Interventions to Improve the Breaking of Bad or Difficult News by Physicians, Medical Students, and Interns/Residents

DOI: 10.1097/ACM.0000000000002308

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
Academic Medicine

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
Academic Medicine

DOI: 10.1097/ACM.0000000000002308

Interventions to Improve the Breaking of Bad or Difficult News by Physicians, Medical Students, and Interns/Residents: A Systematic Review and Meta-Analysis

Judith Johnson, PhD, ClinPsyD, and Maria Panagioti, PhD

J. Johnson is joint lecturer, School of Psychology, University of Leeds, Leeds, and Bradford Institute for Health Research, Bradford, United Kingdom; ORCID: https://orcid.org/0000-0003-0431-013X.

M. Panagioti is senior research fellow, National Institute for Health Research School for Primary Care Research, Manchester Academic Health Science Centre, University of Manchester, Manchester, United Kingdom; ORCID: https://orcid.org/0000-0002-7153-5745.

Correspondence should be addressed to Judith Johnson, School of Psychology, University of Leeds, Lifton Place, Leeds LS29JT, UK; telephone: +44 (0)1133435719; e-mail: j.johnson@leeds.ac.uk; Twitter: @DrJTJohnson.

Supplemental digital content for this article is available at http://links.lww.com/ACADMED/A564.

Funding/Support: This research was supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) Yorkshire and Humber (www.clahrc-yh.nihr.ac.uk; reference: NIHR IS-CLA-0113-10020). In addition, the NIHR School for Primary Care Research and the NIHR Greater Manchester Patient Safety Translational Research Centre (Greater Manchester PSTRC) supported the time and facilities of M.P. The research team members were independent from the funding agency.

Other disclosures: None reported.

Ethical approval: Reported as not applicable.
Disclaimer: The views expressed in this report are those of the authors, and not necessarily those of the National Health Service, the NIHR, or the Department of Health and Social Care. The funders had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; and the preparation, review, or approval of the manuscript.
Abstract

Purpose

To assess the effectiveness of news delivery interventions to improve observer-rated skills, physician confidence, and patient-reported depression/anxiety.

Method

MEDLINE, EMBASE, CINAHL, PsycINFO, and Cochrane Register of Controlled Trials databases were searched from inception to September 5, 2016 (updated February 2017). Eligible studies included randomized controlled trials (RCTs), non-RCTs, and controlled before-after studies of interventions to improve the communication of bad or difficult news by physicians, medical students, and residents/interns. The EPOC risk of bias tool was used to conduct a risk of bias assessment. Main and secondary meta-analyses examined the effectiveness of the identified interventions for improving observer-rated news delivery skills and for improving physician confidence in delivering news and patient-reported depression/anxiety, respectively.

Results

Seventeen studies were included in the systematic review and meta-analysis, including 19 independent comparisons on 1,322 participants and 9 independent comparisons on 985 participants for the main and secondary (physician confidence) analyses (mean [SD] age = 35 [7] years; 46% male), respectively. Interventions were associated with large, significant improvements in observer-rated news delivery skills (19 comparisons: standardized mean difference [SMD] = 0.74, 95% CI = 0.47–1.01) and moderate, significant improvements in physician confidence (9 comparisons: SMD = 0.60, 95% CI = 0.26–0.95). One study reported intervention effects on patient-reported depression/anxiety. The risk of bias findings did not influence the significance of the results.
Conclusions

Interventions are effective for improving news delivery and physician confidence. Further research is needed to test the impact of interventions on patient outcomes and determine optimal components and length.
Physicians frequently break bad or difficult news to patients. Research into the delivery of bad or difficult news originated in oncology services, where it was found that news communication practices can have a strong and lasting impact on patients’ subsequent symptoms of depression, anxiety, and post-traumatic stress disorder and can influence their treatment choices.\(^1\)-\(^5\) Studies have since investigated the delivery of bad and difficult news in a range of health care settings, including pediatrics,\(^6\) emergency medicine,\(^7,8\) and obstetrics services.\(^9,10\) Together, this body of research has identified several challenges that physicians may face in scenarios where they have to deliver bad or difficult news, such as when the news occurs suddenly and without warning (e.g., in emergency settings), when there is limited time for physicians to prepare to deliver the news (e.g., in obstetric ultrasound settings), or when the news itself is uncertain because the diagnosis or prognosis is unclear. It has also highlighted the negative impact that these events can have on the physicians involved, including increased stress and burnout.\(^8,11\)

A range of interventions that aim to improve the communication skills and confidence of physicians in delivering bad or difficult news have been described. These interventions vary in length and format, but share some similar components; for example, most include elements of didactic teaching, role-playing or simulation,\(^9\) group discussions,\(^12\) or the viewing of instructional videos.\(^13\) The interventions are often designed to enhance fidelity to existing guiding frameworks for bad or difficult news delivery. SPIKES\(^14\) is the most widely used of these frameworks; it proposes six steps, from which the acronym is derived, to improve news delivery events. These steps are (1) setting up the interview, (2) assessing the patient’s perception of the situation, (3) obtaining the patient’s invitation to deliver the news, (4) giving knowledge and information to the patient, (5) addressing the patient’s emotions empathically, and (6) providing a summary and discussing prognosis and treatment options.\(^14\)

Similarly, the more recently proposed SHARE\(^15\) protocol suggests that health care staff
should follow four steps, which taken together form the acronym. These steps are (1) create a supportive environment, (2) consider how to deliver the news, (3) discuss additional information that patients would like to know, and (4) provide reassurance and emotional support.\textsuperscript{15}

Divergent methods have been used to evaluate the effectiveness of these interventions, but the most common practice has been the rating of a participant’s news delivery skills in a simulated exercise by an observer.\textsuperscript{12,13,16} Other practices have included measuring physician confidence in breaking bad news\textsuperscript{17} and gathering information on the patient’s experience.\textsuperscript{15}

In the United Kingdom, National Institute for Health and Care Excellence (NICE) guidelines for miscarriages state that staff should be trained in delivering bad or difficult news,\textsuperscript{18} and the NICE list of quality statements state all National Health Service staff should have competency in communication skills.\textsuperscript{19} Similarly, the US Preventive Services Task Force recommendations for prostate and lung cancer screenings emphasize the importance of communicating with patients.\textsuperscript{20,21} However, news delivery interventions are not routinely implemented and patients report poor satisfaction with this aspect of health care.\textsuperscript{22,23} Thus, there is a need to understand whether formal training interventions to improve the delivery of bad or difficult news by health care professionals are effective.

To address this need, we conducted the first systematic review and meta-analysis of studies that have evaluated bad or difficult news delivery interventions. Our first objective was to assess the effectiveness of these interventions for improving news delivery skills in physicians, medical students, or residents/interns, as rated by an observer such as a researcher, instructor, or standardized patient (an individual who is trained to role play a patient in a standardized format). Our second objective was to assess the effectiveness of these interventions for improving physician confidence in news delivery. Our third objective was to assess the impact of the interventions on improving patient-reported depression and/or
anxiety. Our fourth objective was to compare observer ratings of news delivery skills for SPIKES-based interventions with observer ratings of news delivery skills for interventions based on any other or no framework.

Method

Our review followed the PRISMA statement (see eTable 1 in Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/A564). We prospectively registered our protocol (see eMethods 1 in Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/A564) with PROSPERO (CRD42016045892).

Search strategy and data sources

We searched five electronic bibliographic databases—MEDLINE, EMBASE, CINAHL, PsycINFO, and Cochrane Register of Controlled Trials—from inception to September 5, 2016. We searched the same five databases in February 2017 to update our results. Our search strategy included combinations of two key blocks of terms (bad or difficult news and intervention) using a combination of medical subject headings (MeSH terms) and text words (see eMethods 2 in Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/A564). We scanned the reference lists of eligible studies to identify additional potentially includable studies.

Eligibility criteria

Studies that met the criteria in each of the areas listed below were eligible for inclusion in the systematic review and meta-analysis:

- Population: Studies that looked at physicians, medical students, or residents/interns working in any primary, secondary, or intensive clinical health care settings were included.
- Intervention: Studies with interventions that were designed to improve the delivery of bad or difficult news were included. These mainly included person-
directed training programs for improving news communication skills using real or simulated scenarios with or without an underlying theoretical framework (e.g., SPIKES).

- **Comparison**: Studies with any type of control group (e.g., no intervention, alternative intervention, wait list) were included.

- **Outcomes**: Studies in which the outcome of news delivery skill as rated by an observer (e.g., researcher, standardized patient) was measured were included. Where more than one news delivery skill metric was reported, we used the scale pertaining most closely to overall news delivery skill. Studies in which the outcomes of physician confidence in bad or difficult news delivery or patient-reported depression and/or anxiety were measured were also included.

- **Design**: Studies that had quantitative intervention designs, such as randomized controlled trials (RCTs), non-randomized controlled trials (nRCTs), controlled before-after studies (CBAs), and interrupted time series design studies (ITSs), as outlined in the Cochrane handbook, were included.\(^{25}\)

- **Context**: Studies conducted in any health care or educational setting that were English language and in peer-reviewed journals were included.

**Exclusion criteria**

Studies of communication interventions that did not assess outcomes relevant to breaking bad or difficult news were excluded. Studies that tested news delivery skills for “good,” “neutral,” or a range of news types, rather than focusing on “difficult,” “bad,” or “negative” news were also excluded, as were non-English language papers and gray literature.

**Study selection**

We exported the search results from each database and reference list scanning to Endnote version 7.1 (Clarivate Analytics, Philadelphia, Pennsylvania) and removed duplicates. We
undertook study selection in two stages. In the first stage, we screened the titles and abstracts of identified studies, and in the second stage, we accessed the full texts of the remaining studies to further screen them according to the eligibility criteria. We (J.J. and M.P.) independently reviewed a proportion of the titles and abstracts (10%) and our interrater reliability was high (k = 1.0). J.J. screened the remaining titles and abstracts. J.J. and M.P. completed full-text screenings for all eligible articles. We resolved any disagreements (which were rare) by discussion.

**Data extraction**

We devised a data extraction form in Excel 2010 (Microsoft Inc., Redmond, Washington) and piloted it on five randomly selected studies. We extracted quantitative data for the meta-analysis to a separate Excel file. J.J. undertook data extraction, with 10% of articles independently extracted by M.P. as well. We resolved any disagreements by discussion. We extracted the following descriptive information from eligible studies:

- **Study**: research design, recruitment method, and content of the control condition;
- **Participants**: sample size, age, gender, discipline, and setting;
- **Intervention**: content of the intervention, delivery format (group or individual), theoretical underpinning, and measurement time points; and
- **Outcomes**: observer-rated news delivery skill, physician confidence in bad or difficult news delivery, and patient-reported depression and/or anxiety.

**Risk of bias assessment**

We used the Effective Practice and Organisation of Care (EPOC) risk of bias tool\(^26\) to conduct a critical appraisal, as it is appropriate for use across all different types of intervention designs, as described in the Cochrane handbook,\(^25\) including RCTs, nRCTs, CBAs, and ITSs. The EPOC tool contains nine standardized criteria, which are each rated on a three-point scale (where 0 = low risk, 1 = unclear risk, and 2 = high risk). We considered
studies that received a low risk score across at least six of the nine criteria to be less susceptible to risk of bias.

**Data analysis**

We synthesized our results using meta-analysis. We used standardized mean difference (SMD) as the effect size to pool results across the studies. We calculated SMDs and associated confidence intervals (CIs) for the news delivery outcomes of all the studies, and we pooled the results using the metaan command in Stata 15 (StataCorp LLC, College Station, Texas).\(^{27}\) The main meta-analysis examined the effectiveness of the identified interventions for improving news delivery skills as rated by an observer (e.g., researcher, standardized patient). Secondary meta-analyses examined the effectiveness of the interventions for improving physician confidence in delivering bad or difficult news and for improving patient-reported depression and/or anxiety. When studies collected data at more than one follow-up assessment point, we used the first assessment point following the intervention. Pre-specified subgroup analyses\(^{28}\) tested the effectiveness of interventions (via observer ratings of news delivery skills) for whether they were based on the SPIKES framework versus any other or no framework. We performed sensitivity analyses to examine whether results were maintained when only those studies with low risk of bias scores were included.

We used a random effects model to account for heterogeneity in all analyses. We assessed heterogeneity with the \(I^2\) statistic and associated CIs. Conventionally, \(I^2\) values of 25%, 50%, and 75% indicate low, moderate, and high heterogeneity, respectively.\(^{29}\) We inspected the symmetry of funnel plots (asymmetry indicates publication bias), and we conducted the Egger’s test of small-study effects to quantify observations in the funnel plots.\(^{30}\) We constructed funnel plots using the metafunnel command\(^{31}\) and performed Egger’s test using the metabias command in Stata 15.\(^{32}\)
Results

Our search strategy yielded 3,206 records (see Figure 1). An additional 41 records were identified via reference list scanning. Once we removed duplicates, we screened the titles and abstracts of the remaining 2,270 studies; 71 of these were retained for full-text screening. (For a list of the excluded studies, see eMethods 3 in Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/A564.) Seventeen studies were included in the systematic review and meta-analysis, with one study being reported in two papers, which we include separately in Appendix 1 but otherwise treat as a single study in the analyses and reporting. When studies were eligible for the systematic review but the necessary outcome data for the meta-analysis were not reported, we contacted the study authors for this information. However, there were two papers that were eligible for the review but were not ultimately included because we were not able to gather the information needed for the meta-analysis.

Characteristics of the studies and participants

Study characteristics are presented in Appendix 1. These included 19 independent comparisons on 1,322 participants for the main analysis of observer-rated news delivery skill and 9 independent comparisons on 985 participants for the secondary analysis of physician-reported confidence in news delivery (mean [SD] age = 35 [7] years). There was only one study reporting intervention effects on patient-reported depression and/or anxiety. The majority of studies took place in the United States (n = 10). One study was conducted in each of the following: Hong Kong, Israel, Japan, Puerto Rico, the United Kingdom, and Belgium (this study was reported on in two papers). One study was conducted in both Germany and The Netherlands. Overall, the proportion of male to female participants was approximately equal (46% male).
Five studies were conducted with medical students, 1 study was conducted with medical interns, and 11 studies were conducted with physicians or residents. Most studies used a convenience sampling method, and the main eligibility criteria were being a medical student, intern/resident or physician working or training in a specific setting (data not shown). Studies were conducted in a range of settings, including oncology (n = 3), pediatrics (n = 3), university hospitals (n = 3), palliative care (n = 2), obstetrics (n = 2), primary care (n = 1), hospital outpatient departments (n = 1), and intensive care (n = 1). The settings of two studies were unclear. Studies measured the outcome of interventions using a variety of tools, but 16 measured news delivery skills using an observer-rated measure, while 6 measured participant-reported news delivery self-confidence and 1 measured patient-reported depression and/or anxiety (see eTable 2 in Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/A564).

**Characteristics of the interventions**

A range of theoretical models were used in the studies, but the single most commonly used model was SPIKES, which was cited in seven studies (see Appendix 1). Interventions varied in length; whereas some were delivered as a stand-alone intervention (n = 9), others were delivered in the context of a broader medical or communications training course (n = 8). The specific length of interventions delivered as part of wider training courses was not stated in all studies. However, the majority described interventions that lasted less than one working day, with the shortest intervention being 10 minutes. All were participant-directed interventions, and the majority involved a simulation exercise or role-play with feedback. None of the studies tested the impact of making service-level changes (e.g., the introduction of patient information leaflets) on bad or difficult news delivery. Two studies had more than one intervention condition to investigate the impact of the number and type of intervention components.
**Risk of bias characteristics**

For risk of bias characteristics, see eFigure 1 (in Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/A564). Seven studies were RCTs, 2 were nRCTs where participants were allowed to switch their allocation after randomization if they were unable to attend their allocated condition, and 8 were CBAs. Five studies were scored as low on six or seven of the nine risk of bias criteria, indicating low overall risk of bias; 8 studies were scored as low on four or five of the criteria, indicating moderate overall risk of bias; and 5 studies were scored as low on two or three of the criteria, indicating high overall risk of bias.

**Main and secondary meta-analyses**

Interventions were associated with large, significant improvements in observer-rated news delivery skills (19 comparisons: SMD = 0.74, 95% CI = 0.47 to 1.01; $I^2 = 84\%$, 95% CI = 40% to 98%; Figure 2), and moderate, significant improvements in physician confidence in news delivery (9 comparisons: SMD = 0.60, 95% CI = 0.26 to 0.95; $I^2 = 88\%$, 95% CI = 43% to 99%; Figure 3). However, the heterogeneity between the studies was high in both analyses (as indicated by the $I^2$ statistic [shown above]).

As only one study investigated patient-reported depression and/or anxiety, a meta-analysis of this outcome was not possible; however, we did calculate the individual effect size in this study. The intervention group was associated with small, significant reductions in patient-reported depression and/or anxiety symptoms, as compared with the control group (SMD = –0.17, 95% CI = –0.33 to –0.01).

**Subgroup analyses for use of SPIKES framework**

The pooled effect size of interventions on observer-rated news delivery skills was very large and significant across the studies that used SPIKES (7 comparisons: SMD = 1.14, 95% CI = 0.63 to 1.66; $I^2 = 92\%$, 95% CI = 40% to 100%) and was moderate and significant across the studies that used any other or no framework (12 comparisons: SMD = 0.52, 95% CI = 0.21 to
The effects of SPIKES-based interventions were significantly larger than those of other interventions (Q = 4.23, P = .04).

The pooled effect size of interventions on physician confidence was large and significant across the studies that used SPIKES (5 comparisons: SMD = 0.70, 95% CI = 0.14 to 1.26; I² = 94%, 95% CI = 40% to 100%) and was medium and significant across the studies that used any other or no framework (4 comparisons: SMD = 0.47; 95% CI = 0.25 to 0.69; I² = 0%, 95% CI = 0% to 1%; see eFigure 3 in Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/A564). However, this difference was not significant (Q = 0.98, P = .69).

Sensitivity analysis

The treatment effect derived by studies at lower overall risk of bias (i.e., those scoring low on six or more of the nine individual risk of bias criteria) showed a very large effect size for observer-rated news delivery skills (5 comparisons: SMD = 1.18, 95% CI = 0.40 to 1.96; I² = 92%, 95% CI = 40% to 100%; see eFigure 4 in Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/A564), suggesting the significance of results was not explained by risk of bias findings. Only one study scoring low on six or more of the risk of bias criteria included an item on physician confidence, so no sensitivity analysis could be performed for this outcome.

Small-study bias

No evidence of funnel plot asymmetry, which might indicate publication bias, was identified for the main analysis focused on observer-rated news delivery skills (Egger’s test: regression intercept = −0.24, in Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/A564).
Discussion

Summary of the main findings

This systematic review and meta-analysis found that interventions for improving the delivery of bad or difficult news were associated with significant, large improvements in observer-rated news delivery skills, and significant, moderate improvements in physician confidence in news delivery. The evidence base was too limited to test whether interventions improved patient-reported depression and/or anxiety. Interventions based on the SPIKES framework for bad or difficult news delivery were associated with very large and significant improvements in observer-rated news delivery skills, whereas interventions based on any other or no framework were associated with moderate and significant improvements.

Strengths and limitations

Strengths of this systematic review and meta-analysis included a pre-specified protocol registered on the PROSPERO database and inclusion of the EPOC tool to estimate risk of bias. We also endeavored to address the two most important threats of meta-analysis, which are publication bias and heterogeneity. We assessed publication bias using formal statistical tests, with no such bias indicators being observed in the performed tests.\(^{50}\) A limitation of the meta-analysis was that we did find the heterogeneity between the studies to be high because the included studies differed considerably on a range of factors, such as the length of follow-up, outcome measures, and the content and length of the intervention. Most interventions lasted a day or less, with one being only 10 minutes.\(^{9}\) To account for this heterogeneity, we applied random effects models in all analyses and conducted pre-specified subgroup and sensitivity analyses (within the limits of power).\(^{51,52}\) The meta-analysis was also limited by its focus on outcomes immediately post-intervention, which prevented us from drawing conclusions about the long-term effects of interventions.
Comparison with previous meta-analyses

This is the first systematic review and meta-analysis of interventions to improve the delivery of bad or difficult news by physicians, medical students, and interns/residents. However, our findings are in line with two previous meta-analyses of broader communication skills training, which were restricted to oncology settings. A Cochrane review and meta-analysis of RCTs found that communication skills training interventions were effective for improving open-ended questioning and empathy in oncology clinicians. Similarly, another meta-analysis of communication skills training in oncology found a moderate effect between communication skills training and positive communication behaviors. This meta-analysis also reported that more extensive training resulted in greater improvements than shorter interventions and suggested a trend toward interventions leading to improved patient outcomes. As with these previous meta-analyses, our meta-analysis found that a wide range of interventions have been used to train health care practitioners in communication skills, and there is no consensus regarding best practices. Our meta-analysis expands on these previous studies by focusing on one aspect of communication, bad or difficult news delivery, which is known to particularly influence the emotional well-being of both patients and physicians, and by finding evidence of effectiveness in a wider range of health care settings. However, our meta-analysis found comparatively larger overall effect sizes for the main and secondary outcomes (observer-rated news delivery skills and physician self-confidence). This may have been achieved by the tighter focus on news delivery interventions in comparison with generic communication interventions, as news delivery interventions are content-specific and which situations they should be applied to is clearer. This clarity may increase the likelihood of accurate and effective use by physicians over the long term and, therefore, support the use of news delivery interventions over more generic communication interventions.
Implications for clinicians, policymakers, and researchers

The importance of physicians being proficient in communication skills, particularly bad or difficult news delivery skills, is highlighted in guidance from both NICE (United Kingdom) and the US Preventive Task Force.\textsuperscript{18–21} However, research from a range of health care settings suggests a majority of patients are dissatisfied with this element of care.\textsuperscript{22,23} Our results suggest that news delivery interventions are effective in improving news delivery skills and indicate that implementation of such interventions could improve adherence to guidelines for communication in health care. Our results also suggest that interventions are more effective when based on the SPIKES framework for news delivery.\textsuperscript{14} This framework was initially developed in oncology but studies included in the meta-analysis applied it to pediatrics,\textsuperscript{13,35,44} obstetrics,\textsuperscript{9,40} and primary care,\textsuperscript{34} suggesting that it can be adapted for a range of settings. It should be noted, though, that only limited information was available regarding the extent to which curriculum developers within each individual study relied on specific theoretical models, thus, this finding should be viewed tentatively. Our results also suggest that such interventions may be beneficial for the well-being of physicians. Breaking bad or difficult news is often experienced as stressful by physicians,\textsuperscript{8,42,55} but our meta-analysis suggests that interventions can significantly enhance physician confidence in this aspect of health care, which may reduce stress in relation to news delivery events and support physician well-being.

Our meta-analysis highlights two main limitations in the current evidence base, which could be addressed in future research. First, only one of the included studies measured patient outcomes (in this case, patient-reported depression and/or anxiety),\textsuperscript{15} and there is a need for trials to test whether interventions in news delivery skills translate to improved patient satisfaction and reduced patient depression and/or anxiety following news delivery events. The one study that did test this found small, significant reductions in depression and/or
anxiety for patients seeing physicians trained in news delivery;\textsuperscript{15} while this is promising, it also indicates that such studies may require large sample sizes to detect effects. Second, there is a need to further test individual components of news delivery interventions to identify which of these actively impact skills improvement. Only two studies included in the meta-analysis had more than one intervention condition,\textsuperscript{37,38} which prevented any sub-group analysis exploring whether adding more learning elements increased outcomes. Further research could also examine the optimal length of an intervention for producing positive change, as parsimonious interventions are increasingly desired in resource-limited health systems.

**Summary**

This systematic review and meta-analysis suggests that interventions for improving the breaking of bad or difficult news are effective for improving news delivery skills in physicians, medical students, and interns/residents and physician confidence in news delivery. Improvements were greater when interventions were based on the SPIKES framework. Further research is needed to test the impact of interventions on patient outcomes and to determine the optimal components and length.
References


35. Attar MA, Hernandez ME, Mullan PB, Tang TS, Haftel H. Pediatric residents’ competency in communicating bad news and eliciting spiritual needs: Development


References cited in Appendix 1 only


Figure Legends

Figure 1
Flowchart of the study identification process used in a September 2016 systematic review and meta-analysis of the literature on bad or difficult new delivery interventions. The authors updated the search results in February 2017.

Some studies were excluded for more than one reason so the listed studies do not equal 53.

Two papers reported on the data collected from a single study and population. The authors treated these as a single study in the analyses and reporting.

Figure 2
Forest plot of the effects of interventions on observer-rated news delivery skill scores from a September 2016 (updated in February 2017) systematic review and meta-analysis of the literature on bad or difficult new delivery interventions. Each solid line represents one study in the meta-analysis, plotted according to the standardized mean difference (SMD). The solid box on each solid line shows the SMD for that study, and the open box on the dashed line represents the pooled SMD. Study IDs given in the following format Author yearref.

Abbreviations: ES indicates standardized mean difference; CI, confidence interval; Bowyer 2010a, Bowyer intervention group 1 (this group watched a 15-minute video on SPIKES); Bowyer 2010b, Bowyer intervention group 2 (this group received a 45-minute didactic lecture on SPIKES, observed an example of bad or difficult news delivery, and participated in a small-group discussion); Bowyer 2010c, Bowyer intervention group 3 (this group received both interventions received by groups 1 and 2); Daetwyler 2010a, Daetwyler intervention group 1 (this group completed an e-learning course on the theory and practice of delivering bad or difficult news); Daetwyler 2010b, Daetwyler intervention group 2 (this group completed the same e-learning course and a simulated news delivery task with feedback).
**Figure 3**

Forest plot of the effects of interventions on physician confidence in news delivery scores from a September 2016 (updated in February 2017) systematic review and meta-analysis of the literature on bad or difficult news delivery interventions. Each solid line represents one study in the meta-analysis, plotted according to the standardized mean difference (SMD). The solid box on each line solid shows the SMD for that study, and the open box on the dashed line represents the pooled SMD. Study IDs given in the following format Author year ref.

Abbreviations: ES indicates standardized mean difference; CI, confidence interval; Bowyer 2010a, Bowyer intervention group 1 (this group watched a 15-minute video on SPIKES); Bowyer 2010b, Bowyer intervention group 2 (this group received a 45-minute didactic lecture on SPIKES, observed an example of bad or difficult news delivery, and participated in a small-group discussion); Bowyer 2010c, Bowyer intervention group 3 (this group received both interventions received by groups 1 and 2).
Records identified through database searches (n = 3,206)

Additional records identified through reference list scanning (n = 41)

Records after duplicates (n = 977) removed (n = 2,270)

Titles and abstracts screened (n = 2,270)

Records excluded (n = 2,199)

53 full-text articles excluded, with reasons:
• 7 were not in clinical staff or student populations
• 2 were not in physician/resident/interns or medical student populations
• 6 did not test an intervention
• 22 did not include a control group
• 34 did not report a quantitative communication skill outcome measure
• 2 were eligible for the review but the authors were unable to gather the relevant outcomes in the necessary format for meta-analysis

Full-text articles assessed for eligibility (n = 71)

Studies included in the systematic review and meta-analysis (n = 17)
Figure 2

<table>
<thead>
<tr>
<th>Study ID</th>
<th>SMD (95% CI)</th>
<th>% Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexander 2006</td>
<td>0.98 (0.39, 1.57)</td>
<td>5.42</td>
</tr>
<tr>
<td>Amiel 2006</td>
<td>1.10 (0.37, 1.83)</td>
<td>4.78</td>
</tr>
<tr>
<td>Attar 2010</td>
<td>1.37 (0.49, 2.25)</td>
<td>4.14</td>
</tr>
<tr>
<td>Bowyer 2010a</td>
<td>0.30 (0.08, 0.52)</td>
<td>6.94</td>
</tr>
<tr>
<td>Bowyer 2010b</td>
<td>0.38 (0.14, 0.62)</td>
<td>6.88</td>
</tr>
<tr>
<td>Bowyer 2010c</td>
<td>0.54 (0.30, 0.78)</td>
<td>6.88</td>
</tr>
<tr>
<td>Daetwyler 2010a</td>
<td>0.42 (-0.25, 1.09)</td>
<td>5.05</td>
</tr>
<tr>
<td>Daetwyler 2010b</td>
<td>0.61 (-0.08, 1.30)</td>
<td>4.96</td>
</tr>
<tr>
<td>Fujimoto 2014</td>
<td>1.44 (0.64, 2.24)</td>
<td>4.48</td>
</tr>
<tr>
<td>Gorniewicz 2017</td>
<td>1.23 (0.54, 1.92)</td>
<td>4.96</td>
</tr>
<tr>
<td>Karkowsky 2016</td>
<td>0.23 (-0.42, 0.86)</td>
<td>5.15</td>
</tr>
<tr>
<td>Marko 2015</td>
<td>2.83 (2.20, 3.46)</td>
<td>5.24</td>
</tr>
<tr>
<td>Merckaart 2013</td>
<td>0.80 (0.39, 1.21)</td>
<td>6.24</td>
</tr>
<tr>
<td>Morton 2000</td>
<td>0.54 (-0.17, 1.25)</td>
<td>4.87</td>
</tr>
<tr>
<td>Nellis 2017</td>
<td>0.33 (-0.53, 1.19)</td>
<td>4.22</td>
</tr>
<tr>
<td>Silva 2008</td>
<td>2.89 (1.20, 4.58)</td>
<td>1.92</td>
</tr>
<tr>
<td>Szmullowicz 2010</td>
<td>0.57 (0.00, 1.14)</td>
<td>5.52</td>
</tr>
<tr>
<td>Vetto 1999</td>
<td>0.32 (0.01, 0.63)</td>
<td>6.64</td>
</tr>
<tr>
<td>Wijnen-Meijer 2015</td>
<td>-0.77 (-1.30, -0.24)</td>
<td>5.70</td>
</tr>
</tbody>
</table>

Overall (I-squared = 83.5%, P = 0.000) | 0.74 (0.46, 1.01) | 100.00 |
Figure 3
## Appendix 1

### Participant Demographics and Setting and Design Characteristics of the Studies Included in a September 2016 (Updated in February 2017) Systematic Review and Meta-Analysis of the Literature on Bad or Difficult New Delivery Interventions

<table>
<thead>
<tr>
<th>First author, year</th>
<th>Participants (recruitment strategy)</th>
<th>Setting (country)</th>
<th>Male, no. (%)</th>
<th>Age (in years), mean</th>
<th>Research design</th>
<th>Intervention</th>
<th>Theoretical basis of the training</th>
<th>Control</th>
<th>Measurement time points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexander, 200633</td>
<td>Medical residents (unclear)</td>
<td>Palliative care (US)</td>
<td>29/57 (52)</td>
<td>NA</td>
<td>CBA</td>
<td>Ninety-minute breaking bad news session delivered as part of a 5-hour communication skills training program that included small-group lectures and discussions, videos, and role-plays. After the 90-minute session, participants undertook 2 hours of communication role-plays.</td>
<td>Unclear</td>
<td>No intervention</td>
<td></td>
</tr>
<tr>
<td>Amiel, 200634</td>
<td>General practitioners (unclear)</td>
<td>Primary care (Israel)</td>
<td>13/34 (38)</td>
<td>IG: 44 CG: 46</td>
<td>CBA</td>
<td>Fourteen 90-minute small-group sessions addressing (1) learning methods of stress management and crisis intervention, (2) coping with emotions when delivering bad or difficult news, (3) communication skills, and (4) role-plays with simulated patients.</td>
<td>Buckman’s six-step protocol</td>
<td>CG participated in a Balint group</td>
<td>All participants completed baseline and post-intervention measures. Specific time frames unclear.</td>
</tr>
<tr>
<td>Study</td>
<td>Design</td>
<td>Setting</td>
<td>Type</td>
<td>Participants</td>
<td>Intervention</td>
<td>Results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------</td>
<td>---------------</td>
<td>---------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attar, Pediatrics 2010</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>A training curriculum including three 1-hour sessions on delivering bad or difficult news. Training included didactic teaching and peer-to-peer role-playing. Residents who could not attend the training viewed materials via a web platform.</td>
<td>Integrated three frameworks: Buckman’s six-step protocol, SPIKES, and HOPE model for spirituality.</td>
<td>CG participants completed one assessment during their first year as subspecialty pediatric fellows (specific time point unclear). IG participants completed assessments during the first year of their residency and at the end of their third year (specific time point in the first year unclear).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Betson, University setting 1997</td>
<td>NA</td>
<td>NA</td>
<td>RCT</td>
<td>Participants received a 3-hour seminar on breaking bad or difficult news as part of a wider communication training course. During this, IG participants watched a culturally appropriate (Cantonese) news delivery video.</td>
<td>Unclear</td>
<td>Measured at four time points: baseline, immediately after viewing the respective videos, between 3–6 weeks after viewing the videos, and 4 months after viewing the videos.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bowyer, University setting 2010</td>
<td>NA</td>
<td>NA</td>
<td>RCT</td>
<td>There were three intervention levels: (1) a 15-minute video on SPIKES.</td>
<td>No intervention</td>
<td>News delivery skill was measured once after the</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
It seems all students were entered into the study as part of their training.

<table>
<thead>
<tr>
<th>Study Authors, Year</th>
<th>Group Description</th>
<th>Oncology</th>
<th>Participants</th>
<th>n</th>
<th>Study Design</th>
<th>Intervention Details</th>
<th>Sense of Preparedness</th>
<th>Intervention</th>
<th>Study Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daetwyler, 2010&lt;sup&gt;38&lt;/sup&gt;</td>
<td>Medical interns (participants volunteered from a pool of 62 interns)</td>
<td>Oncology (US)</td>
<td>40/55 (73)</td>
<td>29</td>
<td>nRCT</td>
<td>SPIKES; (2) a 45-minute didactic lecture on SPIKES, observation of a faculty facilitator showing a good example of news delivery, and small-group discussion; and (3) a 15-minute video and 45-minute didactic lecture on SPIKES, observation of a faculty facilitator showing a good example of news delivery, and small-group discussion.</td>
<td>Student sense of preparedness to break news was measured before the news delivery task, immediately after the task, and at the end of the 12-week rotation.</td>
<td>Unclear</td>
<td>No intervention</td>
</tr>
<tr>
<td>Fujimori, 2014&lt;sup&gt;15&lt;/sup&gt;</td>
<td>Oncologists (recruited from two hospitals)</td>
<td>Oncology (Japan)</td>
<td>26/30 (87)</td>
<td>39</td>
<td>RCT</td>
<td>Two-day long training course involving group discussion, didactic teaching with text and videos, role-playing, and SHARE&lt;sup&gt;15&lt;/sup&gt;</td>
<td>Baseline and after the intervention (with a 2-week gap between the assessments).</td>
<td>SHARE&lt;sup&gt;15&lt;/sup&gt;</td>
<td>No intervention</td>
</tr>
<tr>
<td>Study</td>
<td>Setting</td>
<td>Participants</td>
<td>Intervention</td>
<td>Methodology</td>
<td>Outcome Measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>--------------</td>
<td>--------------</td>
<td>-------------</td>
<td>-----------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gorniewicz, 2017&lt;sup&gt;39&lt;/sup&gt;</td>
<td>Family medicine and internal medicine residents (participation was part of their usual training)</td>
<td>Unclear, but seems it was held in a university setting (US)</td>
<td>A 60-minute e-learning module delivered as part of a wider communication skills training course. The module included clips of interviews with patients and quiz questions.</td>
<td>RCT</td>
<td>Baseline and 2 weeks later (after the IG had received the intervention).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Karkowsky, 2016&lt;sup&gt;9&lt;/sup&gt;</td>
<td>House staff in obstetrics and gynecology (unclear)</td>
<td>Obstetrics (US)</td>
<td>A 10-minute simulation session that covered similar material to that viewed by participants in the CG but that also reviewed the participant’s baseline simulation performance.</td>
<td>RCT</td>
<td>Baseline and post-assessment (with between 2–12 weeks after the baseline assessment. There was also a 6-month follow up, but by this point, all participants had had undergone both conditions.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marko, 2015&lt;sup&gt;40&lt;/sup&gt;</td>
<td>Medical students (recruited voluntarily over a 6-month period via convenience sampling)</td>
<td>Obstetrics (US)</td>
<td>One hour of testing and 2 hours of training involving didactic teaching, observation of faculty demonstrating good practice of news delivery, role-play with faculty feedback, and provision of course</td>
<td>CBA</td>
<td>Baseline and post-assessment (with between 4–5 weeks between assessments).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Setting</td>
<td>Intervention</td>
<td>Control</td>
<td>Randomization</td>
<td>Research Design</td>
<td>Outcomes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>---------</td>
<td>--------------</td>
<td>---------</td>
<td>---------------</td>
<td>----------------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Merckx, 2013&lt;sup&gt;41&lt;/sup&gt;</td>
<td>Medical residents (recruited via letters and phone calls to Belgian French-speaking hospitals)</td>
<td>Oncology (Belgium)</td>
<td>IG: 16/48 (33)&lt;br&gt;CG: 19/47 (40)</td>
<td>RCT</td>
<td>BIC-CST&lt;sup&gt;58&lt;/sup&gt;</td>
<td>Wait list</td>
<td>Assessments were made before randomization for both groups, after completion of the training program for the IG, and 8 months after the first assessment for the CG.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meunier, 2013&lt;sup&gt;42&lt;/sup&gt;</td>
<td>Medical residents (recruited via letters and phone calls to Belgian French-speaking hospitals)</td>
<td>Oncology (Belgium)</td>
<td>IG: 16/50 (32)&lt;br&gt;CG: 19/48 (40)</td>
<td>RCT</td>
<td>BIC-CST&lt;sup&gt;58&lt;/sup&gt;</td>
<td>Wait list</td>
<td>Assessments were made before randomization for both groups, after completion of the training program for the IG, and 8 months after the first assessment for the CG.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morton, 2000&lt;sup&gt;43&lt;/sup&gt;</td>
<td>Intensive care</td>
<td>Intensive care (UK)</td>
<td>IG: 14/32&lt;br&gt;IG: 38</td>
<td>RCT</td>
<td>Interactive workshop with short presentations,</td>
<td>Unclear</td>
<td>No intervention</td>
<td>Baseline (3–4 weeks before the</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Setting</td>
<td>Intervention</td>
<td>Communication Skills Assessment</td>
<td>Results</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>---------</td>
<td>-------------</td>
<td>---------------------------------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nellis, 2017</td>
<td>Medical residents (all second-year residents based on the pediatric intensive care unit where the study took place were invited to participate over a 2-year period)</td>
<td>Pediatric intensive care (US)</td>
<td>A 40-minute simulation-based training workshop. Participants worked in pairs on a simulation of caring for a critically ill child. One resident discussed the neurologic implications of a prolonged arrest with the parent, and the second resident informed the parent that the child had died after a second arrest. Residents received feedback on their communication skills from a trained faculty facilitator.</td>
<td>Unclear</td>
<td>CG received informational resources focused on communicating with children and their families at the end of life.</td>
<td>Self-perceived competency was measured before and after the intervention. Observer-rated communication skills were measured once (6 weeks after the intervention).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silva, 2008</td>
<td>Pediatric medical residents (participated as part of their required teaching activities)</td>
<td>Pediatrics (Puerto Rico)</td>
<td>Four 1-hour sessions, each involving didactic teaching, viewing a good example of bad or difficult news delivery, viewing videos of news delivered both appropriately and SPIKES</td>
<td>No intervention</td>
<td>Measures were taken 3 weeks after the intervention was completed.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Setting</td>
<td>Intervention</td>
<td>Study design</td>
<td>Outcome Measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Szmulowicz, 2010^45</td>
<td>Internal medicine residents (all second-year residents at the hospital were invited by e-mail, with a follow-up letter at 2 weeks)</td>
<td>Six hours of training involving small-group discussions, modeling of conversations, didactic teaching, and role-plays with feedback.</td>
<td>nRCT (two participant s switched to the CG after they were randomized to the IG)</td>
<td>Baseline and after the intervention. The average amount of time between the intervention and the second assessment was 14 weeks.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vetto, 1999^47</td>
<td>Medical students (unclear, but students were recruited from a university)</td>
<td>Delivered as part of a new curriculum for medical students. It included 6 hours of panel discussions with clinicians, group discussions, and role-playing.</td>
<td>CBA</td>
<td>Assessment was conducted at one time point following the intervention during students’ third-year family practice clerkships.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wijnen-Meijer, 2015^46</td>
<td>Medical students (recruited on a voluntary basis via ads)</td>
<td>A new medical curriculum featuring early clinical experience, integration of science and practice, and progressive increase of clinical responsibility. The curriculum</td>
<td>CBA</td>
<td>Assessment was conducted at one time point following the intervention (for the IG this was 2 weeks prior to their graduation and for</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
allocated more time to training in communication skills, although specific details of training in bad or difficult news delivery were not provided. the CG this was 3 months prior to their graduation).

Abbreviations: US indicates United States; NA, not available; CBA, controlled before-after study; CG, control group; IG, intervention group; SPIKES, six-step framework for bad or difficult news delivery (steps are as follows: setting up the interview, assessing the patient’s perception of the situation, obtaining the patient’s invitation to deliver the news, giving knowledge and information to the patient, addressing the patient’s emotions empathically, and providing a summary and discussing prognosis and treatment options); HOPE, model of spirituality which instructs physicians to elicit information from their patients concerning their sources of hope, organized religion, personal spirituality, and the effects of this on their medical care; RCT, randomized controlled trial; nRCT, non-randomized controlled trial; SHARE, four-step protocol for delivering bad or difficult news (steps are as follows: create a supportive environment, consider how to deliver the news, discuss additional information that patients would like to know, and provide reassurance and emotional support); BIC-CST, Belgian Interuniversity Curriculum - Communication Skills Training; UK, United Kingdom; EDHEP, European Donor Hospital Education Programme.

a Where available, the authors report the overall sample number and percentage or mean but some studies only provided these by IG or CG.

b These papers report data collected from a single study and population. The authors treated these as a single study in the analyses and reporting. That is, only the data reported by Merckaert and colleagues was included in the main meta-analysis, and only the data reported by Meunier and colleagues was reported in the secondary meta-analysis.

c This data is based on the overall sample included in the study, which included 32/64 (50%) nurses. Nurse outcome data is not included in the meta-analyses.