The performance of junior doctors in applying clinical pharmacology knowledge and prescribing skills to standardized clinical cases

Sam Harding,[1] Nicky Britten[2] & David Bristow[3]

[1]Peninsula Medical School, University of Plymouth, Drakes Circus, Plymouth PL4 8 AA, [2]Institute of Health Service Research, Peninsula Medical School, University of Exeter, St Luke's Campus, Exeter EX1 2LU and [3]Reader in Pharmacology and Clinical Education, Director of Undergraduate Medical Studies, Peninsula Medical School, University of Plymouth, Drakes Circus, Plymouth PL4 8 AA, United Kingdom

WHAT IS ALREADY KNOWN ABOUT THIS SUBJECT

- Safe prescribing is a core competency in undergraduate medical education.
- A large proportion of undergraduate medical students and recently graduated doctors in the UK are not confident in their ability to prescribe effectively and safely.
- Errors are common in all healthcare settings and prescribing errors are the most common type.

WHAT THIS STUDY ADDS

- This study produced twelve valid and statistically reliable assessments of clinical pharmacology and therapeutics (CPT) knowledge and prescribing skills in areas that pose a high risk to patient safety.
- The findings show that a large proportion of foundation year 1 (FY1) doctors fail to demonstrate the level of CPT knowledge and prescribing ability judged by a subject matter expert (SME) panel to be required at this stage of their careers.
- We suggest strategies and areas where teaching can be focused to improve the safety and effectiveness of prescribing by FY1 doctors.

AIMS

Recent studies suggest a worryingly high proportion of final year medical students and new doctors feel unprepared for effective and safe prescribing. Little research has been undertaken on UK junior doctors to see if these perceptions translate into unsafe prescribing practice. We aimed to measure the performance of foundation year 1 (FY1) doctors in applying clinical pharmacology and therapeutics (CPT) knowledge and prescribing skills using standardized clinical cases.

METHODS

A subject matter expert (SME) panel constructed a blueprint, and from these, twelve assessments focusing on areas posing high risk to patient safety and deemed as essential for FY1 doctors to know were chosen. Assessments comprised six extended matching questions (EMQs) and six written unobserved structured clinical examinations (WUSCEs) covering seven CPT domains. Two of each assessment types were administered over three time points to 128 FY1 doctors.

RESULTS

The twelve assessments were valid and statistically reliable. Across seven CPT areas tested 51–75% of FY1 doctors failed EMQs and 27–70% failed WUSCEs. The WUSCEs showed three performance trends; 30% of FY1 doctors consistently performing poorly, 50% performing around the passing score, and 20% performing consistently well. Categorical rating of the WUSCEs revealed 5% (8/161) of scripts contained errors deemed as potentially lethal.

CONCLUSIONS

This study showed that a large proportion of FY1 doctors failed to demonstrate the level of CPT knowledge and prescribing ability required at this stage of their careers. We identified areas of performance weakness that posed high risk to patient safety and suggested ways to improve the prescribing by FY1 doctors.

Introduction

The ability to prescribe safely and effectively is a core competency in undergraduate medical education [1]. In 2008 the Medical School Council's Safe Prescribing Working Group identified eight competencies in relation to knowledge and skills in prescribing required by all foundation doctors [2]. Attaining these competencies remains a challenge to medical educationalists, as research suggests that clinical pharmacology and therapeutics (CPT) teaching and assessment appears to leave students poorly prepared to apply their knowledge in a clinical setting [3–6].

Errors are common in all healthcare settings. In a systematic review of the incidence and nature of in-hospital adverse events, over 9% of the patients admitted to hospital were harmed by error [7]. Whilst error can occur at any stage of the medicine-use process [8], prescribing errors are the most common type in both hospital [9, 10] and primary care settings [11, 12].

A number of studies have found that a significant proportion of junior doctors feel unprepared to prescribe safely [4–6], and report unsafe prescribing behaviours [3]. Undergraduate medical students express similar concerns about their preparedness to prescribe, and a proportion of students rate their knowledge of CPT as poor [3]. The work of Coombes *et al.* and Heaton *et al.* suggest that many medical students have little confidence in their prescribing abilities and fear that their abilities in CPT are insufficient to meet the competencies identified by the GMC [1, 4, 13–15], even in prescribing common medicines.

The perceptions of the medical students and junior doctors are supported by the anecdotes of their seniors and other health care professionals [16], although it is also argued that there is insufficient evidence to support a general conclusion of under preparation of UK medical graduates for prescribing [17]. A study of junior doctors' prescribing practices in an Australian medical school did show inappropriate prescribing in two-thirds of new medical graduates at the beginning of their intern year [18–20].

The aim of this new research was to provide objective data on the ability of UK foundation year 1 (FY1) doctors to apply safely and effectively CPT knowledge and prescribing skills to standardized clinical cases.

Methods

Ethics

Ethics approval was granted by the Derriford Hospital Trusts Research Development and Support Unit and the South West Devon Research Ethics Committee (LREC Number: 06/Q2103/99).

Participants

Junior doctors were identified through three postgraduate teaching centres in South West (UK) teaching hospitals: Royal Cornwall Hospital, Truro; Derriford Hospital, Plymouth; and Royal Devon and Exeter Hospital, Exeter. Medical graduates entering the FY1 programme in the South West Deanery and participating in this study had graduated from medical schools throughout the UK (see online supplement for more detail). Each hospital provided 2.5 h of CPT teaching within a month of the FY1 (first year following graduation from medical school) doctors starting work.

Assessors

The assessment tools were constructed by a panel of subject matter experts (SME). The panel consisted of five clinical pharmacists (one with a special interest in education), two medical consultants (one a professor of Clinical Pharmacology and the other with a special interest in Clinical Pharmacology), three specialist registrars (one geriatric medicine registrar with special interest in clinical pharmacology, one acute medicine registrar and one general surgeon), four FY1 doctors, and one psychologist.

An independent examiner panel was also convened with experts having similar credentials to the SME panel, who were not involved in the development of the assessment items, or the first time marking of the assessments. This panel consisted of two clinical pharmacists, one medical consultant, two medical specialist registrars and two senior house officers.

Development of assessment tools

Effectively a FY1 doctor needs to be competent in all areas of CPT as outlined in previous publications [1, 14]. The research described in this paper required the identification of a smaller set of domains to develop the assessment tools. Therefore the SME constructed a blueprint of medical specialties and areas of CPT. Each member of the SME panel was asked, using their expertise, to select and rank in order of greatest importance areas from the blueprint which represented the domains of CPT they believed the junior doctors must be competent in at the start of their clinical (in-hospital) career. They were informed that in this process competence was defined by demonstrating safe and effective application of CPT knowledge to ensure patient safety and minimize drug error. The individual SME choices were then collated and the areas with the greatest number of responses were selected for item development. The blueprint matrix is accessible in the online supplement.

The final assessment topics from the blueprint domains were: (i) renal failure and fluid management, (ii) analgesia in emergency medicine, (iii) anti-coagulants and analgesia in the post-operative patient, (iv) diverticular disease with i.v. administration of antibiotics and fluids, (v) contraindications in the use of anti-coagulants and antibiotics, (vi) respiratory, chronic obstructive pulmonary disease and (vii) diabetes, insulin selection.

Prior to constructing the assessment items the SME panel decided upon a test time of 5 min for each extended matching question (EMQ) and 10 min for each written unobserved structured clinical examination (WUSCE). The SME panel members then used an iterative process to write and review the assessments and the marking schemes. The varying levels of clinical experience brought by the SME allowed them to gauge the level/difficulty of the questions, as well as associating them with current in-hospital practice [21]. Through this process the items were set at a level the panel felt a junior doctor should be able to achieve. We introduced an option in the WUSCE marking schemes for a doctor to fail the question due to a major drug or safety error, even if the majority of the question was correctly answered. Therefore at the end of each item the assessor was asked to note any part of the answer they considered to be suboptimal (i.e. not the 'best' choice of intervention/treatment, route, dose, and timing, but nothing overtly dangerous), dangerous (a choice of intervention/treatment, route, dose and timing, which could cause serious, but not life threatening injury to the patient) or lethal (a choice of intervention/treatment, route, dose, and timing, which could cause the death of the patient).

Six WUSCEs and six R-Type EMQs were constructed during a total of 18 meetings. This meant that not all seven domains were represented by all 'types' of assessment. It also limited the broader application of the results to the areas which the SME believed the junior doctors *must* be competent in at the start of their first clinical year and which posed the greatest risk to patient safety (see online supplement for figure showing development iterations and examples of item formats).

Administration of assessments

Three time points separated by 5 months were identified over 1 year of the FY1 clinical practice where the developed assessments could be administered. The FY1s were recruited at weekly compulsory teaching sessions at each of the teaching hospitals. Each site was visited on 2 consecutive weeks during each time point. This allowed for an EMQ and WUSCE to be administered each week, leading to two of each assessment items at each time point (e.g. time point 1; WUSCE 1 and 2 and EMQ 1 and 2). The items were administered to the FY1 doctors in the order in which the SME panel developed them (see online supplement for figure showing development iterations). There was an allowance of 10 min for the completion of the WUSCE and 5 min for the EMQ. The same assessments were carried out at all sites within a week.

Marking

The correct responses to the EMQ were established at the time of their development, and a pass mark of 70% was set using the Angoff methodology [22]. The WUSCEs were double marked by a clinical pharmacist and a clinician who were both part of the SME panel. The marking scheme

had both numerical and categorical aspects. A categorical judgement was made about whether, in their expert opinion, the response to the whole WUSCE would be classified as a pass or fail. They were also asked to note the absence or presence of anything considered 'suboptimal', 'dangerous' or 'lethal'. A numerical score was obtained by evaluating specific elements and completion of each sub question within the WUSCE. The independent examiner panel was asked to resolve any disagreement between marking decisions of the SME panel members in relation to WUSCE pass/fail decisions, and to review a random selection of 10% of WUSCE scripts (a marking scheme and further details of the standard setting is given in the online supplement).

Data analysis

Collected data were entered and stored in an SPSS data file (SPSS V15). The data were exported into Minitab and Excel for analyses that were not possible in SPSS. The statistical methods used to check reliability and validity of the developed items were Cronbach's alpha, intraclass correlation coefficients (ICC) and analysis of variance (ANOVA). Analysis of the results was achieved through descriptive statistics, mean averages, Spearmans rho correlation coefficient and plotting charts to visually check patterns of attainment on both EMQs and WUSCEs.

Results

One hundred and twenty-eight FY1 doctors completed one or more of the CPT assessments. The average age of these was 26 years, with 49 (41.5%) being male.

Validity of assessment items and marking

The agreement of assessors from different disciplines was assessed using the Kappa statistic [23]. It is a preferable statistical measure to simply using percentage agreement because it corrects for the probability that raters agree by chance alone. When Kappa was calculated for WUSCE 1–6, the coefficients of 1.000, 1.000, 0.876, 1.000, 1.000 and 0.917 were found, respectively. This indicated that a good level of agreement was reached. Any disagreements were resolved by the independent examiner panel.

The Cronbach's alpha of the EMQs ranged from 0.55 to 0.78 and the WUSCEs from 0.74 to 0.80, indicating a 'good' scale had been constructed [24]. The calculated ICC of the scorers for individual WUSCEs ranged from 0.83 to 0.98 indicating a very good level of agreement between the clinical pharmacist and clinician [24]. The ANOVA calculating inter-rater reliability supported the ICC showing no difference between the assessors' marking ($F = 0.03$; $P = 0.86$). The G statistic for WUSCE was 0.85, indicating 'good' reliability [25, 26].

Performance of FY1 doctors in CPT EMQ

Figure 1 shows the spread of the scores obtained on each EMQ. It shows that on five out of the six EMQs the mean
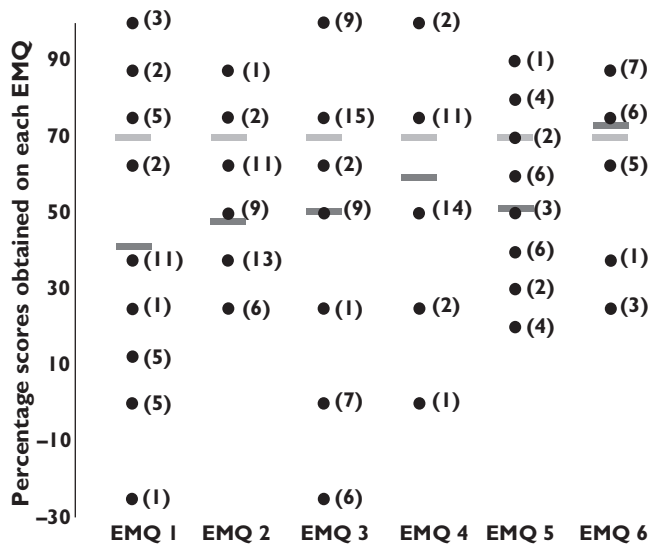
Figure 1

The spread of percentage scores obtained by individual FY1 doctors on each EMQ and the relationship to the mean average percentage score and pass mark (70%). Confidence intervals are not shown in this figure.
Key: ▬▬ = Pass mark of assessment item
▬▬ = Mean average percentage mark obtained by FY1 on each assessment item
EMQ 1 topic = Contra-indications in the use of anti-coagulates and antibiotics
EMQ 2 topic = Anti-coagulants and analgesics in the post operative patient
EMQ 3 topic = Diverticular disease; i.v. administration of antibiotics
EMQ 4 topic = Respiratory; Chronic obstructive pulmonary disease
EMQ 5 topic = Diabetes; insulin selection
EMQ 6 topic = Analgesia in emergency medicine
Numbers in brackets e.g. (1), indicate the number of individual FY1 doctors who obtained the particular score on each question
Number of FY1 doctors completing the assessment item; EMQ 1 = 35, EMQ 2 = 42, EMQ 3 = 49, EMQ 4 = 30, EMQ 5 = 28, EMQ 6 = 26.
Maximum possible scores for each question are: EMQ 1 = 8, EMQ 2 = 8, EMQ 3 = 8, EMQ 4 = 8, EMQ 5 = 10, EMQ 6 = 8.
Minimum possible scores for each question are: EMQ 1 = −8, EMQ 2 = −8, EMQ 3 = −8, EMQ 4 = −8, EMQ 5 = −10, EMQ 6 = −8.
Range of scores on each question (min;max): EMQ 1 = −2 to 8, EMQ 2 = 2 to 7, EMQ 3 = −2 to 8, EMQ 4 = 0 to 8, EMQ 5 = 2 to 9, EMQ 6 = 2 to 7

total score on the question was below the pass mark. The scores plotted for a sample of FY1 doctors indicated that their knowledge varied across each individual EMQ.

No pattern of attainment (e.g. FY1s consistently performing higher or lower than the mean and pass mark scores) could be ascertained from looking at individual or groups of responses. The percentage of FY1 doctors failing the individual EMQs is given in Table 1.

Performance of FY1 doctors in CPT WUSCE

Figure 2 shows the spread of the scores obtained on each WUSCE. Different WUSCEs have different maximum and minimum possible scores. The figure shows that on four out of the six items the mean total score on the question was below the pass mark.

The WUSCE score for FY1 doctors completing the majority of WUSCEs was plotted and suggested that an individual doctor tended to maintain a similar level of performance across all the WUSCEs. The WUSCE data were further analyzed to assess any relationship between WUSCE scores as a percentage of the possible obtainable score for each WUSCE. Three profiles emerged (Figure 3) with under 20% of FY1 doctors consistently performing higher than the mean and pass mark scores. The middle performers represented about 50% of the FY1 doctors usually scoring around the mean, but tending to be just below the pass mark, and just over a 30% consistently performed poorly.

FY1 doctors judged to have passed individual WUSCEs can have made 'suboptimal' errors, but not dangerous or lethal ones. The number of FY1 doctors who completed each question and the percentages that were judged to have failed in each of the topic areas are shown in Table 1.

No significant correlation was found between the FY1 doctors' marks on EMQs and WUSCEs. There was also no significant correlation found when the data were assessed to see if there was a relationship between pass/fail decisions on individual EMQs and the corresponding WUSCEs.

All those failing the WUSCE questions made at least 'suboptimal' drug or safety errors and some candidates who passed an item also recorded 'suboptimal' errors. These were errors in process tasks such as not completing patient details, failing to complete all sections of a drug chart appropriately, or selecting a drug that was appropriate but not deemed the 'best' drug for the scenario. No-one marked with a dangerous or lethal decision passed a WUSCE (Table 2). The types of decisions made by the assessors as 'dangerous' were not reacting to investigation results when presented or not treating a significant part of the medical problem such as atrial fibrillation. From the one hundred and sixty-one completed WUSCEs, there were eight responses (5%) which were considered by assessors as potentially 'lethal'. These errors were in unsafe medication choices, including the inappropriate use of drugs such as β-adrenoceptor blockers, anti-coagulants and prescribing amoxicillin to people with penicillin allergy. Only twenty-three WUSCEs (14%) were categorized by the SME as not having any level of error.

Discussion

There is a widespread perception from juniors doctors, senior healthcare professionals and medical educationalists, that junior doctors are not adequately prepared for prescribing at the outset of their careers [3–6, 15, 16]. A lack of skills and knowledge in CPT has been suggested as contributing to prescribing errors and to compromising patient safety [10, 15, 27–29]. This study has revealed that a large proportion of those FY1 doctors tested were unable to demonstrate safe and effective application of CPT

Table 1

Percentage of FY1 doctors judged to have failed each assessment

	EMQ % failed	n EMQ	WUSCE % failed	n WUSCE WUSCE	Comparison*
EMQ 1 and WUSCE 5 – **Contra-indications in the use anti-coagulants and antibiotics**	65.7%	35	68.0%	20	6
EMQ 2 and WUSCE 3 – **Anti-coagulants and analgesia in the post-operative patient**	66.7%	42	61.1%	25	7
EMQ 3 and WUSCE 4 – **Diverticular disease; i.v. administration of antibiotics and fluids**	51.0%	49	26.9%	21	19
EMQ 4 and WUSCE 6 – **Respiratory; chronic obstructive pulmonary disease**	56.7%	30	40.0%	18	8
EMQ 5 – **Diabetes; insulin selection**	75.0%	28	–	–	–
EMQ 6 and WUSCE 2 – **Analgesia in emergency medicine**	57.7%	26	70.3%	33	5
WUSCE 1 – **Renal failure; fluid management and arteriopathy**	–	–	61.7%	44	–

Table 1 shows how the EMQs and WUSCEs relate to each other in terms of the medical domain they assess. The values are the percentage of the FY1s completing each assessment item who were deemed to have failed the item (EMQ and WUSCE). It also shows the number of FY1s who actually completed each of the items. *The number of individual FY1 doctors who completed both the WUSCE and EMQ and were therefore available for Spearman Rho correlation analysis.
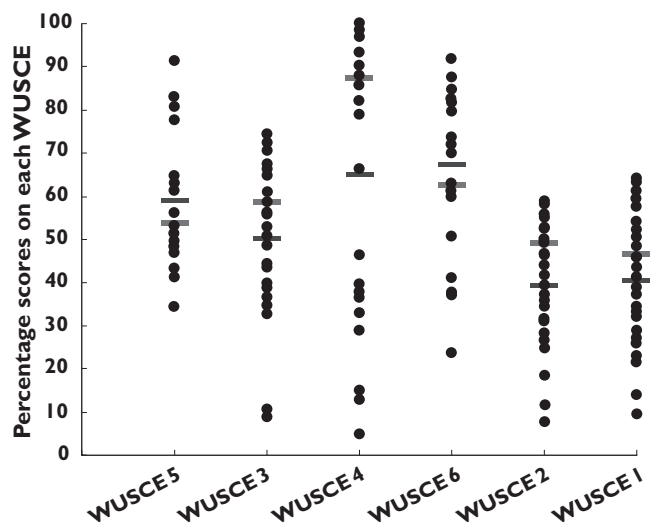
Figure 2

The spread of percentage scores obtained by individual FY1 doctors across each WUSCE. The possible scores in each WUSCE vary and are given in the key (below). The figure also shows the relationship of the mean percentage average scores obtained and the pass percentage for each WUSCE. Confidence intervals are not shown in this figure.

Key: ▬ = Pass mark of assessment item
▬ = Mean average mark obtained by FY1 on each assessment item
WUSCE 5 topic = Contra-indications in the use of anti-coagulates and antibiotics
WUSCE 3 topic = Anti-coagulants and analgesics in the post-operative patient
WUSCE 4 topic = Diverticular disease; i.v. administration of antibiotics
WUSCE 6 topic = Respiratory; Chronic obstructive pulmonary disease
WUSCE 2 topic = Analgesia in emergency medicine
WUSCE 1 topic = Renal failure; fluid management and artiopathy
Number of FY1 doctors completing the assessment item = WUSCE 1 = 44, WUSCE 2 = 33, WUSCE 3 = 25, WUSCE 4 = 21, WUSCE 5 = 20, WUSCE 6 = 18.
Maximum possible scores for each question are: WUSCE 1 = 69.8, WUSCE 2 = 43, WUSCE 3 = 44.5, WUSCE 4 = 14.9, WUSCE 5 = 46.5, WUSCE 6 = 21.6.
Minimum possible scores for each question are: WUSCE 1 = –13, WUSCE 2 = 0, WUSCE 3 = 0, WUSCE 4 = 0, WUSCE 5 = –2, WUSCE 6 = –4.
Range of scores on each question (max;min): WUSCE 1 = 6.0 to 45.5, WUSCE 2 = 3.3 to 25.5, WUSCE 3 = 4.0 to 33.1, WUSCE 4 = 0.8 to 14.9, WUSCE 5 = 16.0 to 42.5, WUSCE 6 = 5.2 to 19.8
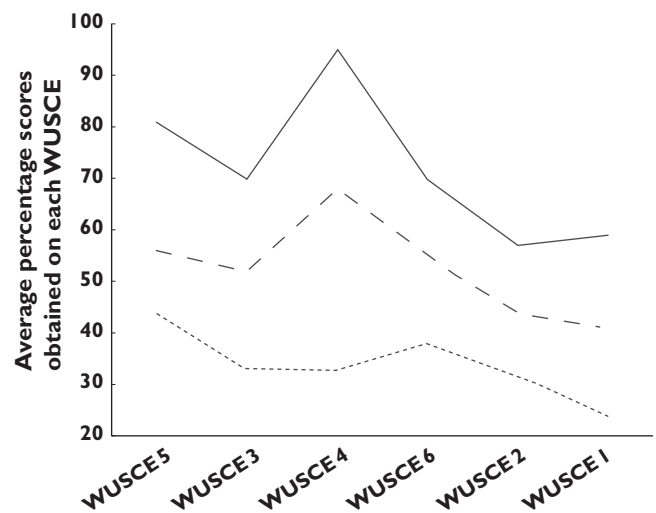
Figure 3

The profiles which emerged from visual inspection of individual FY1 doctors percentage pass marks. 'Best performers' group constituted less than 20% of FY1 doctors. 'Middle performers' represented approximately 50% of the FY1 doctors. 'Worst performers' are made up of slightly more than 30% of FY1 doctors completing each WUSCE. Best performers (—); Middle performers (– –); Worst performers (······)

knowledge and prescribing skills in areas judged by the SME panel to be required at this point in their careers.

Our study recognized from the outset that it was not feasible to assess the performance of FY1 doctors in all domains within the undergraduate medical CPT curriculum. Our approach used an SME panel, familiar with the current expected (authentic) clinical demands placed on a FY1 doctor, to construct a curriculum blueprint of areas deemed essential to the role of FY1 doctor and from this to identify those domains posing the highest risk to patient safety or for harmful drug errors. The iterative assessment development process operationalized the SME panels' judgements on the content of the assessments to be developed, and the methodology allowed for an appropri-

Table 2

Categorical analysis of errors made in WUSCEs

WUSCE	Suboptimal	Dangerous	Lethal
1	31	10	2
2	24	4	3
3	18	6	3
4	7	2	0
5	16	1	0
6	10	1	0

Table 2 shows the number of FY1s who the assessor deemed to have made 'suboptimal', 'dangerous' or potentially 'lethal' errors during the completion of the individual WUSCEs. The values are the number of each level of error identified by the assessors. However, one FY1 doctor can be represented in all three of the levels in WUSCE 1. The total number of WUSCEs taken was 161.

ate sampling of blueprinted domains to assess the competence of the FY1 doctors in applying knowledge in CPT and prescribing skills in these high risk domains.

The expertise from the initial SME group who constructed the assessment items and the independent examiners who graded the responses, provided face and content validation to the format, content and level of difficulty of the assessment items, as well as the outcomes. The assessment tools have been shown to be statistically reliable as well as valid. Cronbach's alpha was calculated to assess the internal consistency of test items that could not solely be answered as being right or wrong. The Cronbach's alpha for this analysis demonstrated that both the EMQs and WUSCEs had good internal consistency [24]. It is important in an assessment where multiple people may be asked to judge the performance of the candidates, that their judgements are comparable. The inter-rater reliability in this study was evaluated by the use of ICC and showed no difference between the assessors. This demonstrated that the methodology used to construct the marking schemes was effective and efficient both in allowing accurate assignment of a score for the assessment items, while allowing markers to demonstrate their expert status [21]. Another important facet of assessment construction is generalizability (G) theory. This evaluates how well the observed scores reflect corresponding true scores [30]. The figure of G = 0.85 obtained in this study indicated that there was a good level of reliability of the test items and that the majority of variance was attributable to the responses of the participating FY1 doctors. The statistical measures used in this study all indicated that the developed items were reliable and that a majority of the variation in the scores was due to the FY1 doctors' performance rather than the assessors.

In this study we assessed performance in a number of key areas of CPT and found that approximately two-thirds of the FY1 doctors were prescribing and applying CPT knowledge inappropriately (Figures 1,2, Table 1). A study performed in Australia looked at some of the same areas of CPT (postoperative analgesia, asthma and COPD, commu-nity acquired pneumonia) and found similar results at the start of their first clinical year [18]. In the Australian study 75% of the junior doctors were prescribing 'appropriately' for all conditions by the end of their first year [18]. Although we have not constructed this study to measure longitudinal effects over the FY1 year, we did observe a trend of improved mean average performance in WUSCEs and EMQs (Figures 1, 2) and, more markedly, an apparent reduction in error making during WUSCEs taken across the FY1 year (Table 2). This finding is consistent with conclusions from other work [15]. Thus, we would cautiously predict that the FY1 doctors appear to behave in a similar way to the Australian interns and that the majority will improve their prescribing performance over the FY1 year. However, specifically designed longitudinal work with the developed EMQs and WUSCEs is needed to confirm this.

The results of this study add to concern raised by previous research that has shown newly qualified doctors taking sole responsibility for around 20–35% of the prescriptions that they chart [31, 32]. The WUSCEs in particular enabled six out of the eight of core drugs and therapeutic problems as outlined by Maxwell & Walley [33] and the Medical School Council Safe Prescribing working group [2] to be assessed. The non-completion of prescribing charts was one error that the SME judged as 'suboptimal' in the WUSCE performance. It is important that the medication is not merely safe and effective for the patient, but that the process of prescribing captures relevant information that is legible, unambiguous and complete [2]. The findings from the WUSCEs showed that the majority of FY1s did not complete the paperwork appropriately (suboptimal), failed to comply with policy and procedure, failed to change current medication following test results and failed to follow the five rights of medication (right patient, right drug, right dose, right time and right route), to a safe level [34]. These latter failures (five rights of medication) were the type judged as 'dangerous', during the marking of the WUSCE. This suggests that in part, the familiarization with all types of paperwork that an FY1 may be expected to complete would be a beneficial area of teaching prior to graduating from medical school. Adopting a systems theory approach would suggest that a positive development may be to standardize nationally hospital prescription charts [2]. A systems intervention that could be introduced more rapidly would be provision of standardized core competency teaching and familiarization of charts, policies and procedures to all medical undergraduates.

Prescribing has been described as consisting of two related but distinct areas, the basic pharmacological knowledge required to understand drug effects and interactions, and the actual mechanics of prescribing [10, 15]. This differentiation of knowledge and application of knowledge was achieved with the use of two different question formats: EMQs assessing knowledge recall (Table 1 and Figure 1), and WUSCEs looking at application

of CPT knowledge and prescribing skills (Table 1 and Figure 2). The finding that competence (and level of knowledge) as defined by the EMQs, was patchy across all FY1 doctors participating in this research (Figure 1) did not help to identify areas of performance weakness, or guide medical schools to provide greater or different teaching in relation to the tested areas. However, the WUSCEs were able to distinguish between the best and worst FY1 doctors at applying their CPT knowledge and prescribing skills (Table 1 and Figures 2,3). The lack of correlation between the EMQs and WUSCEs is explained, at least in part, by the EMQs being less able to discriminate between the application of CPT knowledge than the WUSCEs. The results obtained in this study are in agreement with the experts' perceptions and the junior doctors' own opinions about their level of preparedness, namely, that the majority demonstrate weakness in CPT knowledge and prescribing skills [3–6, 13, 18].

Although many of the FY1 doctors completing the WUSCEs failed them, our results suggest that it is a minority of respondents who consistently perform poorly, and would be designated as following the 'worst performers' profile in Figure 3. It is possible that some of those failing and categorized as a 'middle performer' (about half of the FY1 doctors) would subsequently pass if hospital paperwork was standardized and they were familiarized with it prior to leaving medical school [10], whereas the 'worst performers' are lacking the ability to apply CPT knowledge as well as prescribing skills [2]. It is also of note that a major source of error in the WUSCEs was related to the act of prescribing, although most of the mistakes were minor. However even these minor errors have the potential to do harm and place the patient at risk [7, 8, 35].

The SME panel in this study were asked to make judgements in relation to the level of severity of the errors made in the FY1 doctors' responses to the WUSCEs (Table 2). This means that the study presented here differs from previous research in that the assessments were marked in two ways; a numeric value and the categorical selection of error severity.

Of most concern were those errors categorized by the SME as dangerous (24 scripts) or lethal (8 scripts). These were present in 20% (32 scripts) of the WUSCE scripts (Table 2) and tended to be a failure of CPT knowledge, not reacting to the results of histological or biochemical investigation, or not treating a significant aspect of the patients presenting problems. All of these could lead to significant delay in the patients' recovery or even long-term morbidities. Errors where an unsafe medication choice was made could also lead to unintended morbidities or even death [35]. All FY1 doctors who failed the WUSCEs made at least 'suboptimal' errors, but some who passed the item also recorded 'suboptimal' errors. However, no-one categorized as having made a dangerous or lethal decision passed a WUSCE. It is worth noting that, in the context of patient safety within the UK healthcare system, there are in place successive layers of defences and safeguards that would likely detect and correct most of the prescribing errors detected in this study.

It could be suggested that this categorical method of marking the WUSCE is an appropriate way of identify underperforming FY1 doctors. The factors identified by the categorical marking are not easily overcome with a systematic approach to redesigning the health care system and ensuring familiarization with the documentation. These types of errors require interventions that affect individual prescribers [10]. We suggest that this would be most effective during undergraduate and continuing medical education, and involve patients' medical and drug histories, and the application of knowledge to select and prescribe safe and effective treatment.

In developing assessment tools it is desirable to compare them with an existing external measure. The FY1 doctors in this study must have completed assessments while at medical school. However few medical schools currently have a distinct assessment of CPT [4, 15] and it was therefore not possible to compare the FY1 responses to the assessments in this study with their undergraduate performance. The results of this study and others appear to suggest that targeted teaching and assessment that applies CPT knowledge (e.g. WUSCE) would best discriminate between the best and worst candidates (undergraduates or FY1 doctors) [3–5, 13]. The results of this study suggest that focusing teaching in all the identified domains would result in an immediate impact on improving patient safety.

In conclusion, this study has produced valid and statistically reliable assessment instruments to assess performance in CPT and prescribing. From a curriculum blueprint an SME panel developed assessments in seven CPT domains that were deemed to pose a high risk to patient safety or potential for harmful drug errors. Using these assessment instruments this study has shown that a large proportion of FY1 doctors have a weakness in applying CPT knowledge and prescribing skills in structured clinical cases. However, the research has revealed three performance profiles which indicate that it is actually about a 30% of FY1 doctors who consistently perform poorly. The study has helped to identify specific domains where FY1 doctors are underperforming and committing unsafe prescribing practices and thereby areas of CPT knowledge and prescribing that require targeted applied teaching.

Competing interests

REFERENCES

1 General Medical Council. Tomorrow's Doctors. London: GMC, 2003.

2 Medical Schools Council. Outcomes of the Medical Schools Council Safe Prescribing Working Group. London: Medical Schools Council, 2008.

3 Garbutt JM, Highstein G, Jeffe DB, Dunagan WC, Fraser VJ. Safe medication prescribing: training and experience of medical students and housestaff at a large teaching hospital. Acad Med 2005; 80: 594–9.

4 Heaton A, Webb DJ, Maxwell SR. Undergraduate preparation for prescribing: the views of 2413 UK medical students and recent graduates. Br J Clin Pharmacol 2008; 66: 128–34.

5 Tobaiqy M, McLay J, Ross S. Foundation year 1 doctors and clinical pharmacology and therapeutics teaching. A retrospective view in light of experience. Br J Clin Pharmacol 2007; 64: 363–72.

6 Dean SJ, Barratt AL, Hendry GD, Lyon PM. Preparedness for hospital practice among graduates of a problem-based, graduate-entry medical program. Med J Aust 2003; 178: 163–6.

7 de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. Qual Saf Health Care 2008; 17: 216–23.

8 Leape LL, Bates DW, Cullen DJ, Cooper J, Demonaco HJ, Gallivan T, Hallisey R, Ives J, Laird N, Laffel G. Systems analysis of adverse drug events. ADE Prevention Study Group. JAMA 1995; 274: 35–43.

9 Bates DW, Cullen DJ, Laird N, Petersen LA, Small SD, Servi D, Laffel G, Sweitzer BJ, Shea BF, Hallisey R. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE Prevention Study Group. JAMA 1995; 274: 29–34.

10 Dean B, Schachter M, Vincent C, Barber N. Causes of prescribing errors in hospital inpatients: a prospective study. Lancet 2002; 359: 1373–8.

11 Dovey SM, Phillips RL, Green LA, Fryer GE. Types of medical errors commonly reported by family physicians. Am Fam Physician 2003; 67: 697.

12 Kuo GM, Phillips RL, Graham D, Hickner JM. Medication errors reported by US family physicians and their office staff. Qual Saf Health Care 2008; 17: 286–90.

13 Coombes ID, Mitchell CA, Stowasser DA. Safe medication practice: attitudes of medical students about to begin their intern year. Med Educ 2008; 42: 427–31.

14 General Medical Council. New Doctor. Available at http://www.gmc-uk org/education/foundation/new_doctorasp (last accessed 30 June 2005).

15 Illing J, Marrow G, Kergon C, Burford B, Spencer J, Davies C, Baldauf B, Morrison J, Allen M, Peile E. How prepared are medical graduates to begin practice? A comparison of three diverse UK medical schools. Final Report for the GMC Education Committee, 2008.

16 Aronson JK, Henderson G, Webb DJ, Rawlins MD. A prescription for better prescribing. BMJ 2006; 333: 459–60.

17 Kmietowicz Z. GMC to gather data on prescribing errors after criticism. BMJ 2007; 334: 278.

18 Pearson S, Smith AJ, Rolfe IE, Moulds RF, Shenfield GM. Intern prescribing for common clinical conditions. Adv Health Sci Educ Theory Pract 2000; 5: 141–50.

19 American Education Research Association and American Psychological Association. Standards for Educational and Psychological Testing. Washington, DC: American Education Research Association, 1999.

20 Pace DK. Subject matter expert (SME) Peer use in M & SV & V. Invited paper for session A4 of the Foundations for V & V in the 21st century. Johns Hopkins University AppliedPhysics Laboratory. Available at http://www.cs.clemson edu/~found04/Foundations02/Session_Papers/A4pdf (last accessed 31 January 2008).

21 Lavin RP, Dreyfus M, Slepski L, Kasper CE. Said another way subject matter experts: facts or fiction? Nurs Forum 2007; 42: 189–95.

22 Cizek GJ. Standard setting. In: Handbook of Test Development, eds Downing SM, Haladyna TM. London: Lawrence Erlbaum Associates, 2006; 225–58.

23 Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979; 86: 420–8.

24 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159–74.

25 Cronbach LJ, Nageswari R, Gleser GC. Theory of generalizability: a liberation of reliability theory. Br J Stat Psychol 1963; 16: 137–63.

26 Cronbach LJ, Gleser GC, Nanda HRN. The Dependability of Behavioural Measurements: Theory of Generalizability for Scores and Profiles. New York: John Wiley, 1972.

27 Howe A. Patient safety: education, training and professional development. In: Patient Safety Research into Practice, eds Walshe K, Boaden R. Maidenhead: Open University, 2006; 187–97.

28 Sharpe VA, Faden AI. Adverse Effects of Drug Treatment. Medical Harm. Historical, Conceptual, and Ethical Dimensions of Iatrogenic Illness. Cambridge: Cambridge University Press, 1998; 175–93.

29 Woosley RL. Centers for education and research in therapeutics. Clin Pharmacol Ther 1994; 55: 249–55.

30 Shavelson RJ, Webb NM. Generalizability Theory: a Primer. Thousand Oaks, CA: Sage, 1991.

31 Pearson SA, Rolfe I, Smith T, O'Connell D. Intern prescribing decisions: few and far between. Educ Health 2002; 15: 315–25.

32 Harvey K, Stewart R, Hemming M, Moulds R. Use of antibiotic agents in a large teaching hospital. The impact of antibiotic guidelines. Med J Aust 1983; 2: 217–21.

33 Maxwell S, Walley T. Teaching safe and effective prescribing in UK medical schools: a core curriculum for tomorrow's doctors. Br J Clin Pharmacol 2003; 55: 496–503.

34 Hicks RW, Santell JP, Cousins DD, Williams RL. MEDMARX 5th Anniversary Data Report. A Chart Book 2003 Findings and Trends 1999–2003. Rockville, MD: The United States Pharmacopeia Center for the Advancement of Patient Safety, 2004.

35 Reynard J, Reynolds J, Stevenson P. Clinical error: the scale of the problem. In: Practical Patient Safety, eds Reynard J, Reynolds J, Stevenson P. Oxford: Oxford University Press, 2009; 1–14.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Online supplement
The online supplement provides supporting text and graphics including examples of the developed test items and marking schema.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.