

Article

# Unannounced versus announced hospital surveys: a nationwide cluster-randomized controlled trial

LARS HOLGER EHLERS<sup>1</sup>, KATHERINA BELTOFT SIMONSEN<sup>1</sup>,  
MORTEN BERG JENSEN<sup>1,2</sup>, GITTE SAND RASMUSSEN<sup>3</sup>,  
and ANNE VINGAARD OLESEN<sup>1</sup>

<sup>1</sup>Danish Center for Healthcare Improvements, Department of Business and Management, Fibigerstraede 11, DK-9220, Aalborg University, Aalborg, Denmark, <sup>2</sup>Center for Research in Econometric Analysis of Time Series, Department of Economics and Business, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark, and <sup>3</sup>The Danish Institute for Quality and Accreditation in Healthcare, Olof Palmes Allé 13, 1.th., DK-8200 Aarhus N, Denmark

Address reprint requests to: Lars Holger Ehlers, Danish Center for Healthcare Improvements, Department of Business and Management, Aalborg University, Fibigerstraede 11, DK-9220, Denmark; Tel: +45-60180848; Fax: +45-98153505; E-mail: lehlers@id.aau.dk

Editorial Decision 20 March 2017; Accepted 27 March 2017

## Abstract

**Objective:** To evaluate the effectiveness of unannounced versus announced surveys in detecting non-compliance with accreditation standards in public hospitals.

**Design:** A nationwide cluster-randomized controlled trial.

**Setting and participants:** All public hospitals in Denmark were invited. Twenty-three hospitals (77%) (3 university hospitals, 5 psychiatric hospitals and 15 general hospitals) agreed to participate.

**Intervention:** Twelve hospitals were randomized to receive unannounced surveys (intervention group) and eleven hospitals to receive announced surveys (control group). We hypothesized that the hospitals receiving the unannounced surveys would reveal a higher degree of non-compliance with accreditation standards than the hospitals receiving announced surveys. Nine surveyors trained and employed by the Danish Institute for Quality and Accreditation in Healthcare (IKAS) were randomized into teams and conducted all surveys.

**Main outcome measure:** The outcome was the surveyors' assessment of the hospitals' level of compliance with 113 performance indicators—an abbreviated set of the Danish Healthcare Quality Programme (DDKM) version 2, covering organizational standards, patient pathway standards and patient safety standards. Compliance with performance indicators was analyzed using binomial regression analysis with bootstrapped robust standard errors.

**Results:** In all, 16 202 measurements were acceptable for data analysis. The risk of observing non-compliance with performance indicators for the intervention group compared with the control group was statistically insignificant (risk difference (RD) = -0.6 percentage points [-2.51-1.31],  $P = 0.54$ ). A converged analysis of the six patient safety critical standards, requiring 100% compliance to gain accreditation status revealed no statistically significant difference (RD = -0.78 percentage points [-4.01-2.44],  $P = 0.99$ ).

**Conclusions:** Unannounced hospital surveys were not more effective than announced surveys in detecting quality problems in Danish hospitals.

**Trial Registration number:** ClinicalTrials.gov NCT02348567, <https://clinicaltrials.gov/ct2/show/NCT02348567?term=NCT02348567>.

**Key words:** certification/accreditation of hospitals, measurement of quality, experimental research

## Introduction

Accreditation programs for hospitals are implemented in >70 countries worldwide to support patient safety and high quality performance [1]. A common feature of all programs is an external peer review (or evaluation) process to accurately assess the level of hospital performance in relation to established standards. With a few exceptions, the standard way of planning and conducting this external evaluation is through announced surveys with dates usually set 6–12 months in advance to allow for preparation and improvement [1, 2].

Announced surveys have, however, been criticized for taking up too much time and resources in preparation for the scheduled visits, and that they may induce hospitals to ‘put on a show’ to receive certification, instead of focusing on continuous quality improvement [3–5]. Unannounced (or short notice) surveys have been proposed as an alternative method that differs in at least two important ways. First, the external surveyors on an unannounced visit will observe typical practice, rather than an arranged reality. Thus, unannounced surveys may be more effective in revealing the ‘true’ day-to-day performance level and in detecting non-compliance with standards that hospitals must improve. Second, the intention behind unannounced surveys is that hospitals remain constantly ready to meet accreditation standards, thus making the model of accreditation an operational management tool for working with standards and performance indicators on a continuous basis, rather than a checklist for the preparation of a scheduled visit.

The only empirical study of the difference in effect between announced and unannounced surveys is a pilot study in two Australian accreditation programs. This study indicated that more organizations would fail to reach the accreditation threshold than in previous survey if the surveys were conducted on short notice [5]. In general, the consequences of surveying methods have only been examined in a limited manner [6–9]. Studies of trends in hospital performance have for example failed to offer a clear picture of any effect of the Joint Commission’s move toward unannounced site visits in 2006 [10, 11].

In Denmark, all hospitals have been accredited by the Danish Institute for Quality and Accreditation in Healthcare (IKAS) since 2010 [12]. The Danish Healthcare Quality Program (DDKM) for hospitals is a flexible and mandatory system with a strong quality improvement philosophy and a down-toned control element with no formal financial or organizational consequences resulting from a survey, but the details of the accreditation status are published on the Internet. Denmark is probably a special case in hospital accreditation. With >96% of hospital activity under public governance and with free access to performance data, the need for external inspection and evaluation is less than in many other countries with a large private hospital sector. The DDKM was implemented as a national strategy for quality improvement emphasizing the same quality model for all healthcare organizations, i.e. public and private hospitals, general practitioners, nursing homes and pharmacies [12, 13]. Hospital surveys have been conducted every third year as announced surveys with midterm visits halfway through the period [12, 13]. The DDKM meets the requirements of the International

Society for Quality in Health Care’s (ISQua) international principles for healthcare standards; IKAS is an ISQua-accredited organization [2, 14–16] (a description of DDKM in English is available at <http://www.ikas.dk/IKAS/English>).

Hospital accreditations have been criticized by Danish physicians, who have argued, among other things, that preparation for an external visit takes time away from patients and that the DDKM leads to excessive documentation work [17, 18]. As a response to these criticisms, the IKAS board of directors decided in 2014 to investigate whether or not unannounced surveys should be implemented in the coming version 3 of DDKM, planned for implementation in 2016.

A nationwide cluster-randomized controlled trial (C-RCT) of unannounced versus announced surveys in Danish public hospitals was therefore planned in order to evaluate the effectiveness of unannounced surveys versus announced surveys in detecting non-compliance with accreditation standards in public hospitals. The study was conducted from August 2014 to March 2015 based on the performance indicators listed in an abbreviated set of the national accreditation standards from the DDKM version 2, covering organizational standards, continuity of care standards and patient safety standards. We hypothesized that the hospitals receiving the unannounced surveys would reveal a higher degree of non-compliance with accreditation standards compared with the hospitals receiving announced surveys.

## Methods

### Study design and setting

This study follows the extension to the CONSORT 2010 statement on cluster-randomized trials [19]. The Central Denmark Region Committees on Health Research Ethics decided that this trial does not constitute clinical research and thus does not require notification to the committee. For a more detailed description of the trial design, this paper refers to the trial protocol [12].

All Danish public hospitals ( $n = 30$ ) were invited to participate. Ten hospitals due to receive the periodic IKAS midterm survey were given the opportunity to choose either participation in the trial or receiving the scheduled survey. Private hospitals constitute <4% of the hospital sector and were excluded from the trial. Unannounced hospital survey had not been carried out in Denmark prior to this study.

The intervention group was defined as receiving unannounced surveys and the control group was defined as receiving announced surveys. Nine experienced and trained surveyors were selected to conduct all external surveys in the trial. Each survey was conducted by a team of two surveyors, one nurse and one doctor. A special tracer tool was developed and a two-day seminar with all surveyors and the researchers was conducted prior to the trial to ensure consistent and objective tracer activity at all locations [12, 20].

Each survey lasted 3 days, with 2 days of tracer activity and a subsequent meeting with the management of the hospital on Day 3 to report the findings. All participating hospitals were given a date

for the management meeting. From this information, all hospitals were led to assume that they had been randomly selected for an announced visit and that the external survey would be conducted 2 days prior to the meeting. For the control group, the 2-day tracer activity was conducted 2 days prior to the meeting. For the intervention group, the 2-day tracer activity was executed ~3 months earlier.

### Randomization

A stratified randomization of hospitals was performed with respect to type of hospital. Block randomization was used to ensure the same number of unannounced and announced hospital surveys within each small stratum. Prior to the randomization each region was required to experience at least one unannounced and one announced hospital survey. The dates for the survey, as well as the surveyor team, were also randomly assigned. The randomization was blinded to IKAS and the surveyors until a few days before trial start. All randomization was performed in Microsoft Excel 2007 using the RAND() function.

### Outcome

The outcome measure was the level of compliance with 113 performance indicators from the abbreviated set of the national accreditation standards from DDKM version 2. The level of compliance was assessed through the tracer activities performed by the external surveyors and measured on a 4-point scale as 'consistent implementation', 'consistent implementation with a few deviations', 'weak implementation' or 'failure of implementation'. For a minority of the performance indicators, the level of compliance was assessed dichotomously as either 'consistent implementation' or 'failure of implementation'. All 113 performance indicators were employed for each hospital department chosen for inspection by the surveyors.

### Statistical analysis

Based on previous surveys, we expected ~700 measurements per hospital, with 5% showing inconsistent implementation of the performance indicators for the control group and 10% for the intervention group. Assuming an intra-correlation coefficient of 0.01 and a significance level of 0.05, the study had a power of 98% to detect a minimum of 5% difference in effect. A pooled bivariate analysis of all indicators was performed to evaluate the effect difference between the trial groups. All 113 performance indicators were included as fixed effects in this analysis. In addition, bivariate analysis was performed for eight selected patient safety standards, each of which has the potential to change an organization's accreditation status if these standards are not evaluated as 'consistently implemented'. Both the bivariate and the pooled analyses were performed using binomial regression analysis. Bootstrapping was conducted to achieve a robust standard error through resampling with replacement of the 23 clusters, repeated 100 times. This resampling procedure was made to ensure no exaggeration of variation within a hospital due to clustering. The dichotomous outcome of the binomial regression analyses was defined as 'one' if assessment of a performance indicator was coded 'consistent implementation'; otherwise, the outcome was coded as 'zero'. The decision to dichotomize the 4-point scale was necessitated by methodological constraints (low number of observations in the nonconsistent category cells). Risk difference reported as percentage points was chosen as association measure in the binomial regression analyses with a two-sided significance level of 0.05. All analyses were performed in STATA 13 (StataCorp LP, 2013, College Station, TX, USA).

### Results

The study recruited 23 (77%) Danish public hospitals for participation, including general hospitals ( $n = 15$ ), university hospitals ( $n = 3$ ) and psychiatric hospitals ( $n = 5$ ). Seven hospitals declined participation (see Fig. 1).

This C-RCT provided 16 202 performance measurements for data analysis, with a mean of 704 measurements per hospital. The intervention group obtained 8519 performance ratings, whereas the control group obtained 7683 performance ratings (see Table 1). An intra-cluster correlation coefficient was calculated as 0.11.

The intervention group produced 7728 observations of 'consistent implementation' on the 4-point scale and the control group produced 6997. The variation is due to the unequal number of clusters in the two trial groups. Compared with the total number of observations within each group, the number of consistently implemented standards was larger for the control group (96.36%) than for the intervention group (95.93%).

As can be seen in Table 2, the intervention group revealed fewer consistently implemented standards (97.41%) than did the control group (98.34%) with regard to the dichotomously measured performance indicators.

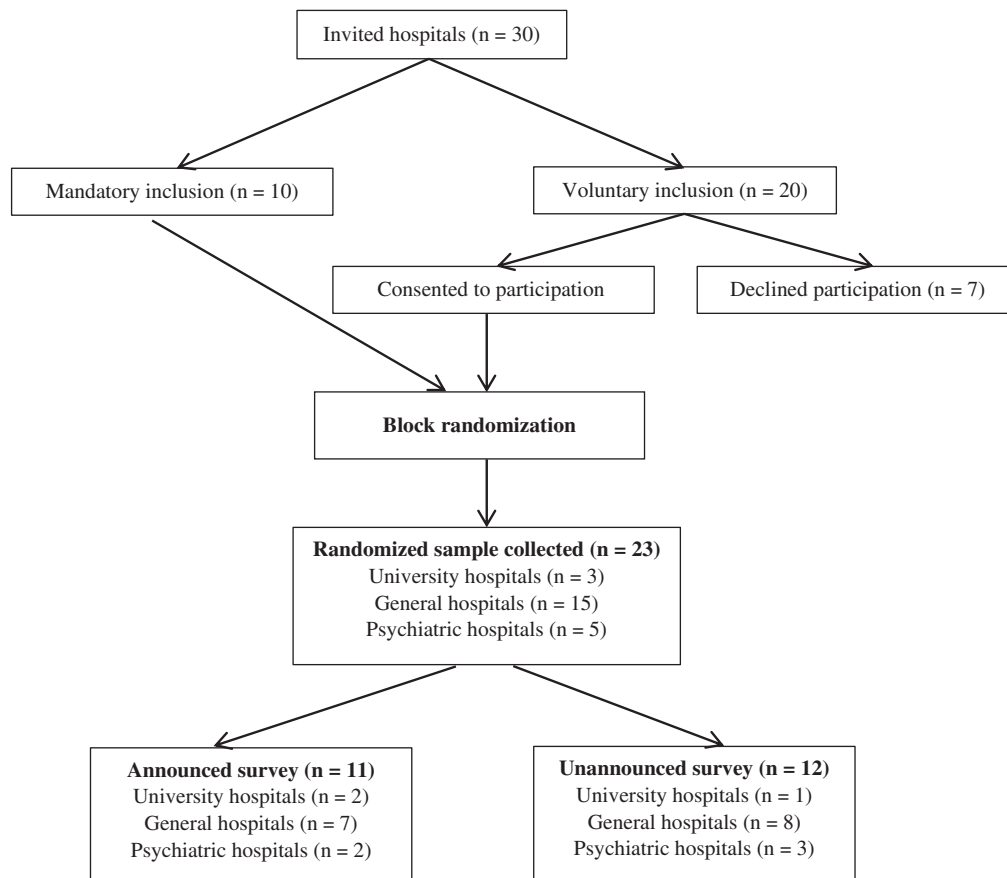
A pooled analysis of the effect of the unannounced surveys showed an absolute difference of -0.6 percentage points in consistent implementation between the intervention group and the control group (see Table 3). This result indicates that the intervention group had 0.6 percentage points fewer consistently implemented standards than did the control group, but the result was not statistically significant ( $P = 0.54$ ). The robust standard error was calculated based on the 78 out of 100 binomial regression analyses that were able to converge.

Eight patient safety critical standards from the DDKM version 2 have the potential to change an accreditation status from 'accredited' to 'accredited with remarks', 'conditional accreditation' or 'not accredited' if assessed as other than 'consistent implementation'. Table 4 demonstrates six patient safety critical standards. Two standards, 'Treatment of cardiac arrest' and 'Observation and follow-up on critical observation results', are removed from the table. These standards were assessed as 100% consistently implemented in all included hospitals and were thereby statistically unable to converge in the binomial regression analysis. Sub-group analyses of the remaining six patient safety critical standards showed no statistically significant difference between the intervention and control groups (see Table 4). The robust standard error was calculated based on the bootstrapped samples for which the binomial regression analysis converged.

### Discussion

This study is the first nationwide and cluster-randomized controlled trial of unannounced versus announced hospital surveys powered to detect a significant difference in effect. Using binomial regression analysis in the pooled analysis of all performance indicators, we found no evidence of an increased effectiveness of unannounced surveys in finding a higher level of non-compliance with the accreditation standard. Secondary bivariate analysis of the findings on each of the eight patient safety standards also failed to find evidence of a difference in effectiveness.

Our study has a high level of internal validity. A particular strength is the randomization of hospitals, survey dates and survey teams, as well as the allocation concealment until a few days before the survey. Cluster randomization was an appropriate design, as findings on wards from the same hospital are correlated.



**Figure 1** Participant flow diagram representing the number of clusters that were randomly assigned to either the control group (announced surveys) or intervention group (unannounced surveys).

**Table 1** Observations from intervention and control groups assessed as a 4-scale measure

	Failure of implementation	Weak implementation	Consistent implementation with a few deviations	Consistent implementation	Total
Intervention group, N (%)	103 (1.28)	120 (1.49)	105 (1.30)	7728 (95.93)	8056
Control group, N (%)	65 (0.90)	93 (1.28)	106 (1.46)	6997 (96.36)	7261
Total	168	213	211	14 725	15 317

**Table 2** Observations from intervention and control groups assessed as a dichotomous measure

	Failure of implementation	Consistent implementation	Total
Intervention group, N (%)	12 (2.59)	451 (97.41)	463
Control group, N (%)	7 (1.66)	415 (98.34)	422
Total	19	866	885

**Limitations**

Several factors could have influenced the results. First, a potential risk of selection bias was present, as the hospitals were included in the study based on voluntary participation. One hospital from the Zealand Region and six hospitals from the Capital Region of Denmark declined participation. However, we find no reason to expect different quality levels in these hospitals, and their decline is more likely to reflect the

fact that their next organization-wide accreditation survey was scheduled immediately after the trial period.

Second, it was agreed that the trial results would not be made available to the public or have any influence on accreditation status and certification. Hospital managers in the control group thus had no formal incentive to prepare for the external evaluation. Considering the current criticism of the DDKM as a large administrative burden, the hospital managers may have chosen to ‘do nothing’ and act as if the hospital was assigned to the group of unannounced surveys, even though this was not the case. However, the political interest in results from accreditation surveys has been very high since the introduction of the DDKM in 2010; the implementation of such passive strategies would probably be feasible only through agreement between all hospitals in a region. Rumors aside, we have not been able to verify the existence of such a strategy in any of the five Danish regions.

Third, the unequal assessment of performance indicators by the nine surveyors could involve a risk of bias, with the argument that their

**Table 3** Binomial regression analyses of the overall effect difference between intervention and control based on all consistently implemented standards

Intervention group (clusters = 12; N = 8519)		Control group (clusters = 11, N = 7683)		Absolute difference (95% CI)	Robust standard error	P-value
Consistent implementation	Nonconsistent implementation	Consistent implementation	Nonconsistent implementation			
No. (%)				Percentage points	Bootstrapped	
8179 (96.0)	340 (4.0)	7412 (96.5)	271 (3.5)	-0.60 (-2.51-1.31)	0.98	0.54

CI, confidence interval. Results are reported with risk difference as association measure and with bootstrapped robust standard errors.

**Table 4** Binomial regression analyses based on consistent implementation of six patient safety critical standards from the DDKM version 2

Patient safety critical standards	Intervention group (12 clusters)		Control group (11 clusters)		Absolute difference (95% CI)	Robust standard error	P-value
	Consistent implementation	Nonconsistent implementation	Consistent implementation	Nonconsistent implementation			
	No. (%)				Percentage points	Bootstrapped	
Patient identification	225 (98.25)	4 (1.75)	198 (97.50)	5 (2.46)	0.41 (-2.84-3.66)	1.66	0.81
Timely reaction to diagnostic tests	438 (95.42)	21 (4.58)	396 (97.30)	11 (2.70)	0.95 (-5.45-7.35)	3.27	0.77
Medical prescription	268 (98.53)	4 (1.47)	241 (98.80)	3 (1.23)	-0.93 (-2.58-0.73)	0.84	0.27
Medical exemption	256 (99.22)	2 (0.78)	235 (97.90)	5 (2.08)	3.90 (-4.77-12.56)	4.42	0.38
Medical administration	352 (97.24)	10 (2.76)	315 (98.40)	5 (1.56)	-1.40 (-5.52-2.72)	2.10	0.51
Safe surgery	67 (93.06)	5 (6.94)	65 (94.20)	4 (5.80)	-0.48 (-14.22-13.26)	7.01	0.95
Overall/converged analysis	1993 (97.41)	53 (2.59)	1809 (98.10)	35 (1.90)	-0.78 (-4.01-2.44)	1.64	0.99

CI, confidence interval. Results are reported with risk difference as association measure and with bootstrapped robust standard errors.

assessment is inevitably rather subjective, even though a tracer tool developed specifically for this study was applied to standardize the surveys. Before the study start, the surveyors had different expectations for the trial results, with some expecting to see a difference between the control and intervention groups, and others believing the contrary. However, the surveyors were randomized into different teams to reduce the risk of bias from any prejudiced approach to conducting the survey.

Finally, the time constraint for the external survey could have resulted in different opportunities to collect measurements of performance ratings in the two groups if unannounced visits were associated with more practical problems and barriers. However, the special tracer tool and the training program for the surveyors were used to ensure that the tracer activity was performed in exactly the same way in each hospital. The surveyors did not report any problems with regard to a lack of time or availability of relevant data or personnel during the survey.

These limitations nevertheless clarify the complexity and challenges of applying traditional clinical research methods in evaluating complex interventions. It has been discussed elsewhere whether or not complex interventions are suitable for RCT designs [7, 21, 22].

### Generalizability

Denmark has a long history of governmental initiatives to ensure quality and safety in healthcare, and a unique unified accreditation system, with impressive quality monitoring [23, 24]. Denmark is

ahead of most OECD countries in efforts to monitor and benchmark healthcare quality and it is unsurprising that hospitals in both groups performed well [23]. This study does not clarify what contextual factors are essential for the generalization of the findings. Some aspects that seem important are the relatively immature national system of accreditation, the relatively low accreditation threshold, and the voluntary participation in a test setting where the results from the external surveys do not impact accreditation status or certification.

### Lessons learnt

This trial provided no indication that unannounced surveys were more inconvenient than announced surveys. The surveyors reported positive feedback from hospital managers and staff in the intervention group, indicating a positive attitude among hospital employees toward the implementation of unannounced surveys.

### Acknowledgements

The authors would like to thank the hospitals for their participation and the surveyors for their engagement and great work in preparing and carrying out the surveys.

### Funding

This work was funded by the Danish Institute for Quality and Accreditation in Healthcare (IKAS). IKAS had no role in the randomization, data analyses

or decision to publish. None of the authors have financial interest in the results presented.

## References

- Greenfield D, Pawsey M, Braithwaite J. Accreditation: a global mechanism to promote quality and safety. In: Sollecito WA, Johnson JK (eds). *McLaughlin and Kaluzny's Continuous Quality Improvement in Health Care*. Jones & Bartlett Learning; Burlington, 2013.
- Fortune T, O'Connor E, Donaldson B. *Guidance on Designing Healthcare External Evaluation Programmes Including Accreditation*. Dublin: The International Society for Quality in Health Care, 2015.
- Greenfield D, Travaglia J, Braithwaite J, *et al*. Unannounced Surveys and Tracer Methodology: Literature Review [Internet]. The University of New South Wales, Centre for Clinical Governance Research. A Report for the Australian Accreditation Research Network: Examining Future Health Care Accreditation Research.; 2007 pp. 1–30. <http://www.webcitation.org/6WSk7BsWk>, ISBN: 9780733425462
- The Australian Council on Healthcare Standards. Piloting Innovative Accreditation Methodologies: Short Notice Surveys. Commonwealth of Australia: The Department of Health and Ageing on behalf of the Australian Commission on Safety and Quality in Health Care. 2009
- Greenfield D, Moldovan M, Westbrook M *et al*. An empirical test of short notice surveys in two accreditation programmes. *Int J Qual Health Care* 2012;**24**:65–71.
- Hinchcliff R, Greenfield D, Moldovan M *et al*. Narrative synthesis of health service accreditation literature. *BMJ Qual Saf* 2012;**21**:979–91.
- Brubakk K, Vist GE, Bukholm G *et al*. A systematic review of hospital accreditation: the challenges of measuring complex intervention effects. *BMC Health Serv Res* 2015;**15**:280.
- Greenfield D, Braithwaite J. Health sector accreditation: a systematic review. *Int J Qual Health Care* 2008;**20**:172–83.
- Grepperud S. Is the hospital decision to seek accreditation an effective one? *Int J Health Plann Manage* 2015;**30**:E56–68.
- Towers TJ, Clarck J. Pressure and performance: buffering capacity and the cyclical impact of accreditation inspections on risk-adjusted mortality. *J Healthc Manag* 2014;**59**:323–35.
- Schmaltz SP, Williams SC, Chassin MR *et al*. Hospital performance on national quality measures and the association with Joint Commission Accreditation. *J Hosp Med* 2011;**6**:454–61.
- Simonsen K, Olesen V, Jensen M *et al*. Unannounced or announced periodic hospital surveys: a study protocol for a nationwide cluster-randomised, controlled trial. *Comp Eff Res* 2015;**5**:1–7.
- Institut for Kvalitet og Akkreditering i Sundhedsvæsenet. 2014, Denmark. <http://www.ikas.dk/> (30 October 2015, date last accessed).
- International Society for Quality in Health Care. <http://isqua.org/accreditation/accreditation> (30 October 2015, date last accessed).
- Shaw CD, Kutryba B, Braithwaite J *et al*. Sustainable healthcare accreditation: messages from Europe in 2009. *Int J Qual Health Care* 2010;**22**:341–50.
- Shaw CD, Braithwaite J, Moldovan M *et al*. Profiling health-care accreditation organizations: an international survey. *Int J Qual Health Care* 2013;**25**:222–31.
- Triantafillou P. Against all Odds? Understanding the emergence of accreditation of the Danish hospital. *Soc Sci Med* 2014;**101**:78–85.
- Ehlers LH, Jensen MB, Simonsen KB *et al*. Attitudes towards accreditation among hospital employees in Denmark: a cross sectional survey. *Int J Qual Health Care*. (under review).
- Campbell M, Piaggio G, Elburne D, Altman DG. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012;**345**:e5661.
- Joint Commission Resources. *Applied Tracer Methodology: Tips and Strategies for Continuous System Improvement*. Oakbrook Terrace, Illinois, USA: The Joint Commission on Accreditation of Healthcare Organizations, 2007.
- Craig P, Dieppe P, Macintyre S *et al*. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;**337**:1655–63.
- Shojania KG. Conventional evaluations of improvement interventions: more trials or just more tribulations? *BMJ Qual Saf* 2013;**22**:881–4. doi:10.1136/bmjqs-2013-002377.
- OECD. *OECD Reviews of Health Care Quality: Denmark*. Paris, France: Health Division, 2013.
- Mainz J, Kristensen S, Bartels P. Quality improvement and accountability in the Danish health care system. *Int J Qual Health Care* 2015;**27**:1–5. doi:10.1093/intqhc/mzv080.