Numeracy
Advancing Education in Quantitative Literacy

Volume 9 | Issue 1 Article 5

2016

Clinician Numeracy: The Development of an Assessment Measure for Doctors

Anne A. Taylor
Keele University, Staffordshire, England, a.taylor1@keele.ac.uk
Lucie M. Byrne-Davis
University of Manchester, Cheshire, England, lucie.byrne-davis@manchester.ac.uk

Recommended Citation
Taylor, Anne A., and Lucie M. Byrne-Davis. "Clinician Numeracy: The Development of an Assessment Measure for Doctors." *Numeracy* 9, Iss. 1 (2016): Article 5. DOI: http://dx.doi.org/10.5038/1936-4660.9.1.5

Clinician Numeracy: The Development of an Assessment Measure for Doctors

Abstract

Low numeracy in doctors poses serious risks to patient safety because inaccurate drug dose calculation may lead to under-treatment or overdose, while erroneous data interpretation affects medical decision making. Most research on numeracy in healthcare focuses on health numeracy in patients, while research on numeracy in doctors, "clinician numeracy", is limited, partly due to the lack of a suitable assessment measure. We developed a new assessment, the Medical Interpretation and Numeracy Test (MINT), to assess clinician numeracy. The MINT tests computational, analytical and statistical constructs, using a combination of questions validated in other studies, and new test material specifically designed for doctors. We recruited 135 recently qualified doctors attending a teaching session on clinical decision making and risk communication to take our test. Psychometric analysis indicates that the MINT is a valid and reliable measure of clinician numeracy, with good internal-consistency reliability. Correlation with other numeracy/health numeracy tests varied greatly: this variation is understandable in view of the limited scope of many existing assessments that test only single constructs of numeracy/health numeracy. We conclude that the MINT provides a broad overview of clinician numeracy and can be a useful new assessment measure. Because of its important implications for patient safety, further research is needed to investigate clinician numeracy in doctors and other healthcare professionals, and to address and remediate deficiencies.

Keywords

health Numeracy, physician numeracy, clinician numeracy, numeracy assessment

Cover Page Footnote

Anne Taylor is an anesthesiologist at the University Hospitals of North Midlands NHS Trust, and a hospital dean at the School of Medicine, Keele University. She has an MSc in Medical Education from the University of Manchester, and is involved in undergraduate and postgraduate medical education. She researches numeracy in medical practice.

Lucie Byrne-Davis is a lecturer in assessment and psychometrics at Manchester Medical School, University of Manchester. She is a registered Health Psychologist, a Chartered Psychologist and a Fellow of the Higher Education Academy. She has an MSc in Health Psychology from the University of Bath and a PhD in Health Psychology from University of Bristol. She researches assessment of constructs related to health professional education, training and practice.

Introduction

Health numeracy, which encompasses the ability to use quantitative data of all kinds in a medical context, is essential to optimizing healthcare. People with low health numeracy are less likely to access healthcare information; they struggle to understand data related to the risks and benefits of medical intervention, are less able to manage chronic health conditions, and are less likely to comply with treatment (Schwartz et al. 1997; Estrada et al. 2004; Apter et al. 2006; Gigerenzer et al. 2007; Huizinga et al. 2008; Woloshin et al. 2008; Reyna et al. 2009). Healthcare professionals must access, interpret and communicate medical information accurately; this necessity is fundamental both to shared decision-making and to informed consent. Moreover, healthcare professionals must take particular care in consultations with patients who have low health numeracy, to ensure that information is understood correctly (Weiss et al. 2005; Rothman et al. 2006; Fagerlin et al. 2007).

Low numeracy affects people of all educational levels (Lipkus et al. 2001; Golbeck et al. 2005; Gazmararian et al. 2005; Peters et al. 2007; Reyna et al. 2009), including healthcare professionals. Nurses and pharmacists are known to struggle with drug dose calculations, posing a threat to patient safety in relation to the preparation and administration of drugs, intravenous (IV) fluids and nutritional supplements (Latif and Grillo 2002; Oldridge et al. 2004; McMullan et al. 2010; Wright 2010; Hegener et al. 2013). The term “health numeracy” is generally used in relation to patients. Numeracy in healthcare professionals, including doctors, is covered by the term “clinician numeracy”, defined by Caverly et al. (2012) as “the ability to use numbers and numeric concepts in the context of taking care of patients”.

Doctors are generally assumed to have high numeracy (Rowe et al. 1998). Evidence from studies across the globe, however, demonstrates that some medical students and doctors have difficulty with drug dose calculation (Rolfe and Harper 1995; Rowe et al. 1998; Selbst et al. 1999; Oldridge et al. 2004; Wheeler et al. 2004; Simpson et al. 2009; Harries et al. 2013) and understanding biostatistical data (Sheridan and Pignone 2002; Ghosh and Ghosh 2005; Gigerenzer et al. 2007; Windish et al. 2007; Gigerenzer and Muir Gray 2011; Wegwarth et al. 2011, Johnson et al. 2014). This reality has important implications for safe, effective patient care. Calculation of the correct dose of drugs or IV fluids for an individual patient requires consideration of several factors including the patient’s age and weight, the pharmacokinetic properties of the drug, the presence of medical conditions such as liver or renal disease which may alter metabolism, and the coprescription of other medications. Particular care is needed with drugs administered by the IV route: these drugs are often provided in solution, and may need dilution. Furthermore, because the concentration of such drugs involves a ratio (epinephrine 1:1000), percentage (lidocaine 2%), or mass per unit volume (atropine 600 μg/ml), calculation of the correct dose may require several steps (Appendix, final sample question). Miscalculations can lead to serious patient harm: a decimal-point error can be life-threatening in healthcare, since it results in a tenfold change in the drug dose. Pediatric patients are particularly susceptible to errors in drug dose calculation (Hughes and Edgerton 2005).

Clinician numeracy (CN) in doctors merits investigation to establish whether, and to what extent, it may have an impact on medical practice. No standard of numerical competence has been set for doctors, nor is there any specific assessment. We researched the literature for a suitable assessment measure, finding several tests designed for the general public, patients, and healthcare professionals (Tables 1 and 2). Some tests aimed to assess basic numeracy in the general public (Williams et al. 1995; Baker et al. 1999; Weiss et al. 2005), while others investigated statistical literacy (Schwartz et al. 1997; Schwartz et al. 2005; Peters et al. 2007; Cokeley et al. 2012, Weller et al. 2013), and a few focused on numeracy and the management of a particular disease (Estrada et al. 2004; Apter et al. 2006; Huizinga et al. 2008). Although the Numeracy Understanding in Medicine Instrument (Schapira et al. 2012) tests a range of constructs, its test material—in common with other assessments developed for patients—was very simple, as observed by Vacher and Chavez (2009). None of these tests were suitable for assessing clinician numeracy. Tests designed for clinicians were limited in content, either assessing ability to calculate drug doses (Rolfe and Harper 1995; Rowe et al. 1998; Oldridge et al. 2004; Wheeler et al. 2004; Simpson et al. 2009; McMullan et al. 2010; Harries et al. 2013); or understanding of biostatistics and data relating to risk (Sheridan and Pignone 2002; Windish et al. 2007; Wegwarth et al. 2011; Johnson et al. 2014). While numeracy tests developed by Sikorskii et al. (2011) for students entering university were of an appropriate level for doctors, many items tested familiarity with advanced mathematical functions unnecessary in clinical practice. Therefore no existing test met the needs of our research, which was to obtain a broad overview of clinician numeracy.

The aim of this study was a) to develop a numeracy test appropriate for doctors, and b) to establish its validity by: (i) careful content development and systematic blueprinting, and (ii) testing theories regarding the properties of a valid test of clinician numeracy, using evidence from existing assessments (Table 3). We describe the development and validation of the Medical Interpretation and Numeracy Test (MINT).

DOI: http://dx.doi.org/10.5038/1936-4660.9.1.5

Table 1.
Health Numeracy Tests in the General Public and Patients

Study	Participants	Test name	Scope of test and important findings
Williams et al. (1995)	2659 Patients	Test of functional health literacy in adults (TOFHLA).	Overall health literacy: 50 literacy and 17 numeracy questions Importance of health literacy to "receiving proper health care"
Schwartz et al. (1997)	287 Female veterans	None, sometimes referred to as the Schwartz test	Understanding probability 3 key probability questions Association between numeracy and risk comprehension
Baker et al. (1999)	211 Patients	Shortened TOFHLA (S-TOFHLA)	Overall health literacy: 36 literacy and 4 numeracy questions Reliability and validity similar to those of TOFHLA
Lipkus et al. (2001)	463 Highly educated general public	Expanded numeracy scale	Risk comprehension 10 questions, including the Schwartz test Highly educated people performed poorly on simple numeracy questions
Sheridan and Pignone (2003)	357 Patients	-	Risk comprehension 4 questions, including the Schwartz test Low numeracy in study group Numeracy affects risk comprehension
Estrada et al. (2004)	143 Patients	-	Numeracy and disease management 6 questions, including the Schwartz test Low numeracy prevalent and associated with poor disease control
Schwartz et al. (2005)	178 Patients General public	Medical data interpretation test (MDIT)	Risk comprehension in healthcare. 20 questions. More numerate participants scored better on the MDIT
Weiss et al. (2005)	500 Patients	Newest vital sign test (NVS)	Short test of health literacy based on understanding a nutritional label. 6 questions: 2 literacy and 4 numeracy
Apter et al. (2006)	73 Patients	Asthma numeracy questionnaire	Numeracy and disease management 4 questions related to asthma control Importance of context in test material
Fagerlin et al. (2007)	287 Patients	Subjective numeracy scale (SNS)	8 questions Correlation between subjective and objective numeracy tests
Peters et al. (2007)	303 General public	-	Numeracy and framing of data 14 questions, including the Schwartz test Importance of clear data presentation
Huizinga et al. (2008)	398 Patients	Diabetes numeracy test (DNT)	Numeracy and disease management 43 items More numerate participants scored better on the DNT
Sikorskii et al. (2011)	3701 Students	-	Numeracy in university students 3 tests, each with 14-17 questions Useful assessment and guide to remediation for university students
Schapira et al. (2012)	1000 General public Patients	NUMi (Numeracy Understanding in Medicine instrument)	Testing numeracy across all constructs 20 question test Broad overview of numeracy
Weller et al. (2012)	1970 General public Psychology students	Abbreviated Numeracy Scale (ANS)	Numeracy and decision making 8 questions, including the Schwartz test ANS useful in predicting risk judgements

Table 2.
Clinician Numeracy Tests

Study	Participants	Scope of test and important findings
Rolfe and Harper (1995)	150 Hospital doctors (all grades)	Understanding of drug ampoule labelling Many doctors have difficulty with calculations involving conversion between different labelling formats (ratio, percentage, mass concentration)
Rowe et al. (1998)	64 Trainee hospital doctors (pediatrics)	Calculation of drug doses in pediatric practice Drug calculation errors common, potentially life-threatening Need for education and assessment
Sheridan and Pignone (2002)	62 Medical students	6 questions including the 3-item Schwartz test Numeracy affects risk comprehension Framing effect evident
Oldridge et al. (2004)	111 Healthcare professionals	6 questions testing drug dose calculation Calculation errors common in all groups, including doctors and medical students
Wheeler et al. (2004)	2975 doctors	6 questions related to drug ampoule labelling Converting between different formats causes difficulty
Windish et al. (2007)	277 Trainee doctors	20 questions testing understanding of biostatistics Competence generally insufficient to interpret research
Simpson et al. (2009)	190 Doctors	6 questions testing drug dose calculation and conversion between different labelling formats including items from Rolfe & Harper and Wheeler Converting between different formats causes difficulty Self-assessment accurate
Hanoch et al. (2010)	100 Medical students and trainee doctors	Participants with higher numeracy on the Lipkus scale were better able to select optimal Medicare plans for patients
Anderson et al. (2011)	203 Senior doctors	11 questions: 3-item Schwartz test, and SNS Association between SNS, but not Schwartz, on use of quantitative data in patient consultations
McMullan et al. (2011)	273 Nurses Student nurses	15-20 questions testing numeracy and drug dose calculation Poor numeracy overall; older participants performed better
Wegwarth et al. (2012)	412 Doctors	8 questions related to screening data Doctors have limited understanding of screening statistics
Cokely et al. (2012)	51 Trainee physician's assistants*	4-item Berlin Numeracy Test of statistical numeracy Importance of statistical numeracy in decision making
Harries et al. (2013)	364 Medical students	4 questions testing drug dose calculation Errors common: only 23% competent on first testing
Johnson et al. (2013)	308 medical students and 50 trainee doctors	4 questions: 3-item Schwartz test, and a risk comprehension question based on a clinical case

* Cokely et al. describe the use of the Berlin Numeracy Test in a total of 5336 participants, 51 of whom have a healthcare background as trainee physician's assistants.

DOI: http://dx.doi.org/10.5038/1936-4660.9.1.5

Table 3.

Assessment of a Clinician Numeracy Test

Properties of a Clinician Numeracy test	Assessed by
Correlation with other Health/Clinician Numeracy tests	Participant scores on test items correlate with their scores on items previously validated in other tests
Score varies with level of education	Participant scores are similar to those of doctors and medical students, but better than those of the general public, schoolchildren and other students
Score varies with level of mathematics instruction	Participants with A-level mathematics have higher scores than those without
Performance depends on item difficulty	Participants score better on questions deemed "easy" than on those considered "difficult" Participants score better on computational questions than on analytical or statistical ones

Methods

We designed a 43-item assessment, the MINT, to measure a doctor's ability to apply mathematical principles to solve common clinical problems. Thirty-one (72%) questions came from existing numeracy tests, and we developed a further 12 (28%) questions.

Development of the MINT

We based our assessment on the health numeracy framework developed by Golbeck et al. (2005), which aligns well with the quantitative skills required in clinical practice: *computational health numeracy* is important for prescribing; *analytical health numeracy* relates to data interpretation and medical decision making; and *statistical health numeracy* is necessary to understand and communicate risk, as well as for treatment selection and to practice evidence-based medicine (Table 4). There is some overlap between constructs: e.g., analytical health numeracy is necessary for many computational and statistical tasks.

The MINT aims to assess a doctor's numeracy in relation to everyday medical practice. Clinical tasks requiring numeracy include prescribing drugs and intravenous fluids, interpreting test results, evaluating different treatment options, and using clinical guidelines. We developed a blueprint that mapped test items according to clinical skill and health numeracy construct (Fig. 1). Most MINT questions tested more than one construct; for blueprinting and analysis, such items were assigned to the single most appropriate category.

Table 4.

Numeracy in Healthcare: constructs, competence and clinical application

Construct	Areas of competence (Golbeck et al. 2005)	Clinical application
Computational numeracy	Basic mathematical skills Simple manipulation of numbers, quantities, items, or visual elements in a health context e.g understanding information on a nutritional label	Calculation of drug doses Management of fluid and nutritional regimens Use of formulae in medicine Advising patients on disease management e.g. anticoagulant therapy, blood glucose control in diabetes
Analytical numeracy	Making sense of information Understanding graphs and other data displays Higher functions e.g. inference, estimation, proportions, percentages, frequencies	Interpreting medical test results and data regarding different treatments Understanding drug pharmacokinetics Estimation (cross-checking) of calculations e.g. drug doses Diagnostic skills Managing disease processes Advising patients on disease management Clinical decision making and treatment selection
Statistical numeracy	Understanding basic biostatistics including probability statements Ability to compare different scales (probability, proportion, percent) Ability to critically analyse quantitative information e.g. life expectancy or risk Understanding concepts such as randomisation and blinding	Understanding information on risk presented in different formats Risk communication Interpreting medical data Clinical decision making and treatment selection Practicing evidence based medicine

Content validity was established by careful selection of test items, by colleague review, and by pilot testing. Although none of the existing health numeracy tests were suitable in their entirety for doctors, we considered that several questions validated in previous tests were appropriate for the MINT. Most of these had been developed by experts in the fields of mathematics, statistics and/or medicine, and all had been subject to rigorous scrutiny. We included 31 such questions in the MINT. In addition, we developed 12 new items to cover specific clinical topics and to ensure that a variety of data display formats was incorporated into the test.

Our aim was to obtain an overview of numeracy in doctors and medical students as it pertains to normal clinical practice. Clinical tasks requiring numeracy vary greatly in their level of difficulty; therefore, so also do MINT items. MINT questions came from diverse sources, ranging from those designed for schoolchildren to those written for doctors. Five questions were for schoolchildren: two from tests for UK 11 year-olds (KS2 2011), and three from the Programme for International Student Assessment (PISA) for 15 year-olds

(OECD 2003). Nine questions were taken from health numeracy tests designed for patients or the general public: four from the Newest Vital Sign (NVS) test (Weiss et al. 2005), an assessment of health literacy that requires interpretation of a nutritional label; three from a study linking numeracy and risk comprehension (Schwartz et al. 1997); one from the Diabetes Numeracy Test (DNT) (Huizinga et al. 2008); and one from a test devised for educated consumers (Peters et al. 2007). Two questions came from a sample admissions test to a UK nursing school (KCL 2013), and six from a study assessing data comprehension in U.S. medical students (Sheridan and Pignone 2002). A further nine questions were taken from tests developed for U.S. university entrants (Sikorskii et al. 2011). Questions not originally based in healthcare were rewritten to a medical setting, given the importance of contextualizing assessments (Schuwirth and Van der Vleuten 2011).

Intended Learning Objective (ILO)	**No. (%) of questions which measure this ILO.**	**Item no. and HN construct tested**		
		Computational	**Analytical**	**Statistical**
Prescribing: calculations related to drug/fluid dose, concentration, and preparation	6 (14)	5, 12, 30, 40, 43	6	
Data interpretation: written information e.g. comparing treatments	8 (19)	27	8, 18, 37	3, 7, 17, 36
Data interpretation: information presented in tables, charts and graphs	13 (30)	20, 21, 22,23	2, 24, 25, 28, 33, 34, 35, 41, 42	
Probability, including conversion between frequency, proportions and percentages	7 (16)			10, 11, 16, 19, 26, 31, 39
Clinical problem-solving	9 (21)	1, 4, 38	9, 13, 29, 32	14, 15
No. of questions	**43 (100)**	**13**	**17**	**13**

Figure 1. MINT blueprint grid

Item difficulty was ranked 1–5, according to the group for whom the questions were originally intended (Table 5). The easiest questions were those designed for primary schoolchildren, and then in ascending order, those aimed for the general public, applicants to nursing school, secondary schoolchildren, and students entering university. However, the 12 new MINT questions and the six items previously used for U.S. medical students varied greatly in difficulty. Both authors analysed these questions independently, reaching immediate agreement on six. One author graded seven of the remaining questions higher, and five lower, than the other rater. We discussed each disputed question until agreement was reached. These questions were all written for doctors or medical students, and differences in grading may be explained by the fact that one of us (AT) is a doctor, while the other (LBD) is not. This argument is supported by the work of Levy et al. (2013) on the importance of context in assessing numeracy.

Table 5.
Source, target group and level of difficulty of MINT items

Source	Number of test items (number included in the MINT)	Target group	Item number	Level of difficulty (1-5)
KS 2	48 (2)	Primary school children	28, 42	1
Schwartz et al.	3 (3)	General public	10, 16, 19	2
Weiss et al.	6 (4)	General public	20 – 23	2
Huizinga et al.	43 (1)	Patients	1	2
KCL	15 (2)	Entrants to nursing	4, 38	3
OECD (PISA)	100 (3)	15 year olds	24, 25, 32	4
Sikorskii et al.	33 (9)	Entrants to university	3, 14, 15, 29, 31, 33, 34, 39, 41	5
Peters et al.	15 (1)	Educated population	26	5
Sheridan and Pignone	8 (6)	First year medical students (U.S.)	7, 8, 17,18, 36	3
			37	5
New MINT questions	12	Doctors	2, 9	1
			27	2
			5, 12, 30, 35, 40, 43	3
			6, 13	4
			11	5

We used an MCQ format with five answer options. When five different plausible answers could not be provided—e.g., treatment comparison questions comparing two treatments—the option "Don't know" was given. We used a standard scoring system with one mark for a correct answer, and zero marks for incorrect answers or unanswered questions. Participants were not allowed to use calculators.

Clinical colleagues reviewed the MINT to ensure that it was readable, understandable, relevant, and clinically accurate. Following their feedback, we made some minor adjustments to the test, and then piloted it on a convenience sample of 14 third-year medical students, under examination conditions. Time taken to complete the test ranged from 35 to 60 minutes. All students agreed that the test material was relevant and representative of clinical practice at their level or slightly higher. They considered it a fair test, at an appropriate level of difficulty. Students thought that the number of questions was reasonable and that one hour was sufficient to complete the test.

A range of supplementary questions covered areas such as the highest examination level to which participants had studied mathematics and the grade achieved, as we hypothesized that this history would influence performance. We also explored attitudes towards mathematics, including whether participants considered mathematical ability important for doctors, or thought it should be used when selecting candidates for medical school. Finally, we asked participants to self-assess their ability in math, because some evidence suggests a correlation with objective measures of numeracy (Fagerlin et al. 2007; Simpson et al. 2009).

Participants

Foundation Trainees (FTs) are doctors within their first two years of practice following graduation, enrolled in the UK Foundation Programme (UKFP). FTs were recruited from four hospitals: University Hospital of South Manchester NHS Foundation Trust, Burton Hospital NHS Foundation Trust, Mid Staffordshire NHS Foundation Trust, and University Hospital of North Staffordshire NHS Trust. Recruitment to the study is shown in Figure 2. Attendance at FT teaching sessions is variable, and evidence suggests that less numerate individuals may self-select out of numeracy tests (Sheridan and Pignone 2003; Simpson et al. 2009). However, attendance at our sessions was similar to mean attendance for FT teaching on each site.

Standard exclusion criteria in numeracy tests include poor vision, lack of fluency in English, and cognitive dysfunction. However, none of these factors was applicable to our study group, all of whom had passed the criteria for entry to the UKFP and practice as doctors.

Procedure

The study took place between November 2013 and May 2014. The test was administered during a training session on “Clinical decision making and Risk communication”, approved as part of the FT education programme by Foundation Programme Directors at all sites. Ethical approval for the study was granted by the University of Manchester Research Ethics Committee panel; NHS Research and Development organizational approval was received from Health Education North West.

Statistical Analysis

Data was analysed using Microsoft Excel and IBM SPSS. We used bivariate analyses to determine the associations between MINT score and performance on different subsets of our test, and to establish the relationship between MINT score, previous achievements in math, and subjective estimations of competence. We used chi-square tests for categorical variables and Fisher’s exact test when numbers in comparison groups were small. We tested the internal consistency reliability of the questions in assessing clinician numeracy using Cronbach's alpha. We assessed the ability of questions to differentiate between higher and lower scoring participants by testing item discrimination (MSU n.d., USF n.d.). Since there is no gold standard measure of clinician numeracy, we created a set of *a priori* hypotheses about a valid assessment of clinician numeracy (Table 3) based on the characteristics of published tests and previous empirical studies of numeracy. We used these hypotheses to assess the construct validity of the MINT.

Results

Most (70%; 135/194) of those invited to participate were recruited to the study, and they represented 45% (135/299) of the foundation trainees employed by the four participating trusts (Fig. 2). The preponderance (83%; 112/135) of the participants were UK graduates, representing 27 different medical schools. First-year students made up 43% (58/135), and the remainder were in their second year of training. Gender breakdown was 53% (72/135) female; 46% (62/135) male; and one undeclared. No one declared a diagnosis of dyslexia. The maximum possible score was 43; the mean score was 32.76 (76%), with a 95% confidence interval of 31.6–33.9. The range of scores was 14–42 (33-98%), with an interquartile range of 29–38 (67–88%).

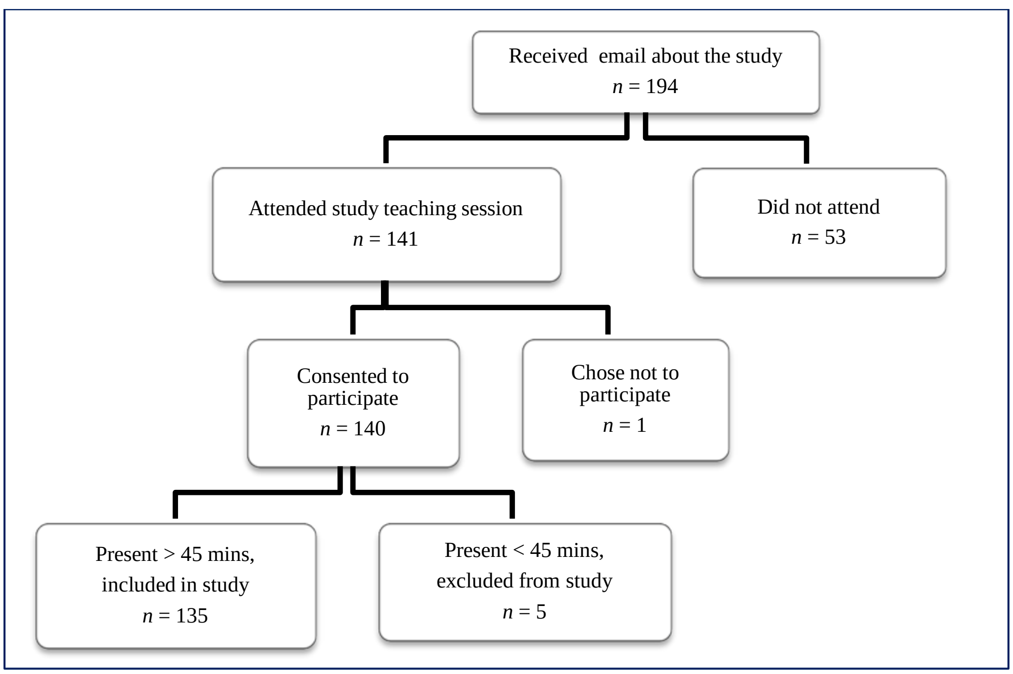


Figure 2. Recruitment to study

Psychometric data

Cronbach's alpha (α) for the MINT was 0.868, demonstrating that although the questions came from a variety of sources and included apparently diverse material (e.g., some questions were based on clinical scenarios, others on interpretation of data displays), all were testing the same construct: clinician numeracy. Further proof of the consistency of test items is indicated by the finding that there was minimal variation in the alpha value on removing any item (range α = 0.860–0.870).

The MINT is not a difficult test: facility of MINT items was high, with a range between 0.4 and 0.97; only eight items had a facility below 0.6. In conjunction with this result, we found that MINT items did not discriminate well between high- and low-performing participants: only three questions (items 15, 26, 34) had discrimination values of $\geq$ 20%. These were among the hardest items, with a facility of 0.50, 0.40 and 0.53 respectively.

Properties of a Clinician Numeracy Test. We assessed construct validity of the MINT using the criteria described in Table 3.

1. Part-whole correlations between performance on the MINT and subsets of questions taken from other numeracy/health numeracy tests were statistically significant (Table 6), which is not surprising given the high Cronbach's alpha.
2. Performance was related to level of education. This property was assessed by comparing FT performance with that of other groups on three subsets of questions.
    a. FTs performed significantly better than patients and the general public, and at a similar level to U.S. medical students and doctors on the Schwartz et al. (1997) test.
    b. FTs outperformed entrants to university in the U.S. on questions devised by Sikorskii et al. (2011).
    c. FTs performed at a similar level to US medical students on the questions used by Sheridan and Pignone (2002).
3. We classified five levels of math instruction: General Certificate of Secondary Education (GCSE) (age 16); non-UK high school; GCSE-Advanced Subsidiary level (age 17); GCSE-Advanced level (age 18); and university degree. Correlation between MINT score and level of math instruction was statistically significant, although weak (r = 0.185, p <0.05). [1]

[1] The main UK high-school exams are General Certificate of Secondary Education (GCSE), taken at age 16, General Certificate of Education - Advanced Subsidiary level (AS), taken at age 17, and General Certificate of Education - Advanced level (A-level) taken at age 18. Criteria for entry to UK medical schools vary, and, while many students will have achieved 'A' grades at A-level

4. We graded difficulty of MINT questions 1–5, according to their source. A strong inverse correlation between performance and item difficulty (r = –0.751, $p < 0.01$) demonstrated that fewer FTs were able to answer the questions we had deemed harder correctly.
5. The MINT included 13 computational, 17 analytical and 13 statistical questions. FTs found computational questions easiest, demonstrated by a mean score of 85% on this subset, compared to mean scores of 72% for both analytical and statistical subsets.

Table 6.
Correlation between MINT Items and Other Numeracy Tests (Spearman's *r*)

Source of questions (no. in the MINT)	Schwartz 1997 probability	Weiss 2005 NVS	Sikorskii questions	Remaining MINT items
Schwartz (n=3)	1	0.096†	0.333*	0.382*
NVS (n=4)	0.096†	1	0.299*	0.318*
Sikorskii (n=9)	0.333*	0.299*	1	0.776*
New MINT (n=12)	0.373*	0.258*	0.547*	0.884*

*p<0.01; †NS.

We observed a gender effect: male participants had higher mean MINT scores than females ($p < 0.01$), and were significantly more likely to be in the top 10% of the cohort ($\chi^2(1) = 11.631$, $p < 0.01$).

Self-Assessment. FTs were asked to estimate how well they would perform compared to other groups. They were also asked to predict their ranking compared to their peers, at both the start and the end of the MINT. Although FTs correctly predicted that they would outperform patients and the general public, they were poor at predicting their own ranking, and only 15.5% pre-test and 21% post-test expected to rank in the lowest third. Correlation between perception of ability and MINT score was weak (r = –0.208, $p < 0.01$).

Discussion

We developed a measure of clinician numeracy, the Medical Interpretation and Numeracy Test (MINT), and implemented it on a cohort of trainee doctors. We provide several arguments to support the validity of our test, using established standards of assessment development. We tested a clearly-defined construct—

math, some may have only a grade 'B' at GCSE. The syllabus for GCSE math includes the important areas of numeracy for medical practice (Department for Education 2013). However, Lee et al. (2010), in their comprehensive guide to math pre entry to UK university, note that students who stop studying math at GCSE have forgotten much of what they had learnt by the time they arrive at university. All applicants to UK medical schools must also sit the UK Clinical Aptitude Test (http://www.ukcat.ac.uk) which includes a section on quantitative reasoning.

clinician numeracy—using a blueprint, and all test items were set in a healthcare context. The use of a blueprint allowed us to select and develop items relevant to the clinical workload of our participants and at an appropriate level of difficulty. Content validity was established by colleague review and with a pilot test. Further evidence of the MINT's validity came from testing our results against various *a priori* hypotheses (Table 3). Firstly, since the MINT has high internal consistency reliability, part-whole correlations were generally strong, particularly between the MINT and questions devised by statisticians to assess numeracy in students entering university (Sikorskii et al. 2011) (Table 6). However, correlation between the MINT and the NVS questions, although significant, was quite weak, as was correlation between the MINT and the Schwartz test. Furthermore we observed very weak correlations between the Schwartz test and the NVS, and between each of these subsets and the Sikorskii et al. questions. We contend that this finding illustrates one of the limitations of short tests of health numeracy.

The optimal length for a numeracy test is unclear. The MINT is a lengthy test and, as such, has the advantage of being able to test across the scope of clinician numeracy; it includes multiple items testing each construct, with a range of data displays, and we consider it to be a comprehensive test. However, its principal drawbacks are that it is time-consuming for participants, and data entry and analysis are laborious for researchers. Although Schuwirth and Van der Vleuten (2011) counsel that short tests are neither reliable nor valid, many short tests of health numeracy have been validated—e.g., the Schwartz et al. (1997) test, the NVS (Weiss et al. 2005), and the Berlin Numeracy Test (Cokely et al. 2012) — and are widely used. Furthermore, tests such as the TOFHLA and DNT which were originally lengthy have been successfully shortened (Baker et al. 1999; Huizinga et al. 2008). Short tests, however, are limited in their ability to measure across a range of difficulty levels and constructs. The Schwartz test comprises three probability questions and, therefore, assesses statistical health numeracy. The six-item NVS is based on interpretation of a nutritional label, and so it tests computational numeracy. Since the Schwartz test and the NVS assess quite different constructs, it is not surprising that correlation between them is weak. Similarly, neither of these tests correlates strongly with the Sikorskii et al. (2011) subset (testing analytical and statistical constructs), or the MINT, which tests all three constructs of clinician numeracy. This observation suggests that some research findings may need to be qualified: e.g., the seminal Schwartz et al. (1997) paper shows an association between *statistical health numeracy* rather than *numeracy* and the ability to interpret screening test results. We are in the process of developing and validating a shorter version of the MINT, as we recognize that this modification would make it more feasible to deliver and more acceptable to participants. Given the MINT's high internal consistency, we aim to reduce its length by half, while maintaining its integrity in testing clinician numeracy.

Secondly, our hypothesis that doctors would perform better than less well-educated individuals proved true for those questions where data was available for comparison (Schwartz et al. 1997; Sikorskii et al. 2011). The expectation that doctors' performance would be similar to that of U.S. medical students and doctors on various subsets (Sheridan and Pignone 2002; Anderson et al. 2011) was also confirmed, suggesting that the performance of FTs on the MINT may be generalisable to other doctors and medical students.

An interesting finding was that correlation between MINT score and level of math instruction was weak (r = 0.185, p <0.05), as was correlation with grade awarded at highest-level math test (r = 0.284, p <0.01). We suggest that higher-level math instruction in the UK may not be relevant to clinician numeracy, being focused on advanced rather than basic math concepts. Previous research has shown that A-level math conferred no advantage to medical students in epidemiology and biostatistics tests (Ben-Shlomo et al. 2004). Although Sikorskii et al. (2011) observed moderate to strong correlations between U.S. school-level mathematics and their numeracy test, they suggest that the tests are not equivalent because numeracy tests "capture more than the level of educational development". More recently, Levy et al. (2014) suggest that numeracy and health numeracy may be different constructs, based on their finding that framing quantitative questions in a healthcare setting adversely affected performance compared to framing the same problem in a pure math or finance context.

Further evidence of validity of the MINT was the strong inverse correlation between predicted level of difficulty and accuracy of response (r = –0.753, p < 0.01). Items correctly answered by more than 75% of candidates (facility > 0.75) can be considered easy, while difficult items are those answered correctly by less than 25% (facility < 0.25) (USF n.d.). The MINT was not a difficult test: 65% of items had facility > 0.75, and none had facility < 0.25. This is explained by the fact that the MINT is a mastery model test: the test content is material with which participants should be highly proficient. Test items are based on the normal workload of a trainee doctor. In tests like this, where all candidates are expected to score highly, the discrimination index is not useful for item analysis (MSU n.d.), and we found that very few MINT items were good discriminators. Our speculation that computational items would be easier than analytical or statistical ones was also correct: all computational items had facility > 0.75. Finally, the observed gender effect is a common finding in tests of numeracy—e.g., Sikorskii et al. (2011) and Weller et al. (2012)—although not conclusive evidence of validity.

Our finding that participants were poor at self-assessment is not surprising; self-assessment is generally unreliable. However, Simpson et al. (2009) observed a similarity between predicted and actual scores in their numeracy test in doctors. Anderson et al. (2011) reported only a weak correlation (r = 0.282) between the

Subjective Numeracy Scale and the Schwartz et al. (1997) test; they suggested that subjective and objective numeracy tests may assess different constructs, with the former indicating “math confidence” rather than ability.

Clinician numeracy is poorly understood and relatively under-explored; we do not yet know why some doctors have limited calculation and data interpretation skills. We are using the MINT to explore these areas further. We plan to implement a blank answer “show your work” version of the MINT, to investigate how and why errors occur. We anticipate that results of this work will be helpful both in setting a standard of clinician numeracy for doctors and in developing educational material for clinicians at undergraduate and postgraduate levels. A further strand of our research is using the MINT to investigate clinician numeracy in other healthcare professionals. In addition to developing a shorter test, we are exploring the implementation of the MINT electronically, both as an online test, and in the form of a slide presentation with audience participation software. Finally, having forbidden participants from using calculators in this study, we are recruiting a cohort who will be given calculators, allowing us to assess what impact this has on performance.

Limitations

This study has some limitations. While we found the MINT to be a valid test, our results are based on testing in a relatively small cohort of junior doctors working in the UK. Our finding that participants performed at a similar level to doctors and medical students in U.S.-based studies leads us to consider that our results may be generalisable, but further research is clearly needed to confirm this finding. We have not yet had the opportunity to use the MINT to test clinician numeracy in other healthcare professionals, although we consider it may be suitable for this purpose.

The level of difficulty for items taken from existing tests was assigned empirically, based on the source of the question. However, it is evident that there is some overlap in level of difficulty between items devised for primary school children, patients/the general public, and aspiring student nurses. We are in the process of addressing this issue by having the level of difficulty for all questions re-evaluated by a panel unfamiliar with the test.

We acknowledge the debate around the notion that performance in a classroom test is indicative of performance in clinical practice (Rowe et al. 1998; Selbst et al. 1999; McMullan et al. 2010; Wright 2010). Yet, we agree with Rowe et al. (1998) that mistakes are more likely to occur in a busy, stressful ward environment with multiple distractions than during a test in examination conditions. Therefore, candidates who perform poorly in classroom testing may have greater problems (and so be at greater risk to patients) in clinical practice.

Conclusions

Our research supports evidence from other studies that concerns about clinician numeracy in doctors are well founded. We believe this concern points to an important patient safety issue that requires further investigation. The MINT is a valid and useful measure of clinician numeracy that will be helpful in future research.

Acknowledgments

We are indebted to Jim Mulherin for his assistance with the analysis of our data. We thank Professors Huizinga, Peters, Pignone, Schwartz, Sikorskii and Weiss and their colleagues, and the OECD, for permission to use their material in our research. We are extremely grateful to our reviewers for their constructive feedback, which was invaluable in completing this manuscript. Finally, we thank Professor Len Vacher for his support and encouragement through this process.

References

Anderson, B. L., N. A. Obrecht, G. B. Chapman, D. A. Driscoll, and J. Schulkin. 2011. Physicians' communication of Down syndrome screening test results: The influence of physician numeracy. *Genet Med* 13(8): 744–9. http://dx.doi.org/10.1097/GIM.0b013e31821a370f (accessed August 22, 2013).

Apter, A. J., J. Cheng, D. Small, I. M. Bennett, C. Albert, D. G. Fein, M. George, and S. Van Horne. 2006. Asthma numeracy skill and health literacy. *Journal of Asthma* 43(9): 705–710. http://dx.doi.org/10.1080/02770900600925585 (accessed May 15, 2013).

Baker, D. W., M. V. Williams, R. M. Parker, J. A. Gazmararian, and J. Nurss. 1999. Development of a brief test to measure functional health literacy. *Patient Education and Counseling* 38(1): 33–42. http://dx.doi.org/10.1016/S0738-3991(98)00116-5 (accessed August 5, 2013).

Ben-Shlomo, Y., U. Fallon, J. Sterne, and S. Brookes. 2004. Do medical students with A-level mathematics have a better understanding of the principles behind evidence-based medicine? *Medical teacher* 26(8): 731–733. http://dx.doi.org/10.1080/01421590400016290

Cokely, E. T., M. Galesic, E. Schulz, S. Ghazal, and R. Garcia-Retamero. 2012. Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making* 7(1): 25–47.

Caverly, T. J., A. Prochazka, I. Binswanger, J.S. Kutner, and D. Matlock. 2012. Getting the gist of health risks. http://www.statlit.org/pdf/2012-Caverly-ASA.pdf (accessed 6 October, 2013).

Department for Education. 2013. *GCSE in mathematics: subject content and assessment objectives.* http://www.gov.uk/government/publications/gcse-mathematics-subject-content-and-assessment-objectives (Accessed October 1, 2015).

Estrada, C. A., M. Martin-Hryniewicz, B. T. Peek, C. Collins, and J. C. Byrd. 2004. Literacy and numeracy skills and anticoagulation control. *The American journal of the medical sciences* 328(2): 88–93. (accessed May 15, 2013). http://dx.doi.org/10.1097/00000441-200408000-00004

Fagerlin, A., B. J. Zikmund-Fisher, P. A. Ubel, A. Jankovic, H. A. Derry, and D. M. Smith. 2007. Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making* 27(5): 672–680. http://dx.doi.org/10.1177/0272989X07304449 (accessed October 2, 2013).

Gazmararian, J. A., J. W. Curran, R. M. Parker, J. M. Bernhardt, and B. A. DeBuono. 2005. Public health literacy in America: An ethical imperative. *American journal of preventive medicine* 28(3): 317–322. http://dx.doi.org/10.1016/j.amepre.2004.11.004 (accessed February 21, 2015).

Ghosh, A. K., and K. Ghosh. 2005. Translating evidence-based information into effective risk communication: Current challenges and opportunities. *Journal of Laboratory and Clinical Medicine* 145(4): 171–180. http://dx.doi.org/10.1016/j.lab.2005.02.006 (accessed May 6, 2013).

Gigerenzer, G., W. Gaissmaier, E. Kurz-Milcke, L.M. Schwartz, and S. Woloshin. 2007. Helping doctors and patients make sense of health statistics. *Psychological science in the public interest* 8(2): 53–96. http://dx.doi.org/10.1111/j.1539-6053.2008.00033.x

Gigerenzer, G., and J.A. Muir Gray, eds. 2011. *Better doctors, better patients, better decisions: Envisioning health care 2020*. Cambridge Mass: MIT.

Golbeck, A. L., C. R. Ahlers-Schmidt, A. M. Paschal, and S. E. Dismuke. 2005. A definition and operational framework for health numeracy. *American journal of preventive medicine* 29(4): 375–376. http://dx.doi.org/10.1016/j.amepre.2005.06.012

Hanoch, Y., T. Miron-Shatz, H. Cole, M. Himmelstein, and A. D. Federman. 2010. Choice, numeracy, and physicians-in-training performance: The case of Medicare Part D. *Health Psychology* 29(4): 454. http://dx.doi.org/10.1037/a0019881 (accessed September 29, 2013).

Harries, C. S. 2013. Can medical students calculate drug doses? *Southern African Journal of Anaesthesia and Analgesia* 19(5): 248–251. http://dx.doi.org/10.1080/22201173.2013.10872934

Hegener, M. A., S. M. Buring, and E. Papas. 2013. Impact of a required pharmaceutical calculations course on mathematics ability and knowledge retention. *American journal of pharmaceutical education* 77(6): 124. http://dx.doi.org/10.5688/ajpe776124

Hughes, R. G. and E. A. Edgerton. 2005. Reducing pediatric medication errors: Children are especially at risk for medication errors. *American Journal of Nursing* 105(5): 79–84. http://dx.doi.org/10.1097/00000446-200505000-00035

Huizinga, M. M., T. A. Elasy, K. A. Wallston, K. Cavanaugh, D. Davis, R. P. Gregory, L. S. Fuchs, R. Malone, A. Cherrington, and D. A. DeWalt. 2008. Development and validation of the Diabetes Numeracy Test (DNT). *BMC Health Services Research* 8(1): 96. http://dx.doi.org/10.1186/1472-6963-8-96 (accessed September 29, 2013).

Johnson, T.V., A. Abbasi, E. D .Schoenberg, R. Kellum, L. D. Speake, C. Spiker, A. Foust, A. Kreps, C. W. M. Ritenour, O. W. Brawley, and V. A. Master. 2014. Numeracy among trainees: Are we preparing physicians for evidence-based medicine? *Journal of surgical education* 71(2): 211–215. http://dx.doi.org/10.1016/j.jsurg.2013.07.013 (accessed October 3, 2015).

KCL. See King's College, London.

Key Stage 2. Year 6 Level 6 Maths SATs Papers 2011. http://www.satspapers.org.uk/SATs_Papers/KS2_Maths/SATS_MATHS_2011%20KS2/ks2-mathematics-2011-level-6-test-b.pdf(accessed September 13, 2013).

King's College, London. 2013. Florence Nightingale School of Nursing and Midwifery. Numeracy Screening Assessment for entry to BSc & PGDip Nursing & Midwifery programmes 2013-14. http://www.kcl.ac.uk/nursing/study/applicants/preregistration/selectiondays/Numeracy-example-2013-14.pdf (accessed September 28, 2013).

KS2. See Key Stage 2.

Latif, D. A., and J. A. Grillo. 2002. Assessing the basic math skills of first-year doctor of pharmacy students. *Journal of Pharmacy Teaching* 9(2): 17–25. http://dx.doi.org/10.1300/J060v09n02_02

Lee, S., R. Browne, S. Dudzic, and C. Stripp. 2010. *Understanding the UK Mathematics Curriculum Pre-Higher Education: a guide for Academic Members of Staff*. Mathematics in Education and Industry. http://www.mei.org.uk/files/pdf/Pre_Uni_Maths_Guide.pdf (accessed October 1, 2015).

Levy, H., P. A. Ubel, A. J. Dillard, D. R. Weir, and A. Fagerlin. 2013. Health numeracy: The importance of domain in assessing numeracy. *Medical Decision Making* 34: 107-115. http://dx.doi.org/10.1177/0272989X13493144

Lipkus, I. M., G. Samsa, and B. K. Rimer. 2001. General performance on a numeracy scale among highly educated samples. *Medical Decision Making* 21(1): 37–44. http://dx.doi.org/10.1177/0272989X0102100105

McMullan, M., R. Jones, and S. Lea. 2010. Patient safety: numerical skills and drug calculation abilities of nursing students and registered nurses. *Journal of advanced nursing* 66(4): 891–899. http://dx.doi.org/10.1111/j.1365-2648.2010.05258.x (accessed July 18, 2012.)

MSU n.d. See Michigan State University.

Michigan State University. Interpreting the index of discrimination. Scoring Office. https://www.msu.edu/dept/soweb/indexdis.html (accessed October 7, 2015)

OECD. See Organisation for Economic Cooperation and Development.

Organisation for Economic Cooperation and Development. 2009. *Take the Test: Sample questions from OECD's PISA Assessments*. Paris, France: OECD Publishing. http://dx.doi.org/10.1787/9789264050815-en (accessed 29 August, 2013).

Oldridge, G. J., K. M.Gray, L. M. McDermott, and C. M. J. Kirkpatrick. 2004. Pilot study to determine the ability of health-care professionals to undertake drug dose calculations. *Internal medicine journal* 34(6): 316–319. http://dx.doi.org/10.1111/j.1445-5994.2004.00613.x

Peters, E., N. Dieckmann, A. Dixon, J. H. Hibbard, and C. K. Mertz. 2007. Less is more in presenting quality information to consumers. *Medical Care Research and Review* 64(2): 169–190. http://dx.doi.org/10.1177/1077558707064002030l (accessed June 5, 2013).

Reyna, V. F., W. L. Nelson, P. K. Han, and N. F. Dieckmann. 2009. How numeracy influences risk comprehension and medical decision making. *Psychological bulletin* 135(6): 943. http://dx.doi.org/10.1037/a0017327 (accessed March 19, 2013).

Rolfe, S., and N. J. Harper. 1995. Ability of hospital doctors to calculate drug doses. *British Medical Journal* 310(6988): 1173. http://dx.doi.org/10.1136/bmj.310.6988.1173

Rothman, R. L., R. Housam, H. Weiss, D. Davis, R. Gregory, T. Gebretsadik, A. Shintani, and T. A. Elasy. 2006. Patient understanding of food labels: The role of literacy and numeracy. *American journal of preventive medicine* 31 (5): 391–398. http://dx.doi.org/10.1016/j.amepre.2006.07.025 (accessed June 10, 2013).

Rowe, C., T. Koren, and G. Koren. 1998. Errors by paediatric residents in calculating drug doses. *Archives of disease in childhood* 79(1): 56–58. http://dx.doi.org/10.1136/adc.79.1.56 (accessed March 10, 2014).

Schapira, M. M., C. M. Walker, K. J. Cappaert, P. S. Ganschow, K. E. Fletcher, E. L. McGinley, S. Del Pozo, C. Schauer, S. Tarima, and E. A. Jacobs. 2012. The Numeracy Understanding in Medicine Instrument: A measure of health numeracy developed using Item Response Theory. *Medical Decision Making* 32(6):851-865. http://dx.doi.org/10.1177/0272989X12447239 (accessed June 19, 2015).

Schuwirth, L. W. T., and C. P. M. van der Vleuten. 2011. General overview of the theories used in assessment: AMEE Guide 57. *Medical teacher* 33(10): 783–797. http://dx.doi.org/10.3109/0142159X.2011.611022 (accessed April 12, 2013).

Schwartz, L. M., S. Woloshin, W. C. Black, and H. G. Welch. 1997. The role of numeracy in understanding the benefit of screening mammography. *Annals of internal medicine* 127(11): 966. http://dx.doi.org/10.7326/0003-4819-127-11-199712010-00003

Schwartz, L. M., S. Woloshin, and H. G. Welch. 2005. Can patients interpret health information? An assessment of the medical data interpretation test. *Medical Decision Making* 25(3): 290–300. http://dx.doi.org/10.1177/0272989X05276860 (accessed May 11, 2013).

Selbst, S. M., J. A. Fein, K. Osterhoudt, and H. Wayne. 1999. Medication errors in a pediatric emergency department. *Pediatric emergency care* 15(1): 1–4. http://dx.doi.org/10.1097/00006565-199902000-00001

Sheridan, S. L., and M. Pignone. 2002. Numeracy and the medical student's ability to interpret data. *Effective clinical practice* 5(1): 35.

Sheridan, S. L., M. P. Pignone, and C.L. Lewis. 2003. A randomized comparison of patients' understanding of number needed to treat and other common risk reduction formats. *Journal of General Internal Medicine* 18(11): 884-892. http://dx.doi.org/10.1046/j.1525-1497.2003.21102.x

Sikorskii, A., V. Melfi, D. Gilliland, J. R. Kaplan, and S. Ahn. 2011. Quantitative literacy at Michigan State University, 1: Development and initial evaluation of the assessment. *Numeracy* 4(2):5. http://dx.doi.org/10.5038/1936-4660.4.2.5 (accessed October 2, 2013)

Simpson, C. M., G. B. Keijzers, and J. F. Lind. 2009. A survey of drug-dose calculation skills of Australian tertiary hospital doctors. *Medical Journal of Australia* 190(3):117.

UK Clinical Aptitude Test. 2015. http://www.ukcat.ac.uk (accessed October 1, 2015).

USF n.d. See University of South Florida.

University of South Florida. Classroom Assessment: Basic Concepts. C. Item Analysis. http://fcit.usf.edu/assessment/selected/responsec.html (accessed October 7, 2015).

Vacher, H. L., and T. Chavez. 2009. Quantitative literacy on the Web of Science, 2 - Mining the health numeracy literature for assessment items. *Numeracy* 2(1): Article 5. http://dx.doi.org/10.5038/1936-4660.2.1.5 (accessed October 23, 2013).

Wegwarth, O., W. Gaissmaier, and G. Gigerenzer. 2011. Deceiving numbers: Survival rates and their Impact on doctors' risk communication. *Medical Decision Making* 31(3): 386–394. http://dx.doi.org/10.1177/0272989X10391469 (accessed July 13, 2013).

Weiss, B. D., M. Z. Mays, W. Martz, K. M. Castro, D. A. DeWalt, M. P. Pignone, J. Mockbee, and F. A. Hale. 2005. Quick assessment of literacy in primary care: the newest vital sign. *The Annals of Family Medicine* 3(6): 514–522. http://dx.doi.org/10.1370/afm.405 (accessed July 13, 2013).

Weller, J. A., N. F. Dieckmann, M. Tusler, C. K. Mertz, W. J. Burns, and E. Peters. 2013. Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making* 26(2): 198–212. http://dx.doi.org/10.1002/bdm.1751 (accessed June 18, 2015).

Wheeler, D. W., D. D. Remoundos, K. D. Whittlestone, T. P. House, and D. K. Menon. 2004. Calculation of doses of drugs in solution. *Drug safety* 27 (10): 729–734. http://dx.doi.org/10.2165/00002018-200427100-00003

Williams, M. V., R. M. Parker, D. W. Baker, N. S. Parikh, K. Pitkin, W. C. Coates, and J. R. Nurss. 1995. Inadequate functional health literacy among patients at two public hospitals. *Journal of the American Medical Association* 274(21): 1677–1682. http://dx.doi.org/10.1001/jama.1995.03530210031026

Windish, D. M., S. J. Huot, and M. L. Green. 2007. Medicine residents' understanding of the biostatistics and results in the medical literature. *Journal of the American Medical Association* 298(9): 1010–1022. http://dx.doi.org/10.1001/jama.298.9.1010 (accessed October 8, 2013).

Woloshin, S., L.M. Schwartz, and H. G. Welch. 2008. *Know your chances: understanding health statistics.* Berkeley: University of California Press.

Wright, K. 2010. Do calculation errors by nurses cause medication errors in clinical practice? A literature review. *Nurse Education Today* 30(1): 85–97. http://dx.doi.org/10.1016/j.nedt.2009.06.009 (accessed July 28, 2012).

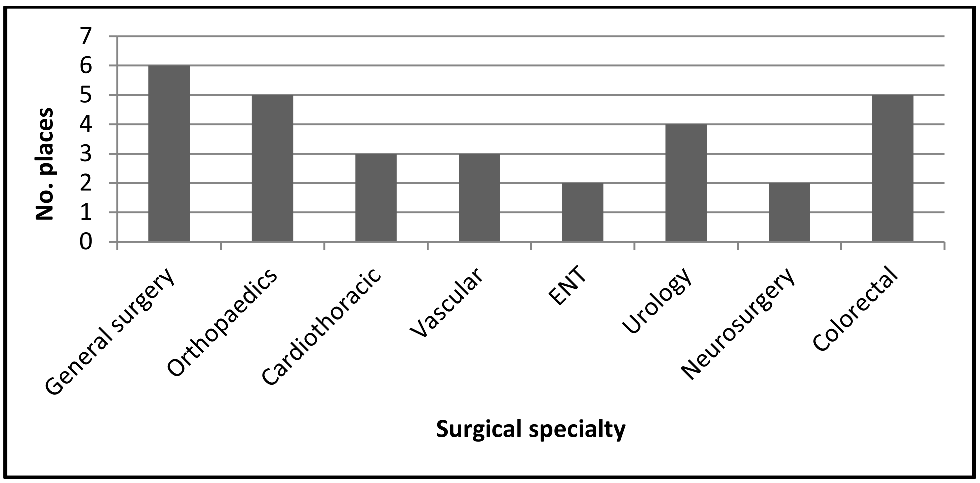
APPENDIX

MINT: sample questions

Questions are listed here in order of source, from those devised for primary schoolchildren, to those devised for medical students and doctors. Many of these questions have been amended slightly from the original: some to improve clarity, and others to conform to a medical context or to a five-answer MCQ format.

QUESTION DESIGNED FOR PRIMARY SCHOOLCHILDREN
KS2 (2011)

The chart below shows the number of training places for FY1 doctors in various surgical specialties in a large teaching hospital.



Sam is an FY1 trainee. Assuming that places are allocated at random, how likely is he to be placed in General Surgery?

A. 50% **B**. 40% **C**. 30% **D**. 20% **E**. 10%

QUESTION DESIGNED FOR PATIENTS
Huizinga et al. (2008)

Maria has diabetes and is planning to exercise in the gym for one hour. She needs to eat 6 g of carbohydrate for every 30 mins she exercises. She has some biscuits in her gym bag. Each biscuit contains 8 g of carbohydrate. How many biscuits should she eat before she exercises?

A. 1/2 biscuit
B. 1 biscuit
C. 3/4 biscuit
D. 2 biscuits
E. 1 and 1/2 biscuits

QUESTION DESIGNED FOR ENTRANTS TO NURSING SCHOOL
KCL (2013)

You are asked to review Mr Brown as the ward sister is worried about his urine output. The chart below shows Mr Brown's urine output over the past four days:

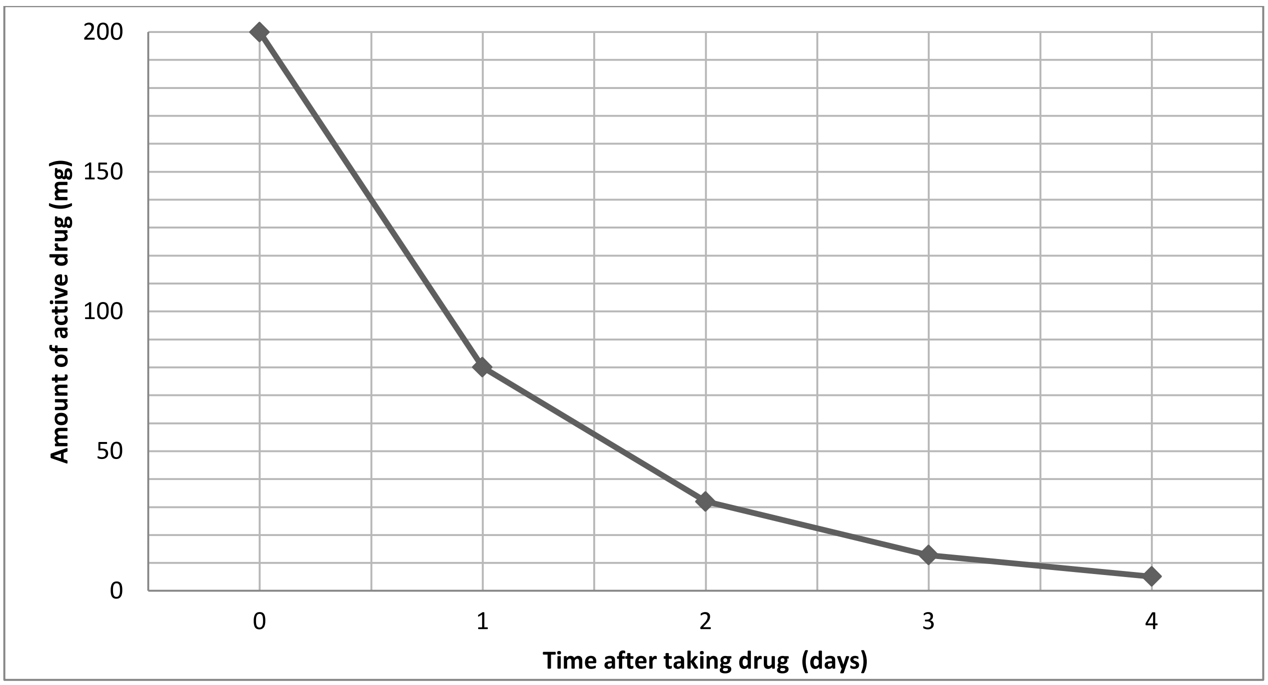
Day	**Urine output (ml)**
Monday	532
Tuesday	472
Wednesday	472
Thursday	364

What is Mr Brown's average urine output per day over this 4-day period?

A. 1460 ml
B. 472 ml
C. 480 ml
D. 460 ml
E. 1840 ml

QUESTIONS SIMILAR TO THOSE DESIGNED FOR SECONDARY SCHOOLCHILDREN (replacing original questions from OECD)

Alex enters a clinical trial, and is given 200 mg of the test drug by IV injection. The following graph shows the initial amount of the drug in Alex's bloodstream, and the amount that remains active in Alex's blood after one, two, three and four days.



1. **Approximately what percentage of the drug remains active after 24 hours?**

   A. 50% **B**. 10% **C**. 40% **D**. 20% **E**. 30%

2. **Approximately how many mg of the drug remains active after 36 hours?**

   A. 80 mg **B**. 13 mg **C**. 33 mg **D**. 55 mg **E**. 5mg

QUESTION DESIGNED FOR ENTRANTS TO US UNIVERSITY

Sikorskii et al. (2011)

There is a 2 in 100 chance of living 5 years or longer without treatment for a type of cancer. Drug X increases the chance of living 5 years or longer to 6%. Drug Y increases the chance of living 5 years or longer by 50%. If a patient wants the best chance of living 5 years or longer, which drug should be prescribed?

A. Drug Y
B. Drug X
C. Either drug, the chance of living longer is the same
D. Neither drug, the chance of living longer is better without treatment
E. Don't know

QUESTION DESIGNED FOR EDUCATED CONSUMERS
Peters et al. (2007)

100 women attend hospital for a mammogram. 10 of these women have a malignant tumour, while 90 do not. Of the 10 patients with malignancy, the mammogram detects the cancer in 9, but misses the tumour in one patient. Of the 90 women who are disease-free, the mammogram indicates correctly that 81 of them are healthy, but wrongly indicates that 9 of them have cancer. Mrs Jones is told that her mammogram is positive. What are the chances that she actually does have cancer?

A. 1 in 2 **B**. 1 in 10 **C**. 1 in 9 **D**. 2 in 9 **E**. 9 in 10

DOI: http://dx.doi.org/10.5038/1936-4660.9.1.5

QUESTIONS DESIGNED FOR MEDICAL STUDENTS
Sheridan and Pignone (2002)

Imagine that 40 out of 1000 people are expected to develop disease Y over the next 5 years. Treatment A reduces the chance of getting disease Y by 10 per 1000 people. Treatment B reduces the chance of getting disease Y by 4 per 1000 people. Select the correct answer.

A. Treatment A is more effective than Treatment B
B. Treatment B is more effective than Treatment A
C. Treatment A and Treatment B are equally effective
D. Don't know
E. Don't know

What is the risk of developing disease Y after receiving Treatment A?

A. 36:1000 **B**. 35:1000 **C**. 39:1000 **D**. 30:1000 **E**. Don't know

NEW MINT QUESTION DESIGNED FOR FOUNDATION TRAINEES

Mo weighs 100 kg, and presents to A&E with a wound in his thigh. You are asked to suture it, using the local anaesthetic bupivacaine which comes in a solution containing bupivacaine 5mg/ml. The maximum dose of bupivacaine that can be safely given is 2 mg/kg. What is the maximum amount of bupivacaine you can use when suturing Mo's wound?

A. 500ml **B**. 20ml **C**. 150ml **D**. 50ml **E**. 40ml