



Establishing a Multicentre Trauma Registry in India: An Evaluation of Data Completeness

Gowri Shivasabesan^{1,2,3} · Gerard M. O'Reilly^{1,2,3} · Joseph Mathew^{1,2,4,5} · Mark C. Fitzgerald^{2,4,5} · Amit Gupta⁶ · Nobhojit Roy^{3,7} · Manjul Joshipura⁸ · Naveen Sharma⁹ · Peter Cameron^{1,2,3} · Madonna Fahey^{2,10} · Teresa Howard^{2,5} · Zoe Cheung² · Vineet Kumar¹¹ · Bhavesh Jarwani¹² · Kapil Dev Soni⁶ · Pankaj Patel¹² · Advait Thakor¹² · Mahesh Misra¹³ · Russell L. Gruen¹⁴ · Biswadev Mitra^{1,2,3} on behalf of the Australia-India Trauma Systems Collaboration (AITSC)

Published online: 20 June 2019
© Société Internationale de Chirurgie 2019

Abstract

Background The completeness of a trauma registry's data is essential for its valid use. This study aimed to evaluate the extent of missing data in a new multicentre trauma registry in India and to assess the association between data completeness and potential predictors of missing data, particularly mortality.

Methods The proportion of missing data for variables among all adults was determined from data collected from 19 April 2016 to 30 April 2017. In-hospital physiological data were defined as missing if any of initial systolic blood pressure, heart rate, respiratory rate, or Glasgow Coma Scale were missing. Univariable logistic regression and multivariable logistic regression, using manual stepwise selection, were used to investigate the association between mortality (and other potential predictors) and missing physiological data.

Results Data on the 4466 trauma patients in the registry were analysed. Out of 59 variables, most ($n = 51$; 86.4%) were missing less than 20% of observations. There were 808 (18.1%) patients missing at least one of the first in-hospital physiological observations. Hospital death was associated with missing in-hospital physiological data (adjusted OR 1.4; 95% CI 1.02–2.01; $p = 0.04$). Other significant associations with missing data were: patient arrival time out of hours, hospital of care, 'other' place of injury, and specific injury mechanisms. Assault/homicide injury intent and occurrence of chest X-ray were associated with *not* missing any of first in-hospital physiological variables.

Conclusion Most variables were well collected. Hospital death, a proxy for more severe injury, was associated with missing first in-hospital physiological observations. This remains an important limitation for trauma registries.

✉ Gowri Shivasabesan
gowri.shiva14@gmail.com

- ¹ Emergency and Trauma Centre, The Alfred Hospital, Melbourne, Australia
- ² National Trauma Research Institute, The Alfred Hospital, 85-89 Commercial Rd, Melbourne, VIC 3004, Australia
- ³ School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
- ⁴ Trauma Service, The Alfred Hospital, Melbourne, Australia
- ⁵ Central Clinical School, Monash University, Melbourne, Australia
- ⁶ Division of Trauma Surgery and Critical Care, All India Institute of Medical Science, New Delhi, India

- ⁷ WHO Collaborating Centre for Research on Surgical Care Delivery in LMICs, Surgical Unit, BARC Hospital (Govt. of India), Mumbai, India
- ⁸ Academy of Traumatology, Ahmedabad, Gujarat, India
- ⁹ Department of Surgery, All India Institute of Medical Sciences, Jodhpur, India
- ¹⁰ Tasmanian Health Service, Hobart, Australia
- ¹¹ Lokmanya Tilak General Hospital and Municipal Medical College, Mumbai, India
- ¹² Smt. NHL Municipal Medical College, Ahmedabad, India
- ¹³ Mahatma Gandhi University of Medical Sciences and Technology, Jaipur, India
- ¹⁴ College of Health and Medicine, Australian National University, Canberra, Australia

Introduction

Each year, 5 million people die from injuries worldwide [1]. Low- and middle-income countries (LMICs) are disproportionately affected by the burden of trauma. More than 90% of deaths due to injury occur in these countries [2]. India faces a high injury burden with injuries causing around 11% of the country's total deaths [3]. The World Health Organization (WHO) acknowledges that to prevent injury and improve care, evidence-based strategies should be country and environment specific. This requires the existence of comprehensive injury data on the many aspects of trauma patient care [4].

Trauma registries document presentation, care, and outcomes of injured patients [5]. They have the potential to drive change on many levels—from pre-hospital care to specific hospital trauma care quality improvement to government legislation and resource allocation [5]. The valid use of a trauma registry relies on the quality of its data [6]. Missing observations can lead to biases in analysis and interpretation of information. However, the overall completeness of variables in trauma registries has not been well described and no standardized benchmark of acceptable levels of missing data exist [7–9]. If data are missing, the validity of research can become uncertain—especially as most manuscripts based on trauma registry data do not appropriately report and manage incomplete observations [10, 11]. Hence, good quality data recording in registries is critical.

The collection and recording of first in-hospital physiological variables [heart rate (HR), systolic blood pressure (SBP), respiratory rate (RR), and Glasgow Coma Scale (GCS)] are particularly important to guide trauma resuscitation and outcomes. Between them, at least one of these first in-hospital physiological variables is a component of almost all physiological scores developed to adjust for injury severity for trauma care benchmarking and quality improvement (e.g. Revised Trauma Score, Kampala Trauma Score) [12, 13]. Yet physiological observations are at a high risk of being missing in trauma registries [14]. Furthermore, death in hospital has been found to be an independent predictor of missing physiological variables in trauma registries, implying more severely injured patients have less complete data [14]. This has substantial implications for studies assessing mortality as an outcome measure as severely injured patients (those at highest risk of death) may be excluded from analysis.

The aim of this study was to assess the prevalence of missing data in a new multihospital Indian trauma registry, the Australia-India Trauma Systems Collaboration Trauma Registry. The study examined whether any variables—especially hospital mortality—were associated with

missing first in-hospital physiological data. By identifying predictors of missing these data points, this study aimed to inform interventions to improve data completeness in the registry.

Methodology

Setting

The Australia-India Trauma Systems Collaboration (AITSC) is a partnership between Australia and India, led by the National Trauma Research Institute (NTRI), a department of Monash University and Alfred Health, and the Jai Prakash Narayan Apex Trauma Center, All India Institute of Medical Sciences (AIIMS), Delhi [15]. The AITSC was funded by the Indian Government (through the Department of Science and Technology) and the Australian Government (through the Department of Industry, Innovation, and Science). Various projects have been underway across four Indian trauma hospitals to develop India's trauma systems, spanning from pre-hospital to post-discharge care [16, 17]. Underpinning and informing projects on the study of clinical care systems was the development of a trauma registry, established in 2016 [15]. This study of registry quality was undertaken prior to the commencement of the other interventions, notably the AITSC pre-hospital notification study [15].

The AITSC trauma registry is a registry developed to collect trauma data, laying the foundation for trauma registries in India. In its first year, the AITSC registry involved four major Indian trauma hospitals—the Jai Prakash Narayan Apex Trauma Center (JPN), All India Institute of Medical Sciences (AIIMS), New Delhi; the Lokmanya Tilak Municipal General (LTMG) Hospital, Mumbai; the Sheth Vadilal Sarabhai (VS) General Hospital, Ahmadabad; and the Guru Teg Bahadur (GTB) Hospital, New Delhi. Data were collected on all trauma patients who presented with a potentially life-threatening or limb-threatening injury. The hospital of presentation was presented in a non-identifiable format for this study.

At each study site, two trained data collectors certified to code injury severity using the Abbreviated Injury Scale recorded patient data. This occurred either through direct observation of healthcare staff and patients, or by extraction from paper medical records (e.g. for patients arriving out of shift hours), as per the AITSC Trauma Registry Data Dictionary (version 1.02, March 2016). Shifts of data collectors were from 9 am to 7 pm, Monday–Saturday. Data were recorded in prepared data sheets, using codes to identify observations/values. When the data collectors were unable to find or observe the data, or when data were inadequately described, they coded it as such.

Data cleaning involved the project officers reviewing the data retrospectively and correcting any obvious mistakes (e.g. inconsistency between admission and discharge dates) or following up on blank observations with data collectors to the best of their ability.

Patient selection and data extraction

Data for all adults (age ≥ 18 years) with a potentially life-threatening or limb-threatening injury entered into the AITSC trauma registry from inception (19 April 2016) until 30 April 2017 were extracted. Variables (59 in total) were grouped as data on hospital arrival, demographics, injury occurrence, emergency department (ED) measures, initial investigations (Ix) and procedures, and hospital stay and outcomes. Pre-hospital variables were excluded from this study as the AITSC pre-hospital notification study had not commenced.

Statistical methods

Completeness levels

The level of completeness (proportion, expressed as percentages) for each variable was quantified. ‘Missing data’ included ‘non-valid’ missing data (variables which were left blank) and ‘valid’ missing data (variables coded as ‘missing’ or ‘unknown’). Free text variables (e.g. place of residence) were determined to be missing if left blank or described as ‘unknown’ or ‘not recorded’.

The presence of observations for certain variables was conditional on specific interventions. In such cases, the number of observations missing, given that event occurred, was quantified—e.g. the proportion of missing data on the time a CT scan occurred was based on data showing that the scan had occurred.

Association between in-hospital mortality and other variables with missing physiological data

Descriptive analyses were conducted on variables based on patient mortality. Variables regarding hospital interventions (chest X-ray, CT scan, neurosurgical consult, blood transfusion, mechanical ventilation, operations) were converted into binary variables representing whether they occurred or not. Arrival date was converted into day of week. Arrival time was classified into binary ‘in-hours’ (07:00–18:00) and ‘out of hours’ (18:00–07:00) as per previous studies [18]. Median and inter-quartile ranges were determined for days on mechanical ventilation, length of stay (LOS) in ICU and LOS in hospital. Free text variables and conditional variables were excluded from this and subsequent analyses.

To represent missing first in-hospital (i.e. in the ED) physiological data, a new binary variable ‘miss_phys’ was created; 1 represented if any of SBP, HR, RR, or GCS were missing for a given case, while 0 represented none of these observations being missing for a given case. Descriptive and univariable logistic regression analyses were conducted to see whether in-hospital mortality or other factors were associated with missing first-recorded physiological data. Results were reported using unadjusted odds ratios (95% confidence intervals).

Multivariable logistic regression was used to investigate the independent association between mortality and missing physiological data, controlling for covariates, and to assess any other predictors of such missing physiological data. Variables that demonstrated an association ($p < 0.1$) with ‘miss_phys’ in the univariable analysis, including any sub-category for categorical variables, and missing less than 20% of data (as per benchmark for trauma registry variables by Ringdal et al.) were included in the multivariable model [9]. Analysis was conducted using manual stepwise selection. Results were reported as adjusted odds ratios (95% confidence intervals). Statistical significance was defined as $p < 0.05$. Missing data on predictors and covariates were managed using list-wise deletion. Data analysis was conducted using STATA version 14.0 (College Station, Texas, USA).

Results

The registry included 4466 adult patients in the specified time period. The majority of patients (2587, 58%) in the registry were from Hospital 1, with 689 (15%) derived from Hospital 2, 473 (11%) from Hospital 3, and 717 (16%) from Hospital 4. Most patients survived (3835, 85.9%) while 582 (13.0%) died. There were 49 (1.1%) patients with unknown survival data.

Of all patients, 2014 were recorded as arrival by ambulance—583 patients (13.1% of total) arrived direct from scene while others were inter-facility transfers. Road traffic incidents accounted for 2485 patient injuries (55.6%).

Completeness of variables

Table 1 outlines the extent of missing observations in the AITSC trauma registry, which spanned from 0 to 67.4%. Four variables had no missing observations (data collector, admission date, sex, and first vital sign date). Most variables ($n = 51$, 86.4%) were missing less than 20% of observations.

Tables 2 and 3 describe the key characteristics of patients based on the primary predictor of hospital

mortality status. They also provide a description of each variable in the registry, as well as the proportion of missing data.

Predictors of missing physiological data

Univariable analysis

Data on any one of first in-hospital SBP, HR, GCS, or RR were missing in 808 patients (18.1%). They showed the following proportions of individual missing observations: HR—72 (1.6%); SBP—102 (2.3%); GCS—168 (3.8%); and RR—721 (16.1%). Table 4 outlines the characteristics of patients based on their status of missing physiological data and the results of the univariable logistic regression between the different variables (potential predictors) of missing physiological data.

Multivariable analysis

Table 5 presents the results of the multivariable logistic regression. The final model included 3382 cases (75.7% of all patients in registry). The following variables were found to be independently associated with missing first physiological data ($p < 0.05$): death, arrival time out of hours (18:00–07:00), hospital of care, ‘other’ place of injury as compared to home/residential institution, and blunt force and sharp force mechanism of injury (i.e. those injuries caused by blunt/sharp objects, people, or animals) as compared to road trauma incidents.

Both an ‘intentional—assault/homicide’ intent of injury and the occurrence of a chest X-ray were found to be associated with not missing first physiological data.

Discussion

This was the first study to examine completeness of variables in the AITSC registry in the context of establishing a new multicentre trauma registry in India. There were low rates of missing data compared to other trauma registries, highlighting a key success [8]. Hospital death was associated with missing physiological data, implying more severely injured patients were more likely to have absent physiological observations in the registry. Other key predictors of missing these observations included out-of-hours arrival time, hospital of care, and mechanism of injury. Respiratory rate was missing much more frequently than SBP, HR, or GCS (RR missing 16.1% of observations while the others were missing less than 5%).

The findings highlight the potential limitations of using a registry with missing data. Managing missing data is critical to ensuring the validity of a trauma registry for

benchmarking and quality improvement, and its subsequent use to improve trauma care. Identifying predictors of missing observations can inform efforts to improve this registry’s data quality. This study also provides important lessons for the development and strengthening of registries in other LMICs. Despite the importance of such registries, few exist in LMICs; similarly, only 1% of trauma registry publications are derived from the least developed countries [19, 20].

The AITSC registry’s low rates of missing data were comparable, if not lower, than those reported from other registries. For example, the proportion of missing data in an established Oregon State Trauma Registry for four reported variables was: intubation attempt, 9%; GCS, 17%; SBP, 22%; and RR, 17% [21]. In other trauma registries, missing data have ranged from 4.9% to 20.4% for GCS, 2.1% to 33.7% for RR, and 2.4% to 28.7% for SBP [8]. While there are no formal guidelines for appropriate levels of missing data for trauma registries, Ringdal et al. used > 80% as the benchmark desired level of completeness—most variables in the AITSC registry achieved this level of completeness [9]. Variables with high rates of missing observations in the registry were mostly those regarding trauma arrival and time of investigations/procedures. These variables contribute to the appraisal of trauma care, allowing for baseline assessments and subsequent evaluations of interventions (e.g. time taken, appropriateness) which can ultimately improve patient outcomes. As such, minimizing the extent to which these variables are missing observations will maximize the impact of the AITSC registry in trauma quality improvement. Further, respiratory rate was missing in a substantial proportion of cases, so an increased attention to the collection of observations for this variable should be emphasized across sites. One key use of physiological data is its use in various trauma scoring systems to estimate probability of survival and benchmark trauma quality care across registries [12]. Hence, scores that require the respiratory rate clearly will face limitations when using this registry.

The association between death in hospital and missing physiological variables has been previously reported and was confirmed in this report. This study echoes previous findings of mortality predicting missing observations for specific physiological variables (GCS, RR) in the Victorian State Trauma Registry [14]. The other key predictors of missing and not missing physiological data have important implications for improving data collection in the AITSC registry (see Fig. 1). Hospital of presentation affected the completeness of such data, and hence, strategies on data collection could be shared between sites. Better data recording occurred when cases had law enforcement and/or legal implications (i.e. assault cases); hence, this rigour should be translated to all trauma cases. This may be

Table 1 Proportion and percentage of patients missing data for variables in the AITSC trauma registry, colour coded using a traffic light system based on % missing (green = low levels of missingness, yellow = moderate levels, red = high levels)

Variable	n*	Missing	Missing (%)	Variable	n*	Missing	Missing (%)
Data Collector		0	0.0	ED cont.			
Hospital arrival				First HR		72	1.6
Arrival Trauma flag		240	5.4	First RR		721	16.1
Trauma call out		2568	57.5	RR qualifier		230	5.2
Date of trauma call out	711	6	0.1	First SpO2		713	16.0
Time of trauma call out	711	38	0.9	Supplementary O2		344	7.7
Trauma Leader present		2992	67.0	GCS-Eye		170	3.8
Trauma Bay Ready		3008	67.4	GCS-Verbal		171	3.8
Referring hospital		29	0.6	GCS-Motor		168	3.8
Mode of arrival		313	7.0	GCS Total		174	3.9
Arrival Date		3	0.1	GCS qualifier		41	0.9
Arrival Time		474	10.6	ED disposition		54	1.2
Admission Date		0	0.0	ED disposition Date		118	2.6
Admission Time		9	0.2	ED disposition Time		368	8.2
Demographic				Hospital Ix/Procedures/Operations (OP)			
Age		9	0.2	OP Type		42	0.7
Sex		0	0.0	OP Date		20	2.5
Place of residence		253	5.7	OP time commenced	1190	589	28.5
Healthcare Record ID		691	15.5	CXR date		32	0.5
Injury Occurrence				CXR time		3427	483
Injury date		53	1.2	CT Type/s		25	0.4
Injury time		317	7.1	CT Date		23	0.6
Injury place		561	12.6	CT Time		3304	490
Mechanism of Injury		31	0.7	Neurosurg consult – Date		27	0.4
Dominant type of injury		8	0.2	Neurosurg consult - Time	2995	782	25.2
Injury intent		39	0.9	Blood transfusion – Date		39	0.7
Primary vehicle	2485	198	4.4	Blood transfusion – Time	651	256	35.6
Vehicular injury patient role	2485	234	9.4	Hospital stay & outcome			
Activity engaged when injured		2177	48.7	In hospital patient events		2720	60.9
ED				In hospital vent. Days		55	1.2
First vital sign Date		0	0.0	In hospital ICU stay days		28	0.6
First vital sign Time		592	13.3	Hospital disposition		51	1.1
First SBP		102	2.3	Hospital disposition date		98	2.2

*n = 4466 unless otherwise specified

KEY	0%	<5%	5-10%	11-20%	21-50%	>50%

achieved by better education of data collectors and clinicians as some data points are rarely recorded by clinical staff and hence cannot be inputted into the registry.

This study also emphasizes the importance of strengthening ‘after-hours’ data collection. In overall mature trauma systems, comparable mortality among injured patients presenting on weeknights versus weekdays and lower mortality among injured patients on weekends versus weekdays have been demonstrated [22, 23]. However, among the most critically ill trauma cohort, the after-hours model of care has been reported to be associated with worse outcomes [18]. This highlights the importance of sustained data collection at all hours. In this registry, after-hours presentation was associated with missing

physiological data, i.e. those times where data collectors were not prospectively recording information. Barriers to after-hours data collection may include the expense and lack of availability of after-hours data collectors (i.e. for concurrent data collection), as well as the lack of robust patient record-keeping and/or storage and retrieval systems for retrospective data collection. Thus, strengthening the quality of data for patients presenting after-hours presents a key challenge for this emerging registry.

While the literature on effective interventions for improving trauma data collection in LMICs is limited, the specific actions mentioned above could also be complemented by implementing general underlying departmental and hospital processes. Organizing regular meetings

Table 2 Demographics and injury event information for patients in the AITSC registry based on mortality status and composite characteristics (all patients in registry—including those with missing mortality status), including proportion of missing data for each variable

Variable	Alive N = 3835	Dead N = 582	Overall N = 4466
Age (years); mean (SD)	36.5 (16.5)	41.1 (15.1)	37.1 (15.3)
Age—missing (%)	6 (0.2)	3 (0.5)	9 (0.2)
Gender/sex (%)			
Male	3264 (85.1)	482 (82.8)	3790 (84.9)
Female	570 (14.9)	100 (17.2)	675 (15.1)
Other	1 (0.0)	0 (0.0)	1 (0.0)
Gender/sex—missing (%)	0 (0.0)	0 (0.0)	0 (0.0)
Hospital of presentation (%)			
1	2314 (60.3)	273 (46.9)	2587 (57.9)
2	578 (15.1)	111 (19.1)	689 (15.4)
3	320 (8.3)	141 (24.2)	473 (10.6)
4	623 (16.3)	57 (9.8)	717 (16.1)
Hospital—missing (%)	0 (0.0)	0 (0.0)	0 (0.0)
Place of injury (%)			
Home or Residential Institution	663 (19.8)	116 (22.6)	782 (17.5)
Road/highway	2265 (67.8)	309 (60.1)	2615 (58.6)
Trade/farm/construction	136 (4.1)	10 (2.0)	147 (3.3)
Railway/trails	146 (4.4)	57 (11.1)	205 (4.6)
Other	132 (4.0)	22 (4.3)	156 (3.5)
Injury place—missing (%)	493 (12.9)	68 (11.7)	561 (12.6)
Mechanism of injury (%)			
Road traffic incident (RTI)	2141 (56.1)	308 (54.1)	2485 (55.6)
Other transport	148 (3.9)	55 (9.7)	205 (4.6)
Falls	832 (21.8)	154 (27.1)	991 (22.2)
Blunt force	210 (5.5)	10 (1.8)	222 (5.0)
Sharp force	107 (2.8)	8 (1.4)	118 (2.6)
Gunshot	63 (1.7)	6 (1.1)	69 (1.5)
Other	316 (8.3)	28 (4.9)	344 (7.7)
Mechanism—missing (%)	18 (0.5)	13 (2.2)	32 (0.7)
Dominant type of injury (%)			
Blunt	3509 (91.6)	562 (96.9)	4117 (92.2)
Penetrating	282 (7.4)	16 (2.8)	301 (6.7)
Other	38 (1.0)	2 (0.3)	40 (0.9)
Dominant injury type—missing (%)	6 (0.2)	2 (0.3)	8 (0.2)
Activity engaged when injured (%)			
Leisure/Sport	145 (7.3)	31 (11.4)	176 (3.9)
Work	682 (34.4)	75 (27.6)	759 (17.0)
Vital activities (sleep/eat)	115 (5.8)	27 (9.9)	144 (3.2)
Other	1038 (52.4)	139 (51.1)	1210 (27.1)
Activity—missing (%)	1855 (48.3)	310 (53.2)	2177 (48.7)
Intent of injury (%)			
Unintentional	3335 (87.5)	536 (94.4)	3915 (87.7)
Intentional: self-harm/suicide	52 (1.4)	5 (0.9)	57 (1.3)
Intentional: assault/homicide	421 (11.1)	27 (4.8)	453 (10.1)
Other	2 (0.1)	0 (0.0)	2 (0.0)
Intent—missing (%)	25 (0.7)	14 (2.4)	39 (0.9)

Table 3 Hospital arrival, processes of care, and outcomes for patients in the AITSC registry based on mortality status and composite characteristics (all patients in registry—including those with missing mortality status), including proportion of missing data for each variable

Variable	Alive N = 3835*	Dead N = 582*	Overall N = 4466*
Arrival on weekend (%)	1064 (27.8)	180 (30.9)	1261 (28.2)
Arrival day—missing (%)	3 (0.1)	0 (0.0)	3 (0.1)
Arrival time out of hours (%)	1824 (52.8)	277 (54.0)	2109 (47.2)
Arrival time—missing (%)	381 (9.9)	69 (11.8)	474 (10.6)
Red arrival trauma flag (%)	1927 (52.9)	468 (86.2)	2422 (54.2)
Trauma flag—missing (%)	193 (5.0)	39 (6.7)	240 (5.4)
Trauma call out occurred (%)	585 (35.2)	(7.4)	697 (15.6)
Trauma call out—missing (%)	2173 (56.6)	387 (66.5)	2568 (57.5)
Leader present on arrival (%)	934 (72.4)	103 (62.1)	1038 (23.2)
Leader present—missing (%)	2545 (66.3)	416 (71.5)	2992 (67.0)
Trauma bay ready (%)	1257 (98.6)	161 (98.2)	1435 (32.1)
Trauma bay ready—missing (%)	2560 (66.8)	418 (71.8)	3008 (67.4)
Referred from another hospital (%)	1767 (46.3)	319 (55.5)	2103 (47.1)
Referral status—missing (%)	22 (0.6)	7 (1.2)	29 (0.6)
Mode of arrival (%)			
Walking	11 (0.3)	1 (0.2)	13 (0.3)
Car/motorbike/auto-rickshaw	1412 (39.2)	112 (22.2)	1528 (34.2)
Ambulance	1659 (46.1)	316 (62.6)	2014 (45.1)
Police	481 (13.4)	72 (14.3)	556 (12.4)
Other	37 (1.0)	4 (0.8)	42 (0.9)
Mode—missing (%)	235 (6.1)	77 (13.2)	313 (7.0)
ED disposition (%)			
Discharged home	4 (0.1)	0 (0.0)	4 (0.1)
Ward	2230 (58.9)	133 (23.2)	2380 (53.3)
Operating theatre	741 (19.6)	122 (21.3)	867 (19.4)
ICU/HDU	635 (16.8)	207 (36.1)	868 (19.4)
Death	0 (0.0)	86 (15.0)	86 (1.9)
Other	179 (4.7)	26 (4.5)	206 (4.6)
ED Disp.—missing (%)	46 (1.2)	8 (1.4)	55 (1.2)
Mechanical ventilation on arrival (%)	448 (12.6)	316 (60.0)	783 (17.5)
Mech. vent—missing (%)	282 (7.4)	55 (9.5)	344 (7.7)
Operating procedure (OP) (%)	1665 (43.4)	252 (43.3)	1938 (43.4)
OP—missing (%)	2 (0.1)	0 (0.0)	4 (0.1)
CXR done (%)	2921 (76.4)	427 (74.3)	3385 (75.8)
CXR—missing (%)	13 (0.3)	7 (1.2)	23 (0.5)
CT scan done (%)	2808 (73.2)	418 (71.8)	3268 (73.2)
CT—missing (%)	0 (0.0)	0 (0.0)	2 (0.0)
Neurosurgical (NS) consultation (%)	2481 (64.9)	403 (69.7)	2922 (65.4)
NS Consult.—missing (%)	11 (0.3)	4 (0.7)	18 (0.4)
Blood transfusion (BT) (%)	453 (1.9)	143(24.9)	612 (3.7)
BT—missing (%)	18 (0.5)	7 (1.2)	28 (0.6)
In-hospital complication (Cx) (%)	29 (2.0)	65 (23.1)	95 (2.1)
Cx—missing (%)	2396 (62.5)	301 (51.7)	2720 (60.9)
Days on mech. vent; median (IQR) **	6 (3–10)	4 (2–8)	5 (2–10)
Days on mech. vent—missing (%)	0 (0.0)	1 (0.2)	23 (2.2)
ICU LOS (days); median (IQR) [#]	4 (2–8)	4.5 (2–9)	3 (2–8)
ICU LOS—missing (%)	0 (0.0)	0 (0.0)	15 (1.5)

Table 3 continued

Variable	Alive N = 3835*	Dead N = 582*	Overall N = 4466*
Hospital LOS (days); median (IQR)	7 (3–14)	3 (1–10)	7 (3–13)
Hosp. LOS—missing (%)	102 (2.7)	12 (2.1)	163 (3.6)

*Unless otherwise specified

**Alive $n = 594$, dead $n = 423$, overall $n = 1039$; #alive $n = 1197$, dead $n = 432$, overall $n = 1644$

between data collectors and investigators at each site, and among data collectors across institutions, would help support them and address specific concerns or problems. Frequently monitoring data collection procedures and conducting data quality audits, including checking inter-rater agreement of data, would help detect key issues [24, 25]. Hospital-wide education on the importance of data recording would be beneficial for data extraction and would emphasize the need for data quality to be the responsibility of all stakeholders.

A key limitation of this study was that it focused solely on data completeness. Other domains of data quality (e.g. accuracy, case capture) were not assessed. These areas are critical and can also impact the extent of missing data. For example, if cases which have most data points recorded are the only ones entered in the registry, observations may not appear to be missing but ‘capture’ is not complete.

Data accuracy impacted the assessment of variable completeness, where variable completeness was conditional—i.e. when assessing ‘times’ variables dependent on specific interventions occurring. The quantification of missing data in these cases was based on a denominator of cases coded for that event occurring and an underlying premise that this data coding was accurate. Similarly, this study considered that data were missing due to a failure to record clinical information, rather than a failure to measure and thereby promptly diagnose patients. While it is unlikely that core physiological data were not measured, it is impossible to properly verify this.

Complete case analysis (exclusion) was used for covariates in analyses of predictors of missing physiological data. In an attempt to minimize the limitations associated with complete case analysis, only those variables which were missing less than 20% of observations were included in the predictor model (as benchmarked by Ringdal et al.) [9]. Furthermore, the key predictor examined (mortality) was almost complete—this information was only missing in 49 (1.1%) of patients.

Lastly, a high proportion of the cases in the registry were derived from one hospital (58%), introducing

institutional bias when considering the whole registry. However, this study’s purpose was to explore the data quality and potential biases introduced when using the complete AITSC registry. It would be beneficial to conduct sub-analysis on each hospital to give more specific, and relevant, data to individual hospitals. This would enable them to better focus efforts on improving their internal records and identifying areas of improvement in their specific contexts.

Future investigation into other aspects of data quality is critical in the AITSC registry. Data should be collected and recorded as thoroughly as possible, and ideally, follow-up evaluation on data completeness would occur annually. However, missing data are ubiquitous and almost impossible to avoid. There remains a need for consensus on acceptable levels of missing data in trauma registries. The absence of observations does not necessarily invalidate a dataset, as long as appropriate measures to manage these are undertaken. This can be achieved by understanding the associations and patterns of missing data, and using suitable advanced statistical techniques such as multiple imputation, a statistical method that has been shown to lead to more valid conclusions when used for missing physiological variables in trauma registries [26].

Conclusion

This first assessment of data quality in a new trauma registry in India found that most variables had low rates of missing data, an important success. Hospital death was found to be associated with missing observations for key first in-hospital physiological variables, as were out-of-hours arrival and hospital of presentation. The completeness of physiological data would likely be improved by: the adoption of data collectors 24 hours a day; the implementation and encouragement of rigorous data recording and collection (as done in assault cases) by all staff; and the sharing of recording processes between sites.

Table 4 Predictors of missing first physiological data—univariable logistic regression (conducted on variables missing < 20% of data on all patients). ‘No missing phys. data’ represent none of SBP/HR/GCS/RR missing, and ‘Missing phys. data’ represent any of SBP/HR/GCS/RR missing

Variable	No missing phys. data (<i>n</i> = 3658)	Missing phys. data (<i>n</i> = 808)	Unadjusted OR (95% CI)	<i>p</i> value
Death (%)	399 (11.0)	183 (23.1)	2.4 (2.0–3.0)	<0.001
Age—mean (SD)	37.3 (15.5)	36.4 (14.8)	1.0 (1.0–1.0)	0.1
Gender/sex (%)				0.003
Male*	3078 (84.1)	712 (88.1)	1	
Female	580 (15.9)	95 (11.8)	0.7 (0.6–0.9)	
Other	0 (0.0)	1 (0.1)	–	
Arrival on weekend (%)	1045 (28.6)	216 (26.7)	0.9 (0.8–1.1)	0.3
Arrival time out of hours (%)	1683 (50.6)	426 (63.9)	1.7 (1.5–2.0)	<0.001
Hospital ID (%)				<0.001
1	2571 (70.3)	16 (2.0)	1	
2	208 (5.7)	481 (59.5)	371.6 (221.4–623.6)	
3	235 (6.4)	238 (29.5)	162.7 (96.4–274.7)	
4	644 (17.6)	73 (9.0)	18.2 (10.5–31.5)	
Red arrival trauma flag (%)	2141 (60.9)	281 (39.5)	0.4 (0.4–0.5)	<0.001
Place of injury (%)				<0.001
Home + residential Inst.*	656 (21.0)	126 (16.1)	1	
Road/highway	2123 (68.0)	492 (63.0)	1.2 (1.0–1.5)	
Trade/farm/construction	117 (3.8)	30 (3.8)	1.3 (0.9–2.1)	
Railway/trails	123 (3.9)	82 (10.5)	3.5 (2.5–4.9)	
Other	105 (3.4)	51 (6.5)	2.5 (1.7–3.7)	
Mechanism (%)				<0.001
RTI*	2067 (56.8)	418 (52.6)	1	
Other transport	131 (3.6)	74 (9.3)	2.8 (2.1–3.8)	
Falls	863 (23.7)	128 (16.1)	0.7 (0.6–0.9)	
Blunt force	127 (3.5)	95 (12.0)	3.7 (2.8–4.9)	
Sharp force	71 (2.0)	47 (5.9)	3.3 (2.2–4.8)	
Gunshot	49 (1.4)	20 (2.5)	2.0 (1.2–3.4)	
Other	332 (9.1)	12 (1.5)	0.2 (0.1–0.3)	
Dominant type of injury (%)				0.001
Blunt*	3396 (93.0)	721 (89.3)	1	
Penetrating	229 (6.3)	72 (8.9)	1.5 (1.1–2.0)	
Other	26 (0.7)	14 (1.7)	2.5 (1.3–4.9)	
Intent of injury (%)				<0.001
Unintentional*	3258 (89.6)	657 (83.1)	1	
Self-harm/suicide	51 (1.4)	6 (0.8)	0.6 (0.2–1.4)	
Assault/homicide	325 (8.9)	128 (16.2)	2.0 (1.6–2.4)	
Other	2 (0.1)	0 (0.0)	–	
Referred from other hospital (%)	1720 (47.2)	383 (48.1)	1.04 (0.9–1.2)	0.7
Mode of arrival (%)				0.7
Walking*	11 (0.3)	2 (0.3)	1	
Car/motorbike/auto	1299 (36.4)	229 (39.4)	1.0 (0.2–4.4)	
Ambulance	1747 (48.9)	267 (46.0)	0.8 (0.2–3.8)	
Police	478 (13.4)	78 (13.4)	0.9 (0.2–4.1)	
Other	37 (1.0)	5 (0.9)	0.7 (0.1–4.4)	
ED disposition (%)				<0.001
Discharged home*	3 (0.1)	1 (0.1)	1	
Ward	1895 (52.5)	485 (60.6)	0.8 (0.1–7.4)	

Table 4 continued

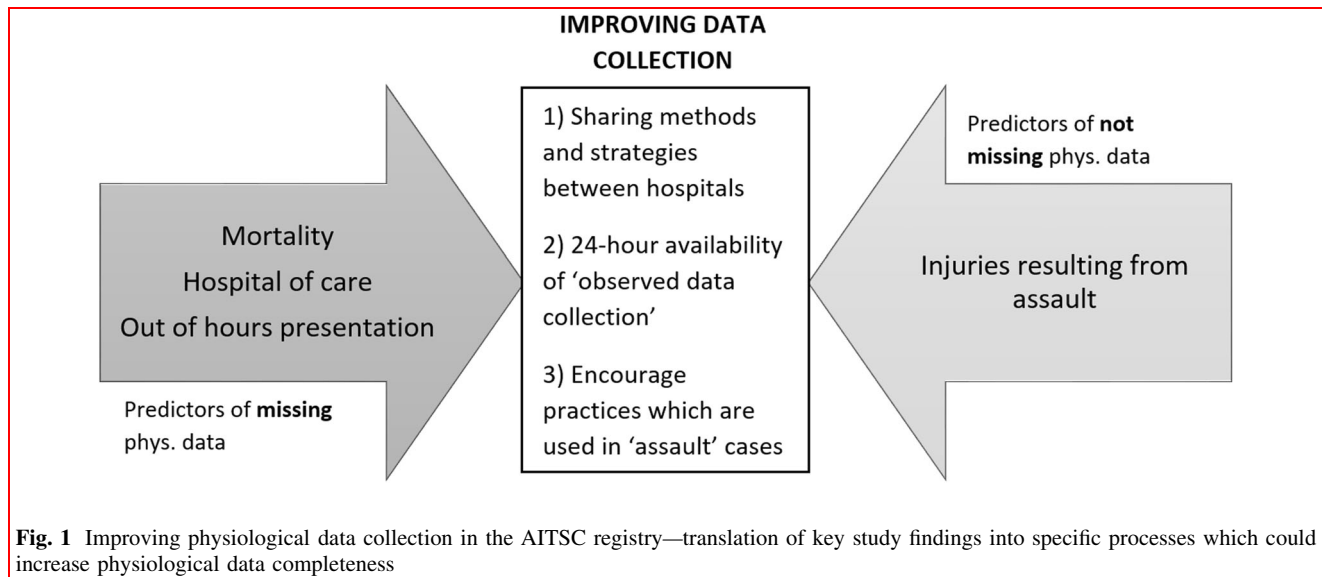
Variable	No missing phys. data (<i>n</i> = 3658)	Missing phys. data (<i>n</i> = 808)	Unadjusted OR (95% CI)	<i>p</i> value
Operating theatre	810 (22.4)	57 (7.1)	0.2 (0.0–2.1)	
ICU/HDU	803 (22.4)	65 (8.1)	0.2 (0.0–2.4)	
Death	29 (0.8)	57 (7.1)	5.9 (0.6–59.2)	
Other	70 (1.9)	136 (17.0)	5.8 (0.6–57.1)	
Mechanical ventilation on arrival (%)	663 (18.7)	120 (20.9)	1.2 (0.9–1.4)	0.2
Operating procedure (%)	1687 (46.2)	251 (31.1)	0.5 (0.4–0.6)	<0.001
CXR done (%)	3053 (83.7)	332 (41.8)	0.1 (0.1–0.2)	<0.001
CT scan done (%)	2799 (76.5)	469 (58.1)	0.4 (0.4–0.5)	<0.001
Neurosurgical consultation (%)	2461 (67.5)	461 (57.6)	0.7 (0.6–0.8)	<0.001
Blood transfusion (%)	507 (13.9)	105 (13.2)	0.9 (0.8–1.2)	0.6
Hospital LOS (days)—median (IQR)	7 (4–15)	3 (1–7)	0.94 (0.93–0.95)	<0.001

*Referent category for determination of odds ratios

Table 5 Predictors of missing first physiological data—adjusted results of stepwise multivariable logistic regression

Variable	Adjusted OR (95% CI)	<i>p</i> value
Death	1.4 (1.02–2.01)	0.04
Arrival time out of hours	1.7 (1.3–2.2)	<0.001
Hospital ID		
1*	1	
2	419.7 (201.6–873.8)	<0.001
3	134.4 (65.8–274.4)	<0.001
4	25.3 (11.8–53.9)	<0.001
Place of injury		
Home + residential institution*	1	
Road/highway	1.4 (0.8–2.3)	0.3
Trade/farm/construction	1.4 (0.7–3.0)	0.4
Railway/rails	1.7 (0.5–5.6)	0.4
Other	2.0 (1.02–3.81)	0.04
Mechanism		
RTI*	1	
Other transport	1.1 (0.3–3.9)	0.9
Falls	0.9 (0.5–1.6)	0.7
Blunt force	2.4 (1.1–5.3)	0.04
Sharp force	2.8 (1.1–7.2)	0.03
Gunshot	2.9 (0.9–8.7)	0.1
Other	1.5 (0.5–5.2)	0.5
Intent of injury		
Unintentional*	1	
Intentional: self-harm or suicide	0.5 (0.1–2.8)	0.5
Intentional: assault/homicide	0.3 (0.1–0.6)	0.001
Other		
CXR	0.7 (0.48–0.99)	0.04

*Referent category for determination of odds ratios



Funding The AITSC was funded by the Indian Government (Department of Science and Technology) and the Australian Government (Department of Industry, Innovation and Science), through the Australia-India Strategic Research Fund (AISRF), Grand Challenge Round 2, AISRF-GA12, Grant Number GCF0200130.

Compliance with ethical standards

Conflict of interest The authors declare they have no conflict of interest.

Ethics approval Ethics approval for this study was granted by the Alfred Hospital Ethics Committee (Project 245/17), the Monash University Human Research Ethics Committee (CF16/1814 - 2016000929), and each hospital site in India (AIIMS—IEC/NP-327/2013; LTMG—IEC/83/14; VS—approved 13/11/2013; GTB—approved 12/2/2015).

References

- Haagsma JA, Graetz N, Bolliger I, Naghavi M, Higashi H, Mullany EC et al (2016) The global burden of injury: incidence, mortality, disability-adjusted life years and time trends from the global burden of disease study 2013. *Inj Prev* 22(1):3–18
- Mock C, Juillard C, Brundage S, Goosen J, Joshipura M, editors (2009) WHO Guidelines for trauma quality improvement programmes Geneva: World Health Organisation 2009
- Dandona L, Dandona R, Kumar GA, Shukla DK, Paul VK, Balakrishnan K et al (2017) Nations within a nation: variations in epidemiological transition across the states of India, 1990–2016 in the global burden of disease Study. *Lancet* 390(10111):2437–2460
- Organization WH (2010) Injuries and violence the facts. 2010 [cited 2017 20/4/17]; Available from: http://apps.who.int/iris/bitstream/10665/44288/1/9789241599375_eng.pdf
- Moore L, Clark DE (2008) The value of trauma registries. *Injury* 39(6):686–695
- Haider AH, Saleem T, Leow JJ, Villegas CV, Kisat M, Schneider EB et al (2012) Influence of the National Trauma Data Bank on the study of trauma outcomes: is it time to set research best practices to further enhance its impact? *J Am Coll Surg* 214(5):756–768
- O'Reilly GM, Gabbe B, Moore L, Cameron PA (2016) Classifying, measuring and improving the quality of data in trauma registries: a review of the literature. *Injury* 47(3):559–567
- Porgo TV, Moore L, Tardif PA (2016) Evidence of data quality in trauma registries: a systematic review. *J Trauma Acute Care Surg* 80(4):648–658
- Ringdal KG, Lossius HM, Jones JM, Lauritsen JM, Coats TJ, Palmer CS et al (2011) Collecting core data in severely injured patients using a consensus trauma template: an international multicentre study. *Crit Care* 15(5):R237
- O'Reilly GM, Jolley DJ, Cameron PA, Gabbe B (2010) Missing in action: a case study of the application of methods for dealing with missing data to trauma system benchmarking. *Acad Emerg Med* 17(10):1122–1129
- Shivasabesan G, Mitra B, O'Reilly GM (2018) Missing data in trauma registries: a systematic review. *Injury* 49:1641–1647
- Lefering R (2002) Trauma Score Systems for Quality Assessment. *Eur J Trauma* 28(2):52–63
- MacLeod JBA, Kobusingye O, Frost C, Lett R, Kirya F, Shulman C (2003) A comparison of the kampala trauma score (KTS) with the revised trauma score (RTS), injury severity score (ISS) and the TRISS method in a ugandan trauma registry is equal performance achieved with fewer resources? *Eur J Trauma* 29:392–398
- O'Reilly GM, Cameron PA, Jolley DJ (2012) Which patients have missing data? An analysis of missingness in a trauma registry. *Injury* 43(11):1917–1923
- Collaboration A-ITS (2016) Australia-India Trauma Systems Collaboration. 2016 [cited 2017 20/4/17]; Available from: <http://aitsc.org/>
- Mitra B, Mathew J, Gupta A, Cameron P, O'Reilly G, Soni KD et al (2017) Protocol for a prospective observational study to improve prehospital notification of injured patients presenting to trauma centres in India. *BMJ Open* 7(7):e014073
- Synnot A, Karlsson A, Brichko L, Chee M, Fitzgerald M, Misra MC et al (2017) Prehospital notification for major trauma patients requiring emergency hospital transport: a systematic review. *J Evid Based Med* 10(3):212–221

18. Mitra B, Cameron PA, Fitzgerald MC, Bernard S, Moloney J, Varma D et al (2014) “After-hours” staffing of trauma centres and outcomes among patients presenting with acute traumatic coagulopathy. *Med J Aust* 201(10):588–591
19. O’Reilly GM, Joshipura M, Cameron PA, Gruen R (2013) Trauma registries in developing countries: a review of the published experience. *Injury* 44(6):713–721
20. O’Reilly GM, Cameron PA, Joshipura M (2012) Global trauma registry mapping: a scoping review. *Injury* 43(7):1148–1153
21. Newgard CD (2006) The validity of using multiple imputation for missing out-of-hospital data in a state trauma registry. *Acad Emerg Med* 13(3):314–324
22. Carr BG, Reilly PM, Schwab CW, Branas CC, Geiger J, Wiebe DJ (2011) Weekend and night outcomes in a statewide trauma system. *Arch Surg* 146(7):810–817
23. Ono Y, Ishida T, Iwasaki Y, Kawakami Y, Inokuchi R, Tase C et al (2015) The off-hour effect on trauma patients requiring subspecialty intervention at a community hospital in Japan: a retrospective cohort study. *Scand J Trauma Resusc Emerg Med* 23:20
24. Zehtabchi S, Nishijima DK, McKay MP, Clay Mann N (2011) Trauma registries: history, logistics, limitations, and contributions to emergency medicine research. *Acad Emerg Med* 18(6):637–643
25. Arts DG, De Keizer NF, Scheffer GJ (2002) Defining and improving data quality in medical registries: a literature review, case study, and generic framework. *J Am Med Inform Assoc* 9(6):600–611
26. Moore L, Hanley JA, Lavoie A, Turgeon A (2009) Evaluating the validity of multiple imputation for missing physiological data in the national trauma data bank. *J Emerg Trauma Shock* 2(2):73–79

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.