survey goals)
- ❑ The same *potential* pitfalls of global questions apply either way:
  - Comprehension: too long or complex
  - Combines disparate elements that are ideally remembered or estimated differently
  - Too large in scope to be reasonably estimated
  - Newly consolidated global questions will likely omit some details from the source questions– will there be sufficient prompts for respondents to consider all of these elements?

# Comparability of responses

---

- ❑ Will global questions formed from a set of specific questions produce the same results?
- ❑ Probably not.
- ❑ Specific questions are *likely* to produce higher estimates in aggregate than global questions (but not always). Possible reasons:
  - More specific questions offer better prompts—more complete reporting
  - Or, specific questions might not be completely distinct (double-reporting)

# Example: cheese questions

---

- ❑ During the last 30 days, how many times did you eat cheese, including cheese as snacks, and cheese in sandwiches, burgers, lasagna, pizza, or casseroles? Do NOT count cream cheese."
  
- ❑ The next questions are about cheese you have eaten in the last 30 days. Please do NOT include any cream cheese you may have eaten.
  - During the last 30 days, how many times have you eaten cheese on a sandwich, including burgers?
  - During the last 30 days, how many times have you eaten cheese in lasagna, pizza, casseroles, or mixed in with other dishes?
  - During the last 30 days, how many times have you eaten cheese as a snack or appetizer?

# Cheese consumption in 30 days, single vs. multiple questions

---

Single question: 13.9 (n=218)

Multiple questions: 19.0 (n=228)

Difference significant at  $p < .01$

- However, we cannot say for certain which version is more accurate

# Other comparisons between single and multiple cheese questions

---

- ❑ In behavior coding, “undesirable” behaviors appeared to be more common with single, global questions:

	Global	Spec1	Spec2	Spec3
Inadequate initial response	15.9	9.9	8.3	3.1
Probes used	13.7	7.8	6.3	2.1
Requested help/repeat	19.1	15.1	3.1	2.1

- ❑ However, when aggregating results of the specific questions, the advantage disappears
- ❑ Furthermore, time for administration is significantly longer for the multiple questions (51 seconds, as opposed to 28 seconds)



# How accurate are responses to global questions?

---

- ❑ How accurate are global questions:
  - In an absolute sense
  - Compared to the specific questions they could replace
- ❑ If specific questions are significantly closer to reality, and the higher accuracy is analytically critical, they might be worth the additional expense.
- ❑ If the global questions are more accurate, or any loss in accuracy is tolerable to us, then it makes sense to take advantage of their efficiency.

# Validation study: question domains

---

## Global:

## Decomposed:

- |                  |   |
|------------------|---|
| 1) Phys activity | (chores, walking, exercise)                                 |
| 2) Cheese        | (sandwich, in a dish, snack)                                |
| 3) Cereal        | (hot, cold)   |
| 4) Pasta & rice  | (pasta, rice)   |
| 5) Oil           | (cooking, add salad, add other)                             |
| 6) Dessert       | (ice cream, cookies/cake,<br>candy/chocolate, donut/muffin) |

# Validation study

---

- ❑ First phase– completion of three-day web diary of food consumption and physical activities
- ❑ Second phase– contacted for participation in split-ballot telephone survey (global and decomposed questions spread across two versions)
- ❑ Incentive of \$45 (later boosted to \$75) offered to those who completed both phases

# Expected data pattern

---

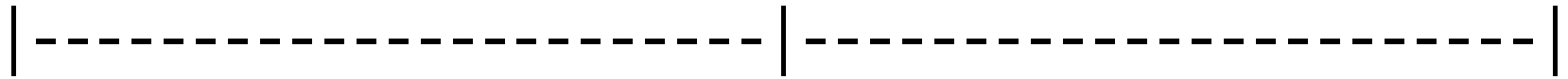
Low freq

High freq

G

D

X



X=diary report

G=global response

D=decomposed response

# Bias of global and decomposed questions

Domain	Question type	Bias to diary (%)	
Cheese	Global	-20.9	(p<.01)
	<b>Decomposed*</b>	16.6	(p<.1)
Physical activity	Global	-19.5	(p<.05)
	<b>Decomposed</b>	-14.6	(p<.05)
Oil	<b>Global*</b>	-16.9	(p<.05)
	Decomposed	-25.1	(p<.01)
Cereal	Global	21.4	(p<.01)
	<b>Decomposed*</b>	6.8	n.s.
Pasta and rice	<b>Global</b>	1.3	n.s.
	Decomposed	10.2	n.s.
Dessert	<b>Global</b>	-9.9	n.s.
	Decomposed	14.3	n.s.

# Bias of global and decomposed questions— second (conservative) coding

Domain	Question type	Bias to diary (%)	
Cheese	<b>Global*</b>	-10.8	n.s.
	Decomposed	22.5	(p<.05)
Physical activity	Global	-9.4	n.s.
	<b>Decomposed</b>	-0.4	n.s.
Oil	<b>Global*</b>	-16.9	(p<.05)
	Decomposed	-25.1	(p<.01)
Cereal	Global	40.6	(p<.01)
	<b>Decomposed</b>	29.5	(p<.01)
Pasta and rice	<b>Global*</b>	3.9	n.s.
	Decomposed	16.7	(p<.1)
Dessert	<b>Global*</b>	5.6	n.s.
	Decomposed	28.7	(p<.01)



# Overall assessment

---

- ❑ Determining the “real values” for validity checks is challenging
- ❑ But whichever version of real values you accept, the results are mixed: sometimes global questions do better and sometimes not as well as multiple questions.
- ❑ Considering all eleven comparisons made, decomposed questions performed better five times; global did better six times

# Making sense of the data

---

- Previous literature suggested the possibility of global questions being better than multiple questions, at least sometimes:
  - Variable effectiveness of global questions, depending upon regularity of the behavior and response strategy– global may be better for regular, estimated behaviors (Menon, 1997)
  - Multiple questions less accurate than global e.g., due to double-counting, for frequent, non-distinct behaviors (Belli et al, 2000)



# We didn't buy it

---

- ❑ For one thing, our decomposition of questions were based on observations of responses in the cognitive lab that suggested logical ways to separate questions
- ❑ Some decompositions in the literature arguably break the question into less memorable events
  - Washing hair in different domains (before a date, before a party, etc.)
  - Local vs. long distance phone calls
- ❑ Multiple questions *should work better* when
  - Constructed to reflect the way that behavior is actually encoded, and
  - Estimation is the likely response strategy
- ❑ So why didn't it always work in our case?

# Two examples of global questions

---

- ❑ From [day] to [day], how much time did you spend doing moderate or strenuous physical activities, including yard work or other chores, walking for exercise or to get somewhere, or other exercise such as running, cycling, working out in a gym, or playing sports?
- ❑ The next question asks about dessert foods, including ice cream, candy, chocolate, cookies, cakes and pies, and other sweet bakery items you might eat at breakfast or as a snack like doughnuts, Pop tarts, Danishes, and muffins. Please include anything that was low-fat or fat-free, but do NOT include sugar-free items. From [day] to [day], how many times did you eat these foods?

# Assessing global questions

---

- Is the accuracy of global questions likely to vary across domains?
  - Definitely
- Can responses to global questions be more accurate than responses to multiple, specific questions?
  - Possibly– depends how well the question lines up with the way information is organized in memory
  - If specific questions are optimally designed, moving to global questions may move to more generic estimation strategies and possible sacrifice of precision
  - But if specific questions are not optimally designed, global questions could theoretically invoke a better estimation strategy than their counterparts.

# Future research directions

---

- Given that the quality of global questions could vary considerably, data are needed to evaluate how well they match what respondents can report.
- Cognitive laboratory data (from probing or think-alouds):
  - What strategies tend to be used by respondents (estimation, counting)
  - Which question(s) match better the way respondents think and remember?
  - How adequate are their estimation strategies given our data needs?

# Future research– validation data

---

- ❑ Necessary for assessing accuracy
- ❑ Often very difficult and expensive to collect
- ❑ Not immune from quality problems and methodological challenges
- ❑ Key concerns with diaries:
  - Making sure that what they produce corresponds with the survey data
  - Well thought out coding procedures
  - Can be difficult to employ for longer reference periods
- ❑ Viable validation data for CES?

# Final thoughts

---

- ❑ Further research on the relationship between bias and frequency of the event being measured would be welcome
- ❑ As global questions cover wider conceptual terrain and longer reference periods, they are more likely to invoke estimation strategies
- ❑ Estimation is not *necessarily* less accurate, but the possibility of less precise data should be explored on a topic-by-topic basis