Methods Report for the Project on the Good Physician

Sampling: Data are from a 2011 survey of third-year medical students. Using a stratified, two-stage cluster sample design we selected 24 of the US allopathic medical schools, and then randomly selected 960 students from those schools in order to obtain a nationally representative sample population of schools and students.

In Sept 2009, samples were drawn from the American Medical Association Physician Master File. The frame is a near complete listing of medical students pursuing MDs and DOs in the US and its territories. We limited our samples to medical students coming from 133 U.S. allopathic (MD) medical schools (mean class size = 140). To ensure a nationally representative sample, we used a systematic sampling strategy based on probability proportional to size of school and implicit stratification. This yielded 24 medical schools. This sampling strategy is summarized below:

Sample Design

We decided to construct a two-stage cluster sample of students within schools in order to have a nationally representative sample of students and the ability to examine school effects. In step 1 we implicitly stratified all medical schools by sorting students them by census region, public/private status, Social Mission ranking and whether the school had a Gold Humanism program. In step 2 we selected schools using a probability proportional to size strategy. We decided in advance that we had resources to survey 960 students. To implement a strategy by which students in the sample would be able to rate each other we decided to select 40 students within each selected school. The number of desired students within schools also took account of estimates of response rates. Thus, the number of selected schools was equal to the number of desired students, 960, divided by the number of desired students within schools, 20, or 24.

We next calculated a sampling interval, k, equal to the 3022 which is number of total students in our population (72528, based on AMA data) divided by the desired number of schools (24). We selected a random number between 1 and k as our start value and selected schools containing the k_ith student, where i is an integer from 1 to 24. This procedure did not result in any duplicate school selections. Finally, we selected 40 students from each selected school using simple random sampling.

Weighting

We constructed a base weight for each student that was a ratio of an estimate of the total number of third year medical students divided by the total number of students selected. Four adjusted weights were constructed and should be applied based on the portion of the sample the analyst is interested in using. The first adjusted weight, r1_all_finwt, is to be used by analysts who are interested in using round 1 questionnaire data for the entire obtained sample, which included mostly third year student but also some second and fourth year students (obtained sample, n=605). This weight was constructed by taking the base rate and multiplying it by the reciprocal of the male and female within-school response rates for the entire round 1 obtained sample. The second adjusted weight, r1_thrdyr_finwt, is to be used by those wishing

only to analyze round 1 data of third year students only (n=564). Although we based our sample selection on an AMA indicator of student third year status this indicator did not match student self-report for a small number of cases. Similar to the procedure used for the first weight, we calculated as an adjustment factor the reciprocal of the male and female within-school response rates and applied it to the base weight of only those students who indicated that they were in their third year.

The third and fourth adjusted weights were constructed for the analysis of round 1-round 2 panel data. For the third weight, the same adjustment factor was constructed and applied to the base weight for all students who had completed both the round and the round 2 questionnaire (n=500). The fourth weight was calculated by applying a similarly constructed adjustment factor to the base weight for students who indicated that they were in their third year and who also completed both the round 1 and round 2 questionnaires (n=474).

Questionnaire Development: Questionnaire topics were developed and pretested through cognitive interviews with a small group of medical students throughout the U.S. who tested the survey instruments. Cognitive interviews on selected questions were conducted with a small group of physicians from around the country who treated terminally ill patients. After IRB approval, the survey was launched in January 2011, according to the following schedule:

Time	Activity – Questionnaire Sample[1] (n=960)
• Week 1:	Pre-notification letters/emails sent along with requests for address verification
• Week 2:	First mailing with survey, incentive and postcard with link to webpage sent
• Week 3:	Reminder post card follow-up sent
• Week 4:	
• Week 5:	Second mailing and email requests with online version of survey sent;
• Week 6:	
• Week 8:	Third mailing and email requests with online version of survey sent;

Participants received a letter in advance informing them of the nature of the study and the importance of their responses and asking for confirmation of mailing address. One week later, they received an envelope containing a more detailed explanation of the study and a paper survey instrument with an upfront incentive of $5 bill. At that time they were also invited to visit our study webpage online, provide their email address and respond to more open-ended questions regarding their experiences through a secure, password-protect webpage. For our third wave, we offered non-respondents an additional $5 gift card incentive to complete the

survey. In addition, respondents also had the option to complete an online version of the mailed questionnaire since some respondents may only be reached via email rather than by postal mail.

The follow-up survey was launched in September 2011 (when the third years became fourth years) among those who completed the first survey. This follow-up survey contained a $10 bill. We also followed a similar procedure as above for the follow-up survey. At the end of this second questionnaire, students were invited to participate in future studies in which they were offered the opportunity to complete peer ratings and/or further qualitative interviews (separate IRB application was submitted for these studies).

At the conclusion of the study, we sent participants a 1 page follow-up mailed/online survey update to confirm medical students' recent residency choices as well as confirm with study participants that they still want to remain in the study after graduation by submitting their updated contact information.

Mode of Data collection: All questionnaires were administered by mail and email.

Data collection Schedule:

First Wave - Advance Letter Mailed: January 17, 2011

First Wave - First Questionnaire Mailed: January 28, 2011

First Wave - Post Card Reminder Mailed: February 4, 2011

First Wave - Second Questionnaire Mailed: March 18, 2011

First Wave - Third Questionnaire Mailed: April 29, 2011

Second Wave - Advance Letter mailed: September 14, 2011

Second Wave - First Questionnaire Mailed: September 21, 2011

Second Wave - Post Card Reminder Mailed: September 28, 2011

Second Wave - Second Questionnaire Mailed: November 4, 2011

Second Wave - Third Questionnaire Mailed: December 12, 2011

In addition, intermediate mailings were conducted as we received better information about mailing addresses from returned letters and questionnaires.

All students in the target sample received a $5 cash incentive in the first questionnaire mailing during wave 1 and $10 during wave 2. Sample members who were nonrespondents by the third mailing received a notification with their third questionnaire that they would receive an additional $5 during wave 1 for completing the questionnaire and returning it to us. Nonrespondent by the third mailing in wave 2 received no additional cash incentive.

Data Processing: Questionnaires were receipted as soon as they were returned. Each questionnaire was then double entered into an Excel spreadsheet. After the data entry was completed the two versions were compared with one another using and Excel function and discrepancies were checked against the hard copy.

Response rates: Of the 960 cases fielded, (4.3%) were declared out of scope because surveys were returned after attempts were made to contact respondents at two different addresses or we had received information that the medical student had dropped out of school, or were inaccurately identified as third years in the AMA masterfile (41 cases). Completed cases, response rates and refusal rates for the in-scope cases are shown in the following table.

1st Survey Completed	564/919	Adjusted Response Rate	61%
2nd Survey Completed	474/564	Adjusted Response Rate	84%

File construction: After the data were cleaned Stata files were constructed with variable names matching the question numbers on the questionnaire and variable and value labels also watching matching the questionnaire.

Case Weighting: Case weights were employed to reflect sources of variance associated with the sample design and to adjust for potential nonresponse bias. Groups were compared using chi-square analysis with significance determined by alpha = .05.

Survey Items: See attached documents for questionnaires (time 1 and time 2) and scales used in our survey.

Qualitative Interviews: We conducted a qualitative interview study on medical students who recently completed their fourth year of medical school (with one exception, who was at the end of their third year). Participants were recruited among respondents of a nationally-representative survey of third-year medical students conducted in 2011. Survey respondents were asked whether they would be willing to participate in a follow-up qualitative phone interview. Those who volunteered were divided into 4 categories based on whether or not they received the Gold Humanism Award (yes/no) and whether they reported high or low "life meaning" on the survey, derived from a short form of the Presence of Meaning Subscale in the Meaning of Life Questionnaire. This resulted in four sampling frames: (1) Low life meaning, Gold award; (2) High life meaning, Gold award; (3) Low life meaning, no Gold award; and (4) High life meaning, no Gold award.

Phone interviews took place during April and May 2012 when the fourth-year students had just matched for residency and were about to start their internships in July 2012. The interviewer, who was blinded to the meaning of the various sampling frames, was instructed to interview all volunteers who fit the criteria for sampling categories 1 and 2, and as many respondents in the remaining two groups as needed until thematic saturation was reached. All subjects provided informed consent at the time of the interview and agreed to be audio-recorded. The study received ethics approval from the University of Chicago Institutional Review Board.

The interviewer used a semi-structured interview guide (see attached), probing for more information when necessary. Of note, respondents were asked whether they felt burned out at any point in their medical training, a question whose validity is consistent with the much longer Maslach Burnout Inventory (MBI). A single research assistant conducted all the interviews, ensuring consistency across respondents. Each interview lasted about an hour, and participants were compensated with a $50 check. A total of 21 phone interviews were conducted with medical students nationwide. Approximately half of respondents were women (57.1%). 17 respondents (81.0%) reported experiencing burnout at least once in their medical school career.

Interviews were transcribed and de-identified. The transcripts were then analyzed using NVIVO 11 (QSR International, Burlington, MA) by a separate group of researchers who had no contact with the participants, nor with their reported level of burnout on the survey. Researchers first coded individually based on the interview guide. An iterative process followed, collectively adding and modifying nodes and coding the transcripts line-by-line. A consolidated list of codes was discussed and finalized by all authors. Researchers met together and re-coded all transcripts through a consensus process.

We approached data analysis with a post-positivist lens, by acknowledging how our personal subjectivities could affect our efforts to produce objective scientific knowledge. One researcher has extensive sociological training and experience in qualitative data analysis, and trained the rest of the research team to use these methods rigorously. Together, our expertise allowed us to approach the data with guided curiosity, and our blinded, iterative approach to analysis helped protect against potential biases. Our research conformed to the standards for qualitative research published by Academic Medicine.

Peer Ratings: See attached document for example of our peer rating system

Comparison of our National Sample with data from AAMC on Total Enrollment in 2010

Project on the Good Physician Survey		
Black or African American	**10.3**	57
American Indian or Alaska Native	**0.4**	2
Asian	**22.0**	136
White	**57.5**	349
Hispanic	**5.4**	32
Other	4.5	27
Total	100.0	603

PGP survey		
Women	46.7	282
Men	53.3	322
Total		604

AAMC 2010 total **(Note: respondents allowed to indicate multiple categories)**		
Black or African American	**7.0%**	5,548
American Indian or Alaska Native	**0.8%**	654
Asian	**22.0%**	17,375
Native Hawaiian or Other Pacific Islander	0.3%	240
White	**60.1%**	47,525
Other Non-Hispanic or Latino Race	0.1%	104
Mexican American	**2.6%**	2,058
Puerto Rican	**2.1%**	1,646
Cuban	**0.7%**	585
Other Hispanic or Latino	**2.8%**	2,219
Foreign	1.7%	1,309
No Race Response or Unknown Citizen	2.8%	2,242
Unduplicated Total Enrollment		79,070

AAMC 2010 total		
Women	47.4%	37,499
Men	52.6%	41,571
Total		79,070

Source: AAMC Data Warehouse: Applicant Matriculant File as of October 4, 2010