

BMJ Open Factors predicting a change in diagnosis in patients hospitalised through the emergency room: a prospective observational study

Stefanie C Hautz,¹ Luca Schuler,² Juliane E Kämmer,^{3,4} Stefan K Schaubert,⁵ Meret E Ricklin,² Thomas C Sauter,² Volker Maier,⁶ Tanja Birrenbach,^{2,6} Aristomenis Exadaktylos,² Wolf E Hautz²

To cite: Hautz SC, Schuler L, Kämmer JE, *et al.* Factors predicting a change in diagnosis in patients hospitalised through the emergency room: a prospective observational study. *BMJ Open* 2016;**6**: e011585. doi:10.1136/bmjopen-2016-011585

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2016-011585>).

Received 19 February 2016
Revised 5 April 2016
Accepted 22 April 2016



CrossMark

For numbered affiliations see end of article.

Correspondence to
Dr Wolf E Hautz;
wolf.hautz@insel.ch

ABSTRACT

Introduction: Emergency rooms (ERs) generally assign a preliminary diagnosis to patients, who are then hospitalised and may subsequently experience a change in their lead diagnosis (cDx). In ERs, the cDx rate varies from around 15% to more than 50%. Among the most frequent reasons for diagnostic errors are cognitive slips, which mostly result from faulty data synthesis. Furthermore, physicians have been repeatedly found to be poor self-assessors and to be overconfident in the quality of their diagnosis, which limits their ability to improve. Therefore, some of the clinically most relevant research questions concern how diagnostic decisions are made, what determines their quality and what can be done to improve them. Research that addresses these questions is, however, still rare. In particular, field studies that allow for generalising findings from controlled experimental settings are lacking. The ER, with its high throughput and its many simultaneous visits, is perfectly suited for the study of factors contributing to diagnostic error. With this study, we aim to identify factors that allow prediction of an ER's diagnostic performance. Knowledge of these factors as well as of their relative importance allows for the development of organisational, medical and educational strategies to improve the diagnostic performance of ERs.

Methods and analysis: We will conduct a field study by collecting diagnostic decision data, physician confidence and a number of influencing factors in a real-world setting to model real-world diagnostic decisions and investigate the adequacy, validity and informativeness of physician confidence in these decisions. We will specifically collect data on patient, physician and encounter factors as predictors of the dependent variables. Statistical methods will include analysis of variance and a linear mixed-effects model.

Ethics and dissemination: The Bern ethics committee approved the study under KEK Number 197/15. Results will be published in peer-reviewed scientific medical journals. Authorship will be determined according to ICMJE guidelines.

Strengths and limitations of this study

- This is a prospective observational study of all patients admitted to internal medicine through the emergency room.
- It includes the collection of patient, physician and encounter factors to predict diagnostic quality.
- It uses linear mixed-effects modelling of dependent variables.
- The study is limited to a single centre.

Trial registration number: The study protocol Version 1.0 from 17 May 2015 is registered in the Inselspital Research Database Information System (IRDIS) and with the IRB ('Kantonale Ethikkommission') Bern under KEK Number 197/15.

BACKGROUND

Diagnostic errors contribute substantially to preventable medical errors.¹ Emergency room (ER) usually assign a preliminary diagnosis to patients who are subsequently hospitalised and may experience a change in their lead diagnosis (cDx) from admission to discharge. This cDx often results from diagnostic errors made in the ER in the first place. Berner and Graber² estimated the rate of diagnostic errors to be at around 15–30% in contexts such as the ER, comparable to what others found.³ A previous Swiss study even found that the diagnosis of patients presenting to the ER with non-specific symptoms changed in 64% of the cases within the following 90 days.⁴

Among the most frequent reasons for diagnostic errors are cognitive slips, which mostly result from faulty data synthesis.⁵ Furthermore, physicians have been repeatedly found to be poor self-assessors and generally overconfident

in the quality of their diagnosis,⁶ which limits their ability to improve. Therefore, some of the clinically most relevant research questions concern how diagnostic decisions are made, what determines their quality and what can be done to improve them.^{2 5} Studies addressing these research questions—especially field studies—are, however, still rare. We thus aim to conduct a single-centre observational field study by collecting diagnostic decision data, physician confidence and a number of influencing factors (see below) in an ER to model real-world diagnostic decisions and investigate the adequacy, validity and informativeness of physician confidence in these decisions.

Specifically, we will address the following research questions:

1. What is the rate of change in diagnosis (cDx) in patients hospitalised in an internal medicine (IM) ward through the ER?
2. How are diagnostic quality and physician confidence in a real-world setting affected by physician, patient and context factors?
3. How are diagnostic quality and physician confidence related?

Since diagnostic error is hard to define and challenging to measure,⁷ as well as perceived as highly judgemental, we chose to assess change in diagnosis. This change is a purely descriptive variable but a necessary (although not sufficient) prerequisite for diagnostic error.

To further guide the study conception, instrument selection and interpretation of results, the project draws on three conceptual frameworks, namely, (1) a procedural model of clinical reasoning,⁸ (2) situated cognition theory⁹ and (3) a contemporary model of physician confidence.^{10 11} They are described in the following.

Clinical reasoning

Of the multiple reasons for diagnostic error, cognitive errors are among the most frequent.⁵ These cognitive errors can be distinguished by the point in the diagnostic process at which they occur:^{5 8}

1. Faulty data acquisition (such as the failure to obtain relevant information);
2. Faulty data synthesis (as the physician integrates findings on a patient with his or her own knowledge).

Most cognitive errors occur during data synthesis,^{2 8 12} although most models of clinical reasoning¹³ regard data acquisition and synthesis as an iterative and circular rather than a linear process of reasoning.¹⁴

Situated cognition theory

It is currently largely unknown how different types of cognitive errors are triggered¹⁵ and it is generally acknowledged that one cannot teach or learn a general problem solving skill.¹⁶ Rather than being rooted in the physician's abilities alone, however, cognitive errors seem to be the result of an interaction between a physician and (1) his or her patient and (2) the context. In

fact, context specificity has been termed 'the one truth in medical education'.¹⁷ Situated cognition theory has been successfully used to explain the recurrent finding of context specificity in experimentally controlled settings.⁹ According to situated cognition theory, the decision quality is determined by three different factors: physician factors (eg, fatigue, experience), patient factors (eg, urgency) and context factors (eg, overall workload, time of day). It also incorporates interactions between these three categories, such as the interaction of physician and patient characteristics resulting in language barriers. Situated cognition theory was thus selected for this study because it suggests a classification of measurable variables into the latent physician, patient and context factors and establishes a relationship between them.

Physician confidence

A current model of physician confidence has been proposed by Eva and Regehr:^{10 11} they distinguish between (1) global self-assessment, which results more from self-conception rather than from an accurate integration of past performance, and (2) self-monitoring, the moment-by-moment reflection in action that is context dependent and equates to situational confidence-regulating behaviour.¹¹ As self-assessment has repeatedly been demonstrated to badly match objective measures of performance,⁶ we will focus on self-monitoring in order to assess the relationship between situational physician confidence and decision quality, physician, patient and encounter factors. Confidence (or lack thereof) prompts surgeons 'to slow down when they should',¹⁸ leads physicians to order further diagnostic tests in difficult cases¹⁹ and helps students to 'know when to look it [sic] up'.²⁰ Situational confidence thus seems to be a strong modulator of physician's actions and therefore has the potential to limit the need for changes in a lead diagnosis. Previous experimental studies have found good indication of adequate self-monitoring in high performing individuals but not in low performing individuals.^{10 21} However, little is known about individual differences in self-monitoring capabilities in medical personnel and the factors influencing them.^{11 22} We thus chose to use situational confidence as a second dependent variable of this study and aim to model it through the collected independent variables.

METHODS

We will collect data in the ER of the Bern University Hospital for all patients of aged 18 years and older admitted to any IM ward. Patient inclusion started on 15 August 2015 and is planned to last until 15 May 2016. The ER at Bern University is a self-contained, interdisciplinary unit that employs around 45 physicians and 120 nurses, and sees more than 40 000 patients per year, of which around 30% are admitted to the hospital.²³

Sample size

A change in lead diagnosis from admittance to discharge and the admitting physician's confidence at the time of admission are the primary dependent variables of this study that we will correlate and aim to model through all other collected data, using a multivariate analysis of variance and a linear mixed-effects model. Sample size is estimated based on $\alpha=0.05$, power=85%, 8 independent predictor variables for the primary outcome (cDx), and $R=0.2$ and a 15% dropout rate, to be 500.

Inclusion/exclusion

Included are patients who are ≥ 18 years old, received an IM lead diagnosis at the ER and are hospitalised to the IM. Excluded are patients who do not fulfil the inclusion criteria, are hospitalised for palliative care, or have IM-physicians involved as consultants only. Physicians included are all physicians from the ER who voluntarily participate in the study. Physicians are reimbursed with 10 CHF (Swiss francs) per questionnaire (see below) they fill in.

Pre-study

In preparation of the main study, a researcher (LS) retrospectively assessed the cDx rate for all patients admitted to the IM via the ER between 1 January and 31 March 2015. This prestudy aimed to assess the quality of the available data from the hospital's electronic patient documentation systems and the number of patients hospitalised to the IM via the ER. Based on these data, the power calculations for the main study were conducted and the expected duration of the data collection phase could be estimated. Also, based on the data collected in the prestudy, four independent medical experts developed and pretested a schema to assess the cDx rate, that is, whether a change in the lead diagnosis made at the ER had occurred or not (see [table 1](#)). Two raters then independently applied the schema to the diagnoses of 90 randomly selected patients to reduce bias. Whenever one rater coded a diagnosis as 'not classifiable' or the two ratings did not agree, the final classification was achieved through a discussion of all four expert raters.

We chose this framework to classify the primary dependent variable (change in diagnosis) over existing classification schemes for diagnostic error for several reasons:

1. By classifying change in diagnosis, we decided on a descriptive and non-judgmental framework because there is no single established definition of diagnostic error and the IM's discharge diagnosis is not necessarily more accurate than the ER's admittance diagnosis.⁴ Labelling a differing admittance diagnosis erroneous may thus not be justified. Furthermore, previous studies have demonstrated only a small overlap between errors during the diagnostic process and patient harm.²⁴ We would argue that a change in lead diagnosis is a clinically more relevant outcome

Table 1 Schema to classify a pair of diagnoses from ER (admission) and IM (discharge)

IM discharge diagnosis compared to ER diagnosis is:	
diagnosis is:	Explanation:
Identical	The two diagnoses are either verbatim or medically identical.
Precise	The IM discharge diagnosis is more precise than the ER diagnosis (eg, by adding an established, disease-specific score or the result of a test that was not available at the ER (eg, microbiological cultivation)) but otherwise identical.
Complicated	The lead discharge diagnosis from the IM was not foreseeable at the time of hospital admission at the ER but became the most prominent during hospitalisation (eg, a pulmonary embolism as a complication of the hospitalisation).
Hierarchically different	The lead ER diagnosis is listed among the IM discharge diagnoses but is not the lead discharge diagnosis.
Diagnostically different	The lead ER diagnosis is not among the IM discharge diagnoses.
Not classifiable	

ER, emergency room; IM, internal medicine.

parameter than any definition of diagnostic error: when a patient who initially requires hospital admission improves to be discharged with an identical diagnosis, treatment was most likely adequate. When, however, a change in diagnosis occurs during hospitalisation, the patient is at risk of missed, delayed or inadequate initial treatment. Thus, studying those factors predicting a change in diagnosis leads to the identification of clinically relevant factors.

2. Assessing diagnostic error may be influenced by factors such as the evolving nature of a disease, hindsight-bias in the rating and the diagnosing clinicians desire to balance the risks of under-diagnosis and over-diagnosis (see ref ⁷ for a detailed discussion of the challenges in defining and measuring diagnostic error). The schema used to classify a change in diagnoses instead (see [table 1](#)) accounts for these factors.

Main study procedure

For the main study, we will collect data in the ER of the Bern University Hospital for all patients admitted to the

IM within a period of approximately 9 months. Of the lead diagnoses, about 10–15% will change within the week after admittance, according to preliminary data from the prestudy that are consistent with the literature.² To get a deeper insight into the reasons for these changes, we will collect data on patient, physician and encounter factors as predictors of the dependent variables at three time points (see figure 1):

1. All physicians working at the ER and participating in the study will be asked to answer demographic questions (physician factors stable over all admittance decisions) once at the time of inclusion comprising age, gender and years, and kind of training. We will further record each physician's performance in a cognitive reflection test²⁵ to assess their tendency to follow intuitive judgements (all questionnaires available on request).
2. At admittance, the lead diagnosis and physician's confidence will be recorded. Further, the following patient, physician and encounter factors will be measured: patient age and gender, triage category, date and time, language barrier, subjective and objective (see below) physician workload, and physician confidence in his or her diagnosis together with the diagnosis made. Data will be collected from (1) physician

questionnaires and (2) the hospital's electronic patient documentation systems.

3. Twenty-eight days after admittance, each admitted patient's current lead diagnosis from the IM or his or her primary discharge diagnosis from the IM, if already discharged, will be recorded from the hospital's electronic patient documentation systems. Twenty-eight days were chosen as follow-up period in respect to commonly used periods to assess morbidity and mortality in many studies in the field of emergency and intensive care medicine.²⁶

Furthermore, we will continuously collect data on the noise level in the ER with a noise level logger and on overall workload, by collecting the number of patients and the triage category, waiting time, total time in the ER and total time to hospitalisation for each patient treated in the ER during the study period. For a detailed description of all data collected, their respective source and scaling, and how these variables might be aggregated to factors, please see online supplementary additional file 1.

Data sets to be analysed

Data will be entered into a web-based database that fulfils the requirements of the Swiss Human Research Act. The databases RedCap and Sharepoint are provided by CTU Bern (clinical trials unit). For statistical analyses, coded data will be exported. The code book that relates each patient to his or her code will be kept strictly under lock and key by the ERs study administrator (MER).

Data processing

In total, we will collect two dependent variables (cDx and admitting physician's confidence) and 23 predictor variables during the study (see online supplementary additional file 1). The cDx rate will be determined using the schema developed in the prestudy (see table 1).

First, we will examine correlations among the predictor variables and aggregate highly correlating variables that are linkable on a theoretical basis. For example, we expect a high correlation between (1) a physician's age, (2) their hierarchical position, (3) their total medical experience and (4) their years of experience in emergency medicine (constituting the factor 'experience'; see last column of online supplementary additional file 1). We further expect a high correlation between (1) a physician's subjective workload, (2) their fatigue, (3) their perceived difficulty with a diagnosis and (4) their self-reported experience with this diagnosis (constituting the factor 'working impairment'). We will calculate indices based on Z-standardised raw data for these variables.

The number of patients per triage category, the number of attending physicians, the number of treatment bays in the ER and the number of admitted patients in the ER at any one time, will be used to

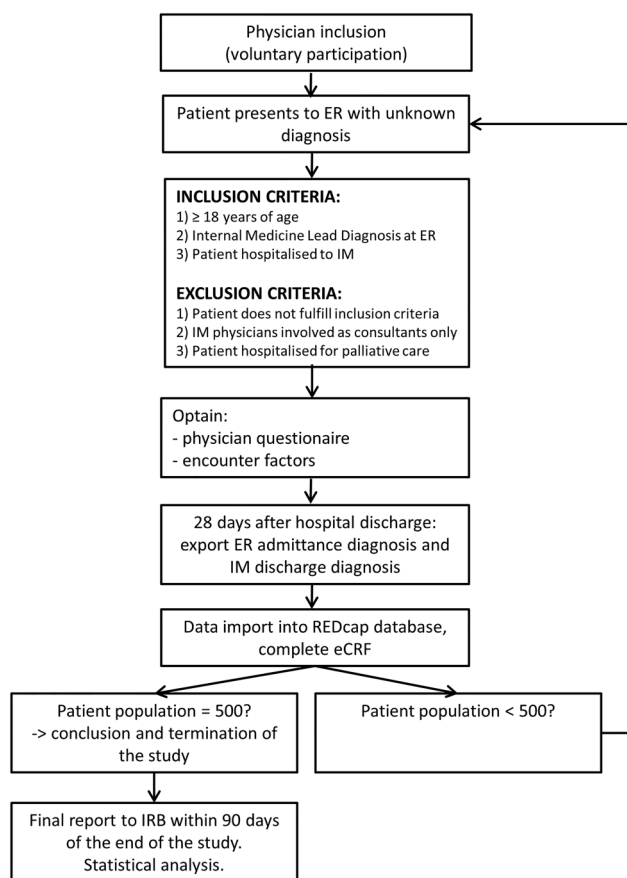


Figure 1 Study design. ER, emergency room; eCRF, electronic care report form; IM, internal medicine; IRB, Institutional Review Board.

calculate an ER overcrowding score according to the EDWIN model.²⁷ For the noise level, which is continuously recorded, we will calculate the frequency and duration of peaks (defined as all noise above 65 dB(A)) within blocks of 2 h length and use these aggregated data as context factor in subsequent analyses.

For all remaining variables (marked as '?' in the online appendix SI_variables_and_sources.xls), we will calculate a principal component analysis and use their factor scores to calculate summary indices. This procedure will be used to combine variables to additional physician, context and patient factors.

Last, we will compute the cognitive reflection test-score (CRT-score) for each physician by counting the number of correct answers to the questions from the CRT.²⁵

Handling of missing data

Data will be screened for missing values. This will include exploration of patterns of missing values dependent on design components such as physicians and time of day. Missing data will be handled either through multiple imputation or the use of full information maximum likelihood estimation, depending on the characteristics of the variable and the pattern of missing values. Drop-out rate and missing data rate will be reported. Furthermore, the duration of the data collection will be held flexible and last until the target number of patients included in the study is achieved.

Planned analyses

Data will be analysed in SPSS, Matlab and R Statistical software. Rater agreement in the classification of IM and ER diagnoses will be reported as coefficient κ .

To assess selection bias, we will report how participating physicians differ from all physicians in the department regarding publicly available variables. These variables include gender, board certification, discipline of board certification and current position. We will further report how included patients differ from all patients of the IM in age, gender and triage category, although selection bias for patients is unlikely because we will include all patients admitted to IM with only minimal exclusion criteria.

For research question 1, descriptive statistics (frequency, frequency per patient subgroup based on gender and age, frequency by diagnostic subgroup) will be reported.

For research question 2, we will conduct a generalised linear mixed-effects model in R. In this approach, we will model the outcome (change in diagnosis/no change) as a Bernoulli distributed variable. The linear component of the model entails the covariates, which will be linked to the outcome variable, using a logit link function. The random effects in the model will include the logistic error term and a residual term for the physician identifier with a mean of 0 and unknown variance.

Research question 3 concerns the question of how well physician confidence is calibrated. We will use methods from cognitive psychology²⁸ to calculate the level and direction of (mis-) calibration. In detail, we will assess how accuracy (ie, change in diagnosis/no change) is aligned with confidence, ranging from 0 (best possible calibration) to 1 (worst possible calibration). Calibration will be calculated as the weighted mean of the squared difference between confidence and proportion correct for each confidence level.²⁸ To compute the direction and magnitude of miscalibration, we will calculate the over-under index (O-U index), ranging from -1 (highest possible level of underconfidence) to +1 (highest possible level of overconfidence). Miscalibration will be computed as the difference between confidence and accuracy.²⁸

Results of the prestudy

During the prestudy period, 186 patients were admitted from the ER to IM. Results of the prestudy thus indicate that we should be able to include, on average, 2 patients admitted to IM through the ER per day, thus determining an 8.3-month study period to include 500 patients.

Of the 90 patients randomly selected from the prestudy to assess the cDx rate, four did not fulfil the main studies inclusion criteria and were thus excluded. One patient's diagnoses (1.2%) were rated as not classifiable by at least one rater. Following discussion with the other raters, these diagnoses were later classified as 'change in diagnosis'. Rater agreement was good ($\kappa=0.61$). Table 2 lists the frequency with which pairs of ER/IM diagnoses were assigned to each category after rater agreement.

ETHICS AND DISSEMINATION

This is a prospective observational study with no interventions performed. Patient data are collected during usual care at the ER and the IM. No additional patient data are recorded for this study. Physicians participate on a voluntary basis. Participant anonymity for both, patients and physicians, will be respected at all times by anonymisation of physician and patient data. According to Swiss research and data privacy laws, no informed consent is required for such studies. Any protocol amendments will be submitted to the Bern ethics committee for approval before

Table 2 Classification of the ER and IM diagnoses of 90 randomly selected patients from the prestudy

IM discharge diagnosis compared to ER diagnosis is	Per (n) cent
Identical	27 31.4
Precise	33 38.4
Complicated	2 2.3
Hierarchically different	9 10.5
Diagnostically different	14 16.3
Not classifiable	0 0

(n)=frequency.

ER, emergency room; IM, internal medicine.

implementation. Termination of the study will be reported to the same committee.

Publication and dissemination of results

Results will be presented to participating physicians within the Departments of Emergency Medicine and Internal Medicine at the University Hospital Bern, and at scientific meetings. Results will be published in peer-reviewed scientific medical journals and authorship will be determined according to ICMJE guidelines.

Author affiliations

¹Medical Faculty, Department of Evaluation and Assessment, Institute of Medical Education, University of Bern, Bern, Switzerland

²Department of Emergency Medicine, Inselspital, University Hospital Bern, Bern, Switzerland

³Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

⁴AG Progress Test Medizin, Charité Medical School, Berlin, Germany

⁵Faculty of Educational Sciences, Centre for Educational Measurement, University of Oslo, Oslo, Norway

⁶Department of General Internal Medicine, Inselspital, Bern, Switzerland

Acknowledgements The authors would like to acknowledge Dr Laura Zwaan, Institute of Medical Education Research Rotterdam, The Netherlands, and Dr Robert El-Kareh, University of California at San Diego, La Jolla, USA, for their valuable and constructive critique of an earlier version of the manuscript.

Contributors LS collected the data for the prestudy. JEK and SKS analysed the data. SCH wrote the manuscript. All the authors designed the study, revised the manuscript critically and gave their approval of the final version of the manuscript, and agreed to be accountable for all aspects of the published work.

Funding The study is funded through internal research funds of the department of emergency medicine, Inselspital University Hospital Bern.

Competing interests None declared.

Ethics approval The Bern ethics committee registered the study as a quality evaluation study under Kantonale Ethikkommission Bern number 197/15 and waived the need for informed consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- Kohn LT, Corrigan JM, Donaldson MS. *To err is human: building a safer health system*. National Academies Press, 2000.
- Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med* 2008;121:S2.
- Schwartz A, Elstein AS. Clinical reasoning in medicine. In: Higgs J, Jones M, Loftus S, Christensen N., eds. *Clinical reasoning in the health profession*; Elsevier, 2008:223–34.
- Peng A, Rohacek M, Ackermann S, et al. The proportion of correct diagnosis is low in emergency patients with nonspecific complaints presenting to the emergency department. *Swiss Med Wkly* 2015;145:w14121.
- Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165:1493–9.
- Davis DA, Mazmanian PE, Fordis M, et al. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA* 2006;296:1094–102.
- Zwaan L, Singh H. The challenges in defining and measuring diagnostic error. *Diagnosis (Berl)* 2015;2:97–103.
- Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Med Educ* 2010;44:94–100.
- Durning SJ, Artino AR, Boulet JR, et al. The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance in experts?). *Adv Health Sci Educ Theory Pract* 2012;17:65–79.
- Eva KW, Regehr G. Exploring the divergence between self-assessment and self-monitoring. *Adv Health Sci Educ Theory Pract* 2011;16:311–29.
- Eva KW, Regehr G. "I'll never play professional football" and other fallacies of self-assessment. *J Contin Educ Health Prof* 2008;28:14–9.
- Kuhn GJ. Diagnostic errors. *Acad Emerg Med* 2002;9:740–50.
- Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ* 2005;39:98–106.
- Elstein AS, Shulman LS, Sprafka SA. *Medical problem solving: an analysis of clinical reasoning*. Cambridge, MA: Harvard University Press, 1978.
- McLaughlin K, Eva KW, Norman GR. Reexamining our bias against heuristics. *Adv Health Sci Educ Theory Pract* 2014;19:457–64.
- Ilgen JS, Humbert AJ, Kuhn G, et al. Assessing diagnostic reasoning: a consensus statement summarizing theory, practice, and future needs. *Acad Emerg Med* 2012;19:1454–61.
- Eva KW. On the relationship between problem-solving skills and professional practice. In: Kanen C, ed. *Elaborating professionalism innovation and change in professional education*. London: Springer, 2011:17–34.
- Moulton CA, Regehr G, Lingard L, et al. 'Slowing down when you should': initiators and influences of the transition from the routine to the effortful. *J Gastrointest Surg* 2010;14:1019–26.
- Meyer AN, Payne VL, Meeks DW, et al. Physicians' diagnostic accuracy, confidence, and resource requests: a vignette study. *JAMA Intern Med* 2013;173:1952–8.
- Eva KW, Regehr G. Knowing when to look it up: a new conception of self-assessment ability. *Acad Med*. 2007;82(10 Suppl):S81–4.
- McConnell MM, Regehr G, Wood TJ, et al. Self-monitoring and its relationship to medical knowledge. *Adv Health Sci Educ Theory Pract* 2012;17:311–23.
- Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. *Acad Med* 2005;80(10 Suppl):S46–54.
- Exadaktylos A, Hautz WE. Emergency medicine in Switzerland. *ICU Manag* 2015;15:160–2.
- Zwaan L, Thijs A, Wagner C, et al. Relating faults in diagnostic reasoning with diagnostic errors and patient harm. *Acad Med* 2012;87:149–56.
- Frederick S. Cognitive reflection and decision making. *J Econ Perspect* 2005;19:25–42.
- Calle P, Cerro L, Valencia J, et al. Usefulness of severity scores in patients with suspected infection in the emergency department: a systematic review. *J Emerg Med* 2012;42:379–91.
- Weiss SJ, Ernst AA, Nick TG. Comparison of the National Emergency Department Overcrowding Scale and the Emergency Department Work Index for quantifying emergency department crowding. *Acad Emerg Med* 2006;13:513–18.
- Weber N, Brewer N. Confidence–accuracy calibration in absolute and relative face recognition judgments. *J Exp Psychol* 2004;10:156–72.