2013; 35: e1181–e1196 MEDICAL TEACHER

WEB PAPER
BEME GUIDE

# A systematic review of the relationship between patient mix and learning in work-based clinical settings. A BEME systematic review: BEME Guide No. 24

JIP DE JONG[1], MECHTELD VISSER[1], NYNKE VAN DIJK[1], CEES VAN DER VLEUTEN[2] & MARGREET WIERINGA-DE WAARD[1]

[1]Department of General Practice/Family Medicine, AMC-University of Amsterdam, Amsterdam, The Netherlands, [2]Department of Educational Development and Research, Faculty of Medicine, Maastricht University, Maastricht, The Netherlands



## Abstract

Background: Clinical workplace-based learning has been the means to becoming a medical professional for many years. The importance of an adequate patient mix, as defined by the number of patients and the types of medical problems, for an optimal learning process is based on educational theory and recognised by national and international accreditation standards. The relationship between patient mix and learning in work-based curricula as yet remains unclear.

Aim: To review research addressing the relationship between patient mix and learning in work-based clinical settings.

Method: The search was conducted across Medline, Embase, Web of Science, ERIC and the Cochrane Library from the start date of the database to July 2011. Original quantitative studies on the relationship between patient mix and learning for learners at any level of the formal medical training/career were included. Methodological quality was assessed and two reviewers using pre-specified forms extracted results.

Results: A total of 10,420 studies were screened on title and abstract. Of these, 298 articles were included for full-text analysis, which resulted in the inclusion of 22 papers. The quality of the included studies, scored with the Medical Education Research Study Quality Instrument (MERSQI), ranged from 8.0 to 14.5 (of 18 points). A positive relationship was found between patient mix and self-reported outcomes evaluating the progress in competence as experienced by the trainee, such as self-confidence and comfort level. Patient mix was also found to correlate positively with self-reported outcomes evaluating the quality of the learning period, such as self-reported learning benefit, experienced effectiveness of the rotation, or the instructional quality. Variables, such as supervision and learning style, might mediate this relationship. A relationship between patient mix and formal assessment has never been demonstrated.

Conclusion: Patient mix is positively related to self-reported learning outcome, most evidently the experienced quality of the learning programme.

## Introduction

Clinical workplace-based learning has played the leading role in educating medical professionals for many years. The importance of an adequate case or patient mix at that workplace for an optimal learning process is intuitively felt by many professionals and is recognised by several national (Australian Medical Council Limited 2010; Liaison Committee on Medical Education (LCME) 2011; Royal College of General Practitioners 2011) and international accreditation standards; the World Federation for Medical Education emphasised this in its Global Standards for Quality Improvement, for Postgraduate Medical Education (Karle 2003). It states 'Training locations must have a sufficient number of patients and an appropriate case-mix to meet training objectives. The training must expose the trainee to a broad range of experience in the chosen field of medicine and, when relevant, include both inpatient and outpatient (ambulatory) care and on-duty activity. The number of patients and the case-mix should allow for clinical experience in all aspects of the chosen specialty, including training in promotion of health and prevention of disease.'

The idea that considerable experience is needed to become a competent doctor fits various theoretical educational frameworks. Feedback, based on contacts with patients, is a central aspect of learning when looking at behavioural learning theories (Hattie & Timperley 2007). Furthermore, experiential learning, or learning from doing, is central to humanist, cognitive and social learning theories (Mann 2011) and elaborated upon in Kolb's experiential learning model (Kolb 1984). In this model, experiences are a central part of the

*Correspondence:* J. de Jong, MD, GP, Department of General Practice, Academic Medical Center-University of Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands. Tel: +31 20 5667189; fax: 0031 20 5669216; email: jipdejong@amc.uva.nl

ISSN 0142–159X print/ISSN 1466–187X online/13/061181–16 © 2013 Informa UK Ltd.
DOI: 10.3109/0142159X.2013.797570

## Practice points

- There is a large semantic overlap between the terms 'clinical experience', 'clinical exposure', 'clinical encounters' or 'patient encounters' and 'case mix' or 'patient mix'.
- 'Patient mix' is mostly presented without definition.
- Operationalisations of patient mix were fairly different, allowing very few inferences from the studies. This is particularly interesting in the light of the emphasis adequate patient mix has gained in the diverse accreditation standards of several countries.
- Educational research would benefit from a standardised approach in patient mix descriptions; volume can always be measured, but diversity should be explicated in relation to the learning outcome. A model for patient mix measurement is suggested.
- We found indications that patient mix is related to self-reported learning outcome, most evidently the experienced quality of the learning programme.
- A relationship between patient mix and formal assessment has not been demonstrated.
- Supervision quality and learning style might mediate the relationship between patient mix and learning.
- Future studies should aim at addressing which parts of patient mix contribute to learning and which parts do not. It would be interesting to see what the effect of tailoring the patient mix to the specific learning goals and needs of individual students would yield.

learning cycle followed by reflection, abstract conceptualisation and active experimentation. Studies showed that participation in meaningful, patient-related, activities is critical to the learning of (postgraduate) learners (Dornan et al. 2007; Teunissen et al. 2007).

According to Ericsson (2004), medical expertise develops by 'deliberate practice'. He argues that expert performance is different from everyday performance, as it continues to improve as a function of more experience, coupled with deliberate practice. Expert performance is reached by actively acquiring and refining a cognitive mechanism to support continued learning and improvement. Becoming a medical expert thus requires engagement in practice and appropriate reflection, which can be stimulated by feedback from coaches or trainers (Ericsson 2004). Based on this, Duvivier et al. recently described medical training programmes as developed to overcome weaknesses and to improve competence. The level of competence must be monitored to provide cues for further improvement. Deliberate practice-based (medical) training is not the repetition of activities but a focused approach aiming for well-defined learning goals (Duvivier et al. 2011). Within this framework, patient mix is an important training condition because it embodies the required representative tasks in the medical domain at issues onto which the desired competence can be practiced. The patient mix offers different experiences on which reflection and assessment can be made by the trainee themselves, by the trainer, or eventually, by an external preceptor.

In addition, other frameworks for medical expertise emphasise the importance of clinical experience for learning, such as theories of cognitive structures (Schmidt et al. 1990) and dual processing (Ark et al. 2007; Evans 2007; Norman 2009). The essence of these theories is that first conscious, intentional learning (deliberate practice) must be established before routines are automated. These automated routines are the basis of adequate medical handling (Pelaccia et al. 2011). Within these frameworks also, the experience needed is provided by an adequate patient mix, so patient mix is an important training condition. Well-supervised learners, exposed to an adequate patient mix, can be assumed to substantially improve medical competence.

This review was carried out in order to evaluate whether this theory could also be confirmed by empirical evidence. Our primary aim was to systematically review research addressing the relationship between patient mix and learning in work-based clinical settings. Our secondary aim was to address the influence of additional variables (e.g. supervision and learning style) on this relationship.

## Definition of patient mix

In order to obtain a view of the patient mix, a clear and workable definition of 'patient mix' has to be formulated. The first article with 'patient mix' in the title in Medline is a commentary by Brandt Jr. (1974) in 1974 on an article of McAllister and Dzur (1974) about the patient population in an acute medical care service. In this commentary, the author states that the 'number of patients' and the 'types of medical problems' are of prime importance and that clinical learning involves both quality and quantity. Numerous papers have been published reporting patient mix based on this 'quality' (diagnosis-diversity) and 'quantity' (patient-volume) approach (Hand et al. 1993; Raghoebar-Krieger et al. 2001; Carney et al. 2002). In these articles, there is a large semantic overlap between the terms 'clinical experience', 'clinical exposure', 'clinical encounters' or 'patient encounters' and 'case mix' or 'patient mix'. 'Case mix' in this respect might be a synonym to patient mix, but may also be used in broader context, for instance referring to the funding of the healthcare system.

'Clinical exposure' can be regarded as the umbrella term for clinical contacts of any kind. The term 'patient mix' inclines towards the description of the diversity of the exposure, focusing not only on variety or diversity but also on quantity or volume. To measure patient mix ('case mix' in the publication), a definition was formulated by Hutchinson 'A system of classifying 'cases' – patients, contacts, episodes, or visits – into groups, which are similar according to some characteristic, such as diagnosis (e.g. International Classification of Diseases), treatment (e.g. Office of Population Censuses and Surveys (OPCS) operation codes), severity, potential for healthcare improvements or costliness (Hutchinson et al. 1991)'. Following Berlowitz, the major difference between a patient mix measure (case mix in the publication) and a classification system is in their application. Classification systems are developed to order patients into groups on the basis of their relationships. Patient mix measures intend to relate these groupings to an outcome. They may include a range of patient

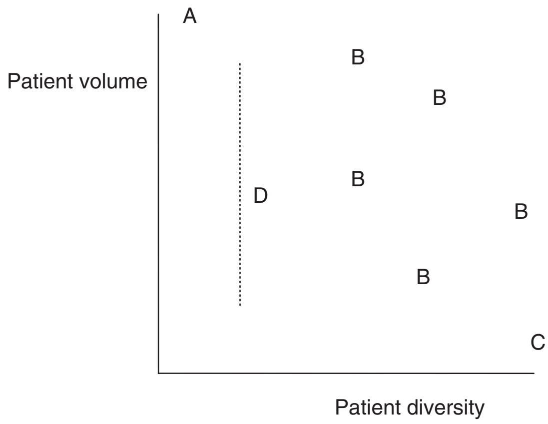


**Figure 1.** Two-dimensional approach of patient mix. Consider a training programme in a large group of patients with the same diagnoses (A). Because there is no diversity, this cannot be considered to be a 'patient mix', but merely the training of a single skill or restricted clinical problem. Situation C expresses a situation in which a large diversity of skills and/or symptoms and diagnoses are theoretically possible, but very few or even no patients are present. This can be regarded as a patient mix, but an extremely meagre one – simply because there are no or very few patients. The line labelled 'D' expresses (an arbitrary) 'cutoff point' where the diversity is rich enough to start calling the population 'patient mix'. All points labelled B are considered to be a patient mix.



characteristics such as diagnoses, disease severity, gender, age, socioeconomic status or functional status. Similarly, the outputs may reflect clinical status, resource utilisation, cost or learning outcome (Berlowitz et al. 1995).

This review focuses on patient mix and its relationship with learning. In order to maximise inclusion possibilities, a sensitive approach was chosen and therefore the original definition was used; the number of patients and the types of medical problems presented to learners (McAllister & Dzur 1974). Patient mix thus is regarded to consist of a number of patients presenting a certain diversity of diseases (Figure 1).

## Methods

### Eligibility criteria (list 1)

As we aimed to assess the strength of the relationship between patient mix and learning, only studies reporting on quantitative data were included, which were conducted with medical students/trainees at any level of the formal medical training/career. Patient mix volume, i.e. the quantity of patients encountered and the diversity of *skills* and/or *symptoms and diagnoses* had to be described. No simple cut-off for the width of this diversity could be given (Figure 1), but studies on the exposure to one restricted clinical problem or skill were excluded as they only described the volume of that skill or problem and no diversity. Learning outcome had to be explicitly assessed. The relationship between patient mix and learning had to be quantified by statistical analysis.

### Information sources and search strategy

The search was conducted across five sources relevant to education in a clinical context: Medline, Embase, Web of Science, ERIC and the Cochrane Library. The search ran from the start date of the database to July 2011 and was not limited by language, geography or research methodology. The search strategy was composed by a clinical librarian. The search strategy had to be able to find a 'reference set' in PubMed. This set contained 38 articles, previously rated as being relevant to the review subject by the authors. The strategy was then translated to the search systems of the other databases.

### Study selection

Two authors (M. W. and J. J., or M. V. and J. J.) individually and independently screened the titles and abstracts of all articles using the inclusion and exclusion criteria. Citations that were selected by one author but not by the other author were discussed in order to achieve negotiated consensus on inclusion or exclusion. In case of doubt or persisting disagreement in this phase, the article was included. The full text of all the potentially relevant articles was retrieved. The full-text articles were screened, again independently, by two authors, using the same criteria and were again compared. In case of disagreement, a decision on inclusion or exclusion was once more reached by negotiated consensus. Most studies that were excluded did not have an adequate description of patient mix or did not statistically address the relationship between patient mix and learning. At each screening phase, each citation was marked as 'yes', 'maybe' or 'no'. Inter-observer agreement of the screening phase was measured by Cohen's Kappa (linearly weighted). Manual searches were conducted across the citations of the papers that were coded, resulting in 17 more citations. These were screened by two authors, but none of them were included.

**List 1.** Inclusion and exclusion criteria.

Inclusion criteria

- Empirical, educational studies with actual patient exposure (no simulations), reporting on quantitative data
- Study population: studies conducted with medical students/trainees at any level of the formal medical training/career
- Patient mix, clinical encounters, or clinical experience in workplace-based learning had to be described. Patient mix had to be described in some detail, thereby addressing the volume as well as the diversity. Studies on the exposure to a restricted clinical problem or skill were excluded. Medical subspecialties were not excluded beforehand, as long as the patient mix was diverse
- Learning outcome measures had to be described by self-reported measures, assessment by trainers, preceptors, or others, or by objective structured clinical examinations (OSCEs) either with real or with standardised patients, multiple choice or other written exercises
- The relationship between patient mix and learning had to be quantified by statistical analysis
- Studies in all languages were included

Exclusion criteria

- Studies on qualitative research
- Dental and veterinary curricula, any paramedical curricula, nursing curricula, physician assistant curricula, nurse practitioner curricula and dietetic curricula
- Theoretical medical curricula (not work-based)
- Complementary/alternative medicine

### Data extraction

A detailed data extraction form was developed using the Best Evidence in Medical Education standard coding sheet and published reviews (Steinert et al. 2006; Colthart et al. 2008; BEME 2012) as a basis. All selected papers were coded by the authors in pairs (M. W. and J. J. or M. V. and J. J.). This form contained a general description of the study design and participants, including the training level and the specialist training area. In addition, patient mix instruments (e.g. electronic logbooks and questionnaires), volume/diversity descriptions (e.g. top 10 skills or diagnoses lists), learning outcome measures and the relationship found between patient mix and learning were recorded (Table 2). We also documented the highest level of the Kirkpatrick hierarchy (Kirkpatrick & Kirkpatrick 2006) on which learning outcomes were assessed. If additional variables were studied in relation to learning outcome (e.g. learning style and supervision), these were also recorded.

### Quality of the studies

To obtain an overview of the quality of the included studies, and thereby the validity of the outcomes, we assessed them with the recently developed Medical Education Research Study Quality Instrument (MERSQI) (Reed et al. 2007). This instrument was chosen because, to our knowledge, it is the only instrument fitted to measure the quality of experimental, quasi-experimental and observational studies considering medical education. The maximum total MERSQI score is 18. Two authors (J. J. and M. V.) independently scored the quality of the included papers. In case of disagreement on item scores, a decision was reached by negotiated consensus.

### Data analyses

The various ways in which patient mix was operationalised were categorised in equivalent approaches of volume and diversity descriptions. Learning outcomes were divided into self-reported outcomes and outcomes using formal assessments. The relationship between patient mix and learning is described in sections based on different learning outcomes, as this allowed for homogeneous reporting of results.

## Results

### Search results

The search resulted in 11,098 titles. After removal of duplicates, 10,420 studies were reviewed based on title and abstract. Of these studies, 298 were identified as potentially relevant and included for full-text analysis, which resulted in the inclusion of 22 papers (Figure 2). The studies identified had insufficient homogeneous or quantitative data to allow meta-analysis or other formal synthesis. During screening of titles and abstracts, the inter-observer agreement kappa (linearly weighted) was 0.34 (M. V.–J. J.) respectively 0.32 (M. W.–J. J.). Table 2 provides a summary of study descriptions and outcomes. This table forms the basis for the inferences from the studies in the following paragraphs.

### Methodological quality of studies

MERSQI sum-scores ranged from 8.0 to 14.5 (median: 11.75). Ten studies reported on the internal structure of their outcome instruments by Cronbach's alpha (Table 2) or principal component analysis. Response rates, if presented, varied from 43% to 100%. Data analyses were appropriate in all but one study (Martin et al. 2000), and all were beyond the descriptive level. Four studies reported outcome only at student reaction (Kirkpatrick level 1) (Dolmans et al. 2002; O'Hara et al. 2002; Saywell et al. 2002; van der Zwet et al. 2010), whereas 10 studies measured knowledge and/or skills (Kirkpatrick 2) (Gruppen et al. 1993; Schwiebert et al. 1993; McLeod et al. 1997; Jacobson et al. 1998; Greenberg & Getson 1999; Boots et al. 2008; Lampe et al. 2008; Nomura et al. 2008; Duke et al. 2011; Yu et al. 2011). Less than half of the studies ($n = 8$) measured outcomes up to the behavioural level (Kirkpatrick 3) (Chatenay et al. 1996; Jolly et al. 1996; McManus et al. 1998; Ahmed & Hughes 1999; Martin et al. 2000; Sorensen et al. 2004; Wimmers et al. 2006a; Fung et al. 2007). None of the included studies explicitly measured patient or healthcare outcome (Kirkpatrick 4).

### Types of studies

In six studies, the mutual dependence of factors related to learning was addressed in a path analysis or structural equation modelling (Jolly et al. 1996; Martin et al. 2000; Dolmans et al. 2002; Sorensen et al. 2004; Wimmers et al. 2006a; van der Zwet et al. 2010; Table 2). Eight studies compared the patient mix of training sites and their contribution to learning. In three of these studies, similar sites were compared (Chatenay et al. 1996; Wimmers et al. 2006a; Yu et al. 2011), three others compared academic vs. non-academic sites (Schwiebert et al. 1993; McLeod et al. 1997; Nomura et al. 2008) and two compared inpatients and outpatients (Jacobson et al. 1998; Duke et al. 2011).

Four studies evaluated the learning effects of an intervention: the introduction of a rotation (Gruppen et al. 1993), a skill-training programme (Boots et al. 2008), identification of 10 preselected complaints (Lampe et al. 2008) and a new internship (Nomura et al. 2008). Two studies compared groups of medical students at a different phase of their training (Ahmed & Hughes 1999; Boots et al. 2008). In four studies, the groups and sites were homogeneous, and no interventions were studied (McManus et al. 1998; O'Hara et al. 2002; Saywell et al. 2002; Fung et al. 2007).

### Operationalisation of patient mix

Patient mix, or any other term intended to describe the exposure of students/trainees at any level, to patients with health problems, was in none of the studies explicitly defined. The terms (clinical) exposure, experiences, encounters or content are used, as are student or learning experiences and of course patient mix. Patient mix was measured with various instruments (Tables 1 and 2), including questionnaires ($n = 11$), interviews ($n = 1$) and logbooks ($n = 13$), the latter hand written ($n = 9$), electronic ($n = 2$) or unspecified ($n = 2$).

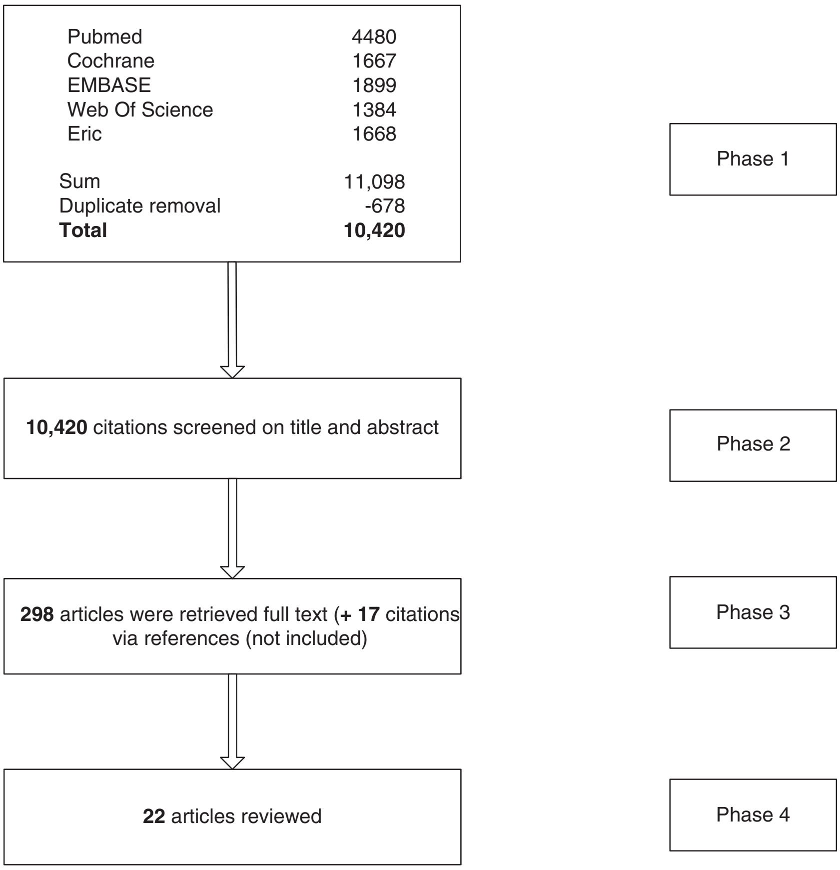


**Figure 2.** Flow chart.

Patient mix was mostly described as the variety of encountered skills and/or diagnoses. Skills were usually technical procedures, such as intubation (Boots et al. 2008) or suturing (Jolly et al. 1996). In some studies ($n = 6$), the patient volume was the most pronounced patient mix characteristic (Schwiebert et al. 1993; Jolly et al. 1996; Greenberg & Getson 1999; Martin et al. 2000; Dolmans et al. 2002; van der Zwet et al. 2010). The diversity of the patient mix in these studies was often additionally addressed by one or two variables, but the reports lacked a detailed insight into the diversity of diagnoses.

In most other studies, the distribution of encountered diagnoses and medical skills was presented. Several authors presented a top 10 or 20 of the conditions the students meet most frequently (Schwiebert et al. 1993; McLeod et al. 1997; Jacobson et al. 1998; Saywell et al. 2002). This method was also used to compare the patient mix of different sites.

## Operationalisation of learning

Learning outcome measures can be divided into self-reported outcomes and formal assessments. Self-reported outcomes are used in 10 studies, in 5 of these the self-estimated competence was measured as self-confidence (Boots et al. 2008; Nomura et al. 2008; Duke et al. 2011) or comfort level (O'Hara et al. 2002; Saywell et al. 2002). In the five other studies, the quality of the learning experience or the educational profit of the experience at issue is asked for; such as the effectiveness, the

**Table 1.** Patient mix and learning outcome instruments and descriptions.

- **Instruments used for measuring patient mix**
  - Questionnaire
  - Hand written logbook
  - Reference cards
  - Electronic logbook
  - Curriculum description
- **Patient mix operationalisations**
  - Volumes of patient encounters
  - 'Top 10'
  - Encounters per organ systems specified
  - Encounters per specialism specified
  - Encounters per predefined skills
  - Content of the course/rotation/apprenticeship
  - Encounters per diagnosis
- **Learning outcomes**
  - 1. Self-assessment
    - Questionnaire/otherwise
      - Self-efficacy
      - Self-perceived knowledge
      - Self-perceived clinical confidence
      - Self-perceived competence
  - 2. Formal assessment:
    - Theoretical examination
      - Multiple-choice exam
      - Other written exams
    - Practical examination
      - OSCE/other simulation(s)
      - Bedside or other patient evaluation
      - Presentation
- **Learning outcomes content**
  - Theoretical knowledge (specifiy)
  - Specific technical skills (id)
  - Clinical competence in practice (id)

**Table 2.** Summary of the included studies.

| Author, year of publication | Country | Design | Time span | Speciality | Number of participants analysed and educational level participants | PM instrument | PM description (volume/diversity) | Learning instrument (self-reported measure is specified, Cronbach's alpha if reported) | Relation patient mix and learning | MERSQI | Highest Kirkpatrick level |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Ahmed & Hughes 1999 | UK | Single group, post test | Seven weeks | Paediatrics | Two-hundred and twenty-six students | Questionnaire | Change in exposure from year 1 to year 2 in 42 conditions and 20 skills | MCQ† and other written exam (0.67–0.71) General and clinical competence assessment | Clinical experience did not relate to MCQ/written exam score | 13 | 3 |
| Boots et al. 2008 | Australia | Pre-post test design with two groups | 10 weeks | Internal medicine | Two-hundred and twenty students and 174 interns | Questionnaire | Encounters (ordinal scale) of 15 specified skills | Questionnaire and self-confidence | There may be a relation between skill exercise and confidence | 9.5 | 2 |
| Chatenay et al. 1996 | Canada | Single cohort randomised to attend one out of four training sites. Post test only | 10 weeks | Surgery | One-hundred and nine students | Hand written logbook | Pat volume of elective/ER admissions, operations scrubs, outpatient clinics and procedures. | MCQ† (0.67), OSCE* (0.48), and clinical performance | Skills were enhanced by increased volume of some, but not all, clinical experience. Complex relation between feedback and OSCE* performance. Quality of feedback seems to mediate this relationship | 14 | 3 |
| Dolmans et al. 2002 | The Netherlands (NL) | Observational cohort study | Three to 12 weeks per rotation | Eight disciplines | One thousand two-hundred and eight residents | Questionnaire | One variable composed of three questionnaire items (sufficient -patients, -diagnostic variety and -patients independently dealt with) | Self-reported effectiveness questionnaire | Self-perceived effectiveness depends on patient mix and supervision. Supervision more strongly influences effectiveness when patient mix is limited | 10 | 1 |
| Duke et al. 2011 | Canada | Non equivalent control group, pre-post test | Four weeks | Family medicine | Seventy-nine residents | (Unspecified) logbook | Percentage of students who attended clinics of 18 subspecialty Number of outpatient clinics attended | MCQ† knowledge test Self-confidence questionnaire - Oral and written feedback assessment | No difference between ambulatory and inpatient sites | 12.5 | 2 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Fung et al. 2007 | USA | Retrospective single group, post test | Not specified (three clerkships) | Inpatient internal, ambulatory and family medicine | One-hundred and sixty-six third-year students | Electronic (PDA**) logbook | Number of patients in six diagnostic categories | OSCE* (0.34–0.65) | No relation between patient exposure and OSCE* score | 12 | 3 |
| Greenberg & Getson 1999 | USA | Single group, post test | Eight weeks | Paediatrics | One hundred and eighteen students | Hand written logbook | Number of patients in four diagnostic domains | Clinical performance, case presentation and NBME‡ | No relation between volume and exam score | 11.5 | 2 |
| Gruppen et al. 1993 | USA | Single group, pre-post test | One month | Internal medicine | Forty-three third-year students | Hand written logbook | – Number of patients, – Percentage of students who encountered 20 diagnostic categories –Top 19 | Written exam | No correlation between the students levels of experience and knowledge | 12 | 2 |
| Sorensen Hoifoidt 2004 | Norway | Cross-sectional | Four months | Psychiatry | Eighty-five preregistration house officers | Questionnaire | Number of subjects having experience in 12 psychiatric skills | Questionnaire, subjective learning benefit (0.84) | Subjective learning benefit was related to amount of experience, competence and formal teaching programme. Supervision and previous experience had no impact on subjective learning | 11.5 | 3 |
| Jacobson et al. 1998 | USA | Observational cohort study | Twelve weeks | Internal medicine | Forty-three students | Hand written logbook | –Top 10 – Percentage of encounters of six diagnostic categories and one skill | Categorisation of self-reported learning points | In and outpatient encounters differed. Learning differences between in- and outpatient apply to pathophysiology, evaluation/work-up and patient education/counselling. | 8.5 | 2 |
| Jolly et al. 1996 | UK | Single group, post test | 4 + 16 weeks | Pathology, then medicine and surgery | One-hundred and fifty-two clinical students | Questionnaire | Questionnaire responses handling patient volume and skills | OSCE* (0.69) | No relation between clinical experience and educational outcome | 13.5 | 3 |
| Lampe et al. 2008 | USA | Non-randomised trial | Six months | Emergency medicine | Thirty-seven senior medical students | Electronic (PDA**) logbook | No of students that met the pre-specified target complaints | MCQ† and written exam | Group seeing a required number of representative patients showed better knowledge | 12.5 | 2 |

(continued)

**Table 2. Continued.**

| Author, year of publication | Country | Design | Time span | Speciality | Number of participants analysed and educational level participants | PM instrument | PM description (volume/diversity) | Learning instrument (self-reported measure is specified, Cronbach's alpha if reported) | Relation patient mix and learning | MERSQI | Highest Kirkpatrick level |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Martin et al. 2000 | UK | Single group, post test | One year | Medicine and surgery | One-hundred and ninety-four (150 returned learning style form) students | Questionnaire | Total number of patients, out-patients and emergencies | OSCE* (0.70) | No association between clinical experience and OSCE* score. Positive association clinical experience with learning style. | 12 | 3 |
| Mcleod et al. 1997 | Canada | Observational cohort study | Eight week | Internal medicine | Forty residents and 29 clinical clerks | Interview (unspecified), logbook, and questionnaire | Patient numbers of 20 diagnoses, ordered as top 20. – (Top) 16 skills, % patient encounters in which skill was relevant | Relevance for learning questionnaire/ interview | Inpatient-based experience is better than ambulatory care experience for learning | 8.5 | 2 |
| McManus et al. 1998 | UK | Prospective study of two cohorts assessed at application to med school and at the end of their final year | About five years | Undergraduate curriculum, not otherwise specified | Six-hundred and eighty-four students (first cohort 301/second 383) | Questionnaire | Single experience score based on 15/20 (first/ second cohort) unspecified acute conditions, 18/18 surgical operations and 17/29 practical procedures | MCQ† and other written exam Clinical performance (0.87, 0.88) | No association between clinical experience and exam score. Study habits predict examination performance | 14.5 | 3 |
| Nomura et al. 2008 | Japan | Pre-post test design with non-equivalent control group | Two years | Multidisciplinary | Two-thousand four-hundred and seventy-four before +1166 after postgraduate 'residents' without clinical experience | Questionnaire | Number of students with 'no experience' with specified diagnosis (grouped) in 22 (sub) specialisms | Questionnaire, self confidence | Clinical experience and confidence levels improved, especially at university hospitals | 10.5 | 2 |
| O'Hara et al. 2002 | USA | Single group, post test | Four weeks + Four days | Women healthcare in a family medicine clerkship | Four-hundred and forty-five students | Hand written logbook | – Volume of patient by ages – Volume and percentage of top 10 diagnosis by age group | Questionnaire, comfort level | Relationship between experience and comfort level between some diagnostic categories | 8 | 1 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Saywell et al. 2002 | USA | Single group, post test | Four weeks | Musculoskeletal medicine in a family medicine clerkship | Four-hundred and forty-five third-year students | Hand written logbook | – Volume of patient by ages – Volume and percentage of top 10 diagnosis by age group | Questionnaire comfort level | Relationship between experience and comfort level between some diagnostic categories | 8 | 1 |
| Schwiebert et al.1993 | USA | Single group, post test | One month | Family medicine | One-hundred and eighty-five third-year students | Hand written logbook | – Proportion of students who encountered each of 20 diagnoses ordered as a top 20. – Number of patients | Written exam, oral exam | Slight differences between university and private practice in patient mix, but no difference in results on oral and written exam. | 11.5 | 2 |
| Wimmers 2006 | NL | Single group, post test | Twelve weeks | Internal medicine | One-hundred and fifty-two students | Hand written logbook | – Number of patients – Number of different diseases encountered | Combined clinical performance assessment and oral examination (0.67) | An increased number of patient encounters did not (directly) lead to improved competence. - Quality of supervision indirectly had impact on student learning and the number of patient encounters | 13 | 3 |
| Yu et al. 2011 | New Zealand | Single group, post test | Six weeks | Surgery | One-hundred and sixty-six fourth-year students | Questionnaire and hand written logbook | Number of patients in seven diagnostic categories | Clinical assessment, critical appraised topic and OSCE* (0.69–0.74) | Heterogeneity of clinical experiment from sites did not translate into heterogeneity of learning outcomes | 13.5 | 2 |
| Van der Zwet et al. 2010 | NL | Single group, post test | Ten weeks | General practice | Two-hundred and eighty four fifth-year students | Questionnaire | Based on factor analysis; patient mix variable composed of three questionnaire items: number of patients, patient variety and quality of patient contacts | Questionnaire, instructional quality | Supervision and patient mix load on instructional quality | 10 | 1 |

*Objective structured clinical examination.
†Multiple-choice questions.
‡ National board of medical examiners.
**Personal digital assistant.

learning benefit, or instructional quality of the rotation (McLeod et al. 1997; Jacobson et al. 1998; Dolmans et al. 2002; Sorensen et al. 2004; van der Zwet et al. 2010).

Formal assessments were more diverse. Usually, knowledge (Gruppen et al. 1993; Schwiebert et al. 1993; Chatenay et al. 1996; McManus et al. 1998; Greenberg & Getson 1999; Lampe et al. 2008; Duke et al. 2011) or skills (Schwiebert et al. 1993; Chatenay et al. 1996; Jolly et al. 1996; McLeod et al. 1997; Ahmed & Hughes 1999; Martin et al. 2000; Fung et al. 2007; Boots et al. 2008; Nomura et al. 2008; Yu et al. 2011) were tested; sometimes including clinical performance (Chatenay et al. 1996; McManus et al. 1998; Ahmed & Hughes 1999; Greenberg & Getson 1999; Wimmers et al. 2006a). Methods used included multiple-choice examinations (MCQ) and other written examinations, clinical assessments, oral examinations, and OSCEs.

## Relationship between patient mix and learning

The relationship between patient mix and learning was tested in two different manners:

(1) By comparing learning outcomes between existing groups of learners (e.g. cohorts) with patient mix described at the group level.
(2) By relating patient mix indices at the level of the individual learner to a learning outcome (Table 2).

## Outcomes based on self-reporting

Of 10 studies, four found a positive relationship between patient mix and self-reported outcomes evaluating the progress in competence as experienced by the students or residents, such as self-confidence and comfort level (O'Hara et al. 2002; Saywell et al. 2002; Boots et al. 2008; Nomura et al. 2008). By contrast, one study found no difference in confidence between residents in a traditional inpatient rotation and a new one in which experience in ambulatory settings was introduced (Duke et al. 2011).

Patient mix was also found to correlate with self-reported outcomes evaluating the quality of the learning period, such as self-reported learning benefit, experienced effectiveness of the rotation, or the instructional quality (McLeod et al. 1997; Jacobson et al. 1998; Dolmans et al. 2002; Sorensen et al. 2004; van der Zwet et al. 2010).

## Outcomes based on formal assessment

*MCQ or other written examinations*: In one trial of the eight studies using MCQ and/or written examinations, students in an intervention group who encountered significantly more often patients with 10 chief prerequisite complaints than the control group (31.8% vs. 6%) outperformed the control group on a general knowledge examination ($p = 0.014$; Lampe et al. 2008). In seven other studies, however, no relationship between patient mix and scores on MCQs or other written examinations was found (Gruppen et al. 1993; Schwiebert et al. 1993; Chatenay et al. 1996; McManus et al. 1998; Ahmed & Hughes 1999; Greenberg & Getson 1999; Duke et al. 2011).

*OSCE*: No association was found between patient mix and performance in four of the five studies using OSCE assessment (Chatenay et al. 1996; Jolly et al. 1996; Martin et al. 2000; Fung et al. 2007; Yu et al. 2011). Fung et al. suggested that the time allotted for students to complete clerkships may not be sufficient to expose them to the number of patients needed to generate a significant effect on clinical performance (Fung et al. 2007). In one study, the OSCE scores even seemed lower in students who attended a higher number of outpatient clinics than those attending fewer outpatient clinics, although experience with emergency admissions and obtaining feedback on these seemed to improve OSCE performance (Chatenay et al. 1996). The authors concluded that the clinical skills were enhanced by an increased volume of some, but not all, clinical experience. Jolly et al. found that students scored higher on OSCEs if they examined patients on their own, if the objectives (presumably the objectives of the rotation, not reported) had been made clear, and a higher number of clinics were attended (Jolly et al. 1996).

*Oral examinations*: In the two studies using oral examinations, no relation between patient mix and students' examination scores was found (Schwiebert et al. 1993; Wimmers et al. 2006b).

## Clinical assessments

Wimmers et al. found that an increased number of patient encounters did not directly lead to improved clinical performance as assessed by supervisors in 227 medical students (Wimmers et al. 2006a), as was the case in two other studies out of six studies using clinical assessments (Chatenay et al. 1996; McManus et al. 1998). They did, however, find a strong relationship between number of patients and number of diseases encountered ($r = 0.89$; Wimmers et al. 2006a). Ahmed and Hughes, in contrast, found that students' exposure rates did correlate with the assessment grades awarded by clinical supervisors, but not with a written exercise (quiz) score (Ahmed & Hughes 1999). Furthermore, Greenberg and Getson (1999) found a weak positive correlation between number of patients seen and the students' clinical performance.

## Variables potentially relevant to the relationship between patient mix and learning

Martin et al. found that students with a deep, strategic, and well-organized learning style reported significantly higher clinical exposure (combined score for three areas of clinical activity). The well-organized style was also associated with OSCE performance (Martin et al. 2000). McManus et al. additionally found that the amount of knowledge students gained from clinical experience was related to strategic and deep learning styles (McManus et al. 1998) as was success in a final examination: positive and significant correlations were found for deep and strategic learning, whereas surface learning correlated negatively.

In the path analysis presented by Wimmers et al. (2006a), supervision quality loaded on patient mix volume and on clinical competence. Sorensen et al., however, did not find supervision to load on patient mix volume or on (subjective) learning benefit (Sorensen et al. 2004), while in the model of Van der Zwet et al. (2010), supervision loaded on both patient mix and instructional quality. In the study by Dolmans et al. (2002), a relationship was found between supervision and the effectiveness of a rotation. Furthermore, a significant two-way interaction was found between patient mix and supervision; the latter more strongly influenced the effectiveness of the rotation than patient mix did. In another study, OSCE score also seemed to be 'modified' by the quality of the feedback (Chatenay et al. 1996).

Sorensen et al. described that the amount of experience of pre-registration house officers, on 12 different psychiatric areas correlated with the quality of the learning environment, which itself was related to the learning benefit (Sorensen et al. 2004).

In the study by Yu et al., the overall quality of the surgical clerkship, as perceived by students, was related to the number of cases seen, although no difference in learning outcome was found (Yu et al. 2011). Jolly et al. found that six of 43 questionnaire variables correlated with OSCE score. Two of these six can be considered to be related to the learning climate, namely 'whether students examined patients on their own' and 'whether objectives were made clear' (Jolly et al. 1996).

## Discussion

In most studies dealing with the relationship between patient mix and student self-assessment (self-confidence and comfort level), indications of a relationship were found. The indications of positive relationships were stronger regarding the quality of the learning experience (learning benefit, instructional quality or effectiveness of a rotation). Supervision quality seems to be a mediating factor, which was repetitively found to improve patient- or education-related outcome (Farnan et al. 2012). This can be regarded to be consistent with the theory of deliberate practice.

The relationship between patient mix and learning outcome was not corroborated with formal assessment outcomes. All but one study dealing with MCQ or other written examinations failed to find any relationship between patient mix and MCQ or written examinations. All the studies relating patient mix to OSCE score found no association, or under some conditions even a negative association (Chatenay et al. 1996). In one study (Ahmed & Hughes 1999), a correlation was found between exposure rate and clinical assessment grades, whereas three other studies did not find such a relationship (Gruppen et al. 1993; McManus et al. 1998; Wimmers et al. 2006a).

The patient mix (also called 'clinical exposure' or 'case mix') in the articles we reviewed was mostly presented without definition. We found studies describing skills, diagnoses, treatments or general ideas about patient mix, within different specialties and measured by logbook or questionnaires and presented differently, making the patient mix descriptions extremely heterogeneous. The heterogeneity we found is particularly interesting. In the light of the emphasis, adequate patient mix has gained in the diverse accreditation standards of several countries (Liaison Committee on Medical Education (LCME) 2011; Royal College of General Practitioners 2011) and internationally (Karle 2003). Due to the heterogeneity, we had difficulty in finding a proper cut-off point for the number of diagnoses or skills that need to be engaged to fulfil the diversity inclusion criterion. This heterogeneity indicates the need for a discussion on the value of the concept. Berlowitz et al. stated that patient mix should describe how patients are distributed along characteristics that may affect specific outcomes of interest (Berlowitz et al. 1995); he thereby stresses that the concept of patient mix in itself is not relevant. The fact that the patient mixes described in the reviewed studies are so diverse, may be partly because they are related to different outcomes in the different settings at different stages of education. It often seemed that the presented patient mix depended on the instrument the authors had to their disposal and not on study-specific operationalisation of the desired patient mix of the attachment. Following Berlowitz, patient mix descriptions should primarily be based on the desired learning outcomes.

An example of a suitable description of the patient mix is found in the study of Gruppen et al. In this study, the patient contact numbers of students are reported for a vast number (>20) of conditions, as are proportions and frequency orders ('top 10'). In this study, students had contact with at least one patient of 61 conditions reported in quartile ranges. The change in knowledge (pre–post test design) is displayed for 14 conditions, as is an overall score.

Besides the relationship with the outcome, more clarity about the relationship between the diversity and volume aspects of patient mix might be strived for. In this review, we found operationalisations of patient mix that were fairly different in that respect, allowing for very few inferences between studies.

The learning outcome measures were classified into self-reported assessment and formal assessment. The precise description of the used formal assessment methods in the studies was often meagre; example questions or exercises were not found. The reliability of clinical assessments is questionable; subjectivity can be a problem. Pulito et al. (2006) found that direct observations of trainees interacting with patients occur too infrequently. Students prepare for assessments, and their results may reflect their preparation more than their real competence (Al-Kadri et al. 2012). Terms such as OSCE or MCQ might suggest that similar instruments were used in different studies. However, the precise content and the number of stations or questions were found to differ, if mentioned at all.

We formulated six possible explanations that could explain why we found so little evidence for the relationship between patient mix and the results of formal assessment.

(1) *Patient mix does not contribute to medical competence development.*

This idea is highly unlikely, although theoretically possible. The positive relationship between patient mix and self-assessment outcome (compared with formal assessment) is not *per se* an indication of a relationship between patient mix and learning. Self-assessments have many limitations, as is discussed in several reviews, all concluding that self-assessment has little validity and reliability (Tracey et al. 1997; Eva &

Regehr 2005; Davis et al. 2006; Sargeant et al. 2008; Colthart et al. 2008). Poor performers were found to overestimate their competence (Kruger & Dunning 1999; Fidler et al. 1999; Violato & Lockyer 2006).

The largely absent relationship between patient mix and formal assessment might indicate that 'clinical experience without training increases confidence but not competence' (Marteau et al. 1990; Bulstrode & Holsgrove 1996; McManus et al. 1998). The idea that one becomes automatically more competent with increasing experience can be illusory. One of the causes for the absence of relationship might be 'arrested development'. Many of our skills stop to improve once we reach a certain level of competence, and a sufficient level of mastery is accepted (Ericsson 2004).

(2) *The relationship between patient mix and learning is more complex and many other variables play a role (such as supervision quality, learning style, learning environment or professionalism).*

Based on deliberate practice and other educational theories, like social learning theory (Bandura 2006) and motivational theories, the importance of other variables in the relationship could be expected. Medical expertise develops by focussing on learning goals and by identifying areas for improvement. Patient mix provides the repetitive tasks needed for deliberate practice. To guide their learning, students also need supervision from preceptors. In several studies, supervision was found to be strongly related to learning outcome (Dolmans et al. 2002; Wimmers et al. 2006a; van der Zwet et al. 2010). It may, therefore, be seen as an important mediator. Supervision quality was, however, not described or measured in the majority of the included studies, and the potentially mediating effect may have been overlooked. This may have been the case with other variables related to learning as well.

(3) *The time span covered in most studies was too short.*

Current educational theories assume a general problem-solving ability, but case-specific competences are considered of predominant importance (Wimmers et al. 2007). This means that competences does not transfer easily (Patel & Cranton 1983; van der Vleuten & Swanson 1990), implying that exposure in many domains and in many different situations is essential for doctors to become fully competent. This takes time. The time-span covered by most of the included studies may have been too short to find positive results.

(4) *The patient mix is inadequately measured.*

Patient mix is usually described by encountered skills or diagnoses and in terms of volume and diversity. Other potentially relevant descriptors are the complexity in relation to the stage of learning and the learning value or benefit of cases. These aspects, with exceptions, are not usually described, so the validity of the instruments might have been imperfect. To what extent the reported patient mix represents the repetitive tasks needed for deliberate practice was never explicitly described in the studies we included.

Furthermore, in several studies, the patient mix was aggregated per training site and comparison was made between sites, not between students. Maybe this is a too coarse comparison to establish the relationship between patient mix and learning, which can also be regarded as a limitation of our inclusion and exclusion criteria.

(5) *The validity of formal assessment is insufficient.*

In a systematic review, Hamdy et al. (2006) found only mild to moderate correlations between measurements obtained in medical schools and future performance in medical practice. McManus et al. (1998) also questioned the clinical validity of OSCEs. OSCEs, MCQs and other assessments may not be appropriate for determining the specific contribution of patient mix on learning.

Self-reported outcome instruments are usually designed especially for the study, whereas the formal assessments used are commonly part of the standard assessment procedures. These standard assessments are not tailored to the study question and may suffer from bias due to the preparation of students for assessments (Newble & Jaeger 1983). Self-report instruments might therefore be a more appropriate fit for the research questions.

(6) *The quality of the studies was insufficient.*

The majority of the included studies had a single-group, post-test-only design, which may be considered inferior to pre–post test designs or trials. Several studies were merely a comparison of training sites or an evaluation of a new curriculum (Ahmed & Hughes 1999; Nomura et al. 2008).

## Limitations

Composing an efficient and sufficient search strategy is complex. Despite our attempts to sharpen the patient mix definition to an accurate and workable one, we made pragmatic choices. We were not able to formulate the *exact* border between 'some disease (or skills)' and 'patient mix'. This resulted in a low inter-rater agreement. A substantial number of papers were included or excluded based on negotiated consensus. This happened more in the beginning of the review process (Phases 1–3) than later (Phase 4). A minority of the studies included were intended to specifically explore the relationship between patient mix and learning for general educational theory purposes. Most of the studies concerned merely an evaluation of a programme or curriculum change, or were a comparison between training sites. Many studies had to be excluded because they lacked a statistical analysis of the relationship between patient mix and learning outcome.

## Future directions

The volume increase in students and junior doctors may lead to problems with training due to restrictions in clinical interactions. We need to be aware of the likely effects of this increase and what the minimal (or optimal) case load, hours and supervision time is needed, to enable, or optimise adequate training at each learning level. This systematic review emphasises the problem with the description of 'patient mix'. Despite its attention in international accreditation standards, the concept itself seems poorly defined. Educational

**Learning aims**

-> explicitation, operationalisation by assessment

↓

**'Adequate' patient mix:**

- **Diversity:** i.e. diagnoses, complexity, age, gender, ethnicity, continuity/follow up
- **Volume**: descriptions for each diversity item
  + absolute, percentages, top 10/20 , explicit which learners had relatively low/high exposure

**Figure 3.** Model for reporting patient mix.



research would benefit from a standardised approach in patient mix descriptions; volume can always be measured, but diversity should be explicated in relation to the outcome. A model for patient mix measurement is suggested in Figure 3.

In this figure, a model for the description of patient mix is proposed. The reported patient mix should be based on the learning aims of the learners in an attachment. An adequate patient mix is based on the learning aims. Depending on the learning aims, patient mix diversity is operationalised in terms of diagnoses age, gender, ethnicity, complexity of the cases, continuity/follow up or other relevant aspects. The volume that is going along with the different values of the diversity should be reported in both absolute and proportional figures. Frequency sorting (top 10/20) is also recommended, as is reporting of the proportion of the learners that had a relatively high or low exposure.

Future studies should aim at addressing which parts of patient mix contribute to learning and which parts do not. A theoretical framework accounting for other relevant parameters in the relationship between patient mix and learning, such as supervision and learning style, may be helpful, and instead of using the standard assessment procedures, objective outcome tailored to the research question should be developed.

Nearly all studies at hand 'accepted' the patient mix that was presented to the participants as a given factor. Interventions on the patient mix were indirect (curriculum change) active influencing of the patient mix was not found. It would be interesting to see what the effect of tailoring the patient mix to the specific learning goals and needs of individual students would yield.

To avoid bias due to preparation for an examination, research question tailored study outcome (assessments) should be unobtrusive (Swanson et al. 2012); for instance, assessments based on a random selection of routinely made video recordings (Ram et al. 1999; Freeman 2007) could be considered as can frequent work-based assessments by different preceptors. If, at second best, a traditional approach is chosen – similar to the designs we found, triangulation should be strived for. This can be reached by measuring study-tailored self-assessment, including the quality of the learning experience and self-confidence or alike, combined with formal assessment derived from knowledge assessment, and assessment of clinical competence. These studies are preferably performed in multi-institution trials.

In a systematic review, Colthart et al. found indications that 'skills may be better self-assessed than knowledge' and 'the accuracy of self-assessment may be enhanced by increasing the learner's awareness of the standard to be achieved' (Colthart et al. 2008). This is specifically mentioned here, because initially, a medical student may need focused feedback from supervisors; however, as they progress, they should become lifelong learners and must develop the ability to self-assess (Duvivier et al. 2011).

An inquiry into the detailed aspects of patient mix and the contribution of these aspects to learning is desirable. This may be done in qualitative studies; trainees and clinical teachers may be interviewed about their ideas of minimal or optimal patient volumes and spread of diagnosis diversity and their benefit for learning.

The MERSQI, in our experience, is a promising, easy usable instrument for assessing the quality of educational studies. However, some annotations must be made. Several MERSQI items were multi-interpretable, like study design, number of institutions or response rate. Other items, like content and appropriateness of the analysis, did not discriminate between the studies we included. Based on the MERSQI sores, one gets an impression of the spread in quality between the studies. Since the experience with the MERSQI is limited, it does not seem advisable yet to use a cut-off score for excluding low-quality studies.

## Conclusions

In the studies we reviewed, patient mix is mostly presented without definition. Based on our set of studies, we found indications that patient mix, defined by us in terms of volume and diversity, is related to self-reported learning outcome, most evidently, the experienced quality of the learning programme. A relationship between patient mix and the results of formal assessment has rarely been demonstrated. Not only supervision in particular but also learning style seem mediating variables of the relationship between patient mix and learning.

## Acknowledgements

The reviewers wish to thank Faraidi van Etten-Jamaludin, clinical librarian, for her help with designing and performing the search strategies.

***Declaration of interest:*** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

## Notes on contributors

JIP DE JONG, MD, PhD, is a researcher at the GP Specialty Training of the Academic Medical Center-University of Amsterdam. His focus is the patient mix of GP trainees.

MECHTELD VISSER, PhD, is a senior researcher at the GP Specialty Training of the Academic Medical Center-University of Amsterdam.

NYNKE VAN DIJK, MD, PhD, is a senior researcher at the GP Specialty Training of the Academic Medical Center-University of Amsterdam.

CEES VAN DER VLEUTEN, PhD, is a professor of education, chair of the Department of Educational Development and Research and scientific director of the Graduate School of Health Professions Education at Maastricht University.

MARGREET WIERINGA-DE WAARD, MD, PhD, is a professor of General Practice at the GP Specialty Training of the Academic Medical Center-University of Amsterdam.

## References

Ahmed SF, Hughes IA. 1999. Clinical experience during the paediatric undergraduate course. J R Soc Med 92(6):293–298.

Al-Kadri HM, Al-moamary MS, Roberts C, van der Vleuten CPM. 2012. Exploring assessment factors contributing to students' study strategies: Literature review. Med Teach 34(s1):S42–S50.

Ark TK, Brooks LR, Eva KW. 2007. The benefits of flexibility: The pedagogical value of instructions to adopt multifaceted diagnostic reasoning strategies. Med Educ 41(3):281–287.

Australian Medical Council Limited. Standards for Assessment and Accreditation of Specialist Medical Education Programs and Professional Development Programs by the Australian Medical Council 2010, Specialist Education Accreditation Committee. Available from http://www.amc.org.au/images/Accreditation/standards-for-specialist-medical-training-2010.pdf

Bandura A. 2006. Guide for constructing self-efficacy scales. In: Pajares F, Urdan T, editors. Self-efficacy beliefs of adolescents. Greenwhich, Connecticut: IAP. pp 307–337.

BEME Collaboration. 2012. Available from http://www.bemecollaboration.org/

Berlowitz DR, Rosen AK, Moskowitz MA. 1995. Ambulatory care casemix measures. J Gen Intern Med 10(3):162–170.

Boots RJ, Egerton W, McKeering H, Winter H. 2009. They just don't get enough! Variable intern experience in bedside procedural skills. Intern Med J 39(4):222–227.

Brandt Jr E. 1974. The relationship of medical education to patient mix. South Med J 67(4):387.

Bulstrode C, Holsgrove G. 1996. Education for educating surgeons. BMJ 312(7027):326–327.

Carney PA, Pipas CF, Eliassen MS, Mengshol SC, Fall LH, Schifferdecker KE, Olson AL, Peltier DA, Nierenberg DW. 2002. An analysis of students' clinical experiences in an integrated primary care clerkship. Acad Med 77(7):681–687.

Chatenay M, Maguire T, Skakun E, Chang G, Cook D, Warnock GL. 1996. Does volume of clinical experience affect performance of clinical clerks on surgery exit examinations? Am J Surg 172(4):366–372.

Colthart I, Bagnall G, Evans A, Allbutt H, Haig A, Illing J, McKinstry B. 2008. The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide no. 10. Med Teach 30(2):124–145.

Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. 2006. Accuracy of physician self-assessment compared with observed measures of competence: A systematic review. JAMA 296(9):1094–1102.

Dolmans DH, Wolfhagen IH, Essed GG, Scherpbier AJ, Van der Vleuten CP. 2002. The impacts of supervision, patient mix, and numbers of students on the effectiveness of clinical rotations. Acad Med 77:332–335.

Dornan T, Boshuizen H, King N, Scherpbier A. 2007. Experience-based learning: A model linking the processes and outcomes of medical students' workplace learning. Med Educ 41(1):84–91.

Duke P, Curran V, Hollett A. 2011. Training family medicine residents to care for children: What is the best approach? Can Fam Physician 57(2):e46–e50.

Duvivier R, van Dalen J, Muijtjens A, Moulaert V, Van der Vleuten C, Scherpbier A. 2011. The role of deliberate practice in the acquisition of clinical skills. BMC Med Educ 11(1):101. Available from http://www.biomedcentral.com/content/pdf/1472-6920-11-101.pdf

Ericsson KA. 2004. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med 79(10 Suppl):S70–S81.

Eva KW, Regehr G. 2005. Self-assessment in the health professions: A reformulation and research agenda. Acad Med 80(10 Suppl):S46–S54.

Evans JS. 2007. Dual-processing accounts of reasoning, judgment, and social cognition. Annu Rev Psychol 59(1):255–278.

Farnan JM, Petty LA, Georgitis E, Martin S, Chiu E, Prochaska M, Arora VM. 2012. A systematic review: The effect of clinical supervision on patient and residency education outcomes. Acad Med 87(4):428–442.

Fidler H, Lockyer JM, Toews J, Violato C. 1999. Changing physicians' practices: The effect of individual feedback. Acad Med 74(6):702–714.

Freeman A. 2007. Video assessment, In: Jackson N, Amieson AJ, Khan A, editors. Assessment in medical education and training: A practical guide. New york: Radcliff publishing. pp 109–113.

Fung CC, Relan A, Wilkerson L. 2007. Demystifying "learning" in clinical rotations: Do immersive patient encounters predict achievement on the clinical performance examination (CPX)? Acad Med 82:S97–S100.

Greenberg LW, Getson P. 1999. Clinical experiences of medical students in a pediatric clerkship and performance on the NBME subject test: Are they related? Med Teach 21:420–423.

Gruppen LD, Wisdom K, Anderson DS, Woolliscroft JO. 1993. Assessing the consistency and educational benefits of students' clinical experiences during an ambulatory care internal medicine rotation. Acad Med 68(9):674–680.

Hamdy H, Prasad K, Anderson MB, Scherpbier A, Williams R, Zwierstra R, Cuddihy H. 2006. BEME systematic review: Predictive values of measurements obtained in medical schools and future performance in medical practice. Med Teach 28(2):103–116.

Hand R, Garg M, Dajani KF. 1993. Patient mix in the primary ambulatory care clinics of an academic medical center. Acad Med 68(10):803–805.

Hattie J, Timperley H. 2007. The power of feedback. Rev Educ Res 77(1):81–112.

Hutchinson A, Parkin D, Philips P. 1991. Case mix measures for ambulatory care. J Public Health Med 13(3):189–197.

Jacobson EW, Keough WL, Dalton BE, Giansiracusa DF. 1998. A comparison of inpatient and outpatient experiences during an internal medicine clerkship. Am J Med 104(2):159–162.

Jolly BC, Jones A, Dacre JE, Elzubeir M, Kopelman P, Hitman G. 1996. Relationships between students' clinical experiences in introductory clinical courses and their performances on an objective structured clinical examination (OSCE). Acad Med 71(8):909–916.

Karle, H. Postgraduate Medical Education WFME Global Standards for Quality Improvement 2003. http://www.wfme.org/standards/pgme

Kirkpatrick DL, Kirkpatrick JD. 2006. Evaluating training programs: The four levels. 2nd ed. San Fransisco: Berret-Koehler Publishers Inc.

Kolb D. 1984. Experiential learning as the science of learning and development. Englewood Cliffs, NJ: Prentice Hall.

Kruger J, Dunning D. 1999. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. J Pers Soc Psychol 77(6):1121–1134.

Lampe CJ, Coates WC, Gill AM. 2008. Emergency medicine subinternship: Does a standard clinical experience improve performance outcomes? Acad Emerg Med 15(1):82–85.

Liaison Committee on Medical Education (LCME). 2011. Functions and structure of a medical school. Standards for accreditation of medical education programs leading to the M.D. degree. Available from http://www.lcme.org/functions.pdf

Mann KV. 2011. Theoretical perspectives in medical education: Past experience and future possibilities. Med Educ 45(1):60–68.

Marteau TM, Wynne G, Kaye W, Evans TR. 1990. Resuscitation: Experience without feedback increases confidence but not skill. BMJ 300(6728):849–850.

Martin IG, Stark P, Jolly B. 2000. Benefiting from clinical experience: The influence of learning style and clinical experience on performance in an undergraduate objective structured clinical examination. Med Educ 34(7):530–534.

McAllister Jr RG, Dzur J. 1974. Medical admissions to a teaching service: A six months' review in a veterans hospital. South Med J 67(4):388–392.

McLeod PJ, Meagher T, Tamblyn RM, Zakarian R. 1997. Are ambulatory care-based learning experiences different from those on hospital clinical teaching units? Teach Learn Med 9(2):125–130.

McManus IC, Richards P, Winder BC, Sproston KA. 1998. Clinical experience, performance in final examinations, and learning style in medical students: Prospective study. BMJ 316(7128):345–350.

Newble DI, Jaeger K. 1983. The effect of assessments and examinations on the learning of medical students. Med Educ 17(3):165–171.

Nomura K, Yano E, Aoki M, Kawaminami K, Endo H, Fukui T. 2008. Improvement of residents' clinical competency after the introduction of new postgraduate medical education program in Japan. Med Teach 30(6):e161–e169.

Norman G. 2009. Dual processing and diagnostic errors. Adv Health Sci Educ 14:37–49.

O'Hara BS, Saywell RM, Zollinger TW, Wooldridge JS, Kelton GM, Ries JS, Burba JL. 2002. Students' experience with women's health care in a family medicine clerkship. Med Educ 36(5):456–465.

Patel VL, Cranton PA. 1983. Transfer of student learning in medical education. J Med Educ 58(2):126–135.

Pelaccia T, Tardif J, Triby E, Charlin B. 2011. An analysis of clinical reasoning through a recent and comprehensive approach: The dual-process theory. Med Educ Online 16. Available from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3060310/pdf/MEO-16-5890.pdf.

Pulito AR, Donnelly MB, Plymale M, Mentzer J. 2006. What do faculty observe of medical students' clinical performance? Teach Learn Med 18(2):99–104.

Raghoebar-Krieger HM, Sleijfer DT, Hofstee WK, Kreeftenberg HG, Bender W. 2001. The availability of diseases for medical students in a university hospital. Med Teach 23(3):258–262.

Ram P, Grol R, Rethans JJ, Schouten B, van der Vleuten C, Kester A. 1999. Assessment of general practitioners by video observation of communicative and medical performance in daily practice: Issues of validity, reliability and feasibility. Med Educ 33(6):447–454.

Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. 2007. Association between funding and quality of published medical education research. JAMA 298(9):1002–1009.

Royal College of General Practitioners. 2011. Available from http://www.rcgp-curriculum.org.uk/PDF/qmts_GP_Specific_Training_Standards_dec08.pdf

Sargeant J, Mann K, van der Vleuten C, Metsemakers J. 2008. "Directed" self-assessment: Practice and feedback within a social context. J Contin Educ Health Prof 28(1):47–54.

Saywell RM, O'Hara BS, Zollinger TW, Wooldridge JS, Burba JL, Mckeag DB. 2002. Medical students' experience with musculoskeletal diagnoses in a family medicine clerkship. Med Teach 24(2):186–192.

Schmidt HG, Norman GR, Boshuizen HP. 1990. A cognitive perspective on medical expertise: Theory and implication. Acad Med 65(10):611–621.

Schwiebert LP, Ramsey Jr CN, Davis A. 1993. Standardizing the clinical content of a third-year family medicine clerkship. Fam Med 25(4):257–261.

Sorensen Hoifoidt T, Sexton H, Olstad R. 2004. Experiences from psychiatric rotation for pre-registration house officers: Contributions to subjective learning. Med Educ 38(4):349–357.

Steinert Y, Mann K, Centeno A, Dolmans D, Spencer J, Gelula M, Prideaux D. 2006. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. Med Teach 28(6):497–526.

Swanson DB, Case SM, Van der Vleuten CP. 2012. Strategies for student assessment. In: Wadhwa S, editor. Teaching learning and studying. New Delhi: Sarup & Son. p 157.

Teunissen PW, Scheele F, Scherpbier AJ, van der Vleuten CP, Boor K, van Luijk SJ, van Diemen-Steenvoorde JA. 2007. How residents learn: Qualitative evidence for the pivotal role of clinical activities. Med Educ 41(8):763–770.

Tracey JM, Arroll B, Richmond DE, Barham PM. 1997. The validity of general practitioners' self assessment of knowledge: Cross sectional study. BMJ 315(7120):1426–8.

van der Vleuten CPM, Swanson DB. 1990. Assessment of clinical skills with standardized patients: State of the art. Teach Learn Med 2(2):58–76.

van der Zwet J, Hanssen VG, Zwietering PJ, Muijtjens AM, van der Vleuten CP, Metsemakers JF, Scherpbier AJ. 2010. Workplace learning in general practice: Supervision, patient mix and independence emerge from the black box once again. Med Teach 32(7):e294–e299.

Violato C, Lockyer J. 2006. Self and peer assessment of pediatricians, psychiatrists and medicine specialists: Implications for self-directed learning. Adv Health Sci Educ Theory Pract 11(3):235–244.

Wimmers PF, Schmidt HG, Splinter TA. 2006. Influence of clerkship experiences on clinical competence. Med Educ 40(5):450–458.

Wimmers PF, Splinter TA, Hancock GR, Schmidt HG. 2007. Clinical competence: General ability or case-specific? Adv Health Sci Educ Theory Pract 12(3):299–314.

Yu TC, Wheeler BR, Hill AG. 2011. Effectiveness of standardized clerkship teaching across multiple sites. J Surg Res 168(1):e17–e23.

## Appendix 1. **Search strategies**

### *1. PubMed*

Set 1: patient mix

((((“Diagnosis-Related Groups"[MeSH] OR diagnosis related group*[tiab] OR case mix*[tiab] OR casemix*[tiab] OR diagnosis cluster*[tiab] OR patient distribution*[tiab] OR clinical exposure*[tiab] OR clinical encounter*[tiab] OR clinical experience*[tiab] OR patient mix*[tiab] OR logbook*[tiab] OR consultation[tiab] OR selected conditions[tiab] OR disease management[tiab] OR clinical method*[tiab] OR diagnosis cluster*[tiab] OR distribution patients[tiab])))) AND

Set 2: learning

(“Curriculum"[MeSH] OR curricul*[tiab] OR “Education, Medical"[MeSH] OR medical education[tiab] OR “Clinical Competence"[MeSH] OR clerkship*[tiab] OR trainee*[tiab] OR training[tiab] OR resident*[tiab] OR residency[tiab] OR ((“work"[MeSH Terms] OR work[tiab]) AND (based[tiab]) AND (“learning"[MeSH] OR learn*[tiab]))) AND

Set 3: population

((“Hospitals, Teaching"[MeSH] OR teaching hospital*[tiab] OR “Specialties, Medical/education"[MeSH] OR “Primary Health Care"[MeSH] OR student*[tiab] OR practice[tiab]))

**2. Embase**

Set 1: patient mix:

case mix/or case mix*.ti,ab. or casemix*.ti,ab. or diagnosis related group/or diagnosis related group*.ti,ab. or clinical exposure*.ti,ab. or clinical encounter*.ti,ab. or clinical experience*.ti,ab. or patient mix*.ti,ab. or logbook*.ti,ab. or consultation.ti,ab. or selected conditions.ti,ab. or disease management.ti,ab. or clinical method*.ti,ab. or (diagnosis adj1 cluster*).ti,ab. or (distribution adj2 patient*).ti,ab

Set 2: learning:

curriculum/or curricul*.ti,ab. or exp Medical Education/or medical education.ti,ab. or exp Clinical Competence/or clerkship*.ti,ab. or trainee*.ti,ab. or training.ti,ab. or work based learning*.ti,ab. or (residency or resident*).ti,ab.

Set 3: population:

exp Teaching Hospital/or teaching hospital*.ti,ab. or exp medicine/or exp Primary Health Care/or student*.ti,ab. or practice.ti,ab

**3. Cochrane Library**

(1) (case mix*):ti,ab,kw
(2) (casemix*):ti,ab,kw
(3) (diagnosis related group*):ti,ab,kw
(4) MeSH descriptor Diagnosis-Related Groups explode all trees
(5) (clinical exposure*):ti,ab,kw
(6) (clinical encounter*):ti,ab,kw
(7) (clinical experience*):ti,ab,kw
(8) (patient mix*):ti,ab,kw
(9) (logbook*):ti,ab,kw
(10) (consultation):ti,ab,kw
(11) (selected conditions):ti,ab,kw
(12) (diseases management):ti,ab,kw
(13) (clinical method*):ti,ab,kw
(14) (diagnosis cluster*):ti,ab,kw
(15) (distribution patient*):ti,ab,kw
(16) (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15)
(17) MeSH descriptor Curriculum explode all trees
(18) (curricul*):ti,ab,kw
(19) MeSH descriptor Education, Medical explode all trees
(20) (medical education):ti,ab,kw
(21) MeSH descriptor Clinical Competence explode all trees
(22) (clerkship*):ti,ab,kw
(23) (trainee*):ti,ab,kw
(24) (training):ti,ab,kw
(25) (work based learning):ti,ab,kw
(26) (residency or resident*):ti,ab,kw
(27) (#17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26)
(28) MeSH descriptor Hospitals, Teaching explode all trees
(29) (teaching hospital*):ti,ab,kw
(30) MeSH descriptor Specialties, Medical explode all trees with qualifier: ED
(31) MeSH descriptor Primary Health Care explode all trees
(32) (student*):ti,ab,kw
(33) (practice*):ti,ab,kw
(34) (#28 OR #29 OR #30 OR #31 OR #32 OR #33)
(35) (#16 AND #27 AND #34)

**4. ERIC**

(1) exp "Case Method (Teaching Technique)"/
(2) exp Clinical Experience/or clinical exposure.mp.
(3) clinical encounter.mp.
(4) logbook*.ti,ab.
(5) exp Patients/
(6) (case mix* or casemix*).ti,ab.
(7) 1 or 2 or 3 or 4 or 5 or 6
(8) exp Curriculum/
(9) curricul*.ti,ab.
(10) exp Medical Education/
(11) medical education.ti,ab.
(12) clinical competence.mp.
(13) exp "Clinical Teaching (Health Professions)"/
(14) clerkship*.ti,ab.
(15) exp Trainees/
(16) trainee*.ti,ab.
(17) residen*.ti,ab.
(18) work based learning.mp.
(19) 11 or 9 or 17 or 12 or 15 or 14 or 8 or 18 or 16 or 10 or 13
(20) exp Medical Education/
(21) teaching hospital*.ti,ab.
(22) exp Primary Health Care/
(23) exp Medical Students/
(24) 22 or 21 or 23 or 20
(25) 24 and 7 and 19
(26) *medical education/
(27) 22 or 21 or 26 or 23
(28) 27 and 7 and 19

**5. Web of Science**

Title = ("family practice" OR "general pract*" OR "family medicine" OR "primary care" OR "internal medicine" of psychiatr* OR "hospital*" or surgery) AND Title = (curriculum or training* OR trainee* OR clerks* OR residen* OR education* OR learn* OR medical student* OR internship* OR work based learning) AND Title = ("case mix" OR "casemix" OR "experience*" or disease* OR logbook* OR "patient mix" OR examination* OR patient* OR diagnos* OR condition*)

**Glossary**

BEME: Best Evidence Medical Education
GP: General Practitioners
MCQ: Multiple-choice questions
MERSQI: Medical Education Research Study Quality Instrument
OSCE: Objective Structured Clinical Examination
PM: Patient mix
RCT(s): Randomised Controlled Trial(s)