APPENDIX 2. STUDY CHARACTERISTICS ANS OUTCOMES

First author; year (country)	Design	Setting	Sample	Intervention characteristics	Findings (effects, psychometrics, feasibility)
Wolff and Bourke 2002 (Australia)[22]	Quasi-experimental (BA)	ED (n=1)	Reviewed patient medical records (n=20050)	Clinical incident reporting in addition to standardized screening of medical records on AEs: retrospective screening of medical records for EAs and assessment using the AE severity scale by a clinical risk manager; creation of weekly AE reports for the unit management describing types and severity of events and improvement actions; presentation of an aggregated quarterly report detailing actions taken and AE rates to the hospital's main quality improvement committee; AEs with substantial impact on the hospital were discussed by a surveillance committee who made recommendations for action; the medical staff group reviewed recommendations which were implemented following acceptance	Effect: AE relative risk reduction over two years=85.3% (95% CI, 62.7% to 100%), p<0.0001
Hendrie et al 2007 (Australia)[23]	Non-experimental	ED (n=1)	Case histories of patients (n=3332)	AE screening method: case records were screened for EAs by an experienced registrar; EA classification was based on a validated method using a 104-item data collection instrument with criteria on management causation, outcome and EA preventability; classifications were analyzed statistically	Reliability: inter-rater agreement on classification of AEs, ƙ=0.15*; on judgments about management causation, ƙ=0.50, and on preventability, ƙ=0.58 Feasibility: time to detect an adverse event was substantial
Patterson et al 2012 (United States of America)[38]	Non-experimental	EMS (n=NR)	Patient care reports (n=250)	Identification and severity rating method for AEs: a consensus definition of an adverse event in EMS; an index for rating AE severity ranging	Face validity: method developed by a panel of EMS medical director physicians (n=5) Reliability: multi-rater agreement on classification of AEs: ƙ=0.24 (95% CI 0.19 – 0.29)
Patterson et al 2014 (United States of America)[24]	Non-experimental	HEMS (n=NR)	Expert clinicians (n=10)	Identification and severity rating method for AEs: a consensus definition of an adverse event in HEMS: a consensus definition of an AE and four-step protocol for AE detection: 1) a trigger tool to operationalise AE detection in patient care reports using key words or phrases contained within a PCR that have a high probability of being linked to patient harm, 2) a method for rating AE severity, 3) a method for rating proximal cause	Validity: S-CVI (the average I-CVI) for trigger tool items: 0.94 Validity: S-CVI (the average I-CVI) for proximal cause items: 0.95 Validity: S-CVI (the average I-CVI) for severity items: 0.95 Face validity: draft framework developed by a panel of experienced clinicians both in emergency medicine and HEMS
Clunas et al 2009	Non-	ED (n=1)	Reviewed patient	Audit of all deaths that occurred within 48 hours of ED presentation	Usability: a major external hospital review was recommended in

(Australia)[25]	experimental		deaths (n=303)	in addition to auditing all deaths that occurred in the ED itself	5% of deaths within the ED and 5% in deaths within 48 hours of ED presentation
					Usability: internal review was recommended in 1.3% of deaths within the ED and 3.5% in deaths within 48 hours of ED presentation
					Usability: 25% of the death cases within the ED and resulting in an external review were not identified by the hospital IIMS;
					Usability: 36% of the death cases within 48 hours of ED presentation and resulting in an external review were not identified by the hospital IIMS
van Noord et al 2010 (the Netherlands)[26]	Non-experimental	ED (n=31)	Closed and settled claim files (n=47)	Retrospective RCA method (PRISMA-medical): incident description by causal tree based on the information gathered from the claim files; classification of root-causes according a modified model; development of a classification-action matrix	Reliability: inter-rater agreement on classification of root-causes, k=0.78
					Validity: risk managers confirmed that identified root-causes are commonly seen
					Feasibility: delay between incident occurrence, detection and reporting made it difficult to draw firm conclusions from RCAs
					Feasibility: RCAs were time consuming
Patterson et al 2010 (United States of America)[27]	Non-experimental	EMS agencies (n=3)	EMTs and paramedics (n=71)	Safety Attitudes Questionnaire (EMS-SAQ): modified version of de ICU-SAQ; 30 items; 5-point Likert-type scale ranging from 'strongly agree' to 'strongly disagree'; 6 domains	Feasibility: response rate=85%
					Feasibility: respondents who missed or skipped items=27%
					Feasibility: positive feedback on instrument utility from EMS chief administrators
					Validity: CSDFr=1.2; CFI=.95; NNFI= .92
					Reliability: Cronbach's α for each domain varied between 0.65 – 0.88
Patterson et al 2010 (United States of America)[28]	Non-experimental	EMS agencies (n=61)	Care providers (n=1595)	Safety Attitudes Questionnaire (EMS-SAQ): 60 items; modified version of the validated ICU-SAQ; 5-point Likert-type scale ranging from 'strongly agree' to 'strongly disagree'; 6 domains; administered on paper forms and via internet; anonymised and voluntary	Reliability: Cronbach's α for each domain varied between 0.68 – 0.83
Flowerdew et al 2011 (United Kingdom)[29]	Non-experimental	ED (n=2)	U	Observational physician skill assessment: behavioural marker system to assess 12 emergency medicine–specific nontechnical skills required by emergency care physicians; 9-point rating scale to assess skills divided into 'unacceptable', 'acceptable' and 'exemplary'	Validity: provisional assessment tool was developed according to published literature and curricula
					Validity and feasibility: staff interviews and field observations were held to determine completeness of skill list and whether

					skills were observable
					Validity: reported content validity†
Jaynes et al 2013 (United States of America)[30]	Non-experimental	EMS (n=NR)	EMS care providers (n=380)	EMS and HEMS working relationship satisfaction questionnaire: measures overall EMS satisfaction with the quality of EMS/HEMS patient care coordination; 22 items; 5-point Likert scale ranging from 'never/very poor' to 'always/very good'	Validity: providers, medical directors and administrators (n=12) defined working relationship activities; generated items and reviewed the questionnaire content
					Reliability: Cronbach's α for each domain varied between 0.85-0.88
Evans et al 2007 (Australia)[31]	Quasi experimental (NEG)	ED (n=4)	Attendances (n=66669) in EDs (n=2) with intervention versus attendances (n=78264) in EDs (n=2) with usual procedure	Incident reporting program: display of posters and manuals in clinical areas describing what types of incidents staff should report; informing staff on the possibility to report anonymously and the importance of reporting near-misses; replacement of the three-page report form (usual procedure) by one-page report form; introduction of a call service enabling staff to report an incident at any time; initial assessment of incident reports by the patient safety manager; anonymous reports were validated and managed only by the patient safety manager, and identified reports were validated and managed by medical nursing unit heads; newsletters with statistics, de-identified RCA findings and recommendations were distributed to all ED staff, for example at scheduled departmental meetings; individual feedback was provided for serious incidents	Effect: overall increase of 39.5 incident reports per 10000 ED attendances (95% CI 17.0 to 62.0; p<0.001)
					Effect: increase of 9.5 incident reports per 10000 ED attendances by ED doctors ((95% CI 2.2 – 16.8; p=0.001)
Zwart et al 2011 (Netherlands)[32]	Quasi experimental (NEG)	GPOHS (n=3)	GPOHS with intervention (n=1); GPOHS with usual procedure (n=2; control)	Local incident-reporting procedure: a local multidisciplinary committee was trained to screen and analyze incident reports within two weeks instead of a central assessment of incidents every two months performed by an advisory committee of the board of directors of the GP OHSs collaboration (usual procedure); the local committee was responsible for feedback to reporters and to the organization, and for development of improvement measures when appropriate	Effect: number of reported incidents in intervention GP OHS increased 16-fold compared with the control GP OHSs
					Effect: the type of incidents reported did not alter compared with the control GP OHSs
					Effect/feasibility: improvements were implemented in a shorter time frame in intervention GP OHS compared to the control GP OHSs
					Feasibility: Implementation of a LIRP was associated with extra costs for administration and analysis
Reznek and Barton 2014	Quasi experimental	ED (n=1)	Incident reports (n=314)	Standardized non-punitive peer review process of incident reports: incidents were submitted electronically via the hospital incident	Effect: increase of monthly frequencies of incident reports by ED practitioners; p=0.0019‡ and p=0.0025§

(United States of America)[33]	(ITS)			reporting system, or directly via electronic, written or verbal communications with ED leadership; for each report, a screening review was performed by the ED clinical director; if errors or near misses could not be excluded, the case report progressed to a full peer review evaluation; in monthly peer review meetings a special committee and attending staff reviewed the de-identified medical record and responses of involved practitioners, and voted on the presence or absence of errors; practitioners were regularly reminded that peer review was undertaken to guide quality improvement and not for punitive purposes	Effect: increase of monthly frequencies of reports by non-ED practitioners within the hospital; p< 0.0001
Schull et al 2011 (Canada)[39]	Non-experimental	ED (n=NR)	Candidate indicators (n=170)	Patient safety indicators	Feasibility: four safety indicators are classified as feasible‖; two indicators are classified as feasible if quality of data in current data fields is enhanced
					Face validity: indicators are assessed and selected by experts (n=21)
Pham et al 2011 (United States of America)[34]	Non-experimental	ED (n=1)	Patients seen in the ED within 72 hours of prior visit (n=6858) and patients not seen in the ED within 72 hours (n=211321)	Safety indicator: patient returns to an ED within 72 hours of their initial visit	Effect: total recourses utilized of patients seen within 72 hours, mean ± SE=5.0 ± 0.08 versus patients not seen within 72 hours, mean ± SE=5.5 ± 0.10, p<0.05
					Effect: level I triage acuity of patients seen within 72 hours, %=17 (95% CI 15 - 19) versus patients not seen within 72 hours, %=20 (95% CI 19 - 22), p<0.05
					Effect: admission rate of patients seen within 72 hours, %=13 (95% CI 12 – 15) versus patients not seen within 72 hours. %=13 (95% CI 13 – 14)
Jones et al 2013 (United States of America)[35]	Non-experimental	ED (n=2)	Care providers (n=60)	Teamwork training on patient safety (TeamSTEPPS): course in a period of 4 weeks educating employees on how to communicate safety concerns, report errors and system failures; use of video vignettes illustrating good communication and barriers to communication that facilitated group discussion; use of handouts with communication techniques that participants practiced both in class and after the sessions	Effect: no statistical difference in care provider perception of the culture of safety in the ED pre and post training (p>0.05)¶
Patterson et al 2013 (United States of	Quasi experimental	Paediatric ED (n=1)	Care providers (n=151)	Multidisciplinary simulation-based training: a two-day program; review of information on the magnitude of risk from medical error,	Effect: increase in ED personnel safety knowledge from baseline to re-evaluation** (p< 0.001)

America)[36]	(ITS)			error theory and principles of CRM prior to training; introduction of techniques to prevent medical error, improve critical communications, increase situation awareness, develop resilience, and improve sharing of mental models and closed loop communication with mini-lectures; presentation of five simulations and team participation in reproducible simulated scenarios, followed by immediate video assisted debriefing	Effect: increase in overall SAQ attitudes median score from baseline to re-evaluation** (p< 0.001)
					Effect: attitude changes seen following the intervention were not significantly diminished at time of re-evaluation** (p>0.017)
					Feasibility: time required in initial simulation training condensed from 12 to 4 h.
Shaw et al 2006 (United States of America)[37]	Non-experimental	Paediatric ED (n=1)	Staff (n=99)	Unit-based Patient Safety WalkroundsUnit-based Patient Safety Walkrounds: scheduled 30-minute rounds; performed twice a months by a physician and two staff nurses; data collection on two clinical improvement topics followed by a general discussion in the conference room with ED staff; a patient safety committee reviewed recorded results and incident reports; ED staff is informed via a short summary containing salient results, celebration points and areas for improvement	Effect: 44% increase of medication near-miss incident reports over one year compared with the two years before the program was implemented.
					Effect: 23% overall increase in hand hygiene compliance

ED=Emergency Department; BA=Before After; AE=Adverse Event; NR=Not Reported; EMS=Emergency Medical Service; HEMS=Helicopter Emergency Medical Service; S-CVI: Scale Content Validity Index; I-CVI=Item Content Validity Index; IIMS=Incident Investigation and Monitoring System; RCA=Root Cause Analysis; PRISMA=Prevention and Recovery Information System for Monitoring and Analysis; EMT=Emergency Medical Technician; EMS-SAQ=Emergency Medical Service Safety Attitudes Questionnaire; CSDFr=Chi-Square/Degrees of Freedom ratio; CFI=Comparative Fit Index; NNFI=Non-Normed Index; NEG=Non Equivalent Group; GPOHS=General Practice Out-of-Hours Service; LIRP=Local Incident Reporting Procedure; ITS=Interrupted Time Series; IC-SAQ=Intensive Care Safety Attitudes Questionnaire; RMSEA=Root Mean Squared Error of Approximation; CRM=Crew Resource Management.

* After discussion and reassignment the kappa for intra-observer agreement was 0.82.

† Content validity was evaluated using a survey in which experts were asked to rate 36 statements on exemplary behavioural marker statements on a scale of 1 to 5. 75% of items achieved the recommended content validity index greater than 0.75.

‡ ED practitioners directly involved in the care of the patient when the perceived incident occurred.

§ ED practitioners not directly involved in the care of the patient when the perceived incident occurred compared with a control group of practitioners from outside the hospital.

‖ Feasibility of measuring indicator using current administrative data sets.

¶ Based on combined ED results. Care giver perception of patient safety culture in the ED were measured with the Agency for Healthcare Research and Quality's (AHRQ) patient safety culture survey (PSCS) before and after the training.

** Pre- and post (approximately 6 months) training assessment of safety attitudes and knowledge by each individual participant with the SAQ Teamwork and Safety Climate version.