
2017

Clinician Numeracy: Use of the Medical Interpretation and Numeracy Test in Foundation Trainee Doctors

Anne A. Taylor

School of Medicine, Keele University, cairde1@icloud.com

Lucie M. Byrne-Davis

University of Manchester, Cheshire, England, lucie.byrne-davis@manchester.ac.uk

Follow this and additional works at: <https://digitalcommons.usf.edu/numeracy>



Part of the [Medical Education Commons](#)

Recommended Citation

Taylor, Anne A., and Lucie M. Byrne-Davis. "Clinician Numeracy: Use of the Medical Interpretation and Numeracy Test in Foundation Trainee Doctors." *Numeracy* 10, Iss. 2 (2017): Article 5. DOI: <http://doi.org/10.5038/1936-4660.10.2.5>

Authors retain copyright of their material under a [Creative Commons Non-Commercial Attribution 4.0 License](#).

Clinician Numeracy: Use of the Medical Interpretation and Numeracy Test in Foundation Trainee Doctors

Abstract

Patient safety is a priority in healthcare today. Good clinician numeracy in doctors contributes to patient safety, since it is essential for accuracy in prescribing, and in data interpretation. Evidence, however, suggests that although doctors are assumed to be highly numerate, they often make errors in drug dose calculation and struggle to interpret medical statistics. Having developed a new assessment measure, the Medical Interpretation and Numeracy Test (MINT), we describe its use to evaluate clinician numeracy in 135 recently qualified doctors in the UK ("foundation trainees," i.e., in their first two years post-graduation). The maximum possible test score was 43; the range of scores was 14 – 42 (33 – 98% correct), with an interquartile range of 29 – 38 (67 – 88%). Mean score was 32.76 (76%), with a 95% confidence interval of 31.6 to 33.9. Drug dose calculation errors were common, and potentially hazardous. Two thirds of participants had difficulty with simple data interpretation tasks designed for patients. Almost a quarter could not select the better of two treatment options when data were presented as relative risk reduction, absolute risk reduction or number needed to treat. Our study suggests that a significant proportion of medical graduates have poor clinician numeracy. Such doctors may harm patients by making mistakes in drug dose calculation and through flawed medical decision making. While further research is needed to investigate clinician numeracy in doctors, we believe that there is sufficient evidence to incorporate clinician numeracy into undergraduate and postgraduate medical education programs as a matter of patient safety.

Keywords

clinician numeracy, physician numeracy, health numeracy, numeracy assessment

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial 4.0 License](https://creativecommons.org/licenses/by-nc/4.0/)

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 senior lecturer in assessment and psychometrics at Division of Medical Education, University of Manchester. She is a registered Health Psychologist, a Chartered Psychologist and a Principal 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

Healthcare can be hazardous. Despite the best intentions of healthcare providers and medical staff, things often go wrong, with evidence that approximately 10% of patients suffer an adverse incident related to healthcare interventions (Neale et al. 2001; Vincent et al. 2001), and that over 40% of medication orders contain errors (Seden et al. 2013). Not all adverse events and errors are serious, but some are catastrophic, leading to permanent harm or death. It is well recognised that multiple factors contribute to adverse events, and also that a substantial proportion are preventable (Brennan et al. 1991; Vincent et al. 2001; Leape et al. 2005; Vincent et al. 2014). As a consequence, patient safety has become the priority in healthcare, supported by healthcare providers, professional bodies and national patient safety organizations.¹ Patient safety is also a priority in medical education, underpinning the skills and attributes of doctors (Kirch and Boysen 2010; GMC 2015).

Clinician numeracy (numeracy related to patient care) in healthcare professionals is important to patient safety, since low clinician numeracy can lead to medical error (Table 1). Concerns have been raised about numeracy in nurses and pharmacists (Latif and Grillo 2002; Wright 2004; McMullan et al. 2010; Hegener et al. 2013); however, the focus of this paper is clinician numeracy in doctors. Doctors with low computational numeracy may make errors in drug-dose calculations; those with low analytical or statistical numeracy may misinterpret medical data, including test results, leading to inappropriate and ineffective treatment selection. These errors harm patients and waste valuable healthcare resources (Gigerenzer and Grey 2011; Frontier Economics 2014; Malhotra et al. 2015). Although it is known that medical students and doctors may have difficulty with drug-dose calculation and interpreting medical statistics (Rowe et al. 1998; Gigerenzer et al. 2007; Windish et al. 2007; Simpson et al. 2009; Harries and Botha 2013; Johnson et al. 2014), it is generally assumed that they are sufficiently numerate for safe medical practice. The standard of clinician numeracy that doctors require for safe medical practice, however, has not been defined. Further research in this area is needed, starting with an evaluation of the current level of clinician numeracy in doctors.

¹ For example: National Patient Safety Agency (UK), see NPSA 2010; General Medical Council (UK), see GMC 2015; American Medical Association, see AMA nd; National Patient Safety Foundation (U.S.), see NPSF nd.

Table 1.
Numeracy in Healthcare: constructs, competence and clinical application

Construct	Areas of competence*	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

* Golbeck et.al. (2015)

We have previously described our development and validation of an assessment measure of clinician numeracy - the Medical Interpretation and Numeracy Test (MINT) (Taylor and Byrne-Davis 2016). Sample questions of MINT were included in an appendix of the original paper and are reproduced in the appendix here for convenience. A copy of MINT for research purposes may be provided by application to the first author. The purpose of the study reported here is to use MINT, which is undergoing further study and refinement, to gather exploratory baseline data on MINT-measured clinician numeracy in recently qualified doctors in the UK.

Methods

Participants

In the first two years post-graduation, doctors in the UK are known as “foundation trainees” (FTs). FTs from four hospitals in Staffordshire and Cheshire were recruited to the study. There were no exclusion criteria.

Materials

Participants completed the MINT, a 43-item test assessing computational, analytical and statistical numeracy constructs. (Of the 43 items, 13 assess computational, 17 analytical and 13 statistical numeracy). MINT questions vary in

level of difficulty, and all are set in a clinical context appropriate to FTs. The test structure was multiple-choice, single-best-answer format, with five answer options. Correct answers scored one point, and incorrect and unanswered questions scored zero points, for a maximum possible score of 43. We also collected some demographic data about the participants; information on their attitudes about, and previous attainment in, math at high school level; and whether they had dyslexia.

Procedure

Between November 2013 and May 2014, 194 FTs in the four study sites were scheduled to attend a training session on *Clinical decision-making and risk communication*. All were provided with information about the study in advance of the session, and invited to participate. Following the training session, those who chose to participate took the test. Participants were given paper for rough work, but they were not allowed to use calculators. A maximum of one hour was allocated for the test.

Data analysis

We used Microsoft Excel and IBM SPSS programs and an online statistical tool² to analyse our data. Bivariate analyses were used to assess: a) correlation between the scores achieved by FTs on a subset of questions and the scores achieved by other participants tested on this material in previous studies; and b) the relationship between MINT score and the outcome of national math tests taken in high school at ages 16 and 18.

Ethical implications.

Both University Research and Ethics Committee (UREC) approval and NHS Research and Development organizational approval were received for this project. The UREC advised that completion of the MINT test paper and submission of answer sheets by participants would indicate their consent to take part in the study.

Results

All 141 FTs who attended the teaching session were recruited to the study. The attendance rate (141/194, 73%) at this session was similar to attendance at other teaching sessions on all sites. Data from five FTs were excluded as those FTs were not present for sufficient time to complete the test, and one FT withdrew from the study, leaving 135 participants. Participants were similar to FTs

² https://www.medcalc.org/calc/comparison_of_proportions.php MedCalc software bvba (BE)

nationally in terms of gender and ethnic background (Table 2) (GMC 2014). The majority of FTs (112/135 (83%)) were UK graduates, representing 27 of its 32 medical schools. No participant declared a diagnosis of dyslexia. The range of scores and median scores were similar in gender groups and in graduates from different geographical areas (Table 3).

Table 2.
Demographic data of study participants compared to national data on gender and ethnicity

		No. (%)	% trainees in England 2013	% medical students in England 2013*
All FTs		135 (100)		
Year of training	FY1	58 (43)		
	FY2	77 (57)		
Medical School	UK	112 (83)	82	100
	Other	23 (17)	18	
Gender	Male	62 (46)	43	46
	Female	72 (53)	57	54
Ethnicity	White	64 (47)	60	63
	BME		40	
	Asian	30 (22)		22
	Black	7 (5)		3
	Mixed race	3 (2)		4.6
	Chinese	8 (6)		4.5
	Other/Unknown	23 (17)		3.5

*GMC 2014

Table 3.
Demographic data and scores

		Range of scores	Median score
Gender	Male	14 - 42	35
	Female	16 - 42	33
Medical School	UK	18 - 42	34
	Other	14 - 41	33

Overall outcome

No candidate answered all questions correctly, and no question was answered correctly by all candidates, not even those questions based on items originally designed for primary schoolchildren. Test score distribution was skewed, with a range of 14–42 (33–98% correct) as shown in Figure 1. The median score was 34/43 (79%) correct answers, with an interquartile range of 29–38 (67–88%). FTs performed better on computational questions than on questions testing either analytical or statistical constructs (Table 4).

We observed little correlation between previous achievements in math at high-school and MINT score ($r = 0.185$ for level of math instruction, and $r = 0.284$ for math grade); this result is consistent with the finding of Ben-Shlomo et al. (2004) in relation to the performance of medical students in epidemiology and biostatistical tests.

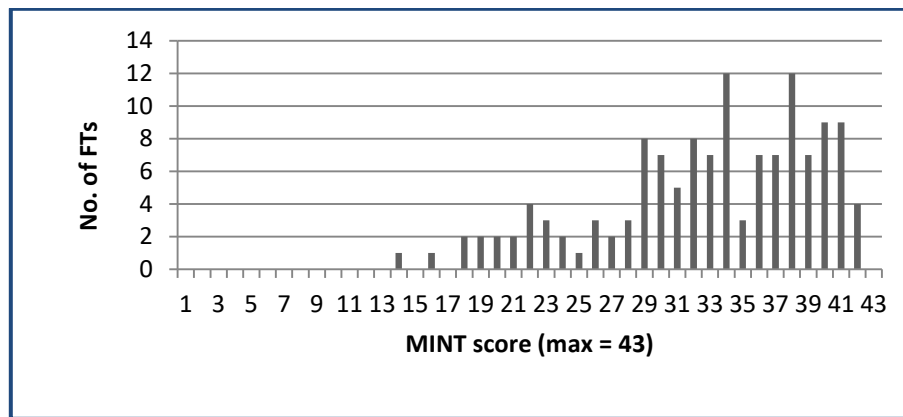


Figure 1. MINT scores

Table 4.
Performance (mean facility) on different numeracy constructs

Numeracy construct (no. items testing this construct)	Facility* (n=135)
Computational (n = 13)	0.85
Analytical (n = 17)	0.75
Statistical (n = 13)	0.72
Overall (n = 43)	0.76

*Facility indicates the proportion of questions answered correctly: a question answered correctly by all has a facility of 1.0; by 50%, a facility of 0.5

On subsets of the MINT where data were available for comparison purposes, FTs performed significantly better than non-medical populations (Schwartz et al. 1997; Sheridan et al. 2003; Sikorskii et al. 2011), and at a similar level to U.S. medical students and doctors (Sheridan and Pignone 2002; Gigerenzer et al. 2007; Anderson et al. 2011) (Tables 5 and 6).

Table 5.
Performance of FTs on probability questions compared to that of patients, U.S. first-year medical students and U.S. doctors

Source	Target group	No.	Coin toss	1% risk (convert to 10:1000)	1:1000 risk (convert to 0.1%)	All 3 correct
Schwartz et al. 1997	FTs	135	93%	89%	82%	70%
	Patients	287	54%*	54%*	20%*	16%*
Sheridan and Pignone 2003	Patients	357	-	-	-	2%*
Sheridan and Pignone 2002	Medical students	62	96% [‡]	90% [‡]	90% [‡]	77% [‡]
Gigerenzer et al. 2007	Physicians	85	100% [‡]	91% [‡]	75% [‡]	72% [‡]
Anderson et al. 2011	Obstetrician/ gynaecologists	203	-	-	-	66% [‡]

*N-1 chi-square test, $p < 0.0001$

[‡] N-1 chi-square test, NS at 0.01 level

Table 6.
Performance of FTs on questions devised for US school leavers at entry to university

Question no.	Question content	% correct (US students) [41]	% correct (Study FTs)	p value*
3	Drug comparison	25	61	< 0.0001
14	Screening test results	23	60	< 0.0001
15	Screening test results	21	50	< 0.0001
29	Problem solving	35	58	< 0.0001
31	Risk prediction	45	70	< 0.0001
33	Data interpretation	79	90	< 0.01
34	Data interpretation	38	53	< 0.001
39	Risk prediction	51	67	< 0.001
41	Data interpretation	48	64	< 0.01

* N-1 chi-square test

Psychometric analysis of the MINT is shown in Table 7.

Computational CN

Thirteen questions tested computational numeracy, including drug-dose calculations, use of medical formulae and interpretation of a nutritional label. The mean score on computational questions was 85%, and 33/135 (24%) of the FTs answered all computational questions correctly.

A question taken from the Diabetes Numeracy Test (Huizinga et al. 2008) had been designed to ask diabetic patients to work out how much carbohydrate (number of cookies) was needed to cover a period of exercise, and 19/135 (14%) of the FTs answered this question incorrectly. This result raises concerns not just about the FTs' basic computational skills, but also about their ability to advise patients on diabetes management.

Four MINT questions were adapted from the Newest Vital Sign test, which was developed for patients (Weiss et al. 2005). These questions test understanding of data provided on a nutritional label. On the MINT, 76/135 (56%) of the FTs answered all four of these questions correctly.

Five questions involved the calculation of doses of drugs commonly used in hospital. Only 33/135 (24%) of the FTs answered all five correctly, while 102/135 (76%) of them made errors that could have caused serious harm in clinical practice. Specifically, 29/135 (21%) of the FTs made decimal-place errors leading to miscalculations of 10 and 100 times the correct dose. An example of this error is a question in which candidates were asked to calculate the volume of insulin containing 8 units, from a solution of 100 units per ml. Although 115/135 (85%) of the FTs selected the correct answer, 13/135 (10%) made decimal-place errors leading to doses of 80 and 800 units of insulin; in clinical practice, such errors could be fatal. FTs also had difficulty calculating the correct volume of local anaesthetic to administer to a patient on a question that required candidates to convert the dose from mg/kg body weight to a volume in ml. Although 108/135

(80%) of the FTs calculated the correct volume, five (3.7%) of them selected a lethal dose of almost four times the therapeutic maximum.

Table 7.
Psychometric data of MINT items (questions are listed in order of construct and then facility)

Item no. and CN construct	Answer given (correct answer in bold, X=not answered)						Facility	Item Discrimination ^ψ	Cronbach's α if item deleted*
	A	B	C	D	E	X			
<i>Computational</i>									
<i>Mean value</i>							0.85	0.073	0.624**
27	0	0	125	6	1	3	0.93	0.066	0.864
23	0	5	124	3	1	2	0.92	0.074	0.864
12	3	0	9	0	123	0	0.91	0.014	0.869
5	0	13	0	119	3	0	0.88	0.029	0.869
4	6	3	118	4	3	1	0.87	0.059	0.865
22	118	11	0	1	4	1	0.87	0.074	0.865
1	4	1	13	0	116	1	0.86	0.074	0.868
20	1	17	2	0	115	0	0.85	0.059	0.868
30	2	5	5	6	115	2	0.85	0.074	0.866
40	8	4	112	0	6	5	0.83	0.125	0.864
43	1	11	5	1	108	9	0.80	0.141	0.862
38	2	6	11	107	4	5	0.79	0.096	0.865
21	10	102	1	18	0	4	0.76	0.066	0.868
<i>Analytical</i>									
<i>Mean value</i>							0.722	0.101	0.714**
2	131	1	0	2	1	0	0.97	0.029	0.867
9	2	128	1	3	1	0	0.95	0.007	0.869
28	124	3	0	2	3	3	0.92	0.037	0.866
8	122	2	3	5	3	0	0.90	0.088	0.865
33	121	3	0	2	1	8	0.90	0.081	0.863
18	10	3	4	114	4	0	0.84	0.052	0.867
24	0	11	5	109	10	0	0.81	0.044	0.868
42	5	3	4	110	3	10	0.81	0.125	0.862
32	3	2	6	17	103	4	0.76	0.118	0.866
35	6	18	91	7	6	7	0.67	0.163	0.862
41	7	85	14	20	0	9	0.63	0.149	0.862
13	5	81	10	31	8	0	0.60	0.088	0.868
29	11	12	14	78	13	7	0.58	0.170	0.862
6	7	28	76	23	0	1	0.56	0.066	0.870
34	13	7	12	14	71	18	0.53	0.2	0.861
25	45	5	57	3	24	1	0.42	0.141	0.865
37	6	57	9	22	34	7	0.42	0.163	0.865
<i>Statistical</i>									
<i>Mean value</i>							0.719	0.124	0.729**
17	126	4	0	4	1	0	0.93	0.052	0.867
16	4	3	1	125	1	1	0.93	0.029	0.868
7	8	123	1	2	1	0	0.91	0.037	0.868
19	1	3	117	0	14	0	0.87	0.059	0.867
10	4	1	11	3	116	0	0.86	0.104	0.865
36	114	7	4	2	2	6	0.84	0.111	0.862
31	13	94	14	3	9	2	0.70	0.170	0.863
39	5	91	8	21	3	7	0.67	0.148	0.862
3	52	82	0	1	0	0	0.61	0.170	0.863
14	38	3	81	1	11	1	0.60	0.192	0.863
11	10	9	29	72	14	1	0.53	0.111	0.868
15	15	68	7	37	4	4	0.50	0.229	0.860
26	53	19	13	5	44	1	0.40	0.207	0.862

*Cronbach's alpha is 0.868 for the full test.

**Mean value of Cronbach's alpha for this construct.

^ψ Item discrimination using the upper and lower 27% of the cohort

Analytical CN

Seventeen questions tested analytical numeracy. These questions were based primarily on the interpretation of data presented in charts, tables and graphs of various kinds. The questions also included computational elements. The mean score on analytical questions was 75%, with 11/135 (8.1%) of the FTs answering all 17 analytical questions correctly.

The simplest analytical question was based on a pie chart, which was answered correctly by 131/135 (97%) of the FTs. In contrast, a straightforward question based on the interpretation of a table from the latest National Institute for Health and Care Excellence (NICE) clinical guidance for prescribing intravenous (IV) fluids (NICE 2013) was answered correctly by only 76/135 (56%) of the FTs.

Statistical CN

Thirteen questions tested statistical numeracy with a focus on understanding probability data and screening test results. The mean score on statistical questions was 72%, with 10/135 (7.4%) answering all 13 statistical questions correctly.

Probability questions varied in difficulty: the easiest asked approximately how often a coin tossed 1000 times was likely to land heads up. The available answer options were: 25, 50, 250, 500 and 1000. Remarkably, 10/135 (7.4%) of the FTs selected incorrect answers.

Three questions asked candidates to select the better of two hypothetical treatment options. The same information was framed in three ways: as relative risk reduction (RRR), where one treatment reduced risk by 25% and the other by 10%; as absolute risk reduction (ARR), where one treatment reduced risk by 10 per 1000 people and the other by 4 per 1000; and as the number needed to treat (NNT), where either 100 or 250 people would need treatment for 5 years for one person to benefit. Although 104/135 (77%) of the FTs answered all three questions correctly, almost a quarter 31/135 (23%) answered at least one format incorrectly. This result confirms a framing effect, whereby the way in which the information is presented affects understanding. FTs found NNT most difficult to understand: only 84% of the FTs answered this format correctly compared to 93% for ARR, and 91% for RRR.

Three questions related to screening test results, including interpreting data on false positive and false negative rates. FTs found these questions difficult. The questions were answered correctly by 81/135 (60%), 68/135 (50%), and 53/135 (40%) respectively.

Discussion

Safe prescribing; accurate medical data interpretation; good, well-informed clinical decision-making—all are core skills for doctors. The errors made in numeracy-based test questions by FTs suggest a numeracy level that may cause these core skills to be insufficient for safe medical practice. FTs are an important cohort of doctors to study since they are on the front line of in-patient care.

FTs are also important as an educational window: as recent graduates, they represent the output of our medical schools. Participants in this study included 112 graduates from 27 UK medical schools, and so we believe that our results may be representative of current UK medical graduates.

Additionally, our findings appear to be generalisable, since in subsets of the MINT where results were available for comparison, FTs' performance was similar to that of medical students and doctors in previous U.S. studies (Sheridan and Pignone 2002; Gigerenzer et al. 2007; Anderson et al. 2011). Although other researchers have shown that medical students and doctors are prone to error in drug-dose calculation (Rowe et al. 1998; Simpson et al. 2009; Harries and Botha 2013), and in interpreting medical statistics (Sheridan and Pignone 2002; Windish et al. 2007; Gigerenzer et al. 2007; Wegwarth et al. 2012; Johnson et al. 2014), we are not aware of any other study that has explored clinician numeracy in comparable to depth to ours. Almost all previous studies have been based on short tests, with between four and eight questions: a size that calls into question the value of the tests, since short tests tend to be unreliable (Schuwirth and Van der Vleuten 2011). By contrast, the MINT is a 43-item assessment measure, and it has been developed specifically for healthcare professionals (Taylor and Byrne-Davis 2016).

We found that FTs had deficiencies in all three numeracy constructs tested: computational numeracy, analytical numeracy, and statistical numeracy. As shown in Table 1, deficiencies in each area can have important consequences for patients. The most frequent application of computational numeracy in medical practice is for calculating drug dosages; prescribing IV drugs and IV fluids is an everyday task for junior doctors in hospital practice. Our results show that 102/135 (75%) FTs made errors in questions related to drug-dose calculation. Decimal-place errors were common, leading to 10- and 100-fold errors in drug-dose calculation. In clinical practice, such errors could cause serious harm to patients.

The MINT also used four questions based on understanding nutritional labels to test computational numeracy. Poor performance in nutritional label tests is strongly associated with poor health literacy in patients (Weiss et al. 2005; Rothman et al. 2008). Nutritional label interpretation in doctors has not previously been assessed. We were surprised to find that only 76/135 (56%) FTs answered

all four questions requiring interpretation of a nutritional label correctly and, moreover, that three did not answer any correctly. This finding suggests that some doctors have deficiencies in health literacy. The finding warrants further investigation.

Many FTs had difficulty with questions testing data interpretation. We found evidence that this difficulty was related partly to deficiencies in analytical and statistical numeracy, but also to variations in data presentation or framing. Framing is important in relation to risk perception (Gigerenzer et al. 2007), and we observed a framing effect in the three treatment-selection questions presented as RRR, ARR, and NNT. Although over 90% answered the question correctly when framed as either ARR or RRR, and 84% answered NNT correctly, 31/135 (23%) of the FTs answered at least one format incorrectly. Such errors would lead to inappropriate treatment selection causing patient harm.

Screening patients for various diseases is common medical practice. Doctors are often confused by the statistics used to report the results of screening (Windish et al. 2007; Gigerenzer et al. 2007; Wegwarth et al. 2012). Three MINT questions were based on screening test results, where the presence or absence of a disease is expressed in terms of true and false positives and negatives; only 45/135 (33%) FTs answered all three correctly. Translated into clinical practice, such errors would cause these doctors to mislead patients regarding the likely presence or absence of disease, leading to needless anxiety or inappropriate reassurance.

Many FTs were confused by a question based on data taken from national guidance on IV fluid prescribing (NICE 2013). Poor performance on this question appeared to be related to poor data presentation: a table showing IV supplement requirements uses different units – while electrolytes are expressed in mg/kg/day, glucose is shown in g/day (NICE 2013). Many FTs calculated glucose requirement as 3500–7000 g/day, based on a g/kg/day calculation. We acknowledge that this answer raises questions about FTs' common sense, but it also demonstrates the confusion that may arise due to a lack of standardization in data presentation.

Although we recognize that errors in prescribing and in data interpretation are multifactorial, we believe that poor clinician numeracy is a possible contributory factor. Moreover, it is preventable. The onus is on the medical profession to determine a level of clinician numeracy that is commensurate with safe medical practice. Formal assessments of numeracy in medical students and doctors could then be required, including assessment at entry to, and exit from, medical school. We believe that education in clinician numeracy should be provided for all medical undergraduates and doctors in training, since lack of practice in mathematical skills leads to a decline in performance (Lee et al. 2010; McMullan et al. 2010).

Limitations

This study has some limitations, principally that our assessment of clinician numeracy is classroom rather than ward-based. However, we consider that FTs are more likely to demonstrate optimum performance in a quiet, classroom assessment with no distractions, than when dealing with multiple competing demands in a stressful clinical environment. Another limitation is that we did not allow participants to use calculators, even though calculators are commonly used in the clinical setting. However, calculators may not improve scores since they cannot overcome a) the participant's misinterpretation of a question; or b) the failure to apply the correct mathematical function. Furthermore, errors may be made when entering data (Bliss Holtz 1994). Nonetheless, we have planned a randomized controlled trial to determine whether calculators affect performance. Finally, we acknowledge that errors in maths tests may be unrelated to numeracy *per se*: we recognize the effect of framing, and that candidates may make careless mistakes, or misinterpret questions, and, moreover, that the lack of clarity in the text or data display of test questions may confuse candidates and cause errors. Analysis of our data suggests that two MINT questions fall into this latter category. In one analytical question, the graph provided was small, and this size appeared to hamper its interpretation by some participants; the graph has now been considerably enlarged. The wording of one question based on nutritional label interpretation appeared to confuse some participants; therefore, the text of this question has now been amended.

Conclusion

Many doctors have poor clinician numeracy, and are at risk of harming patients through errors in prescribing and data interpretation. Our results are consistent with, and build upon, existing evidence of poor numeracy in medical students and doctors worldwide. Further investigation into clinician numeracy is needed, and our current focus is investigating how and why medical students and doctors are making errors in our test. We consider this further study is essential in order to develop appropriate and effective educational interventions for those with deficiencies in numeracy. Furthermore, we consider that an appropriate standard of numeracy competence for doctors should be set: such a standard would ensure that greater attention is paid to clinician numeracy in both undergraduate and postgraduate medical curricula. We believe that such measures would contribute to reducing the hazards of healthcare for our patients.

Acknowledgments

We are grateful to Jim Mulherin for help with data analysis. We thank Professor Doug Corfield and our reviewers for their very helpful feedback on the first draft of this manuscript. We are indebted to Professor Len Vacher for his support, advice and encouragement.

References

- AMA. See American Medical Association
- American Medical Association. n.d. "Improving Patient Safety." Accessed May 29, 2017. <https://www.ama-assn.org/delivering-care/improving-patient-safety>.
- 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." *Genetics in Medicine* 13(8): 744-749. <https://doi.org/10.1097/GIM.0b013e31821a370f>.
- 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. <https://doi.org/10.1080/01421590400016290>.
- Bliss-Holtz, J. 1994. "Discriminating Types of Medication Calculation Errors in Nursing Practice." *Nursing Research* 43(6): 373-375.
- Brennan, T. A., L. L. Leape, N. M. Laird, L. Hebert, A. R. Localio, A. G. Lawthers, J. P. Newhouse, P. C. Weiler, and H. H. Hiatt. 1991. "Incidence of Adverse Events and Negligence in Hospitalized Patients: Results of the Harvard Medical Practice Study I." *New England Journal of Medicine* 324: 370-376. <http://doi.org/10.1136/qshc.2002.003822>.
- Campbell, I. 2007. "Chi-Squared and Fisher-Irwin tests of Two-By-Two Tables with Small Sample Recommendations." *Statistics in Medicine*, 26: 3661-3675. <https://doi.org/10.1002/sim.2832>
- Frontier Economics. 2014. *Exploring the Costs of Unsafe Care in the NHS: A Report Prepared for the Department of Health*. London: Frontier Economics Ltd. Accessed October 2015. <http://www.frontier-economics.com/publications/exploring-the-costs-of-unsafe-care-in-the-nhs/>.
- General Medical Council. 2014. *The State of Medical Education and Practice in the UK 2014*. Accessed October 2015. www.gmc-uk.org/publications/25452.asp.
- General Medical Council. 2015. *Promoting Excellence: Standards for Medical Education and Training*. Accessed October 2015. www.gmc-uk.org/education/standards.asp.

- 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 Public Interest* 8(2): 53–96. <https://doi.org/10.1111/j.1539-6053.2008.00033.x>.
- Gigerenzer, G., and M. Gray, eds. 2011. *Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020*. Cambridge, Mass: MIT Press. <https://doi.org/10.7551/mitpress/9780262016032.001.0001>.
- GMC. See General Medical Council.
- 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. <https://doi.org/10.1016/j.amepre.2005.06.012>
- Harries, C. S., and J. H. Botha. 2013. "Can Medical Students Calculate Drug Doses?" *Southern African Journal Anaesthesia and Analgesia* 19(5): 248–251. <http://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): Article 124. <https://doi.org/10.5688/ajpe776124>.
- 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): Article 96. <https://doi.org/10.1186/1472-6963-8-96>.
- 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. <https://doi.org/10.1016/j.jsurg.2013.07.013>.
- Kirch, D. G., and Boysen P.G. 2010. "Changing the Culture in Medical Education to Teach Patient Safety." *Health Affairs* 29(9): 1600–1604. <http://doi.org/10.1377/hlthaff.2010.0776>.
- 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. <https://doi.org/10.5688%2Fajpe776124>.
- Leape, L. L., and D. M. Berwick. 2005. "Five Years after *To Err Is Human*: What Have We Learned?" *Journal of the American Medical Association* 293(19): 2384–2390. <https://doi.org/10.1001/jama.293.19.2384>.
- 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." Report prepared by Mathematics in Education and

- Industry. Published by The Higher Education Academy Engineering Subject Centre. <http://mei.org.uk/files/pdf/pre-university-maths-guide-2016.pdf>
- Malhotra A., D. Maughan, J. Ansell, R. Lehman, A. Henderson, M. Gray, T. Stephenson and S. Bailey. 2015. "Choosing Wisely in the UK: the Academy of Medical Royal Colleges' Initiative to Reduce the Harms of too Much Medicine." *British Medical Journal*. BMJ 2015;350,h2308. Accessed May 16, 2015. <https://doi.org/10.1136/bmj.h2308>.
- McMullan, M., R. Jones, and S. Lea. 2010. "Patient Safety: Numerical Skills and Drug Calculation Abilities of Nursing Students and Registered Nurses." *Journal Advanced Nursing* 66(4): 891–899. <https://doi.org/10.1111/j.1365-2648.2010.05258.x>.
- National Institute for Health and Care Excellence. 2013–2017. "Intravenous Fluid Therapy in Adults in Hospital." *Clinical Guideline* CG174. Accessed October 2015. <https://www.nice.org.uk/guidance/cg174>.
- National Patient Safety Agency. 2010. "About Us." Accessed May 29, 2017. <http://www.npsa.nhs.uk/corporate/about-us/>.
- National Patient Safety Foundation. n.d. "NPSF: Creating a World Where Patients and Those Who Care for Them Are Free From Harm." Accessed May 29, 2017. <http://www.npsf.org/>.
- Neale, G., M. Woloshynowych, and C. Vincent. 2001. "Exploring the Causes of Adverse Events in NHS Hospital Practice." *Journal of the Royal Society of Medicine* 94(7): 322–330.
- NICE. See National Institute for Health and Care Excellence.
- NPSA. See National Patient Safety Agency.
- NPSF. See National Patient Safety Foundation.
- 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 Preventive Medicine* 31 (5): 391–398. <https://doi.org/10.1016/j.amepre.2006.07.025>.
- Rowe, C., T. Koren, and G. Koren. 1998. "Errors by Paediatric Residents in Calculating Drug Doses." *Archives Disease Childhood* 79(1): 56–58.
- 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. <https://doi.org/10.3109/0142159X.2011.611022>.
- 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 Internal Medicine* 127(11): 966–972.
- Seden, K., J. J. Kirkham, T. Kennedy, M. Lloyd, S. James, A. Mcmanus, A. Ritchings, J. Simpson, D. Thornton, A. Gill, C. Coleman, B. Thorpe, and S. H. Khoo. 2013. "Cross-Sectional Study of Prescribing Errors in Patients Admitted

- to Nine Hospitals across North West England.” *BMJ Open* 3(1), e002036. Accessed October 17, 2015. <https://doi.org/10.1136/bmjopen-2012-002036>.
- Sheridan, S. L., and M. Pignone. 2002. “Numeracy and the Medical Student’s Ability to Interpret Data.” *Effective Clinical Practice* 5(1): 35–40.
- Sheridan, S. L., M. Pignone, and C.L. Lewis. 2003. “A Randomized Comparison of Patients’ Understanding of Number Needed to Treat and Other Common Risk Reduction Formats.” *Journal General Internal Medicine* 18(11): 884–892.
- 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): Article 5. <https://doi.org/10.5038/1936-4660.4.2.5>.
- Simpson, C. M., G. B. Keijzers, and J. F. Lind. 2009. “A Survey of Drug-Dose Calculation Skills of Australian Tertiary Hospital Doctors.” *The Medical Journal of Australia* 190(3): 117–120.
- Taylor A. A., and L. M. Byrne-Davis. 2016. “Clinician Numeracy: The Development of an Assessment Measure for Doctors.” *Numeracy* 9(1): Article 5. <https://doi.org/10.5038/1936-4660.9.1.5>.
- Vincent, C., G. Neale, G. and M. Woloshynowych. 2001. “Adverse Events in British Hospitals: Preliminary Retrospective Record Review.” *British Medical Journal* 322(7285): 517–519.
- Vincent, C., S. Burnett, and J. Carthey. 2014. “Safety Measurement and Monitoring in Healthcare: A Framework to Guide Clinical Teams and Healthcare Organisations in Maintaining Safety.” *BMJ Quality & Safety* 23(8): 670–677. <http://doi.org/10.1136/bmjqs-2013-002757>.
- Wegwarth, O., L. M. Schwartz, S. Woloshin, W. Gaissmaier, and G. Gigerenzer. 2012. “Do Physicians Understand Cancer Screening Statistics? A National Survey of Primary Care Physicians in the United States.” *Ann Intern Med.* 156: 340–349. <https://doi.org/10.7326/0003-4819-156-5-201203060-00005>.
- 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.” *Annals Family Medicine* 3(6): 514–522. <https://doi.org/10.1370/afm.405>.
- Windish, D. M., S. J. Huot, and M. L. Green. 2007. “Medicine Residents’ Understanding of the Biostatistics and Results in the Medical Literature.” *Journal American Medical Association* 298(9): 1010–1022. <https://doi.org/10.1001/jama.298.9.1010>.
- Wright, K. 2004. “An Investigation to Find Strategies to Improve Student Nurse Maths Skills.” *British Journal of Nursing* 13(21): 1280–1284. <https://doi.org/10.12968/bjon.2004.13.21.17114>.

APPENDIX (from Taylor and Byrne-Davis 2016)

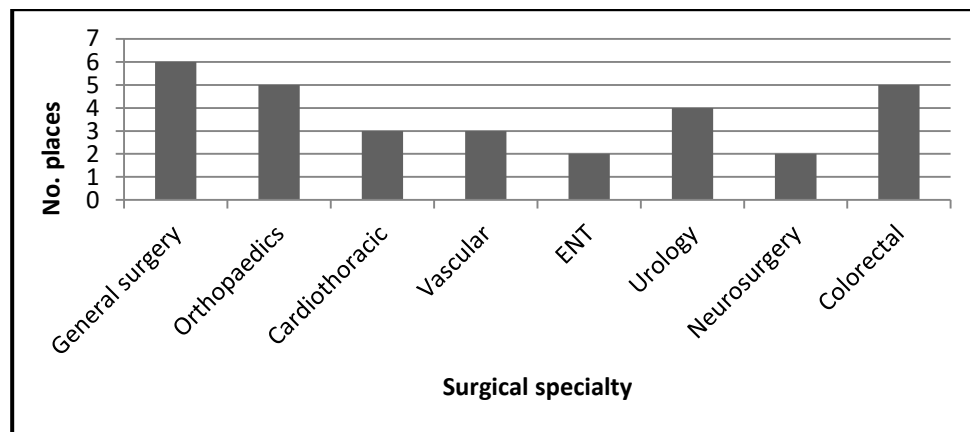
MINT: sample questions

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 6g of carbohydrate for every 30 mins she exercises. She has some biscuits in her gym bag. Each biscuit contains 8g 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:

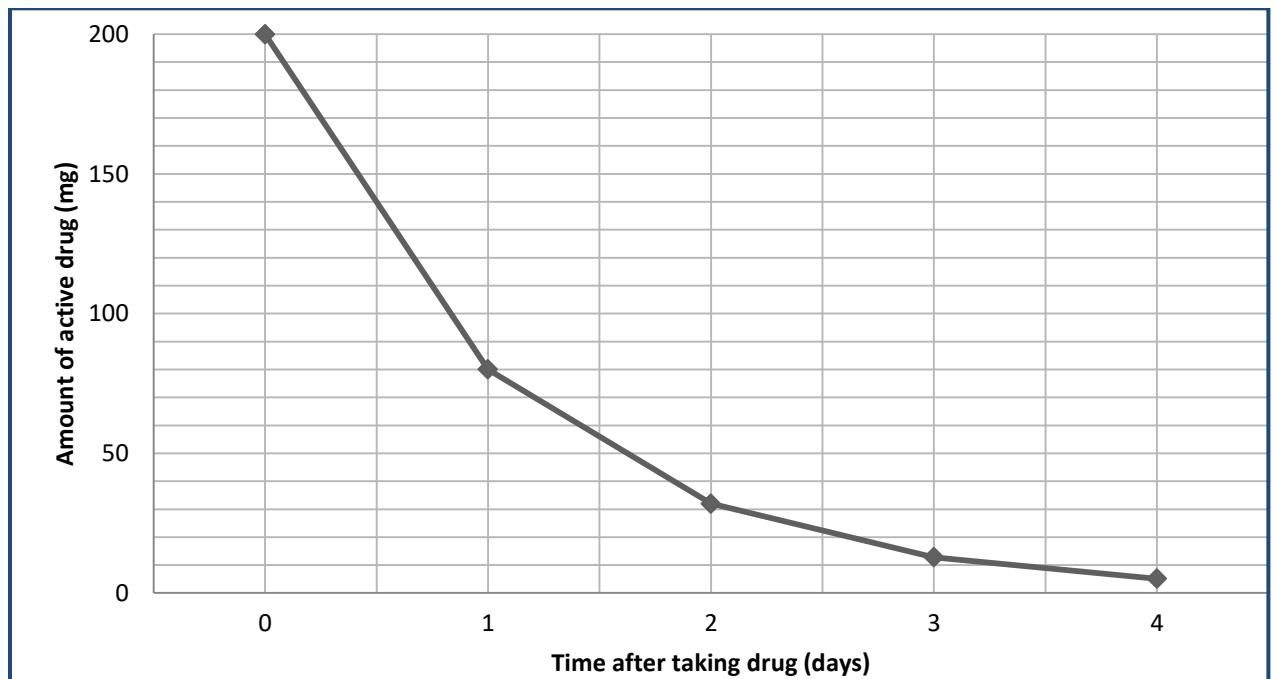
Day	Urine output (ml)
Monday	532
Tuesday	472
Wednesday	472
Thursday	364

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 200mg 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

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 100kg, 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 2mg/kg. What is the maximum amount of bupivacaine you can use when suturing Mo's wound?

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