METHODS

Using clinical databases to evaluate healthcare interventions

Sheila Harvey
London School of Hygiene & Tropical Medicine

Kathy Rowan, David Harrison
Intensive Care National Audit & Research Centre

Nick Black
London School of Hygiene & Tropical Medicine

Objectives: The aim of this study was to test the feasibility of conducting rigorous, nonrandomized studies (NRSs) of healthcare interventions using existing clinical databases in terms of the following: recruiting a large representative sample of hospitals, identifying eligible cases, matching cases to controls to achieve similar baseline characteristics, making meaningful comparisons of outcomes, and carrying out subgroup analyses.
Methods: Data were extracted from the Intensive Care National Audit & Research Centre's Case Mix Programme Database to investigate the impact of management with a pulmonary artery catheter (PAC) in intensive care unit (ICU) patients. Participating ICUs were invited to collect additional data for the analysis. Patients managed with a PAC were matched to control patients on their propensity score. Hospital mortality was then compared between the two groups.
Results: Of 117 eligible ICUs, 68 (58 percent) agreed to participate, of which 57 (84 percent) collected additional data. Although a slightly higher proportion of larger ICUs in university hospitals participated, the patient case-mix was similar to that in nonparticipating ICUs. Almost all patients managed with a PAC (98 percent) were successfully matched to patients managed without a PAC. The two groups were similar for baseline characteristics. However, hospital mortality was worse for PAC patients than for non-PAC patients (odds ratio, 1.28; 95 percent confidence interval, 1.06–1.55). Subgroup analysis suggested that the impact of management with a PAC was modified by severity of illness.
Conclusions: Rigorous NRSs are feasible if they are based on data from high-quality clinical databases. However, the reliability of estimated treatment effects from such studies requires further investigation.

Keywords: Nonrandomized studies, Case-mix adjustment, Propensity scores, Pulmonary artery catheter, Clinical databases

While randomized controlled trials (RCTs) provide the method with the greatest internal validity for evaluating healthcare interventions (3;5;21), most trials pay little or no attention to their external validity, as they are often conducted in single centers which are unrepresentative of the way most heath care is delivered. This, combined with the difficulties or impossibility of conducting RCTs of many healthcare interventions (4;16;17), underlines the importance of investigating whether rigorous nonrandomized studies (NRSs) can provide an alternative or at least complementary method (9;11;16;17;19).

Concern continues to be expressed about the value of NRSs, particularly when the data used are extracted from existing databases (5). While such criticisms are well-founded with regard to many routine administrative databases, maybe high-quality clinical databases should not be so readily dismissed. Many of the latter share some of the methodological strengths of RCTs such as standard rules and definitions for recording data, complete and prospective data collection, and comprehensive patient follow-up (6;10;13;16). In addition, unlike most RCTs, such databases have high external validity as they usually include data from large numbers of centers.

Our aim was to test the feasibility of conducting a NRS of a healthcare intervention using data from an existing clinical database. To do this, we selected the use of the pulmonary artery catheter (PAC) for managing critically ill patients in UK intensive care units (ICUs). The reason for this selection was that in 2002, when this study was designed, there was increasing concern about the clinical effectiveness of management with a PAC. An NRS from the United States, published in 1996 (8), had reported that management with a PAC was associated with a significantly higher hospital mortality compared with management without a PAC (odds ratio [OR], 1.39; 95 percent confidence interval [CI], 1.15 – 1.67). Although two further NRSs also reported higher mortality, in neither was the OR statistically significant ($p < .05$) (1;18). Further uncertainty arose because all three NRSs were restricted to data collected from between one and five U.S. teaching hospitals, thereby challenging their external validity. Despite the widespread concerns that these NRSs had generated about the use of PACs in critical care, at the time, only two small RCTs had been published (12;22), which both found statistically nonsignificant higher mortality associated with use of the PAC.

Against this background, our objectives were to investigate the feasibility of nesting a NRS in an existing, high-quality clinical database in terms of being able to recruit a large representative sample of ICUs, supplement routine datasets with additional ad hoc variables, identify eligible cases, adjust for differences in case-mix to enable meaningful comparisons of outcomes, and to conduct subgroup analyses.

METHODS

Design

The outcomes of a cohort of patients managed with a PAC were compared with a matched cohort of patients managed without a PAC using data from the Case Mix Program Database (CMPD) run by the Intensive Care National Audit & Research Centre (ICNARC) (13). At the time, the CMPD had Section 60 (of the Health and Social Care Act) support from the Patient Information Advisory Group, which allowed for the use of patient identifiable data without their consent. An information poster about the CMPD is displayed in all participating critical care units and an information sheet explaining the purpose of the CMPD is made available for all competent patients and for all next-of-kin.

Participants

General ICUs participating in the CMPD were invited to participate in an additional audit of selected treatments and cardiovascular monitoring. Between May 2003 and December 2004, participating ICUs recorded whether or not the patient had received mechanical ventilation, renal replacement therapy, or had been managed with a PAC during their stay in the ICU. Specialist ICUs, such as neurosciences ICUs, and high dependency units (HDUs) were excluded from the study.

For the PAC group, consecutive admissions to participating ICUs between May 2003 and December 2004 were eligible for inclusion if they were aged 16 years and older and the PAC had been placed after admission to the ICU. Patients admitted electively for preoperative optimization or who had been declared brain dead and were having a PAC placed for hemodynamic optimization before organ donation were excluded. For the non-PAC group, the same inclusion criteria were applied except management with a PAC.

Data Collection and Preparation

The following data were abstracted from the CMPD: patient identifier – sex and postcode; raw clinical data for the ICNARC risk prediction model (Table 1) (14), treatments (vasoactive drugs and mechanical ventilation) received during the first 24 hours in the ICU; mechanical ventilation and renal replacement therapy at any time during the patient's stay in the ICU; status (alive/dead) at discharge from the ICU and the date and time of discharge or death; status (alive/dead) at discharge from hospital and the date of discharge or death. Data for the CMPD are validated locally according to the ICNARC CMP Dataset Specification and undergo extensive validation for completeness, illogicalities, and inconsistencies on pooling centrally. The validation process is repeated until all queries have been resolved. The validated data are then incorporated into the CMPD.

Table 1. Description of the Intensive Care National Audit & Research Centre (ICNARC) risk prediction model

Components of the model
Scores based on data collected during the first 24 hours in the ICU are used to produce an acute physiology score and a predicted risk of hospital death

1. Physiology measured during the first 24 hours in the ICU

Highest heart rate	Highest serum urea
Lowest systolic blood pressure	Highest serum creatinine
Highest central temperature	Highest serum sodium
Lowest respiratory rate	Urine output (24 hr)
PaO$_2$/FIO$_2$ ratio / ventilation	Lowest white blood cell count
Lowest arterial pH	Glasgow Coma Scale score or sedated & paralysed

2. Source of admission

Elective surgery (direct from theater)	Critical care transfer (admissions from an ICU or HDU in the same or another hospital – direct or via accident & emergency)
Emergency surgery (direct from theater)	Accident & emergency (same hospital) or from any location in another hospital (except ICU or HDU)
Ward (same hospital)	Clinic or home

3. Cardiopulmonary resuscitation during the 24 hr prior to admission to the ICU (yes/no)

4. Diagnostic category

Respiratory	Endocrine
Cardiovascular	Hematological
Gastrointestinal	Muscular-skeletal
Neurological	Dermatological
Genitourinary	

5. Age

HDU, high dependency unit; ICU, intensive care unit.

Primary and Secondary Outcome Measures

The primary outcome measure was hospital mortality, defined as death from any cause before ultimate discharge from an acute hospital. The secondary outcomes were ICU mortality, and length of stay in the ICU and in an acute hospital.

Sample Size

The sample size calculation was based on data from previous NRSs (1;8;18) and on data kindly supplied by the Scottish Intensive Care Society Audit Group (Fiona MacKirdy; personal communication) in 1999 from their database containing details of all patients admitted to Scottish ICUs. These data suggested a hospital mortality of around 50 percent for patients managed with a PAC. It was calculated that a sample of 1,076 patients (538 in each group) would have 90 percent power to detect a 10 percent change in hospital mortality based on a 5 percent level of significance for a two-sided test.

Analyses

The analyses were conducted according to a predetermined statistical analysis plan in Stata 10 (StataCorp, College Station, Texas, USA).

Participating ICUs were compared with eligible non-participating ICUs and to all general ICUs in the CMPD according to hospital type (university, university-affiliated, nonuniversity), size of unit (number of beds), activity (mean number of admissions per year), and case-mix, as determined by the ICNARC model predicted risk of hospital mortality (Table 1) (14).

Re-admissions to the ICU during the same hospital stay were identified. For patients admitted more than once during the period, data from the first admission or from the first admission during which the patient was managed with a PAC were used. Patients with missing intervention or outcome data were excluded from the analysis.

Univariate analyses were conducted to examine the independent variables for distribution of data and to identify implausible values, outliers, etc. To assess whether any of the independent variables were strongly related to one another, Pearson's correlation was used to examine variables in pairs. Variables correlated at less than 0.8 were considered to be acceptable for inclusion in the model.

The main analysis was a risk-adjusted comparison of patients managed with and without a PAC using the ICNARC risk prediction model, which has been developed and validated for UK critical care and is based on the best elements of existing models as well as further research into modeling techniques (14). The propensity for being managed with a

PAC was calculated by fitting a logistic regression model with PAC as the outcome and the following independent variables, which were identified by a review of previous studies (1;8;18) and consultation with 10 senior critical care doctors: all variables collected for the ICNARC risk prediction model (Table 1), sex, presence of chronic cardiovascular, respiratory, renal, liver or immune system disease, and arterial base deficit/excess. Before constructing the logistic regression model, the association between the use of PAC and the continuous independent variables, age and arterial base deficit/excess, were examined for linearity. Restricted cubic splines were used to allow for nonlinear relationships. Physiological variables were entered into the model as categorical variables based on scores (Table 1). Missing physiological data were scored as normal (zero weighting).

Patients in the PAC group were then matched to patients in the non-PAC group (1:2 without replacement) based on their propensity score. A caliper size of a quarter of a standard deviation of the logit of the propensity score was used (23). First, a patient was randomly selected from the PAC group and then all the patients in the non-PAC group were searched for patients with the closest propensity score—to within 0.01 on a scale of 0 to 1. Where possible, two non-PAC patients were matched to every PAC patient. The matched patients were then removed from the pool and the process repeated. Patients in either group who could not be matched to patients in the other group were eliminated from the analysis. The two groups were then compared to ensure that they were similar with respect to baseline characteristics. In addition, selected treatments given during the first 24 hours and at any time during the ICU stay were also compared in the two groups. Given the number of comparisons, tests of statistical significance were not conducted.

The OR for death in hospital comparing patients managed with a PAC to those managed without a PAC was derived in a conditional logistic regression model using bootstrapped standard errors. The Wald test was used to assess whether the model accounted for hospital mortality better than would be expected by chance.

The numbers of deaths before discharge from the ICU were reported for PAC and non-PAC groups and the OR calculated. Differences between the two groups in the distributions of lengths of stay in the ICU and in hospital were compared using the Wilcoxon rank sum test.

RESULTS

ICU Characteristics

Of the 117 eligible ICUs invited, 68 agreed to participate. Of these, eleven were excluded from the analysis for failing to collect sufficient validated data on the use of PACs. Data from fifty-seven ICUs were, therefore, included in the analysis. On average, ICUs collected data for 18.8 months (standard deviation 4.24), ranging from 10 to 24 months. Participating ICUs were more likely to be located in a university hospital (30 percent versus 17 percent), to have 11 or more beds (25 percent versus 18 percent) and to have over 400 admissions per year (54 percent versus 34 percent). The case-mix, as determined by the ICNARC model predicted risk of hospital death for all admissions was, however, similar for both participating and nonparticipating ICUs (Supplementary Table 1, which is available at www.journals.cambridge.org/thc2010006.

Patients and Data Quality

There were 42,939 patients admitted to the fifty-seven ICUs of which, 40,880 (95 percent) met the inclusion criteria. Of the 1,068 patients who were managed with a PAC, 19 were lost to follow-up, leaving 1,049 available for the analysis. Of the 38,597 patients managed without a PAC, 795 were lost to follow-up, leaving 37,802 available for matching with PAC group patients (Figure 1).

As the relationship between management with a PAC and the continuous variables age and arterial base deficit/excess was not linear, restricted cubic splines were used to model these variables in the logistic regression model. A propensity score was derived for 33,821 (87 percent) of the 38,851 patients available for analysis. Of the 1,049 PAC group patients, 1,024 (98 percent) were successfully matched to 2,048 non-PAC patients (Figure 1).

Comparison of Baseline Characteristics and Treatment of Groups

The two groups were similar for all baseline characteristics (Table 2) except a higher proportion of patients in the PAC group had evidence of an infection on admission to ICU (36 percent versus 29 percent), although the proportion of patients whose infection was subsequently confirmed was similar for the two groups (9.2 percent versus 9.0 percent).

Some treatment differences were observed. A higher proportion of patients in the PAC group received intravenous vasoactive drug treatment during the first 24 hours (64 percent versus 53 percent), were mechanically ventilated at some stage (91 percent versus 84 percent), or had renal replacement therapy during their stay in the ICU (44 percent versus 25 percent).

Outcomes

A higher proportion of patients in the PAC group (60 percent) died in hospital than in the non-PAC group (54 percent) (Table 3). The OR comparing patients managed with a PAC with those managed without a PAC was 1.28 (95 percent CI, 1.06 – 1.55, bootstrapped standard error, 0.13). The effect of management with a PAC on hospital mortality was examined according to the predicted risk of hospital death. In the lower risk groups, management with a PAC was associated with increased odds of hospital death (Table 3).
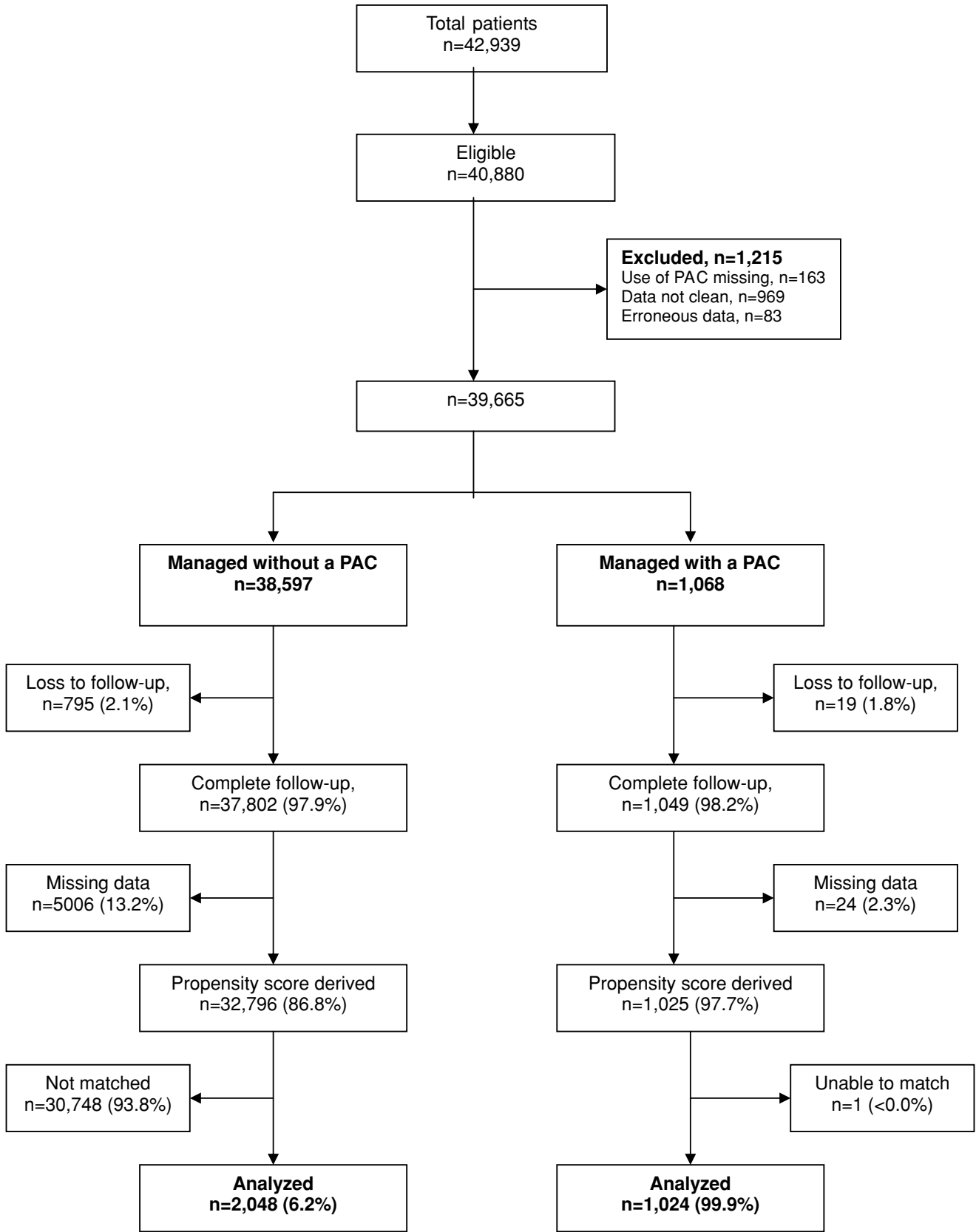
Figure 1. Patient recruitment, based on the CONSORT Statement (2).

Table 2. Comparison of Baseline Characteristics and Management of Patients by Treatment Group

Characteristic	Treatment group	
	Non-PAC ($n = 2,048$)	PAC ($n = 1,024$)
Propensity score, mean (SD)	0.08 (0.07)	0.08 (0.07)
Age, mean (SD)	62.3 (15.6)	61.9 (15.8)
Sex:, n (%):		
Male	1,252 (61.1)	624 (60.9)
Female	796 (38.9)	400 (39.1)
Chronic disease present, n (%):		
Cardiovascular	79 (3.9)	34 (3.3)
Respiratory	41 (2.0)	26 (2.5)
Liver	53 (2.6)	23 (2.3)
Renal	25 (1.2)	13 (1.3)
Immune	149 (7.3)	68 (6.6)
Likely infection on ICU admission, n (%):		
No	679 (33.2)	290 (28.3)
Laboratory confirmed	185 (9.0)	94 (9.2)
Strongly suggestive from evidence	598 (29.2)	371 (36.2)
Arterial base deficit, median (IQR)	−8.0 (−12.7 to −4.1)	−8.1 (−12.6 to −4.2)
APACHE II (1st 24 hrs):		
Acute physiology score, mean (SD)	17.2 (7.0)	17.7 (6.9)
Total score, mean (SD)	21.7 (7.5)	22.1 (7.4)
Predicted risk of death, median (IQR)	0.38 (0.21 – 0.57)	0.39 (0.23 – 0.59)
ICNARC model (1st 24 hours):		
Acute physiology score, mean (SD)	28.2 (10.4)	27.9 (9.6)
Predicted risk of death, median (IQR)	0.61 (0.34 – 0.81)	0.60 (0.37 – 0.79)
Source of admission to the ICU, n (%):		
Theatre and/or recovery	687 (33.3)	346 (33.8)
Ward same hospital	596 (29.1)	299 (29.2)
A&E same hospital	259 (12.7)	122 (11.9)
ICU/HDU same hospital	222 (10.8)	115 (11.2)
Other same hospital (including clinic/home)	125 (6.1)	64 (6.3)
Other hospital	159 (7.8)	78 (7.6)
Surgical status, n (%):		
Non-surgical	1.382 (67.5)	695 (67.9)
Elective surgery	191 (9.3)	94 (9.2)
Emergency surgery	475 (23.2)	235 (23.0)
CPR within 24 hrs prior to admission to the ICU	142 (6.9)	74 (7.2)
During 1st 24 hours in the ICU, n (%):		
Continuous intravenous vasoactive drug treatment for ≥hr	1,084 (52.9)	653 (63.8)
Mechanical ventilation	1,768 (86.3)	891 (87.0)
Never sedated/paralysed & sedated at any time	335 (16.4)	138 (13.5)
Sedated for whole of 1st 24 hrs	994 (48.5)	493 (48.1)
Paralysed & sedated for whole of 1st 24 hrs	88 (4.3)	37 (3.6)
Either sedated/paralysed & sedated for part of 1st 24 hrs	631 (30.8)	356 (34.8)
At any time during stay in the ICU, n (%):		
Mechanical ventilation, n (%)	1,728 (84.4)	927 (90.5)
Renal replacement therapy, n (%)	503 (24.6)	454 (44.3)

PAC, pulmonary artery catheter; SD, standard deviation; IQR, interquartile range; A&E, Accident & Emergency; HDU, high dependency unit; ICU, intensive care unit; ICNARC, Intensive Care National Audit & Research Centre; APACHE, Acute Physiology and Chronic Health Evaluation.

A difference was also observed when ICU mortality was considered: 53 percent in the PAC group versus 44 percent in the non-PAC group; OR 1.50 (95 percent CI, 1.24 – 1.81, bootstrapped standard error 0.15). Finally, hospital survivors and nonsurvivors in the PAC group stayed longer in the ICU and in hospital compared with those in the non-PAC group (Table 3).

DISCUSSION

Main Findings

This study has demonstrated that it is feasible to carry out a rigorous NRS of a healthcare intervention using an existing clinical database. We achieved high levels of agreement from ICUs to participate (58 percent) of which most

Table 3. Hospital Mortality Overall and by Predicted Risk, and Length of Stay in the ICU and in Hospital by Treatment Group

	Non-PAC	PAC	
Hospital mortality	N (%)	N (%)	p[a]
Survivors	940 (45.9)	415 (40.5)	.005
Non-survivors	1108 (54.1)	609 (59.5)	
	OR (PAC vs. no PAC) 1.28, 95% CI 1.06 – 1.55		
Predicted risk[b]	OR	95% CI	SE
≤0.33	1.79	0.89 – 3.61	0.64
0.34 – 0.66	1.96	1.25 – 3.09	0.45
0.67 – 1.00	0.89	0.58 – 1.38	0.20
Length of stay (days)	Median (IQR)	Median (IQR)	p[c]
ICU:			
Survivors	5.0 (2.0 – 12.3)	10.3 (4.9 – 20.4)	<.001
Non-survivors	1.7 (0.7 – 5.7)	3.2 (1.3 – 9.4)	<.001
Hospital:			
Survivors	28 (15 – 50.5)	38 (21 – 65)	<.001
Non-survivors	8 (2 – 20)	9 (3 – 22)	.001

[a]p value from Fisher's exact test.
[b]Intensive Care National Audit & Research Centre (ICNARC) model predicted risk of death in hospital.
[c]p value from Wilcoxon rank sum test for difference in distribution.
PAC, pulmonary artery catheter; IQR, interquartile range; CI, confidence interval; SE, standard error (bootstrapped); OR, odds ratio.

(84 percent) successfully collected the required additional data. Although there was a slightly higher likelihood for larger ICUs in university hospitals to participate, the case-mix of patients was similar to that in nonparticipating ICUs. Eligible patients were successfully identified and almost all of those in the intervention group (98 percent) were successfully matched on their propensity score to patients in the control group. This created two groups with similar baseline characteristics, allowing meaningful comparisons of outcomes to be made. These suggested that the outcome for PAC patients was significantly worse than for non-PAC patients with regard to ICU mortality (OR 1.50; 95 percent CI, 1.24 – 1.81) and hospital mortality (OR 1.28; 95 percent CI, 1.06 – 1.55). The large numbers of patients available meant that subgroup analyses were possible and these suggested poorer outcomes for those with predicted risk of hospital death at admission to the ICU of less than 67 percent.

Strengths and Limitations

This study has three particular strengths compared with many NRSs. First, it used data from a high-quality clinical database that performs well against the quality criteria defined by the Directory of Clinical Databases (10;13). The particular strengths of the CMPD are its wide coverage, making it highly representative of the patient population, the use of explicit definitions for all variables and rules for data collec-

tion, and employment of rigorous data validation methods. Furthermore, collection of raw data enables risk adjustment models to be derived using standard algorithms across all ICUs to calculate scores and risks, allowing for better comparability of risk-adjusted outcomes (14). Second, unlike the three previous NRSs that had been published at the time (1;8;18), this study included a large number of ICUs, located in both university and nonuniversity ICUs, ensuring high external validity. Third, the validity and predictive power of the model used to adjust for case-mix was excellent (14). A major strength of the model is that there are no exclusion criteria. For many risk prediction models, these are often ill-defined, poorly reported, and may exclude up to 15 percent of admissions (25). Furthermore, inconsistent application of criteria may introduce biases into risk-adjusted analyses (14). The risk adjustment carried out in this study also included other important prognostic factors: presence of chronic cardiovascular, respiratory, renal, liver, or immune system disease; sex; and arterial base deficit/excess.

The principal limitation, as with all NRSs, was uncertainty as to whether all significant confounders had been taken into account. All confounders identified by the literature review and through consultation with critical care doctors were measured and included in the model with one exception—the trajectory of a patient's illness. Severity of illness, as determined by the ICNARC risk prediction model, was measured using data collected during the first 24 hours in the ICU. However, although two patients may have the

same predicted risk of death at admission to the ICU, they may be at different points in their illness trajectory. One may have stabilized and be at their worst state of health, whereas the other may be in decline leading toward a more severe condition. Clinical deterioration or lack of response to treatment was identified by critical care doctors, consulted before the analysis, as one of the factors that influences their decision to insert a PAC. Given that an arterial base deficit (when there is increased acidity) is considered to be an indicator of shock, injury, and inadequate resuscitation, they suggested that the arterial base deficit/excess might indicate how a patient was responding to treatment. This variable was therefore included in the model as a covariate, and although it may have been partly successful in adjusting for clinical deterioration, a limitation was that only a single value of arterial base deficit/excess was available, which was the worst value measured during the first 24 hours in the ICU, and the time of the measurement was not recorded. For some patients their worst arterial base deficit might have occurred at admission to the ICU, but have improved later during the first 24 hours as they responded to treatment. For other patients, the arterial base deficit might have worsened after admission to the ICU as their clinical condition deteriorated. It was hard, therefore, to make a distinction between these patients based on a single measure.

Although matching on propensity score produced two groups that were similar with respect to their baseline characteristics, patients in the PAC group were more likely to have received other additional interventions including vasoactive drugs, mechanical ventilation, and renal replacement therapy, indicating that they were a sicker group. It was not appropriate, however, to include these interventions as covariates in the propensity score model as it was not known when they were introduced relative to insertion of the PAC. It does mean, though, that any difference in outcomes between PAC and non-PAC groups should consider these concurrent interventions when attributing causality.

Implications of the Findings

While this study has demonstrated the feasibility of conducting a rigorous NRS, the question remains as to the accuracy of the findings given the inevitable uncertainty with regard to its internal validity (threatened by unadjusted confounding). A limitation of using secondary data is that detailed information about the intervention of interest may not have been recorded. In the case of this study, detailed information relating to PAC use, such as the date and time of insertion and the patient's clinical condition at the time of insertion would have been helpful and might have allowed for more accurate adjustment for differences in baseline prognostic factors between the PAC and non-PAC groups. The matching strategy was also limited because patients were matched on patient characteristics during the first 24 hours in the ICU rather than their characteristics up to and including the point at which

the PAC was inserted. Ideally, if the time of PAC insertion had been known, PAC patients would have been matched to controls who had spent at least as long in the ICU as the PAC patients had before receiving a PAC. In other words, matching would have included the time at risk of having a PAC inserted. Investigators using data from clinical databases should carefully consider therefore the additional data that may be needed to enable a rigorous case-mix adjusted evaluation of the intervention of interest. However, this does have to be weighed against the available resources and practicality of collecting additional data.

The findings of this study are consistent with the range of ORs reported by the three previous NRSs (1.08 to 1.54), which had all suggested that management with a PAC is associated with higher hospital mortality compared with management without a PAC in general ICU populations (1;8;18). In addition, two of these studies also reported that poorer outcome was restricted to patients at lower baseline risk (8;18). Since 2002, three further NRSs have been published, two based on data from single North American university hospitals (7;20) and one on 198 European hospitals (24). While one study (20) reported a significantly higher risk of death with PAC use (OR 1.5; 95 percent CI, $1.1 - 2.0$), the other two found statistically insignificant differences (ORs 1.05; 95 percent CI, $0.92 - 1.21$ and 1.04, 95 percent C $0.79 - 1.38$, respectively) (7;24). The only study to report subgroup analyses found that the risk associated with PAC use was confined to less sick patients (7), consistent with this and earlier studies.

An important question, given the limitations relating to their internal validity, is whether NRSs can provide reliable estimates of treatment effect. One approach to investigating the accuracy of a NRS is to compare its results with one or more RCTs (15). This is currently under way and will be reported elsewhere. Meanwhile, this study has demonstrated the feasibility and methodological rigor that can be brought to bear in NRSs. Given the relatively low cost of a study design that is nested in an existing database (thereby incurring little cost for data collection and processing as most of the data have already been collected for clinical audit) compared with a RCT, plus the many occasions when a randomized design is not possible (4), this study serves to demonstrate the great potential NRSs could offer. This will only be realized if high-quality clinical databases, such as the one used in this study, are established and supported. The need for ongoing development in the identification and measurement of confounding factors (as demonstrated in this study by the need to try and incorporate a measure of the patient's illness trajectory) must continue to be addressed.

SUPPLEMENTARY MATERIALS

Supplementary Table 1
www.journals.cambridge.org/thc2010006

CONTACT INFORMATION

Sheila Harvey, PhD (sheila.harvey@lshtm.ac.uk), Lecturer, Epidemiology & Population Health, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Kathy Rowan, DPhil (kathy.rowan@icnarc.org), Director, **David Harrison**, PhD (david.harrison@icnarc.org), Senior Statistician, Intensive Care National Audit & Research Centre, Tavistock House, Tavistock Place, London WC1H 9HR, UK

Nick Black, MD (nick.black@lshtm.ac.uk), Professor, Public Health & Policy, London, School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK

REFERENCES

1. Afessa B, Spencer S, Khan W, et al. Association of pulmonary artery catheter use with in-hospital mortality. *Crit Care Med*. 2001;29:1145-1148.
2. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med*. 2001;134:663-694.
3. Barton S. Which clinical studies provide the best evidence? The best RCT still trumps the best observational study. *BMJ*. 2000;321:255-256.
4. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ*. 1996;312:1215-1218.
5. Byar DP. Why data bases should not replace randomized clinical trials. *Biometrics*. 1980;36:337-342.
6. Canto JG, Kiefe CI, Williams OD, Barron HV, Rogers WJ. Comparison of outcomes research with clinical trials using pre-existing data. *Am J Cardiol*. 1999;84:923-927, A6.
7. Chittock DR, Dhingra VK, Ronco JJ, et al. Severity of illness and risk of death associated with pulmonary artery catheter use. *Crit Care Med*. 2004;32:911-915.
8. Connors AF Jr, Speroff T, Dawson NV, et al. The effectiveness of right heart catheterization in the initial care of critically ill patients. SUPPORT Investigators. *JAMA*. 1996;276:889-897.
9. Cook D, Heyland D, Marshall J. On the need for observational studies to design and interpret randomized trials in ICU patients: A case study in stress ulcer prophylaxis. *Intensive Care Med*. 2001;27:347-354.
10. DocDat.Org. http://www.docdat.org (accessed October 4, 2007).
11. Feinstein AR, Horwitz RI. Problems in the "evidence" of "evidence-based medicine". *Am J Med*. 1997;103:529-535.
12. Guyatt G. A randomized control trial of right-heart catheterization in critically ill patients. Ontario Intensive Care Study Group. *J Intensive Care Med*. 1991;6:91-95.
13. Harrison DA, Brady AR, Rowan K. Case mix, outcome and length of stay for admissions to adult, general critical care units in England, Wales and Northern Ireland: The Intensive Care National Audit & Research Centre Case Mix Programme Database. *Crit Care*. 2004;8:R99-R111.
14. Harrison DA, Parry GJ, Carpenter JR, Short A, Rowan K. A new risk prediction model for critical care: The Intensive Care National Audit & Research Centre (ICNARC) model. *Crit Care Med*. 2007;35:1091-1098.
15. Harvey S, Harrison DA, Singer M, et al. for the PAC-Man study collaboration. Assessment of the clinical effectiveness of pulmonary artery catheters in management of patients in intensive care (PAC-Man): A randomised controlled trial. *Lancet*. 2005;366:472-477.
16. Hlatky MA, Califf RM, Harrell FE Jr, et al. Comparison of predictions based on observational data with the results of randomized controlled clinical trials of coronary artery bypass surgery. *J Am Coll Cardiol*. 1988;11:237-245.
17. Horwitz RI, Viscoli CM, Clemens JD, Sadock RT. Developing improved observational methods for evaluating therapeutic effectiveness. *Am J Med*. 1990;89:630-638.
18. Murdoch SD, Cohen AT, Bellamy MC. Pulmonary artery catheterization and mortality in critically ill patients. *Br J Anaesth*. 2000;85:611-615.
19. Padkin A, Rowan K, Black N. Using high quality clinical databases to complement the results of randomised controlled trials: The case of recombinant human activated protein C. *BMJ*. 2001;323:923-926.
20. Peters SG, Afessa B, Decker PA, et al. Increased risk associated with pulmonary artery catheterization in the medical intensive care unit. *J Crit Care*. 2003;18:166-171.
21. Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? *N Engl J Med*. 2000;342:1907-1909.
22. Rhodes A, Cusack RJ, Newman PJ, Grounds RM, Bennett ED. A randomised, controlled trial of the pulmonary artery catheter in critically ill patients. *Intensive Care Med*. 2002;28:256-264.
23. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41-55.
24. Sakr Y, Vincent JL, Reinhart K, et al. Use of the pulmonary artery catheter is not associated with worse outcome in the ICU. *Chest*. 2005;128:2722-2731.
25. Wunsch H, Brady AR, Rowan K. Impact of exclusion criteria on case mix, outcome, and length of stay for the severity of disease scoring methods in common use in critical care. *J Crit Care*. 2004;19:67-74.