Original Article

Using trends in electronic recordings of vital signs to identify patients stable for transfer from acute hospitals

CP Subbe, J Weichert & B Duller

Abstract

Patients who are stable might not be required to remain in hospital. We aimed to create objective criteria to indicate stability based on vital signs. An index based on NEWS (NBI) was compared to a Patient Stability Index (PSI) algorithm created by random forest analysis.

Data from the VITAL II study was used to train the algorithm and data from the VITAL III study to validate it. Failure rate of the algorithms was set close to the rate of readmission to UK hospitals at 15%. After a training period of two days the NBI identified stability with acceptable failure rates only after a further 96 hours with a subsequent release of 2143 bed days compared to the PSI which identified stability after only 12 hours leading to potential earlier release of 2652 bed days.

Vital sign-based algorithms might be able to predict safe transfer from hospital and inform management of flow.

Keywords

vital signs, transfer of care, hospital, Early Warning Score, risk management, internal medicine

Key points

- Vital signs can be used to predict instability and stability.
- Machine learning trained algorithm might be able to indicate stability.
- Objective criteria for transfer need to be tested in prospective studies.

Christian P Subbe
Consultant in Acute, Respiratory & Critical Care Medicine, Bangor University & Ysbyty Gwynedd, Bangor, LL57 2PW, Wales, UK

Jochen Weichert
External Innovation Manager Philips, Patient Monitoring & Analytics, Clinical Marketing 3000 Minuteman Road, M/S 4307, Andover, MA 01810

Bernd Duller
Clinical Studies & Research Consultant
Perpet Production, Kolumbusstr. 29, 70439 Stuttgart, Germany

Correspondence
Christian P Subbe
Consultant in Acute, Respiratory & Critical Care Medicine, Bangor University & Ysbyty Gwynedd, Bangor, LL57 2PW, Wales, UK
Email: csubbe@hotmail.com

Background

Discharge decisions are complex[1] but involve the diagnosis of 'stability' prior to transfer from an acute hospital to the community, be it the patients' own home or a rehabilitation facility.

At a time when Western hospitals are bursting at the seams qualified staff are often struggling to deliver current care and plan what lies ahead. Decision making around transfers from acute care rely largely on subjective assessments by senior decision makers,[2] but these might only review patients in detail two or three times during the week. In the meantime, patients are being cared for by more junior staff with both less experience and less confidence in releasing patients from the perceived 'safe haven' of hospital wards and their assessments might clash with perceptions of patients and their carers.[3] At a time when there is increasing knowledge about the detrimental effects of hospitalisation in relation to infection risk,[4] deconditioning[5] and mental health problems as part of a 'post-hospital syndrome'[6] attention needs to focus on objective, effective and efficient ways to predict when a patient is likely to be safely transferred from hospital.

We aimed to test whether the following hypothesis is true: patients with low or unchanging NEWS scores are unlikely to deteriorate in future and could therefore be considered for transfer.

Methods

Setting

Two wards of a university-affiliated district general hospital serving a population of 220,000 inhabitants in the United Kingdom. The hospital covers all hospital specialties including a haematology and oncology unit and a dialysis unit but does not provide neuro-surgical, cardio-thoracic and transplant surgery care. The intensive care unit (ICU) has 8 beds and there are a further 5 high dependency and 5 coronary care beds.

The studies were undertaken on two general medical wards with 30 and 24 beds respectively. Both wards also look after patients with general medical conditions and consultant staff are qualified in General Internal Medicine. One ward provided additional expertise for patients with respiratory disease and the other for patients with gastroenterological conditions. The hospital operates an Acute Medical Unit where low-risk patients can be observed for up to 72 hours. Only patients thought to require longer recovery periods are transferred to general wards. The two study wards were selected for the studies due to their high proportion of patients with abnormal and complex physiology as evidenced by high rates of critical care transfers and cardio-pulmonary arrests prior to the commencement of the studies.

Study patients

We included patients from two separate studies using electronic vital sign recordings: Methods of the VITAL II study have been described elsewhere.[7] VITAL II is an interventional study pre- and post-installation of an advanced monitoring system (Intellivue Guardian Solution) for vital signs in two cohorts in 2012/13 and 2014/15. The intervention phase used electronic vital sign recordings: Spot-check monitors requested the nurse to manually enter respiratory rate (RR) or to confirm a RR measurement from a wireless patient sensor. Information on mental status was entered manually. Early Warning Scores (EWS) were automatically calculated. Algorithms were trained with patients from the intervention phase (ClinicalTrials.gov, NCT01692847, ethics reference 12/WA/0050, Protocol number SD-05163-BBN-IGS A.2). VITAL III is an observational study of monitoring modalities for respiratory rate.[8] The study compared different modalities of monitoring respiratory rate and ran from October 2017 to February 2018 (ClinicalTrial.gov, NCT02973243, ethics reference: 16/WA/0288). The generation of algorithms to predict stability for discharge was one of the planned outcomes of VITAL III. Algorithms were validated with this dataset.

Inclusion and exclusion criteria

Inclusion and exclusion criteria for the respective VITAL II and VITAL III studies have been published elsewhere.[7,8] The datasets from the VITAL II and VITAL III studies were fully anonymised by removing all identifiers such as names, registration numbers and date of birth. For our analyses we included only patients with a minimum length of 48 hours for which at least two complete NEWS scores could be defined. From the VITAL II intervention phase 1485 patient episodes had a minimum length of 48 hours. From the VITAL III study 791 patient episodes had a minimum length of 48 hours.

Parameters included

For all patients, age, gender, main diagnosis (ICD 10 code), data of admission, date of discharge, the degree of physiological abnormality as described by the National Early Warning Score (NEWS)[9] on admission and survival status were collected.

Operational definitions

In line with the Welsh protocol for escalation of care instability was defined as one or more NEWS of 6 or more. The Welsh protocol stipulates an urgent review by a medical practitioner or a Rapid Response Team with critical care skills in these patients. There is no universally accepted definition for 'clinical stability': As primary study outcome we defined 'stability to transfer to a lower acuity area of care' as a NEWS score of equal or less than 3 for a minimum of 48 hours period, allowing for a single score of 4 or 5 and no score of 6 or more till discharge.

We defined clinical acceptability of 'safe for transfer' criteria at a failure rate of 15% with 85% of patients having no further high NEWS after the criteria are met. While this is a deliberate cut-of it is grossly in line with readmission rates for medical emergency admissions of 12-13% (10). A positive predictive value of 85% was hence considered acceptable for safe transfers.

We created two predictive algorithms: The NEWS Based Index (NBI) uses raw NEWS values to define 'stable' (see above). The Patient Stability Index (PSI) analyses trends to predict stability. The PSI takes a value of zero to 10 with zero indicating instability and 10 indicating complete stability.

Machine learning technique used to generate the PSI

As a method of classification we chose random forest analysis[11] because of the ability to interpret its results, the easiness to incorporate categorical and numerical results as well as its ability to learn complex relationships. In random forest analysis the learned representation takes the form of a function combining the sub-score-results of a collection of classification and regression trees (CART). Classification and regression trees have been used in artificial intelligence and machine learning long before random forest analysis was invented.[12] Starting from a root node each node poses an exclusive condition on a feature in the input vector for each branch to a child node. Leaf nodes, those without any children, classify the input in CARTs and provide a sub-score result in random forests. Training a single decision tree for classification may result in very complex tree structures that tend to over fit the data and generalize badly.[13] To remove the overfitting random forests combine multiple - commonly hundreds to thousands - of simpler CARTs into a forest. Each tree gives a classification, the tree "votes" for that class. The random forest prediction is simply the majority votes of the trees (over all the trees in the forest). If the random forest model executes on a window of data it outputs a classification as well as the probability of the classification being true. These probabilities were filtered using an unevenly spaced moving average[14] to reduce variability in the output. Scaling to a range of zero (unstable) to ten (stable) results in the Patient Stability Index.

Since we were able to apply the operational criteria of stability to each sliding window of 48 hour length the random forest classifier learned from 31091 training samples generated by 1485 patients.

Statistical methods and definitions

Data from patients who survived to discharge was analysed by calculating the PSI and NBI for all episodes at each new NEWS. Applying the learned model for the PSI and executing the NBI resulted in numerical values between 0 (unstable) and 10 (stable) for each NEWS score starting after the 48-hour ramp up phase, during which the algorithms gather data needed to analyse the trends.

Once the specific PSI or NBI was in the predefined range for the given time duration (i.e. 12, 24, 36, 48, 60, 72 and 96 hours) the timestamp of this NEWS became the reference point of the following analysis. Starting from this point of reference we scanned the remaining times series of NEWS for further deteriorations (NEWS>5). If no deterioration was detected the episode was marked as true positive else as false positive. For the case without a reference point, when the PSI or NBI never detected stability, the time series was scanned for deteriorations. A time series without deteriorations was accounted for as false negative while episodes with deteriorations counted as true negatives.

Assessing and summarizing the episodes into true positives, false positives, true negatives and false negatives allowed us to compute all established performance measures: sensitivity, specificity, positive predictive value (PPV) as defined in.[22] In addition, we defined the predicted prevalence of stable patients as all true and false positives divided by the entire population. We used the predicted prevalence of stable patient as an approximation of the expected discharge rate, a measure we used to maximize while maintaining the same level of accuracy for discharging clinically stable patients.

Besides these static measures of performance we calculated summary statistics for each algorithm for discrete times (day 3 – day 15) after admission to the two study wards, where we treated the first NEWS as the time of admission. For each day we provide the number of patients classified as stable during that day for a specific duration, the number of patients for which the data ends without an algorithm reaching stability, the number of patients remaining in the units.

Most of the analysis and machine learning for this paper was done in Python 3.6.[15] Converting the raw data from Vitals II and III into features for learning was done using numpy[16] and the pandas package[17] while for learning we utilized the random forest classifier implementation of scikit-learn.[18] Evaluation of the learned classifier was prepared using the pandas package and visualizations rendered using ggplot2[19] with ggthemes[20] using the R language.[21]

Funding

The VITAL II and VITAL III studies were funded by Philips Healthcare. CPS received funding as a Bevan Technology Exemplar to explore the evaluation of the vital signs for safe transfer of patients from the Bevan Commission.

Ethics

The hospital human research ethics committee (Reference 12/WA/0050, Protocol number SD-05163-BBN-IGS A.2) approved the study. Monitoring remained in line with hospital policy in control and intervention phase. The ethics committee confirmed that no patient consent was required.

Results

Patient characteristics

After application of inclusion and exclusion criteria data from two cohorts of patients were included for analysis: The VITAL II intervention phase cohort comprised 1485 patients for analysis (Training Cohort-TC), the cohort from the VITAL III study comprised 791 patients for analysis (Validation Cohort – VC).

The proportion of female patients was 53% in the training and 58% in the validation cohort (Chi-Square 5.18, $p = 0.02$). Median age of patients was 74 (IQR 62 to 80) years in the training and 73 (IQR 60.5 to 80) years in the validation cohort ($z = -2.498$, $p = 0.013$). Median length of stay was unchanged in the validation cohort compared to the training cohort (6 days and 15 hours). Median NEWS on admission to the study wards was 2 (IQR 1 to 5) in the training cohort and 2 (IQR 1 to 4) in the validation cohort. 706 patients (47.5%) had a NEWS of 6 or more in the training cohort and 384 (48.5%) in the validation cohort. For the 791 patients included in Vital III who stayed 48 hours or more 722 patients went home, to other low acuity wards or care home, 62 patients died, 3 patients were associated with an ICU transfer, 4 patients with undefined outcomes. 680 patients stayed 72 hours or more and 578 patients more than 96 hours. The following results are for the validation cohort unless otherwise stated.

'Safe for transfer' rates of NBI and PSI algorithms

Comparing the failure rates of NBI and PSI a PPV of 85% or more was defined as 'safe for transfer'. NBI didn't reach the PPV of 85% but was best if a NEWS of 0 to 3 was present for 96 hours after the 48 hour 'learning phase'. At this stage 40.1% of the patients would be indicated as 'safe for transfer'. For the PSI a PPV of over 85% was seen after 12 hours with a 'safe for transfer' rate of 50.9% (Figure 1).

Breeches in stability after 'stability' criteria are met

After 12 hours of NBI stability criteria deteriorations occurred a mean of 3.5 days (5 days SD) later. After 48 h of NBI stability it took a mean of 5.75 days (6.5 days SD) and after 96 hours of NBI stability it took a mean of 7 days (7 days SD) for the deterioration to occur.

In patients with PSI stability criteria fulfilled for 12h deteriorated occurred a mean of 6,75 days (7.25 days SD) later. After 48 h of PSI stability it took a mean of 8.5 days (8.5 days SD) and after 96 hours of PSI stability an average 9.25 days (8.25 days SD) for the deterioration to occur.

Implication for patient flow

Transfer of care of patients fulfilling NEWS criteria of 'stability' with a PPV of 85% occurred after 96 hours and would have resulted in release of 2143 bed days compared to actual bed-day usage.

Transfer of care of patients fulfilling PSI criteria of 'stability' with a PPV of 85% occurred after 12 hours would have resulted in release of 2652 bed days, and a further 1041 bed days for 96 hours compared to actual bed-day usage. A simulation with the validation cohort of VIII demonstrated gains from algorithm led transfer compared to actually utilised bed days with the PSI indicating 'stability' earlier during the course of admission. Figure 2 shows how many patients could have been transferred under these conditions compared to the patients actually transferred from the wards on the first 15 days.

Transfer of patients based on Positive Predictive Value of the National Early Warning Score Based Index (NBI) and Patient Stability Index (PSI) compared to actual transfers. Because of the training time of 48 hours no algorithm advised discharges are expected at day 0 and 1.

Discussion

To our knowledge this is the first attempt to explore the potential for mathematical algorithms to predict stability as a pre-requisite for transfer from an acute hospital.

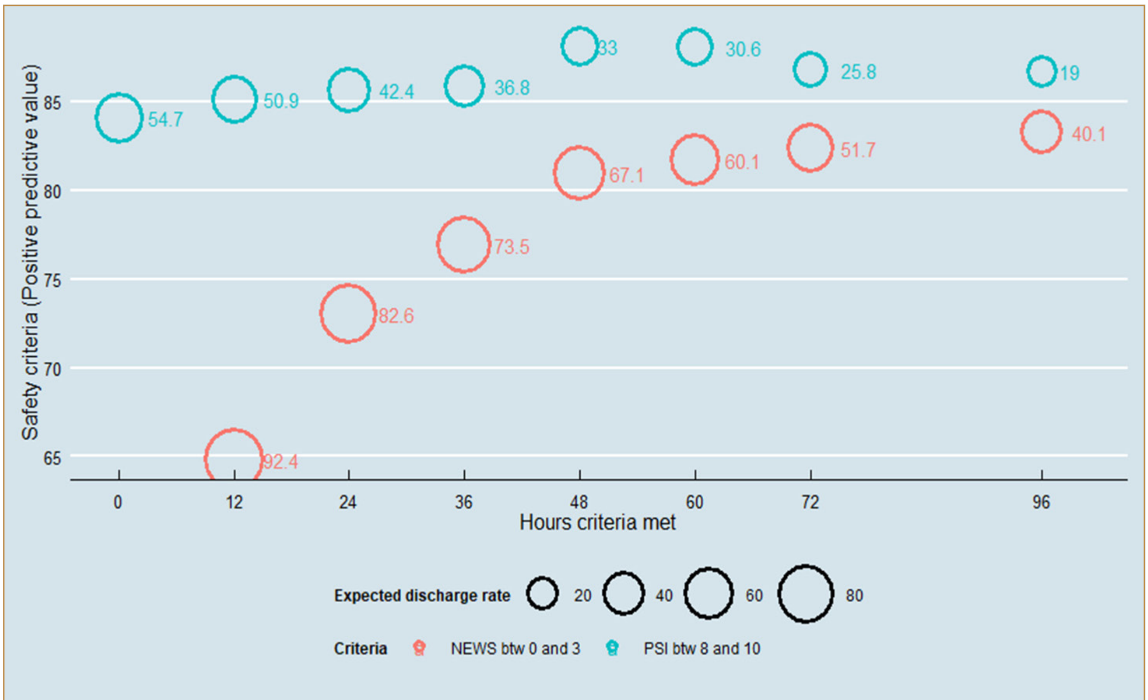

Figure 1. Positive Predictive Value depending on observation period for different algorithms National Early Warning Score Based Index and Patient Stability Index fulfilled for different durations of time (X-axis) against the Positive Predictive Value (Y-axis). The number of patients fulfilling criteria is indicated by the size of each circle. A Positive Predictive Value of 85% was set as likely clinically acceptable 'safety' threshold.

Day on ward:	2	3	4	5	6	7	8	9	10	11	12	13
Hospitalized	680	578	504	428	377	323	278	251	226	201	176	154
Patients leaving	111	102	74	76	51	54	45	44	27	25	25	22
Hospitalized	680	578	321	262	204	155	126	102	82	72	67	60
Patients leaving	111	102	257	59	58	49	29	24	20	10	5	7
Stable patients	0	0	220	22	25	14	10	8	6	3	0	2
Hospitalized	515	380	294	232	177	138	110	91	76	67	62	55
Patients leaving	276	135	86	64	55	39	28	19	15	9	5	7
Stable patients	192	67	47	33	25	6	13	4	5	2	0	2

Figure 2. Simulation of safe transfers

We used a planned secondary analysis of data from two studies from the same institution in the UK. Algorithms based on infrequent sets of vital signs and summary NEWS scores might miss future deteriorations in more than 20% of patients even in patients who have been stable for 48 hours or more. The Patient Stability Index as an algorithm trained on vital signs using random forests allows to specify failure rates in the set-up of the algorithm and identifies stability after a significantly shorter observation period. Algorithms identified a significant number of future days of stability and might have supported earlier transfer of patients to their own home or a rehabilitation facility compared to actual transfer dates.

Discharge planning is a contentious area and criteria led discharges have been successfully tested in defined pathways such as elective surgery, epilepsy and other conditions but often with little effect on clinical or financial outcomes beyond patient satisfaction.[22] Criteria led discharge planning allows nurses or more junior medical team members to discharge patients who would otherwise wait for a senior decision maker. Three different meta-analysis have divergent results: one suggested that nurse led discharge planning in elderly patients with acute illness might reduce readmissions while not affecting length of stay of the index admission,[23] or increase length of stay of index admission and not affect other health related outcome measures[24] but improve readmission rates in those with rehabilitation needs.[25] Nearly always this model of care requires a confirmed diagnosis with a well-known risk profile and trajectory: i.e. in patients who have had no temperature for 24 hours further observation is unlikely to show new abnormalities. Our study differs from this in that we are using pure physiological data. Given the location of the study we know that many of the patients in the study cohort had respiratory and gastroenterological conditions buy many suffered from several and other medical conditions.

Predicted dates of discharge are a tool to drive discharge decision making[26] with the evidence for effectiveness of the approach being predominantly in studies from the United States and in surgical admissions with single diagnosis. Predictions are usually vetted against scheduled hall marks of recovery or based purely on 'expert opinion' of clinicians.

We have thus so far only undertaken a retrospective review of data. Reasons for admission to hospital are complex. Social circumstances[27] and personal preferences of patients, carers and clinicians[3] all modify decision making. These could not be modelled from our data set. Their influence requires testing in a prospective study.

For the purpose of this study we have taken a conservative approach to our assumptions: stability of 48 hours allows to observe the development of potential complications. One of the practical reasons for this is that trend analysis requires a minimum number of data-sets. UK standard for vital sign observations on acute wards varies between two and six-times per day.[9]

While we have been able to create a model that indicates 'stability' we believe that this will not be a replacement for shared decision making between patients, carers and clinicians but might provide assurances when in some cases and could alert clinicians to successful effects of medical treatments allowing patients to return to their own home. Prospective testing of the algorithms in both our and other institutions is needed to test face validity and potential for impact. Framing of the information as part of device interfaces might be a related area that requires attention. More frequent observations and analysis of individual vital signs might unveil reassuring or worrying trends at an earlier stage. This will require further investigation.

Conclusions

Algorithms based on readily available vital signs might be able to predict stability. Clinical application might facilitate earlier transfers of patients from hospital to their own home or post-acute care facilities.

Acknowledgements

The authors would like to thank the clinical teams from Moelwyn and Tryfan ward at the Ysbyty Gwynedd, Bangor for their enthusiasm and professionalism and Evangelos Nikolaropoulos, Jürgen Fischbach and Sabine Funke-Schaeff for their vision and technical expertise.

Competing interests

Jochen Weichert is employed by Philips Healthcare, Christian P Subbe has worked as a principal investigator on studies for Philips Healthcare and has undertaken consultancy work for Philips Healthcare. Bernd Duller is a consultant for Philips Healthcare. The authors have no other competing interest in relation to this study. The studies were funded by Philips Healthcare and undertaken by Betsi Cadwaladr University Health board. Additional funding was received by the Bevan Commission, Swansea.

References

1. Fluitman KS, Van Galen LS, Merten H, Rombach SM, Brabrand M, Cooksley T, *et al.* Exploring the preventable causes of unplanned readmissions using root cause analysis: Coordination of care is the weakest link. *Eur J Intern Med.* 2016;**30**:18–24.
2. Van Galen L, Cooksley T, H. Merten H et al . Physician consensus on preventability and predictability of readmissions based on standard case scenarios. *Neth J Med.* 2016;**74(10)**:434–42.
3. van Galen LS, Brabrand M, Cooksley T, van de Ven PM, Merten H, So RK, *et al.* Patients' and providers' perceptions of the preventability of hospital readmission: a prospective, observational study in four European countries. *BMJ Qual Saf.* 2017 Dec;**26(12)**:958–69.
4. De Angelis G, Murthy A, Beyersmann J, Harbarth S. Estimating the impact of healthcare-associated infections on length of stay and costs. *Clin Microbiol Infect.* 2010 Dec;**16(12)**:1729–35.
5. Jennings J. How can I help older people in hospital get home sooner and prevent deconditioning? *Nurs Older People.* 2017 Jun 30;**29(6)**:21–21.
6. Krumholz HM. Post-hospital syndrome--an acquired, transient condition of generalized risk. *N Engl J Med.* 2013 Jan 10;**368(2)**:100–2.
7. Subbe CP, Duller B, Bellomo R. Effect of an automated notification system for deteriorating ward patients on clinical outcomes. *Crit Care.* 2017;**21(1)**.
8. Subbe CP, Duller B. Continuous monitoring of respiratory rate on general wards what might the implications be for clinical practice? *Acute Med.* 2018;**17(1)**:5–9.
9. Jones M. NEWSDIG: The national early warning score development and implementation group. *Clinical Medicine, Journal of the Royal College of Physicians of London.* 2012. p. 501–3.
10. Society for Acute Medicine. Society for Acute Medicine's Benchmarking Audit (SAMBA) - Annual Report. 2016.
11. Breiman L. Random Forrests. Mach Learn. 2001;**45(1)**:5–32.
12. Breiman L, Friedman J, Stone C, Olshen R. Classification and Regression Trees. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis; 1984.
13. Hastie T, Tibshirani R, Friedman JH (Jerome H. The elements of statistical learning : data mining, inference, and prediction. 745 p.
14. Eckner A. Algorithms for Unevenly Spaced Time Series: Moving Averages and Other Rolling Operators. 2018.
15. v. Rossum G. Python tutorial, technical report cs-r9526. Amsterdam; 1995.
16. Oliphant TE. *Guide to NumPy.* volume **1**. Trelgol Publishing, USA; 2006.
17. McKinney W *et al.* Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference, volume 445. Austin, TX; 2010. p. 51–6.
18. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011.
19. Wickham H. ggplot2: Elegant Graphics for Data Analysis.Use R! Springer International Publishing; 2016.
20. Arnold J. ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'. R package version 4.0.1. 2018.
21. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing. R Found Stat Comput Vienna, Austria. 2012;
22. Gonçalves-Bradley DC, Lannin NA, Clemson LM, Cameron ID, Shepperd S. Discharge planning from hospital. Cochrane Database Syst Rev. 2016 Jan **27;(1)**:CD000313.
23. Fox MT, Persaud M, Maimets I, Brooks D, O'Brien K, Tregunno D. Effectiveness of early discharge planning in acutely ill or injured hospitalized older adults: a systematic review and meta-analysis. *BMC Geriatr.* 2013 Dec 6;**13(1)**:70.
24. Mabire C, Dwyer A, Garnier A, Pellet J. Effectiveness of nursing discharge planning interventions on health-related outcomes in discharged elderly inpatients. *JBI Database Syst Rev Implement Reports.* 2016 Sep;**14(9)**:217–60.
25. Zhu Q-M, Liu J, Hu H-Y, Wang S. Effectiveness of nurse-led early discharge planning programmes for hospital inpatients with chronic disease or rehabilitation needs: a systematic review and meta-analysis. *J Clin Nurs.* 2015 Oct;**24(19–20)**:2993–3005.
26. Rowley S, Sitton-Kent L. Igniting Innovation: The role of expected date of discharge in the significant reduction in length of stay. Northampton; 2015.
27. Subbe CP, Goulden N, Mawdsley K, Smith R. Anticipating care needs of patients after discharge from hospital: Frail and elderly patients without physiological abnormality on day of admission are more likely to require social services input. *Eur J Intern Med.* 2017;**45**:74–7.