REVIEWS

The Empirical Basis for Determinations of Medical Futility

Ezra Gabbay, MD¹, Jose Calvo-Broce, MD¹, Klemens B. Meyer, MD¹, Thomas A. Trikalinos, MD², Joshua Cohen, PhD², and David M. Kent, MD²

¹Division of Nephrology, Tufts Medical Center, Boston, MA, USA; ²Clinical and Translational Research Institute, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Sackler School of Graduate Biomedical Sciences, Tufts University School of Medicine, Boston, MA, USA.

BACKGROUND: Decisions to limit treatment in critically ill patients often rely on publications that make claims of futility based on outcome data. Our objective was to systematically review the criteria for futility and the strength of empirical evidence across clinical studies that purport to support or refute claims of futility.

METHODS: The MEDLINE database was searched for relevant articles published between 1980 and 2008. Selected studies reported original outcome data in critically ill or cardiac arrest patients and claimed that these data can support or refute decisions to limit treatment in comparable patients. Two authors independently abstracted data on patient characteristics, intervention, outcomes, cost, and design.

RESULTS: Forty seven studies supporting a claim of futility and 45 refuting it were reviewed. Median point estimate for adverse outcome in studies supporting claims of futility was 100% (range 75% to 100%); median lower 95% confidence limit was 91% (range 48% to 99%). Explicit thresholds for futility were missing in 88% of articles. The original criteria for quantitative futility were fulfilled by only 28% of data, and almost exclusively in studies of cardiopulmonary resuscitation (CPR) for cardiac arrest. Substantial statistical overlap was observed between data brought in support of futility claims and data brought to refute them.

CONCLUSIONS: Most studies that purport to guide determinations of futility are based on insufficient data to provide statistical confidence for clinical decision-making. They usually lack explicit a priori thresholds for determination of futility. Many studies draw disparate conclusions based on statistically similar data. In most circumstances these problems preclude confident determinations of futility.

KEY WORDS: critically ill; futility; empirical evidence; cardiac arrest.

J Gen Intern Med 25(10):1083-9

DOI: 10.1007/s11606-010-1445-3

© Society of General Internal Medicine 2010

INTRODUCTION

Many factors influence medical decisions regarding critically ill patients, including the patient's wishes, underlying and acute medical conditions, and treatment setting¹⁻⁵. With the advent of new intensive care and resuscitation modalities, physicians are often faced with demands for treatments that are technically feasible but which they may regard as clinically inappropriate because the prognosis is dismal^{6,7}.

The emergence of the concept of medical futility in the 1980s reflected, at least in part, an attempt to re-assert professional judgment in an era when patient autonomy had achieved primacy in ethics and law⁸. Efforts to curb healthcare costs have focused public attention on resource-intensive treatments that are unlikely to help patients whose prognosis is very poor⁹⁻¹¹. However, the concept of futility has proven to be very difficult to define and apply.

In an oft-cited quantitative definition of futility, Scheinderman and colleagues proposed that "when physicians conclude (either through personal experience, experiences shared with colleagues or consideration of reported empiric data) that in the last 100 cases, a medical treatment has been useless, they should regard that treatment as futile"⁸. The statistical basis for this determination is that given 100 successive treatment failures "the clinician can be 95% confident that no more than three successes would occur in every 100 comparable trials"⁸. This quantitative approach was meant to promote decisions to limit treatment that are based on empirical evidence, rather than on subjective opinion alone. Lack of consensus and criteria, and inconsistency in clinical thinking have been major obstacles to a systematic approach to prognosis-based futility determinations¹². Adopting a quantitative approach might appear to address these issues, but without consensus about what constitutes quantitative futility, it may only disguise subjectivity with statistics.

The following were our objectives: 1) to explore the degree to which studies that use quantitative data to support or refute claims of therapeutic futility rely on explicit criteria, and 2) to systematically evaluate variability in statistical confidence conferred by reported data in this context. The implications of our findings for clinical decision-making are discussed.

METHODS

Study Selection and Eligibility

A systematic MEDLINE search for articles with original outcome data pertaining to futility or treatment limitation

Received December 17, 2009

Revised May 26, 2010

Accepted June 14, 2010

Published online July 20, 2010

based on poor prognosis was conducted using the PUBMED search engine. Search limits were: English language, human subjects, and publication date 1/1/1980–11/1/2008. No geographical or age restrictions were set.

Search terms included the following Medical Subject Heading (MeSH) terms: intensive care; critical illness, intubation, artificial respiration, mechanical ventilation (entry term), resuscitation; dialysis, vasoconstrictor agents, mortality, death, survival, medical futility, and the following free text terms: futile, ineffective, withdraw, and withhold.

Review of abstracts, determination of eligibility and data extraction were conducted by two independent reviewers, who subsequently resolved differences by consensus. At the first stage of selection, all abstracts were reviewed. Those that were potentially eligible based on the criteria detailed below were retrieved as full articles. Qualifying articles had primary data (or were meta-analyses based on systematic reviews), included more than 10 patients overall, and pertained to critically ill or cardiac arrest patients in whom maximal curative (rather than only palliative) treatment effort was made. An additional requirement for inclusion was a clear statement or direct implication that the decision to withdraw or withhold treatment in a specifically defined group of patients could be accepted or rejected on the basis of outcome data provided.

Excluded articles included opinion pieces, editorials, case reports and reviews. Studies of outcome *after* a decision to withhold or withdraw treatment were excluded. Additional excluded categories were studies examining attitudes, perceptions and practices of medical teams, patients, or families, studies examining the decision to withdraw or withhold treatment as the main outcome, and studies in which either the size of the group on which the futility claim is based or the number of adverse outcomes could not be determined.

The consensus process consisted of a face-to-face meeting of both reviewers to discuss each abstract about which there was discrepancy in the decision to exclude. The reviewers together went through each of the inclusion and exclusion criteria listed above. If a clear reason for exclusion was not found, and a potential for meeting the inclusion criteria was thought to exist, the abstract was selected for full review of the article. The process was repeated after review of the full article in all instances of discrepancy regarding inclusion in the final analysis. The reviewers met a third time to ensure agreement on each element abstracted from the articles that were ultimately included in the analysis.

Data Extraction and Analysis

Extracted elements included bibliographic information and study characteristics based on the PICOD (population, intervention, comparator, outcome, design) schema for question formulation¹³. Additional extracted data included the main intervention, outcome upon which the conclusion was drawn (e.g. hospital survival, survival with good neurological outcome) and study design. Additional extracted elements included the principal conclusion (i.e. support or refute a futility claim), criteria for defining futile treatment. Analysis of more than one prognostic stratum per article was done if the abstract stated clearly that conclusions were based on more than one patient

group. In articles that explicitly reported a predictive model, the model's reported sensitivity and specificity were extracted.

Descriptive Analysis

Studies were divided into two main categories: those in which the principal conclusion was that treatment in a given situation was futile (studies supporting a futility claim), and those in which the principal conclusion was that treatment might potentially be effective (studies refuting a futility claim). The focus of the analysis was the group of patients on which the futility claim was based. For example, in a study on the effectiveness of CPR in 865 cardiac arrest patients, the focus was on the subset of 131 patients whose cardiac arrest had specific characteristics associated with 100% mortality, and who were the basis for the conclusion regarding futility¹⁴. The proportion of studies providing explicit criteria for futile treatment (defined as a direct reference to a threshold of futility for determining if the observed data satisfy the definition of futile treatment) was determined. The proportion of studies meeting Schneiderman's threshold was calculated. Two interpretations were used to define this threshold: a stricter interpretation requiring a 100% adverse outcome rate in at least 100 patients, and a more relaxed threshold in which the adverse outcome rate could be slightly lower than 100%, as long as the lower 95% confidence limit was greater than 96.4%, which is the lower 95% confidence limit for a proportion of 100/100. Reporting of results was guided by recent guidelines for systematic reviews¹⁵.

Statistics

The 95% CI for the proportion of adverse outcome in the prognostic stratum of interest was calculated using the exact binomial method. Rounding was performed to the nearest integer percentage; all point estimates and confidence limits between 99 and 100 were rounded to 99%. Positive and negative predictive values (PPV and NPV, respectively) were calculated based on outcome prevalence, and model sensitivity and specificity, using Bayes' rule. Microsoft Excel® was used for data extraction and basic calculation, Stat-Pages binomial calculator to calculate the 95% CI for outcome proportions <http://statpages.org/confint.html>.

FINDINGS

Article Selection (Fig. 1)

The search yielded 1735 abstracts, of which 1525 were rejected, that either clearly met the exclusion criteria, or were obviously irrelevant to the subject matter (e.g. discussed futility of treatment for obesity); after resolving disagreements over 67 of the 1735 abstracts, 210 references were selected for review of the full article. Of these, 118 were rejected based on one or more of the inclusion and/or exclusion criteria. The principal reason for exclusion fell into one of six categories (Fig. 1). Ninety-two articles were selected for full data extraction and analysis. The reviewers

Figure 1. Study selection flow diagram. The principal reason for rejection of articles in the full articles review phase fell into one of the following six categories: 1) Article does not directly address the issue of futility of intensive care or resuscitation; 2) Data presentation and/or methodology of the article render the data presented therein not amenable to the analysis specific to this review (e.g. the article does not explicitly mention the absolute number of adverse outcomes and/or the total number of patients in the group upon which the conclusion is based); 3) Article lacks a clear conclusion as to whether treatment should be withheld or withdrawn based on the data presented; 4) Article examines the *decision* to give or withdraw care as the primary outcome; 5) Article primarily examines the efficacy of specific diagnostic or therapeutic interventions (e.g. steroids, pulmonary artery catheters, antibiotics), rather than the futility of the overall life-saving effort; 6) Article is not strictly a clinical study of patient outcomes. These categories were not necessarily mutually exclusive. However, in the interest of simplicity, the articles are presented by the primary category for exclusion.

initially disagreed about inclusion or exclusion of 7 of the 210 articles, but reached agreement in all cases after discussion. In 47 articles, the principal conclusion was that treatment thought to be possibly effective is actually futile, and in 45 the conclusion was that treatment thought possibly to be futile may be potentially effective. Median year of publication was 1997 for the first category and 2000 for the second. Table 1 summarizes the study population and intervention characteristics, and study design in the abstracted articles.

Point Estimates and 95% CIs in the Prognostic Strata of Interest (Fig. 2)

Since a few articles had more than one group of patients whose outcomes directly supported the conclusion, we analyzed outcomes for 54 groups in studies supporting futility claims and 46 groups in studies refuting futility claims. While many studies supporting futility claims had adverse outcome rates of 100%, some had rates as low as 75% (median 100%, interquartile range –IQR– 97–100%). Lower 95% confidence limit for the adverse outcome rates varied substantially, ranging from 48% to 99% (median 91%, IQR 80% to 97%). Upper 95% confidence limit ranged from 78% to 100% (median 100%, IQR 99–100%). The size of the groups (n) in this category ranged from 5 to 2592 patients (median 95, IQR 28–230).

Table 1. Study Characteristics

	Studies supporting claims of futility	Studies refuting claims of futility
Number of studies	47	45
Clinical setting *:		
ICU§	24 (51%)	16 (36%)
Cardiac arrest requiring CPR§	23 (49%)	10 (22%)
Pediatric or neonatal patients	3 (6%)	7 (16%)
Elderly	1 (2%)	14 (31%)
Cancer patients or BMT§ recipients	7 (15%)	7 (16%)
Trauma, burns and other surgical setting	10 (21%)	16 (36%)
Design:		
Retrospective observational	34 (72%)	29 (64%)
Prospective	10 (21%)	11 (24%)
Other**	3 (6%)	5 (11%)

*Categories not mutually exclusive or exhaustive

** Retrospective analysis of prospectively collected data, studies with both retrospective and prospective elements, meta-analyses, and studies the design of which could not be determined

§ICU: intensive care unit, CPR: cardiopulmonary resuscitation, BMT: bone marrow transplantation

Studies refuting futility claims had adverse outcome rates that varied substantially, ranging from 0% to 96% (median 56%, IQR 37% to 85%). Lower 95% confidence limit ranged from 0% to 94% (median 37%, IQR 24% to 73%). Upper 95% confidence limit ranged from 13% to 99% (median 69%, IQR 45% to 90%). The size of the groups (n) in this category ranged from 11 to 2093 (median 90, IQR 50 to 145).

As shown in Figure 2, there was considerable overlap in point estimates and CIs. Of the 54 groups whose outcome was used to support a futility claim, 12 (22%) had point estimates lower than or equal to 96% (the highest point estimate used to refute a claim); 24 (44%) had point estimates lower than or equal to 99% (the largest upper 95% confidence limit for outcomes used to refute such a claim). Of the 46 groups whose outcomes were used to refute claims of futility, 14 (30%) had point estimates greater than or equal to 75% (the smallest point estimate used to support a futility claim); 26 (57%) had point estimates above 48% (the smallest lower 95% confidence limit in studies supporting futility claims).

Thirteen articles supporting a claim of futility employed predictive modeling. The range of reported specificity for predictive models was 98% to 100%. Corresponding sensitivity for these models as predictors of poor outcomes was very low (median 37%, range 12.5–58%).

Schneiderman's Threshold (Table 2)

Fifteen of 54 groups (28%) with outcome rates that purportedly supported claims of futility had more than 100 patients with a very high adverse outcome rate, resulting in a lower 95% confidence limit of 96.4% mortality or more. Fourteen of these (93%) were groups of patients with cardiac arrest requiring cardiopulmonary resuscitation (CPR). The one other group was

Figure 2. Data point estimates in studies supporting claims of futility (black diamond) and those refuting such claims (blue square), with 95% confidence interval (CI). Data supporting claims of futility were arranged in descending order of lower 95% confidence limit, data refuting such claims were arranged in ascending order of upper 95% confidence limit.

of intensive care unit (ICU) patients. If a stricter definition of 100% adverse outcome in a group of 100 patients or more is used, only 13 groups meet Schneiderman's threshold, 12 of which are cardiac arrest patients receiving CPR.

Explicit Definition of Futility (Tables 3 and 4)

Only six of 47 articles that make a claim of futility (13%) had explicit definitions of futility or other criteria for limiting treatment. Five of the six articles referred to Schneiderman's original article, but only one has explicit reference to the minimum denominator of at least 100 required to reach the statistical confidence implied by Schneiderman's definition¹⁶.

Among studies that refute claims of futility, five (11%) gave a definition of futile therapy. Three articles cited Schneiderman's article, but only one specifically mentions the need for a sufficiently large denominator¹⁷.

DISCUSSION

Decisions on limiting curative treatment usually begin with a clinical impression regarding prognosis. Since personal experience

alone is often insufficient to substantiate this impression, clinicians often rely on published data on outcomes in comparable patients. Our aim was to examine the degree of variation or consistency in quantitative data and criteria used to support or refute claims of futility.

We found that most articles which support withholding or withdrawing treatment on empirical grounds do so based on data that show an adverse outcome rate of 100% or very close to it. The degree of statistical confidence, however, is highly variable. Most articles purporting to support claims of futility do not meet the widely-cited quantitative definition of futility, which specifies a lower 95% confidence limit for the probability of adverse outcome between 96% and 97%. This is often the result of small sample size. The range of specificity for predictive models, where reported, was 98% to 100%. The models' corresponding sensitivity as predictors of poor outcomes was very low (median 37%, range 12.5–58%). Given these models' discriminative characteristics, the cost of high specificity is relatively low sensitivity. The low sensitivity results in the identification of relatively few patients in the poor prognostic category upon which the futility claim is based. The small number of patients, in turn, results in the relatively large CI

Table 2. Studies Meeting Schneiderman's Criteria⁽¹⁾

Author	Intervention/ Population	x/n	P (%)	Lower 95% confidence limit
Chang 1989 ²⁹	ICU§/ adult ICU patients	109/109**	100	97
Cone 2005 ¹⁶	CPR§/ OOHCA§	285/285**	100	99
Ewer 2001 ³⁰	CPR§/cancer patients	171/171**	100	98
Falcone 1995 ²⁴	CPR + flight evacuation/ OOHCA	314/314**	100	99
Marsden 1995 ³¹	CPR/OOHCA	240/240**	100	98
O'Brien 2007 ¹⁴	CPR/OOHCA	131/131**	100	97
Richman 2008 ³²	CPR/OOHCA	1159/1160*	99	99
Rosemurgy 1993 ²⁵	CPR/traumatic OOHCA	138/138**	100	97
Sasson 2008 ³³	CPR/OOHCA	2587/2592*	99	99
Sasson 2008 ³³	CPR/OOHCA	1192/1192**	100	99
Stockinger 2004 ²⁶	CPR/traumatic OOHCA	195/195**	100	98
Van-der Hoeven 1993 ³⁴	CPR/OOHCA- adult	165/165**	100	98
van Walraven 1999 ³⁵	CPR/in hospital cardiac arrest	119/119**	100	97
van Walraven 2001 ³⁶	CPR/in- hospital cardiac arrest	266/269*	99	97
Väyrynen 2008 ³⁷	CPR/a- systolic OOHCA	291/291**	100	99

(1) Schneiderman's criteria: futility can be determined based on treatment failure in 100 consecutive cases in which it was attempted
x/n : number of adverse outcomes (x) in the patient group on which the conclusion is based (n), P(%): the percentage of adverse outcomes

*Data meeting threshold based on lower 95% confidence limit >96.4

**Data meeting threshold based on additional criterion of 100% adverse outcome

§ICU: intensive care unit, CPR: cardiopulmonary resuscitation, OOHCA: out of hospital cardiac arrest

Table 3. Explicit Definitions of Futility in Studies Supporting Claims of Futility

Author	Intervention/ population	Definition	Reference to Schneiderman's threshold	Type of definition	Comment
Cone 2005 ¹⁶	CPR§/ OOHCA§	A 1/100 chance of helping the patient	Yes	A posteriori	Addresses issue of the confidence interval & need for denominator ≥ 100
Hamel 1997 ³⁸	Dialysis + ICU§/ ICU	CE ratio of \$50,000/ QALY§	No	A priori	Does not discuss futility as much as cost- effectiveness
Rodriguez 1997 ³⁹	ICU/ ICU	Less than 2% chance of survival with favorable function	Yes	A priori	Refers to Schneiderman's threshold but does not necessarily accept it
Sasson 2008 ³³	CPR/OOHCA	Less than 1% success	Yes	A priori	Does not refer to denominator, but refers to power to detect misclassification
Teno 1994 ⁴⁰	ICU or hospital treatment/ seriously ill adult patients	No more than 1% chance of 2 month survival	Yes	A priori	Refers to Schneiderman's threshold but does not necessarily accept
Väyrynen 2008 ³⁷	CPR/a-systolic OOHCA	Cut off of 1% chance of survival	Yes	A priori	No reference to denominator

§ICU: intensive care unit, CPR: cardiopulmonary resuscitation, OOHCA: out of hospital cardiac arrest, CE: cost effectiveness, QALY: quality adjusted life year.

surrounding the point estimate of the outcome rate, and the failure to satisfy Schneiderman's criterion.

Explicit definitions or criteria for futility were only given in 12% of articles. Schneiderman's original quantitative definition appears to be the most commonly cited reference point for futility. Most articles that refer to it, however, generally disregard the need for a high degree of statistical confidence conferred by a sample size of 100 patients or more. Only a few authors fully acknowledge that wide CIs in small groups preclude confident conclusions regarding futility^{16–19}. Data that meet Schneiderman's criteria are reported almost entirely in the context of CPR for certain types of cardiac arrest, indicating that it sets a relatively conservative standard. The main implication of our findings, therefore, is that in most clinical scenarios, clinicians cannot base decisions to withhold or withdraw treatment based on data-proven futility, since empirical evidence that provides a high level of statistical confidence is lacking.

The range of quantitative definitions for futility usually falls between 0% and 2% of success. While this range may

appear fairly narrow, the difference between 0% chance of success and *any* chance of success, even smaller than 1% is a paradigm shift, since any evidence of treatment effectiveness can undermine a claim of futility. The general lack of reference to the minimal number of treatment failures required to determine futility reflects a failure to recognize the importance of statistical confidence. Other authors use economic definitions or terms such as "rates of success approaching zero" that elude consensus. Only one author provides an element of time in the definition.

There is substantial overlap in both point estimates and 95% CIs between data used to support futility and data used to refute it. This finding underscores the subjective, value-laden nature of decisions to limit treatment in this context. It is also congruent with classical analyses of decision psychology, in which framing can lead to disparate conclusions on the basis of the same data. For example, the degree to which clinical scenarios are perceived as hopeless may alter the threshold for refuting futility, suggesting the influence of "framing effects". The framing effect refers to the influence of

Table 4. Explicit Definitions of Futility in Studies Refuting Claims of Futility

Author	Intervention/ population	Definition	Reference to Schneiderman's threshold	Type of definition	Comment
Azoulay 2001 ⁴¹	ICU§ + MV§/ cancer	100% mortality	Yes	A posteriori	No mention of denominator component of Schneiderman's threshold
Finer 1999 ⁴²	CPR§/neonates	"Intact survival is (im)possible"	No	A posteriori	
Huaringa 2000 ⁴³	ICU + MV/BMT§	1% survival rate	Yes	A posteriori	No mention of denominator component of Schneiderman's threshold
Swor 2000 ⁴⁴	CPR/OOHCA§, elderly	rate of success approaching zero	No	A posteriori	
Velmahos 1998 ¹⁷	Blood transfusion/ trauma	100% mortality	Yes	A posteriori	Does mention importance of denominator as a determinant of uncertainty

§ ICU: intensive care unit, CPR: cardiopulmonary resuscitation, OOHCA: out of hospital cardiac arrest, CE: cost effectiveness, QALY: quality adjusted life year, MV: mechanical ventilation, BMT: bone marrow transplantation

psychological principles that govern the perception of decision problems and the evaluation of probabilities and outcomes to produce predictable shifts of preference when the same problem is framed in different ways²⁰. An example of a framing effect is when people are presented with a problem with two potential solutions (such as strategies to avert a disease outbreak) they tend to make a risk averse choice, when its certain, albeit smaller, potential gains are made explicit (e.g. certainly saving only 200 of 600 lives is preferred to the 1/3 chance of saving 600 and 2/3 of saving none). When the potential losses of the very same options are emphasized the respondents' preferences are reversed (1/3 chance of saving 600 and 2/3 of saving none is preferred to the certain loss of exactly 400 lives)²⁰. In traumatic cardiac arrest, CPR during transport and mechanical ventilation for bone marrow transplant recipients, even mortality rates of approximately 95% can lead to a conclusion that treatment is not futile²¹⁻²³. Because the prevailing perceptions of prognosis in these scenarios are so dismal, the impact of any reports of success, however limited, may lead to a conclusion that refutes futility, whereas the same very low rates of success may lead to the opposite conclusion in less extreme settings (e.g., sepsis). These issues may represent a significant source of bias in the conclusions of studies, and mandate caution in their application to clinical practice.

Another important factor in the inherently provisional nature of data analysis in this context is the potential for future improvement in patient outcomes. For example, the three studies claiming futility of CPR for traumatic cardiac arrest were published in 1993, 1995 and 2004 based on mortality rates of 100%²⁴⁻²⁶, whereas two of three studies refuting the claim that this treatment is futile, were published in 2006 and one in 2003^{21,27,28}. Mortality rates in the latter three articles were 44% to 94%.

Thus, the problems in basing clinical decisions on futility extend beyond the lack of sufficient data and statistical confidence, to include bias, lack of uniform criteria, and the risk of ignoring potential for improved outcome with technological advances and the accumulation of clinical experience. Regardless of the problematic nature of the concept of futility and its implementation, most physicians will face dilemmas in deciding how vigorously to treat patients whose prognoses they believe to be poor. From a societal perspective, the challenge of curbing healthcare costs may prompt discussion of limiting futile, expensive treatments. Empirical data can often help to inform such decisions, but correct interpretation of these data and awareness of their statistical and methodological limitations is crucial. An approach that purports to be strictly evidence-based may distort the decision-making process for an individual patient as well as the process of formulating healthcare policy.

Limitations

The search may have failed to retrieve all relevant data. Some authors may not use the term "futility", due to the debate surrounding the concept. The terms "ineffective," "withhold," and "withdraw" were added to the search to capture articles discussing futile treatment without referring to it explicitly.

Much of the extraction and analysis depended on subjective interpretation and reconstruction of the authors' logic. The prognostic stratum and outcome upon which the conclusions

were based and other aspects of the authors' deductive reasoning, were often very difficult to define.

Finally, the analysis was focused on the statistical and methodological aspects of the determination of futility. The fundamental ethical, legal, and psycho-social issues that have been central to end of life decision making in recent decades were not systematically addressed.

CONCLUSIONS

Applying empirical outcome data to decisions about limiting treatment in critically ill patients is fraught with statistical and methodological problems. Explicit criteria for futility are usually missing, which may increase the degree of subjectivity in the authors' conclusions. Our findings imply that in most circumstances physicians cannot confidently rely on published outcome data to make determinations of medical futility. Most articles that support limitations of treatment do so based on a point estimate of 100% adverse outcome rate or very close to it. However, data are often obtained in small groups of patients, compromising the statistical confidence around the point estimates. Wide CIs reflect the risk that patients will be misclassified as unsalvageable and be denied appropriate curative effort. Only a few articles provide data meeting Schneiderman's classical definition for quantitative futility. Those that meet this standard are almost exclusively reports of outcome in patients undergoing CPR. Reports of patients who are critically ill, but not yet in cardiac arrest, almost always fail to meet this standard. Significant overlap exists in point estimates and CIs between data used to support claims of futility and data used to refute such claims, suggesting a strong influence of subjective impression in the interpretation of data and in the conclusions presented to the reader.

These limitations in the empirical basis for determinations of futility argue against a strictly empirical approach to decisions on limiting treatment in patients in very poor prognosis. Decisions that acknowledge these problems will better reflect the standards of cogent reasoning to which clinicians aspire and which our patients deserve.

Acknowledgement: This work was done with the support of the Clinical and Translational Science Center, Tufts Medical Center. Grant number: NIH/NCRR 1UL1 RR025752

Conflict of Interest: In the last 3 years, Dr. Kent consulted for Eli Lilly and received research funding from Pfizer.

Corresponding Author: Ezra Gabbay, MD; Division of Nephrology, Tufts Medical Center, 800 Washington St., Boston, MA 02111, USA (e-mail: Ezra.gabbay@tufts.edu; gabbayez@hotmail.com).

REFERENCES

1. Zingmond DS, WN. Regional and Institutional Variation in the Initiation of Early Do-Not-Resuscitate Order. *Arch Intern Med*. 2005;165:1705-12.
2. Hakim RB, Teno JM, Harrell FE Jr, et al. Factors associated with do-not-resuscitate orders: patients' preferences, prognoses, and physicians' judgments. SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment. *Ann Intern Med*. 1996;125(4):284-93.

3. A controlled trial to improve care for seriously ill hospitalized patients. The study to understand prognoses and preferences for outcomes and risks of treatments (SUPPORT). The SUPPORT Principal Investigators. *JAMA*. 1995; 274(20):1591-8.
4. **Manara AR, Pittman JA, Braddon FE.** Reasons for withdrawing treatment in patients receiving intensive care. *Anaesthesia*. 1998;53(6):523-8.
5. **Moss AH.** 'At least we do not feel guilty': Managing conflict with families over dialysis discontinuation. *Am J Kidney Dis*. 1998;31(5):868-83.
6. **Jecker NS.** Knowing when to stop: the limits of medicine. *Hastings Cent Rep*. 1991;21(3):5-8.
7. **Burns JP, Truog RD.** Futility: a concept in evolution. *Chest*. 2007;132(6):1987-93.
8. **Schneiderman LJ, Jecker NS, Jonsen AR.** Medical futility: its meaning and ethical implications. *Ann Intern Med*. 1990;112(12):949-54.
9. **Jecker NS, Schneiderman LJ.** Futility and rationing. *Am J Med*. 1992;92(2):189-96.
10. **Talmor D, Shapiro N, Greenberg D, et al.** When is critical care medicine cost-effective? A systematic review of the cost-effectiveness literature. *Crit Care Med*. 2006;34(11):2738-47.
11. **Leonhardt D:** After the Great Recession. *New York Times Magazine*, 2009.
12. **Curtis JR, Park DR, Krone MR, Pearlman RA.** Use of the medical futility rationale in do-not-attempt-resuscitation orders. *JAMA*. 1995;273(2):124-8.
13. **Counsell C.** Formulating questions and locating primary studies for inclusion in systematic reviews. *Ann Intern Med*. 1997;127(5):380-7.
14. **O'Brien E, Hendricks D, Cone DC.** Field termination of resuscitation: analysis of a newly implemented protocol. *Prehosp Emerg Care*. 2008;12(1):57-61.
15. **Moher D, Liberati A, Tetzlaff J, Altman DG.** Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009;151(4):264-9. W264.
16. **Cone DC, Bailey ED, Spackman AB.** The safety of a field termination-of-resuscitation protocol. *Prehosp Emerg Care*. 2005;9(3):276-81.
17. **Velmahos GC, Chan L, Chan M, et al.** Is there a limit to massive blood transfusion after severe trauma? *Arch Surg*. 1998;133(9):947-52.
18. **Meesters RC, van der Graaf Y, Vos A, Eikelboom BC.** Ruptured aortic aneurysm: early postoperative prediction of mortality using an organ system failure score. *Br J Surg*. 1994;81(4):512-6.
19. **Dobkin JE, Cutler RE.** Use of APACHE II classification to evaluate outcome of patients receiving hemodialysis in an intensive care unit. *West J Med*. 1988;149(5):547-50.
20. **Tversky A, Kahneman D.** The framing of decisions and the psychology of choice. *Science*. 1981;211(4481):453-8.
21. **Willis CD, Cameron PA, Bernard SA, Fitzgerald M.** Cardiopulmonary resuscitation after traumatic cardiac arrest is not always futile. *Injury*. 2006;37(5):448-54.
22. **Eisenburger P, Havel C, Sterz F, et al.** Transport with ongoing cardiopulmonary resuscitation may not be futile. *Br J Anaesth*. 2008;101(4):518-22.
23. **Huaringa AJ, Leyva FJ, Giralt SA, et al.** Outcome of bone marrow transplantation patients requiring mechanical ventilation. *Crit Care Med*. 2000;28(4):1014-17.
24. **Falcone RE, Herron H, Johnson R, et al.** Air medical transport for the trauma patient requiring cardiopulmonary resuscitation: a 10-year experience. *Air Med J*. 1995;14(4):197-203. discussion 204-195.
25. **Rosemurgy AS, Norris PA, Olson SM, et al.** Prehospital traumatic cardiac arrest: the cost of futility. *J Trauma*. 1993;35(3):468-73.
26. **Stockinger ZT, McSwain NE Jr.** Additional evidence in support of withholding or terminating cardiopulmonary resuscitation for trauma patients in the field. *J Am Coll Surg*. 2004;198(2):227-31.
27. **Cera SM, Mostafa G, Sing RF, et al.** Physiologic predictors of survival in post-traumatic arrest. *Am Surg*. 2003;69(2):140-4.
28. **Lockey D, Crewdson K, Davies G.** Traumatic cardiac arrest: who are the survivors? *Ann Emerg Med*. 2006;48(3):240-4.
29. **Chang RW.** Individual outcome prediction models for intensive care units. *Lancet*. 1989;2(8655):143-6.
30. **Ewer MS, Kish SK, Martin CG, Price KJ, Feeley TW.** Characteristics of cardiac arrest in cancer patients as a predictor of survival after cardiopulmonary resuscitation. *Cancer*. 2001;92(7):1905-12.
31. **Marsden AK, Ng GA, Dalziel K, Cobbe SM.** When is it futile for ambulance personnel to initiate cardiopulmonary resuscitation? *BMJ*. 1995;311(6996):49-51.
32. **Richman PB, Vadeboncoeur TF, Chikani V, Clark L, Bobrow BJ.** Independent evaluation of an out-of-hospital termination of resuscitation (TOR) clinical decision rule. *Acad Emerg Med*. 2008;15(6):517-21.
33. **Sasson C, Hegg AJ, Macy M, et al.** Surveillance Group. Prehospital termination of resuscitation in cases of refractory out-of-hospital cardiac arrest. *JAMA*. 2008;300(12):1432-8.
34. **van der Hoeven JG, Waanders H, Compier EA, van der Weyden PK, Meinders AE.** Prolonged resuscitation efforts for cardiac arrest patients who cannot be resuscitated at the scene: who is likely to benefit? *Ann Emerg Med*. 1993;22(11):1659-63.
35. **van Walraven C, Forster AJ, Stiell IG.** Derivation of a clinical decision rule for the discontinuation of in-hospital cardiac arrest resuscitations. *Arch Intern Med*. 1999;159(2):129-34.
36. **van Walraven C, Forster AJ, Parish DC, et al.** Validation of a clinical decision aid to discontinue in-hospital cardiac arrest resuscitations. *JAMA*. 2001;285(12):1602-6.
37. **Väyrynen T, Kuisma M, Määttä T, Boyd J.** Medical futility in asystolic out-of-hospital cardiac arrest. *Acta Anaesthesiol Scand*. 2008;52(1):81-7.
38. **Hamel MB, Phillips RS, Davis RB, et al.** Outcomes and cost-effectiveness of initiating dialysis and continuing aggressive care in seriously ill hospitalized adults. SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. *Ann Intern Med*. 1997;127(3):195-202.
39. **Rodriguez RM, Wang NE, Pearl RG.** Prediction of poor outcome of intensive care unit patients admitted from the emergency department. *Crit Care Med*. 1997;25(11):1801-6.
40. **Teno JM, Murphy D, Lynn J, et al.** Prognosis-based futility guidelines: does anyone win? SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment. *J Am Geriatr Soc*. 1994;42(11):1202-7.
41. **Azoulay E, Alberti C, Bornstain C, et al.** Improved survival in cancer patients requiring mechanical ventilatory support: impact of noninvasive mechanical ventilatory support. *Crit Care Med*. 2001;29(3):519-25.
42. **Finer NN, Tarin T, Vaucher YE, Barrington K, Bejar R.** Intact survival in extremely low birth weight infants after delivery room resuscitation. *Pediatrics*. 1999;104(4):e40.
43. **Huaringa AJ, Leyva FJ, Giralt SA, et al.** Outcome of bone marrow transplantation patients requiring mechanical ventilation. *Crit Care Med*. 2000;28(4):1014-7.
44. **Swor RA, Jackson RE, Tintinalli JE, Pirrallo RG.** Does advanced age matter in outcomes after out-of-hospital cardiac arrest in community-dwelling adults? *Acad Emerg Med*. 2000;7(7):762-8.