AUTHORS’ REPLY:
Drs Beasley and Hirooka’s attentive reads of our study report1 and insightful comments provide additional evidence that the health of family medicine research training is a priority area for academic family physicians, emeritus professor and fellow, alike. The partnership between the University of Wisconsin’s Departments of Family Medicine and Industrial and Systems Engineering is one example of the creative entrepreneurship that we believe will bolster family medicine research capacity and quality. Further, it is a model that may be readily adopted in a variety of settings.

We disagree with Dr Hirooka’s suggestion that we purport to qualify the research training efficacy described by our survey respondents. Rather, we attempt to quantify elements that represent how research is currently taught in family medicine fellowships. We do not compare the quality of training offered in research and non-research oriented fellowships, nor do we comment on the elements’ appropriateness or value. We do not generalize the financial challenges revealed by the data to the broader context of family medicine. We do cite recent studies showing the funding difficulties experienced by the family medicine research community2 and discuss the challenges of supporting research training. Fellowships purposefully offer unique educational opportunities. We agree with Dr Hirooka’s concern that this variance may be an important factor not explained in our study and attempted to address this in the limitations section.

Due to the exploratory nature of our research, it is premature to give specific research training recommendations. Instead, we chose to apply our data on research training in fellowships to further the conversation on how to foster research in family medicine, which has traditionally been devalued.3 In the discussion, we hoped to introduce new ideas that would fuel this important discussion. We are pleased, based on Drs Beasley’s and Hirooka’s responses, to have accomplished that aim and hope to engage in further productive dialogues.

AUTHORS’ REPLY:

More Reliability and Validity Data Needed for Resident Evaluation Tools

TO THE EDITOR:
I applaud the authors on their development of the Global Procedural Skills Evaluation (GPSE) for evaluation of resident performance on moderately complex office procedures for the process of instrument development through literature review and use of faculty focus groups, utilization of competencies (medical knowledge, communication, technical skills, etc), and development of an easily understood rating scale.1

The authors did not evaluate the validity of the instrument, however, which is a critical component of instrument design, especially those intended to evaluate resident skill performance. For example, can the instrument be reliably utilized by multiple evaluators, does the instrument measure what it is intended to measure, and does the instrument accurately distinguish novice and expert performance? This approach requires significant effort on the part of educators but also allows us to determine the value of our measurement tools and has been utilized in other studies published in Family Medicine.2,3

As educators, we commonly believe that we can accurately judge resident performance, but multiple studies show that distinguishing novice versus expert performance is dependent on the instrument used, methods of measurement, and faculty training to uniformly apply the instrument. Although the GPSE was developed through an iterative process, it does not yet have evidence to support its use over other instruments or global ratings. I would welcome a validity study of the GPSE from the authors to support its use as we are in short supply of reliable and valid tools for the evaluation of resident performance.

James Cooke, MD
Department of Family Medicine
University of Michigan

REFERENCES

AUTHORS’ REPLY:
Dr Cooke raises an important point that we addressed only briefly in our paper, namely the dearth of performance assessment tools with well-established psychometric features such as reliability and validity.

We do plan to examine some traditional psychometric features of the Global Procedural Skills Evaluation (GPSE). For example, we can use videorecorded